

Mechanism Design

Ron Lavi

The Technion – Israel Institute of Technology

<http://ie.technion.ac.il/~ronlavi>

1 Article Outline

Glossary

Definition of the subject and its importance

Introduction

Formal Model and Early Results

Quasi-linear Utilities and the VCG Mechanism

The Importance of the Domain's Dimensionality

Budget Balancedness and Bayesian Mechanism Design

Interdependent Valuations

Future Directions

References

2 Glossary

A social choice function: A function that determines a social choice according to players' preferences over the different possible alternatives.

A mechanism: a game in incomplete information, in which player strategies are based on their private preferences. A mechanism implements a social choice function f if the equilibrium strategies yield an outcome that coincides with f .

Dominant strategies: An equilibrium concept where the strategy of each player maximizes her utility, no matter what strategies the other players choose.

Bayesian-Nash equilibrium: An equilibrium concept that requires the strategy of each player to maximize the expected utility of the player, where the expectation is taken over the types of the other players.

VCG mechanisms: A family of mechanisms that implement in dominant strategies the social choice function that maximizes the social welfare.

3 Definition of the Subject and its Importance

Mechanism design is a sub-field of economics and game theory that studies the construction of social mechanisms in the presence of rational but selfish individuals (players/agents). The nature of the players dictates a basic contrast between the social planner, that aims to reach a socially desirable outcome, and the players, that care only about their own private utility. The underlying question is how to incentivize the players to cooperate, in order to reach the desirable social outcomes. A *mechanism* is a game, in which each agent is required to choose one action among a set of possible actions. The social designer then chooses an *outcome*, based on the chosen actions. This outcome is typically a coupling of a physical outcome, and a payment given to each individual. Mechanism design studies how to design the mechanism such that the *equilibrium* behavior of the players will lead to the socially desired goal. The theory of mechanism design has greatly influenced several sub-fields of micro-economics, for example auction theory and contract theory, and the 2007 Nobel prize in Economics was awarded to Leonid Hurwicz, Eric Maskin, and Roger Myerson “for having laid the foundations of mechanism design theory”.

4 Introduction

It will be useful to start with an example to a mechanism design setting, the well-known “public project” problem (Clarke [8]): a government is trying to decide on a certain public project (the common example is “building a bridge”). The project costs C dollars, and each player, i , will benefit from it to an amount of v_i dollars, where this number is known only to the player herself. The government desires to build the bridge if and only if $\sum_i v_i > C$. But how should this condition be checked? Clearly, every player has an interest in over-stating its own v_i , if this report is not accompanied by any payment at all, and most probably agents will understate their values, if asked to pay some proportional amount to the declared value. Clarke describes an elegant mechanism that solves this problem. His mechanism has the fantastic property that, from the point of view of every player, no matter what the other players declare, it is always *in the best interest* of the player to declare his true value. Thus, truthful reporting is a dominant-strategy equilibrium of the mechanism, and under this equilibrium, the government’s goal is fully achieved. A more formal treatment of this result is given in Section 6 below.

Clarke’s paper, published in the early 70’s, was part of a large body of work that started to investigate mechanism design questions. Most of the early works used two different assumptions about the structure of players’ utilities. Under the assumption that utilities are general, and that the influence of monetary transfers on the utility are not well predicted, the literature have produced mainly impossibilities, which are described in section 5. The assumption that utilities are quasi-linear in money was successfully used to introduce positive and impressive results, as discussed in detail in sections 6 and 7. These mechanisms apply the solution concept of dominant-strategy equilibrium, which is a strong solution concept that may prevent several desirable properties from being achieved. To overcome its difficulties, the weaker concept of a Bayesian-Nash equilibrium is usually employed. This concept, and one main possibility result that it provides, are described in section 8. The last important model that this entry covers aims to capture settings where the players’ values are not fully observed by the each player separately. Rather, each player receives a *signal* that gives a partial indication to her valuation. Mechanism design for such settings is discussed in section 9.

One of the most impressive applications of the general mechanism design literature is auction theory. An auction is a specific form of a mechanism, where the outcome is simply the specific allocation of the goods to the players, plus the prices they are required to pay. Vickrey [28] initiated the study of auctions in a mechanism design setting, and in fact perhaps the study of mechanisms itself. After the fundamental study of general mechanism design in the 70's, in the 80's the focus of the research community returned to this important application, and many models were studied. The reader is referred to the entry on auction theory for a detailed discussion on this subject. The entry on Implementation Theory is also very much related to the subjects discussed here.

5 Formal Model and Early Results

A social designer wishes to choose one possible outcome/alternative out of a set A of possible alternatives. There are n players, each has her own preference order \succeq_i over A . This preference order is termed the player's "type". The set (domain) of all valid player preferences is denoted by V_i . The designer has a social choice function $f : V_1 \times \dots \times V_n \rightarrow A$, that specifies the desired alternative, given any profile of individual preferences over the alternatives. The problem is that these preferences are private information of each player – the social designer does not know them, and thus cannot simply invoke f in order to determine the social alternative. Players are assumed to be strategic, and therefore we are in a game-theoretic situation.

To *implement* the social choice function, the designer constructs a "game in incomplete information", as follows. Each player is required to choose an action out of a set of possible actions \mathcal{A}_i , and a target function $g : \mathcal{A}_1 \times \dots \times \mathcal{A}_n \rightarrow A$ specifies the chosen alternative, as a function of the players' actions. A player's choice of action may, of-course, depend on her actual preference order. Furthermore, we assume an incomplete information setting, and therefore it cannot depend on any of the other players' preferences. Thus, to play the game, player i chooses a strategy $s_i : V_i \rightarrow \mathcal{A}_i$.

A strategy $s_i(\cdot)$ dominates another strategy $s'_i(\cdot)$ if, for every tuple of actions a_{-i} of the other players, and for every preference $\succeq_i \in V_i$, $g(s_i(\succeq_i), a_{-i}) \succeq_i g(s'_i(\succeq_i), a_{-i})$, for any $a_i \in \mathcal{A}_i$. In other words, no matter what the other players are doing, the player cannot improve her situation by using an action other than $s_i(\succeq_i)$.

A mechanism *implements* the social choice function f in dominant strategies if there exist dominant strategies $s_1(\cdot), \dots, s_n(\cdot)$ such that $f(\succeq_1, \dots, \succeq_n) = g(s_1(\succeq_1), \dots, s_n(\succeq_n))$, for any profile of preferences $\succeq_1, \dots, \succeq_n$. In other words, a mechanism implements the social choice function f if, given that players indeed play their equilibrium strategies (in this case the dominant strategies equilibrium), the outcome of the mechanism coincides with f 's choice.

The theory of mechanism design asks: given a specific problem domain (an alternative set and a domain of preferences), and a social choice function, how can we construct a mechanism that implements it (if at all)? As we shall see below, the literature uses a variety of "solution concepts", in addition to the concept of dominant strategies equilibrium, and an impressive set of understandings have emerged.

The concept of implementing a function with a dominant-strategy mechanism seems at first too strong, as it requires each player to know exactly what action to take, regardless of the actions the others take. Indeed, as we will next describe in detail, if we do not make any further assumptions then this notion yields mainly impossibilities. Nevertheless, it is not completely empty, and it may be useful to start with a positive example, to illustrate the new notions defined above.

Consider a voting scenario, where the society needs to choose one out of two candidates. Thus,

the alternative set contains two alternatives (“candidate 1” and “candidate 2”), and each player either prefers 1 over 2, 2 over 1, or is indifferent between the two. It turns out that the majority voting rule is dominant strategy implementable, by the following mechanism: each player reports her top candidate, and the candidate that is preferred by the majority of the players is chosen. This mechanism is a “direct-revelation” mechanism, in the sense that the action space of each player is to report a preference, and g is exactly f . In a direct-revelation mechanism, the hope is that truthful reporting (i.e. $s_i(\succeq_i) = \succeq_i$) is a dominant strategy. It is not hard to verify that in this two candidates setting, this is indeed the case, and hence the mechanism implements in dominant-strategies the majority voting rule.

An elegant generalization for the case of a “single-peaked” domain is as follows. Assume that the alternatives are numbered as $A = \{a_1, \dots, a_n\}$, and the valid preferences of a player are single-peaked, in the sense that the preference order is completely determined by the choice of a peak alternative, a_p . Given the peak, the preference between any two alternatives a_i, a_j is determined according to their distance from a_p , i.e. $a_i \succeq_i a_j$ if and only if $|j - p| \leq |i - p|$. Now consider the social choice function $f(p_1, \dots, p_n) = \text{median}(p_1, \dots, p_n)$, i.e. the chosen alternative is the median alternative of all peak alternatives.

Theorem 1 *Suppose that the domain of preferences is single-peaked. Then the median social choice function is implementable in dominant strategies.*

PROOF (SKETCH): Consider the direct revelation mechanism in which each player reports a peak alternative, and the mechanism outputs the median of all peaks. Let us argue that reporting the true peak alternative is a dominant strategy. Suppose the other players reported p_{-i} , and that the true peak of player i is p_i . Let p_m be the median index. If $p_i = p_m$ then clearly player i cannot gain by declaring a different peak. Thus assume that $p_i < p_m$, and let us examine a false declaration p'_i of player i . If $p'_i \leq p_m$ then p_m remains the median, and the player did not gain. If $p'_i > p_m$ then the new median is $p_{m'} \geq p_m$, and since $p_i < p_m$, this is less preferred by i . Thus, player i cannot gain by declaring a false peak alternative if the true peak alternative is smaller or equal to the median alternative. A similar argument holds for the case of $p_i > p_m$. ■

In a voting situation with two candidates, the median rule becomes the same as the majority rule, and the domain is indeed single-peaked. When we have three or more candidates, it is not hard to verify that the majority rule is different than the median rule. In addition, one can also check that the direct-revelation mechanism that uses the majority rule does not have truthfulness as a dominant strategy.

Of-course, many times one cannot order the candidates on a line, and any preference ordering over the candidates is plausible. What voting rules are implementable in such a setting? This question was asked by Gibbard [12] and Satterthwaite [27], who provided a beautiful and fundamental impossibility. A domain of player preferences is unrestricted if it contains all possible preference orderings. In our voting example, for instance, the domain is unrestricted if every ordering of the candidates is valid (in contrast to the case of a single-peaked domain). A social choice function is dictatorial if it always chooses the top alternative of a certain fixed player (the dictator).

Theorem 2 ([12, 27]) *Every social choice function over an unrestricted domain of preferences, with at least three alternatives, must be dictatorial.*

The proof of this theorem, and in fact of most other impossibility theorems in mechanism design, uses as a first step the powerful direct-revelation principle. Though the examples we have seen above use a direct revelation mechanism, one can try to construct “complicated” mechanisms with “crazy” action spaces and outcome functions, and by this obtain dominant strategies. How should one reason about such vast space of possible constructions? The revelation principle says that one cannot gain extra power by such complex constructions, since if there exists an implementation to a specific function then there exists a direct-revelation mechanism that implements it.

Theorem 3 (The direct revelation principle) *Any implementable social choice function can also be implemented (using the same solution concept) by a direct-revelation mechanism.*

PROOF: Given a mechanism M that implements f , with dominant strategies $s_i^*(\cdot)$, we construct a direct revelation mechanism M' as follows: for any tuple of preferences $\succeq = (\succeq_1, \dots, \succeq_n)$, $g'(\succeq) = g(s^*(\succeq))$. Since $s_i^*(\cdot)$ is a dominant strategy in M , we have that for any fixed $\succeq_{-i} \in V_{-i}$ and any $\succeq_i \in V_i$, the action $a_i = s_i^*(\succeq_i)$ is dominant when i 's type is \succeq_i . Hence declaring any other type $\tilde{\succeq}_i$ that will “produce” an action $\tilde{a}_i = s_i^*(\tilde{\succeq}_i)$, cannot increase i 's utility. Therefore the strategy \succeq_i in M' is dominant. ■

The proof uses the dominant-strategies solution concept, but any other equilibrium definition will also work, using the same argumentation. Though technically very simple, the revelation principle is fundamental. It states that, when checking if a certain function is implementable, it is enough to check the direct-revelation mechanism that is associated with it. If it turns out to be truthful, we still may want to implement it with an indirect mechanism that will seem more natural and “real”, but if the direct-revelation mechanism is not truthful, then there is no hope of implementing the function.

The proof of the theorem of Gibbard and Satterthwaite relies on the revelation principle to focus on direct-revelation mechanisms, but this is just the beginning. The next step is to show that *any* non dictatorial function is non-implementable. The proof achieves this by an interesting reduction to Arrow’s theorem, from the field of social choice theory. This theory is concerned with the possibilities and impossibilities of social preference aggregations that will exhibit desirable properties. A social welfare function $F : V \rightarrow \mathcal{R}$ aggregates the individuals’ preferences into a single preference order over all alternatives, where \mathcal{R} is the set of all possible preference orders over A . Arrow [3] describes few desirable properties from a social welfare function, and shows that no social choice function can satisfy all:

Definition 1 (Arrow’s desirable properties)

1. A social welfare function satisfies “weak pareto” if whenever all individuals strictly prefer alternative a to alternative b then, in the social preference, a is strictly preferred to b .
2. A social welfare function is “a dictatorship” if there exists an individual for which the social preference is always identical to his own preference.
3. A social welfare function F satisfies the “Independence of Irrelevant Alternatives” property (IIA) if, for any preference orders $R, \tilde{R} \in \mathcal{R}$ and any $a, b \in A$,

$$a >_{F(R)} b \text{ and } b >_{F(\tilde{R})} a \Rightarrow \exists i : a >_{R_i} b \text{ and } b >_{\tilde{R}_i} a$$

(where $a >_{R_i} b$ iff a is preferred over b in R_i). In other words, if the social preference between a and b was flipped when the individual preferences were changed from R to \tilde{R} , then it must be the case that some individual flipped his own preference between a and b .

Arrow’s impossibility theorem holds for the *unrestricted* domain of preferences, i.e. when all preference orders are possible:

Theorem 4 ([3]) *Assume $|A| \geq 3$. Any social welfare function over an unrestricted domain of preferences that satisfies both weak pareto and Independence of Irrelevant Alternatives must be a dictatorship.*

Gibbard and Satterthwaite’s proof reveals an interesting and important connection between the concept of implementation in dominant strategies, and Arrow’s condition of IIA. The proof shows how to construct, from a given implementable social choice function f , a social welfare function, F , that satisfies IIA and weak pareto. In addition, F always places the alternative chosen by f as the most preferred alternative. By Arrow’s theorem, the resulting social welfare function must be dictatorial. In turn, this implies that f is dictatorial. The construction of F from f is the straight-forward one: the top alternative is f ’s choice to the original preferences, say a . Then a is lowered to be the least preferred alternative in all the preferences, and f ’s new choice is placed second, etc. etc. The interesting exercise is to show that the implementability of f implies that F satisfies Arrow’s conditions. In fact, as the proof shows that any implementable social choice function f entails a social welfare function F that “extends” f and satisfies Arrow’s conditions, it actually provides a strong argument for the reasonability of Arrow’s requirement – they are simply implied by the implementability requirement.

In view of these strong impossibility results, it is natural to ask whether the entire concept of a mechanism can yield positive constructions. The answer is a big yes, under the “right” set of assumptions, as discussed in the next sections.

6 Quasi-linear Utilities and the VCG Mechanism

The model formalization of the previous section ignores the existence of money, or, more accurately, the fact that it has a more or less predictable effect on a player’s utility. The quasi-linear utilities model takes this into account, and players are assumed to have monetary value for each alternative.

Formally, the type of a player is a valuation function $v_i : A \rightarrow \mathfrak{R}$ that describes the monetary value that the player will obtain from each chosen alternative (as before v_i is taken from a domain of valid valuations V_i and $V = V_1 \times \dots \times V_n$). Note that the value of a player does not depend on the other players’ values (this is termed the private value assumption). The mechanism designer can now additionally pay each player (or charge money from her), and the total utility of player i if the chosen outcome is a and in addition she pays a price P_i is $v_i(a) - P_i$. A direct mechanism for quasi-linear utilities includes an outcome function $f : V \rightarrow A$ (as before), as well as price functions $p_i : V \rightarrow \mathfrak{R}$ for each player i . (the definition of an indirect mechanism is the natural parallel of the definition of the previous section; the revelation principle holds for quasi-linear utilities as well, and we focus here on direct mechanisms). The implicit assumption is that a player aims to maximize her resulting utility, $v_i(f(v_i, v_{-i})) - p_i(v_i, v_{-i})$, and this leads us to the definition of a truthful mechanism, that parallels that of the previous section:

Definition 2 (Truthfulness, or Incentive Compatibility, or Strategy-proofness) *A direct revelation mechanism is “truthful” (or incentive-compatible, or strategy-proof) if the dominant strategy of each player is to reveal her true type, i.e. if for every $v_{-i} \in V_{-i}$ and every $v_i, v'_i \in V_i$,*

$$v_i(f(v_i, v_{-i})) - p_i(v_i, v_{-i}) \geq v_i(f(v'_i, v_{-i})) - p_i(v'_i, v_{-i})$$

Using this framework, we can return to the example from section 4 (“building a bridge”), and construct a truthful mechanism to solve it. Recall that, in this problem, a government is trying to decide on a certain public project, which costs C dollars. Each player, i , will benefit from it to an amount of v_i dollars, where this number is known only to the player herself. The government desires to build the bridge if and only if $\sum_i v_i \geq C$. Clarke [8] designed the following mechanism. Each player reports a value, \tilde{v}_i , and the bridge is built if and only if $\sum_i \tilde{v}_i \geq C$. If the bridge is not built, the price of each player is 0. If the bridge is built then each player, i , pays the minimal value she could have declared to maintain the positive decision. More precisely, if $\sum_{i' \neq i} \tilde{v}_{i'} \geq C$ then she still pays zero, and otherwise she pays $C - \sum_{i' \neq i} \tilde{v}_{i'}$.

Theorem 5 *Bidding the true value is a dominant strategy in the Clarke mechanism.*

PROOF (SKETCH): Consider the truthful bidding for player i , v_i , vs. another possible bid \tilde{v}_i (fixing the bids of the other players to arbitrarily be \tilde{v}_{-i}). If with v_i the project was rejected then $v_i < C - \sum_{i' \neq i} \tilde{v}_{i'}$. In order to change the decision to an accept, the player would need to declare $\tilde{v}_i \geq C - \sum_{i' \neq i} \tilde{v}_{i'}$. In this case i 's payment will be $C - \sum_{i' \neq i} \tilde{v}_{i'}$ which is smaller than v_i , as observed above. Thus, i 's resulting utility will be negative, hence bidding \tilde{v}_i did not improve her utility.

On the other hand, assume that with v_i the project is accepted. Therefore the player's utility from declaring v_i is non-negative. Note that the price that the player pays in case of an accept does not depend on her bid. Thus, the only way to change i 's utility (if at all) is to declare some \tilde{v}_i that will cause the project to be rejected. But in this case i 's utility will be zero, hence she did not gain any benefit. ■

Subsequently, Groves [13] made the remarkable observation that Clarke's mechanism is in fact a special case of a much more general mechanism, that solves the welfare maximization problem on *any* domain with private values and quasi-linear utilities. For a given set of player types v_1, \dots, v_n , the *welfare* obtained by an alternative $a \in A$ is $\sum_i v_i(a)$. A social choice function is termed a *welfare maximizer* if $f(v)$ is an alternative with maximal welfare, i.e. $f(v) \in \operatorname{argmax}_{a \in A} \{ \sum_{i=1}^n v_i(a) \}$.

Definition 3 (VCG mechanisms) *Given a set of alternatives A , and a domain of players' types $V = V_1 \times \dots \times V_n$, a VCG mechanism is a direct revelation mechanism such that, for any $v \in V$,*

1. $f(v) \in \operatorname{argmax}_{a \in A} \{ \sum_{i=1}^n v_i(a) \}$.
2. $p_i(v) = - \sum_{j \neq i} v_j(f(v)) + h_i(v_{-i})$, where $h_i : V_{-i} \rightarrow \mathfrak{R}$ is an arbitrary function.

Ignore for a moment the term $h_i(v_{-i})$ in the payment functions. Then the VCG mechanism has a very natural interpretation: it chooses an alternative with maximal welfare according to the reported types, and then, by making additional payments, it equates the utility of each player to that maximal welfare level.

Theorem 6 ([13]) *Any VCG mechanism truthfully implements the welfare maximizing social choice function.*

PROOF: We argue that $s_i(v_i) = v_i$ is a dominant strategy for i . Fix any $v_{-i} \in V_{-i}$ as the declarations (actions) of the other players, any $v'_i \neq v_i$, and assume by contradiction that $v_i(f(v_i, v_{-i})) - p_i(v_i, v_{-i}) < v_i(f(v'_i, v_{-i})) - p_i(v'_i, v_{-i})$. Replacing $p_i(\cdot)$ with the specific VCG payment function, and eliminating the term $h_i(v_{-i})$ from both sides, we get: $v_i(f(v_i, v_{-i})) + \sum_{j \neq i} v_j(f(v_i, v_{-i})) < v_i(f(v'_i, v_{-i})) + \sum_{j \neq i} v_j(f(v'_i, v_{-i}))$. Therefore it must be that $f(v_i, v_{-i}) \neq f(v'_i, v_{-i})$. Denote $f(v_i, v_{-i}) = a$ and $f(v'_i, v_{-i}) = b$. The above equation is now $v_i(a) + \sum_{j \neq i} v_j(a) < v_i(b) + \sum_{j \neq i} v_j(b)$, or, equivalently, $\sum_{i=1}^n v_i(a) < \sum_{i=1}^n v_i(b)$, a contradiction to the fact that $f(v_i, v_{-i}) = a$, since $f(\cdot)$ is a welfare maximizer. ■

Thus, we see that the welfare maximizing social choice function can be always be implemented, no matter what the problem domain is, under the assumption quasi-linear utilities. The VCG mechanism is named after Vickrey, whose seminal paper [28] on auction theory was the first to describe a special case of the above mechanism (this is the second price auction; see the entry on auction theory for more details), after Clarke, who provided the second example, and after Groves himself, that finally pinned down the general idea.

Clarke’s work can be viewed, in retrospect, as a suggestion for one specific form of the function $h_i(v_{-i})$, namely $h_i(v_{-i}) = \sum_{j \neq i} v_j(f(v_{-i}))$ (this is a slight abuse of notation, as f is defined for n players, but the intention is the straight-forward one – f chooses an alternative with maximal welfare). This form for the $h_i(\cdot)$ ’s gives the following property: if a player does not influence the social choice, her payment is zero, and, in general, a player pays the “monetary damage” to the other players (i.e. the welfare that the others lost) as a result of i ’s participation. Additionally, with Clarke’s payments, a truthful player is guaranteed a non-negative utility, no matter what the others declare. This last property is termed “individual rationality”.

7 The Importance of the Domain’s Dimensionality

The impressive property of the VCG mechanism is its generality with respect to the domain of preferences – it can be used for *any* domain. On the other hand, VCG is restrictive in the sense that it can be used only to implement one specific goal, namely welfare maximization. Given the possibility that VCG presents, it is natural to ask if the assumption of quasi-linear utilities and private values allows the designer to implement many other different goals. It turns out that the answer depends on the “dimensionality” of the domain, as is discussed in this section.

7.1 Single-dimensional domains

Consider first a domain of preferences for which the type $v_i(\cdot)$ can be completely described by a single number v_i , in the following way. For each player i , a subset of the alternatives are “losing” alternatives, and her value for all these alternatives is always 0. The other alternatives are “winning” alternatives, and the value for each “winning” alternative is the same, regardless of the specific alternative. Such a domain is “single dimensional” in the sense that one single number completely describes the entire valuation vector. As before, this single number (the value for winning), is private to the player, and here this is the *only* private information of the player. The public project

domain discussed above is an example to a single-dimensional domain: the losing alternative is the rejection of the project, and the winning alternative is the acceptance of the project.

A major drawback of the VCG mechanism, in general, and with respect to the public project domain in particular, is the fact that the sum of payments is not balanced (a broader discussion on this is given in section 8 below). In particular, the payments for the public project domain may not cover the entire cost of the project. Is there a different mechanism that always covers the entire cost? The positive answer that we shall soon see crucially depends on the fact that the domain is single-dimensional, and this turns out to be true for many other problem domains as well.

The following mechanism for the public project problem assumes that the designer can decide not only if the project will be built, but also which players will be allowed to use it. Thus, we now have many possible alternatives, that correspond to the different subsets of players that will be allowed to utilize the project. This is still a single-dimensional domain, as each player only cares about whether she is losing or winning, and so the alternatives, from the point of view of a specific player, can be divided to the two winning/losing subsets. The following cost-sharing mechanism was proposed by Moulin [20] in a general cost-sharing framework. The mechanism is a direct-revelation mechanism, where each player, i , first submits her winning value, v_i . The mechanism then continues in rounds, where in the first round all players are present, and in each round one or more players are declared losers and retire. Suppose that in a certain round x players remain. If all remaining players have $v_i \geq C/x$ then they are declared winners, and each one pays C/x . Otherwise, all players with $v_i < C/x$ are declared losers, and “walk out”, and the process repeats. If no players remain then the project is rejected.

Clearly, the cost sharing mechanism always recovers the cost of the project, if it is indeed accepted. But is it truthful? One can analyze it directly, to show that indeed the dominant strategy of each player is to declare her true winning value. Perhaps a better way is to understand a characterization of truthfulness for the general abstract setting of a single-dimensional domain. For simplicity, we will assume that we require mechanisms to be “normalized”, i.e. that a losing player will pay exactly zero to the mechanism. Now, a mechanism is said to be “value-monotone” if a winner that increases her value will always remain a winner. More formally, for all $v_i \in V_i$ and $v_{-i} \in V_{-i}$, if i is a winner in the declaration (v_i, v_{-i}) then i is a winner in the declaration (v'_i, v_{-i}) , for all $v'_i \geq v_i$. Note that a value-monotone mechanism casts a “threshold value” function $v_i^*(v_{-i})$ such that, for every v_{-i} , player i wins when declaring $v_i > v_i^*(v_{-i})$, and loses when declaring $v_i < v_i^*(v_{-i})$. Quite interestingly, this structure completely characterizes incentive compatibility in single-dimensional domains:

Theorem 7 *A normalized direct-revelation mechanism for a single-dimensional domain is truthful if and only if it is value monotone and the price of a winning player is $v_i^*(v_{-i})$.*

PROOF: The first observation is that the price of a winner cannot depend on her declaration, v_i (only on the fact that she wins, and on the declaration of the other players). Otherwise, if it can depend on her declaration, then there are two possible bids v_i and v'_i such that i wins with both bids and pays p_i and p'_i , where $p'_i < p_i$. But then if the true value of i is v_i then bidding v'_i instead of v_i will increase i 's utility, contradicting truthfulness.

We now show that a truthful mechanism must be value-monotone. Assume by contradiction that a declaration of (v_i, v_{-i}) will cause i to win, but a declaration of (v'_i, v_{-i}) will cause i to lose, for some $v'_i > v_i$. Suppose that i pays p_i for winning (when the others declare v_{-i}). Since we assume a normalized mechanism, truthfulness implies that $p_i \leq v_i$. But then when the true type

of a player is v'_i , her utility from declaring the truth will be zero (she loses), and she can increase her utility by declaring v_i , which will cause her to win and to pay p_i , a contradiction.

Thus a truthful mechanism must be value-monotone, and there exists a threshold value $v_i^*(v_{-i})$. To see that this defines p_i , let us first check the case of $p_i < v_i^*(v_{-i})$. In this case, if the type of i is v_i with $p_i < v_i < v_i^*(v_{-i})$, she will lose (by the definition of $v_i^*(v_{-i})$), and by bidding some false large enough v'_i she can win and get a positive utility of $v_i - p_i$. On the other hand, if $p_i > v_i^*(v_{-i})$ then with type v_i such that $p_i > v_i > v_i^*(v_{-i})$ a player will have negative utility of $v_i - p_i$ from declaring the truth, and she can strictly increase it by losing, again a contradiction. Therefore it must be that $p_i = v_i^*(v_{-i})$.

To conclude, it only remains to show that a value-monotone mechanism with a price for a winner $p_i = v_i^*(v_{-i})$ is indeed truthful. Suppose first that with the truthful declaration i wins. Then $v_i > v_i^*(v_{-i}) = p_i$ and i has a positive utility. If she changes the declaration and remains a winner, her price does not change, and if she becomes a loser her utility decreases to zero. Thus a winner cannot increase her utility. Similarly, a loser can change her utility only by becoming a winner, i.e. by declaring $v'_i > v_i^*(v_{-i}) > v_i$, but since she will then pay $v_i^*(v_{-i})$ her utility will now decrease to be negative. Thus a loser cannot increase her utility either, and the mechanism is therefore truthful. ■

This structure of truthful mechanisms is very powerful, and reduces the mechanism design problem to the algorithmic problem of designing monotone social choice functions. Another strong implication of this structure is the fact that the payments of a truthful mechanism are completely derived from the social choice rule. Consequently, if two mechanisms always choose the same set of winners and losers, then the revenues that they raise must also be equal. Myerson [21] was perhaps the first to observe that, in the context of auctions, and named this the “revenue equivalence” theorem.

As a result of this characterization, one can easily verify that the above mentioned cost-sharing mechanism is indeed truthful. It is not hard to check that the two conditions of the theorem hold, and therefore its truthfulness is concluded. This is just one example to the usefulness of the characterization.

7.2 Multi-dimensional domains

In the more general case, when the domain is multi-dimensional, the simple characterization from above does not fit, but it turns out that there exists a nice generalization. We describe two properties, cyclic monotonicity (Rochet [26]) and weak monotonicity (Bikhchandani et. al. [7]), that achieve that. The exposition here also relies on [14]. It will be convenient to use the abstract social choice setting described above: there is a finite set A of alternatives, and each player has a type (valuation function) $v : A \rightarrow \mathfrak{R}$ that assigns a real number to every possible alternative. $v_i(a)$ should be interpreted as i 's value for alternative a . The valuation function $v_i(\cdot)$ belongs to the domain V_i of all possible valuation functions.

Our goal is to implement in dominant strategies the social choice function $f : V_1 \times \dots \times V_n \rightarrow A$. As before, it is not hard to verify that the required price function of a player i may depend on her declaration only through the choice of the alternative, i.e. that it takes the form $p_i : V_{-i} \times A \rightarrow \mathfrak{R}$, for every player i . For truthfulness, these prices should satisfy the following property. Fix any $v_{-i} \in V_{-i}$, and any $v_i, v'_i \in V_i$. Suppose that $f(v_i, v_{-i}) = a$ and $f(v'_i, v_{-i}) = b$. Then it is the case

that:

$$v_i(a) - p_i(a, v_{-i}) \geq v_i(b) - p_i(b, v_{-i}) \quad (1)$$

In other words, player i 's utility from declaring his true v_i is no less than his utility from declaring some lie, v'_i , *no matter what the other players declare*. Given a social choice function f , the underlying question is what conditions should it satisfy to guarantee the existence of such prices.

Fix a player i , and fix the declarations of the others to v_{-i} . Let us assume, without loss of generality, that f is onto A (or, alternatively, define A' to be the range of $f(\cdot, v_{-i})$, and replace A with A' for the discussion below). Since the prices of Eq. 1 now become constant, we simply seek an assignment to the variables $\{p_a\}_{a \in A}$ such that $v_i(a) - v_i(b) \geq p_a - p_b$ for every $a, b \in A$ and $v_i \in V_i$ with $f(v_i, v_{-i}) = a$. This motivates the following definition:

$$\delta_{a,b} \doteq \inf\{v_i(a) - v_i(b) \mid v_i \in V_i, f(v_i, v_{-i}) = a\} \quad (2)$$

With this we can rephrase the above assignment problem, as follows. We seek an assignment to the variables $\{p_a\}_{a \in A}$ that satisfies:

$$p_a - p_b \leq \delta_{a,b} \quad \forall a, b \in A \quad (3)$$

By adding the two inequalities $p_a - p_b \leq \delta_{a,b}$ and $p_b - p_a \leq \delta_{b,a}$ we get that a necessary condition to the existence of such prices is the inequality $\delta_{a,b} + \delta_{b,a} \geq 0$. Note that this inequality is completely determined by the social choice function. This condition is termed the non-negative 2-cycle requirement. Similarly, for any k distinct alternatives a_1, \dots, a_k we have the inequalities

$$\begin{aligned} p_{a_1} - p_{a_2} &\leq \delta_{a_1, a_2} \\ &\vdots \\ p_{a_{k-1}} - p_{a_k} &\leq \delta_{a_{k-1}, a_k} \\ p_{a_k} - p_{a_1} &\leq \delta_{a_k, a_1} \end{aligned}$$

and we get that any k -cycle must be non-negative, i.e. that $\sum_{i=1}^k \delta_{a_i, a_{i+1}} \geq 0$, where $a_{k+1} \equiv a_1$. It turns out that this is also a sufficient condition:

Theorem 8 *There exists a feasible assignment to 3 if and only if there are no negative-length cycles.*

One constructive way to prove this is by looking at the ‘‘allocation graph’’: this is a directed weighted graph $G = (V, E)$ where $V = A$ and $E = A \times A$, and an edge $a \rightarrow b$ (for any $a, b \in A$) has weight $\delta_{a,b}$. A standard basic result of graph theory states that there exists a feasible assignment to 3 if and only if the allocation graph has no negative-length cycles. Furthermore, if all cycles are non-negative, the feasible assignment is as follows: set p_a to the length of the shortest path from a to some arbitrary fixed node $a^* \in A$.

With the above theorem, we can easily state a condition for implementability:

Definition 4 (Cycle monotonicity) *A social choice function f satisfies cycle monotonicity if for every player i , $v_{-i} \in V_{-i}$, some integer $k \leq |A|$, and $v_i^1, \dots, v_i^k \in V_i$,*

$$\sum_{j=1}^k [v_i^j(a_j) - v_i^j(a_{j+1})] \geq 0$$

where $a_j = f(v_i^j, v_{-i})$ for $1 \leq j \leq k$, and $a_{k+1} = a_1$.

Theorem 9 *f satisfies cycle monotonicity if and only if there are no negative cycles.*

Corollary 1 *A social choice function f is dominant-strategy implementable if and only if it satisfies cycle monotonicity.*

This interesting structure implies, as another corollary, the fact that the prices are uniquely determined by the social choice function, for every connected domain (this was discussed above for the special case of single-dimensional domains). Very briefly, from the above, it follows that any two alternatives with $\delta_{ab} + \delta_{ba} = 0$ have $p_a - p_b = \delta_{ab} = -\delta_{ba}$. Thus, determining the price of one alternative completely determines the price of the second alternative. A short argument that we omit shows that the connectedness of the domain implies that for any two alternatives a and b , there's a path a_1, \dots, a_k (with $a_1 = a$ and $a_k = b$) such that $\delta_{a_i, a_{i+1}} + \delta_{a_{i+1}, a_i} = 0$ for every $1 \leq i < k$. Thus, fixing the price of one alternative completely determines the prices of all other alternatives. In particular, if there exists one alternative whose price is normalized to be (always) zero, then all other prices have also been completely determined by the δ_{ab} 's weights (who in turn are completely determined by the function f).

Cycle monotonicity satisfies our motivating goal: a condition on f that involves only the properties of f , without existential price qualifiers. However, it is quite complex. k could be large, and a “shorter” condition would have been nicer. “Weak monotonicity” (W-MON) is exactly that:

Definition 5 (Weak monotonicity) *A social choice function f satisfies W-MON if for every player i , every v_{-i} , and every $v_i, v'_i \in V_i$ with $f(v_i, v_{-i}) = a$ and $f(v'_i, v_{-i}) = b$, $v'_i(b) - v_i(b) \geq v'_i(a) - v_i(a)$,*

In other words, if the outcome changes from a to b when i changes her type from v_i to v'_i then i 's value for b has increased at least as i 's value for a in the transition v_i to v'_i . W-MON is equivalent to cycle monotonicity with $k = 2$, or, alternatively, to the requirement of no negative 2-cycles. Hence it is necessary for truthfulness. As it turns out, it is also a sufficient condition on many domains. Very recently, Monderer [19] shows that weak monotonicity must imply cycle monotonicity if and only if the closure of the domain of valuations is convex. Thus, for such domains, it is enough to look at the more simple condition of weak monotonicity.

7.2.1 The implementability of non-welfare-maximizing social goals

Now that the conditions for implementability are completely understood, it should be asked what forms of social choice functions satisfy them. We already saw that the welfare-maximizer function satisfies them, for any domain, and we ask what *other* implementable functions exist? For the single-dimensional case, we saw another example of a truthful mechanism, and the literature contains many more. For the multi-dimensional case, “interesting” examples are more rare, and a beautiful result by Roberts [25] shows that when the domain has full dimensionality then only weighted welfare maximizers are implementable. In other words, weak monotonicity implies welfare maximization. More precisely, a function f is an “affine maximizer” if there exist weights k_1, \dots, k_n and $\{C_x\}_{x \in A}$ such that, for all $v \in V$,

$$f(v) \in \operatorname{argmax}_{x \in A} \{ \sum_{i=1}^n k_i v_i(x) + C_x \}$$

Roberts [25] shows that, if $|A| \geq 3$ and $V_i = \mathfrak{R}^A$ for all i , then f is dominant-strategy implementable if and only if it is an affine maximizer.

However, most interesting domains are restricted in some meaningful way, and for this wide intermediary range of domains the current knowledge is rather scarce. One impossibility result that extends the result of Roberts to a restricted multi-dimensional case is given by Lavi et. al. [16], who study multi-item auctions. In a multi-item auction, one seller (the mechanism designer) wants to allocate items to players (i.e. an alternative is an allocation of the items to the players). [16] show that every social choice function for multi-item auctions, that additionally satisfy four other social choice properties, must be an affine maximizer.

Before concluding the discussion on dominant-strategy implementation, we demonstrate the necessity for non-welfare-maximizers by considering the following “scheduling domain”. A designer wishes to assign n tasks/jobs to m workers, where worker i needs t_{ij} time units to complete job task j , and incurs a cost of t_{ij} for its processing time (one dollar per time unit). Importantly, this cost is private information of the worker, and workers are assumed to be strategic, each one selfishly trying to minimize its own cost. The load of worker i is the sum of costs of the jobs assigned to her, and the maximal load over all workers (in a given schedule) is termed the “makespan” of the schedule. The welfare maximizing social goal would put each task on the most efficient worker (for that task), which may result in a very high makespan. For example, consider a setting with two workers and n tasks. The first worker incurs a cost of 1 for every task, and the second worker incurs a cost of $1 + \epsilon$ for every task. The social welfare is the minus of the sum of the costs of the two workers, and the VCG mechanism will therefore assign all tasks to the first worker. This is a very highly unbalanced allocation, which takes twice the time that the workers optimally need in order to finish all tasks (roughly splitting the work among them).

Thus one may wish to consider a social goal different from welfare maximization, namely *makespan* minimization. This goal aims to construct a balanced allocation, in order to minimize the completion time of the last task. Such an allocation can also be viewed as being a more “fair” allocation, in the sense of Rawls’ max-min fairness criteria. Due to the strategic nature of the workers, we wish to design a truthful mechanism. While VCG is truthful, its outcome may be far from optimal, as demonstrated above. Nisan and Ronen [23], who have first studied this problem in the context of mechanism design, observed that VCG provides only an “ m -approximation” to the optimal makespan, meaning that VCG may sometimes produce a makespan that is m times larger than the optimal makespan. More importantly, they have shown that *no truthful deterministic mechanism can obtain an approximation ratio better than 2*. To date, the question of closing this gap between m and 2 remains open.

Archer and Tardos [1], on the other hand, considered a natural restriction of this domain, that makes it single-dimensional, and showed with this they can construct many possibilities (for example, a truthful *optimal* mechanism). Thus, here too we see the contrast between single-dimensionality and multi-dimensionality. Lavi and Swamy [17] suggest a multi-dimensional special case, and give a truthful 2-approximation for the special case where the processing time of each job is known to be either “low” or “high”. This special case keeps the multi-dimensionality of the domain. The construction of this result do not rely on explicit prices, but rather use the cycle-monotonicity condition described above, to construct a monotone allocation rule.

8 Budget Balancedness and Bayesian Mechanism Design

The previous sections portray a concrete picture of the advantages and the disadvantages of the solution concept of truthfulness in dominant strategies. On the one hand, this is a strong and

convincing concept, which admits many positive results. However, there are several problems to all these results, that cannot be solved by a truthful mechanism. Among these, the budget-imbalance problem was briefly mentioned, and this section looks again at this problem, as a motivation to the definition of the Bayesian-Nash solution concept.

To recall the budget-imbalance problem of the VCG mechanism, let us consider a specific input to the Clarke mechanism from section 6: suppose the cost of the project is \$100, and there are 102 players, each values the project by \$1. It is a simple exercise to check that the Clarke mechanism will indeed choose to perform the project, and that each player will pay a price of zero (since the project would have been conducted even if a single player is removed). Thus, the mechanism designer does not cover the project’s cost. As described above, this problem, for this specific domain, can be fixed by considering the cost sharing mechanism discussed in section 7. However, this mechanism may sometimes choose not to perform the project although the society as a whole will benefit from performing it (i.e. it is not “socially efficient”), and, even more importantly, it is a solution only for the concrete domain of a public project. Is there a general mechanism (in the sense that VCG is general) that is both socially efficient and budget-balanced? In this section we describe such a mechanism, that was independently discovered by d’Aspremont and Ge’rard-Varet [10] and by Arrow [2]. Its incentive compatibility will not be in dominant strategies. Instead, it is assumed that player types are drawn i.i.d. from some fixed and known cumulative distribution function F (the assumption that the types are drawn from the same distribution is not important, and is made here only for the ease of notation; the assumption that types are not correlated is important and cannot be removed in general). The solution concept of a Bayesian-Nash equilibrium is a natural extension of the regular Nash equilibrium concept, for a setting in which the distribution F is known to all players (this is termed the “common-prior” assumption), and where players aim to maximize the expectation of their quasi-linear utility.

Definition 6 *A direct mechanism $M = (f, p)$ is Bayesian incentive compatible if for every player i , and for every $v_i, v'_i \in V_i$,*

$$E_{v_{-i}}[v_i(f(v_i, v_{-i})) - p_i(v_i, v_{-i})] \geq E_{v_{-i}}[v_i(f(v'_i, v_{-i})) - p_i(v'_i, v_{-i})]$$

In other words, Bayesian incentive compatibility requires that a player will maximize her expected utility by declaring her true type. An alternative formulation is that truthfulness in a Bayesian incentive compatible mechanism should be a “Bayesian-Nash equilibrium” (where the formal equilibrium definition naturally follows the above definition). This is an “ex-interim” equilibrium: the type of the player is already known to her, and the averaging is over the types of the others. A weaker equilibrium notion would be an “ex-ante” notion, where the player should decide on a strategy before knowing her own type, and so the averaging is done over her own types as well. A stronger notion would be an “ex-post” notion, where no-averaging is done at all, and the above inequality is required for every realization of the types of the other players. It can be shown that this stronger ex-post condition is equivalent to the requirement of dominant-strategy incentive compatibility. As a Bayesian-Nash equilibrium only considers the average over all possible realizations, it is clearly a weaker requirement than dominant-strategy implementability.

We will demonstrate the usefulness of this weaker notion by describing a general mechanism that is both ex-post socially efficient and ex-post budget balanced, and is Bayesian incentive-compatible. Define,

$$x_i(v_i) = E_{v_{-i}}[\sum_{j \neq i} v_j(f(v_i, v_{-i}))]$$

The “budget-balanced” (BB) mechanism asks the players to report their types, and then chooses the welfare-maximizing allocation according to the reported types (as VCG does). It then charges some payment $p_i(v_i, v_{-i}) = -x_i(v_i) + h_i(v_{-i})$, for some function $h_i(\cdot)$ that will be chosen later on in a specific way that balances the budget. But let us first verify that the mechanism is Bayesian incentive compatible, regardless of the choice of the functions $h_i(\cdot)$. Note that, for any realization of v_{-i} , we have that,

$$v_i(f(v_i, v_{-i})) + \sum_{j \neq i} v_j(f(v_i, v_{-i})) \geq v_i(f(v'_i, v_{-i})) + \sum_{j \neq i} v_j(f(v'_i, v_{-i}))$$

as the mechanism chooses the maximal-welfare alternative for the given reports. Clearly, taking the expectation on both sides will maintain the inequality. Therefore we get:

$$\begin{aligned} E_{v_{-i}}[v_i(f(v_i, v_{-i})) - p_i(v_i, v_{-i})] &= \\ &= E_{v_{-i}}[v_i(f(v_i, v_{-i}))] + E_{v_{-i}}[\sum_{j \neq i} v_j(f(v_i, v_{-i}))] + E_{v_{-i}}[h_i(v_{-i})] \\ &\geq E_{v_{-i}}[v_i(f(v_i, v_{-i}))] + E_{v_{-i}}[\sum_{j \neq i} v_j(f(v_i, v_{-i}))] + E_{v_{-i}}[h_i(v_{-i})] \\ &= E_{v_{-i}}[v_i(f(v'_i, v_{-i})) - p_i(v'_i, v_{-i})] \end{aligned}$$

which proves Bayesian incentive compatibility. To balance the budget, consider the specific function, $h_i(v_{-i}) = \frac{1}{n-1} \sum_{j \neq i} x_j(v_j)$. Notice that the term $x_j(v_j)$ appears $(n-1)$ times in the sum $\sum_{i=1}^n h_i(v_{-i})$ for any $j = 1, \dots, n$. Therefore $\sum_{i=1}^n h_i(v_{-i}) = \frac{1}{n-1} \sum_{j=1}^n (n-1)x_j(v_j) = \sum_{i=1}^n x_i(v_i)$. To conclude, we have $\sum_{i=1}^n p_i(v_i, v_{-i}) = \sum_{i=1}^n h_i(v_{-i}) - \sum_{i=1}^n x_i(v_i) = 0$, and the budget balancedness follows.

It is worth noting that such an exercise cannot be employed for the VCG mechanism, as there the “parallel” $x_i(\cdot)$ term should depend on the entire vector of declarations, not only on i ’s own declarations. This is the exact point where the averaging of the others’ valuations is crucial.

In addition to the difference in the solution concept, one other important advantage of VCG, in comparison with the BB mechanism, is the fact that VCG (with the Clarke payments) is ex-post “individually rational”: if a player declares her true valuation, it is guaranteed that she will not pay more than her value, no matter what the others will declare. Here, on the contrary, there is no reason why this should be true, in general. Can the solution concept of Bayesian incentive compatibility be used to construct a general budget-balanced and individually rational mechanism? In an important and influencing result, Myerson and Satterthwaite [22] have shown that this is impossible: there is no general mechanism that satisfies the four properties (1) Bayesian incentive compatibility, (2) budget balancedness, (3) individual rationality, and (4) social efficiency. The proof uses a simple, natural exchange setting, where two traders (one buyer and one seller) wish to exchange an item. The seller has a cost c of producing the item, and the buyer obtains a value v from receiving it. Myerson and Satterthwaite show that there is no Bayesian incentive compatible mechanism that decides to perform the exchange if and only if $v > c$, such that Bayesian incentive compatibility and individual rationality are maintained, and the price that the buyer pays exactly equals the payment that the seller gets. In particular, VCG violates this last property, while BB satisfies it, but violates individual rationality (i.e. for some realizations of the values, a buyer may pay more than her value, or the seller may get less than her cost).

Besides this disadvantage of the BB mechanism, there are also additional disadvantages that result from the underlying assumptions of the solution concept itself. In particular, Bayesian incentive compatibility entails two strong assumptions about the characteristics of the players. First,

it assumes that players are risk-neutral, i.e. care only about maximizing the expectation of their profit (value minus price). Thus, when players dislike risk, for example, and prefer to decrease the variance of the outcome, even on the expense of lowering the achieved expected profit, the rational of the Bayesian-Nash equilibrium concept breaks down. Second, the assumption of a common-prior, i.e. that all players agree on the same underlying distribution, seems strong and somewhat unrealistic. Often, players have different estimations about the underlying statistical characteristics of the environment, and this concept does not handle this well. Note that the solution concept of dominant-strategies does not suffer from any of these problems, which strengthens even more its importance. Unfortunately, the classical economics literature mainly ignores these disadvantages and problems. A well-known exception is the critique known as Wilson’s critique [29], who raises the above mentioned problems, and argues in favor of “detail-free” mechanisms. Recently, this critique gained more popularity, and detail-free solution concepts are re-examined. For some examples, see [6, 11, 5].

9 Interdependent Valuations

Up to now, this entry described “private value” models, i.e. models where the valuation (or the preference relation) of a player does not depend on the types of the other players. There are many settings in which this assumption is unrealistic, and a more suitable assumption is that the valuation of a specific player is affected by the valuations of the other players. This last statement may entail two interpretations. The first is that the distribution over the valuations of a specific player is correlated with the distribution over the valuations of the other players, and, thus, knowing a player’s actual valuation gives partial knowledge about the valuations of the other players. This first interpretation is still termed a private value model (but with correlated values instead of independent values), since after the player becomes aware to the actual realization of her valuation, she completely and fully knows her values for the different outcomes.

In contrast, with interdependent valuations, the actual valuation of a player depends on the actual valuations of the other players. Thus, a player does not fully know her own valuation. She only partially knows it, and can determine her full valuation only if given the others’ valuations as well. A classic example is a setting where a seller sells an oil field. The oil, of-course, is not seen on the ground surface, and the only way to exactly determine how much oil is there (and, by this, determine the actual worth of the field) is to extract it. Before buying the field, though, the potential buyers are only allowed to make preliminary tests, and by this to determine an estimation to the value of the field, which is not completely accurate. If all the buyers that are interested in the field have the same technical capabilities, it seems reasonable to assume that the *true value* of the field is the average over all the estimations obtained by the different oil companies. Intuitively, a player that participates in an auction mechanism that determines who will buy the field, and at what price, has to act somehow as if she knows the value of the field, although she doesn’t. Clearly, this creates different complications. Such a model is very natural in auction settings, and indeed the entry on auctions handles the subject of interdependent valuations more broadly. Since this issue is also very relevant to general mechanism design theory, we describe here one specific, rather general result for mechanisms with interdependent valuations, to exemplify the definitions and the techniques being employed.

In the formal model of interdependent valuations, player i receives a signal $s_i \in S_i$, which may be multi-dimensional. Her valuation for a specific alternative $a \in A$ is a function of the signals

s_1, \dots, s_n , i.e. $v_i : A \times S_1 \times \dots \times S_n \rightarrow \mathfrak{R}$. The case where $v_i(a, s_1, \dots, s_n) = v_j(a, s_1, \dots, s_n)$ for all players i, j and all a, s_1, \dots, s_n is termed the “common value” case, as the actual values of all players are identical, and only their signals are different (as in the oil field example). The other extreme is when i ’s valuation depends only on i ’s signal, i.e. $v_i(a, s_1, \dots, s_n) = v_i(a, s_i)$, which is a return to the private value case. The entire range in general is termed the case of interdependent valuations. All the results described in the previous sections fail when we move to interdependent valuations. For example, in the VCG mechanism, a player is required to report her valuation function, which is not fully known to her in the interdependent valuation case. It turns out that the straight-forward modification of reporting the players’ signals does not maintain the truthfulness property, and, in fact, some strong impossibilities exist (Jehiel et. al. [15]). However, interdependent valuations may also enable possibilities, and the classic result of Cremer and McLean [9] will be described here to exemplify this. This result shows how to use the interdependencies in order to increase the revenue of the mechanism designer, so that the entire surplus of the players can be extracted.

[9] study an auction setting where there is one item for sale. n bidders have interdependent values for the item, and it is assumed that the signal that each player receives is single-dimensional, i.e. each player receives a single real number as her signal. The valuation functions are assumed to be known to the mechanism designer, so that the only private information of the players are their signals. It is also assumed that the valuation functions are monotonically non-decreasing in the signals. For simplicity, it is assumed here that the signal space is discretized to be $S_i = \{0, \Delta, 2\Delta, \dots\}$. The last (and crucial) assumption is that the valuation functions satisfy the “single-crossing” property: if $v_i(s_i, s_{-i}) \geq v_j(s_i, s_{-i})$ then $v_i(s_i + \Delta, s_{-i}) \geq v_j(s_i + \Delta, s_{-i})$. This says that i ’s signal affects i ’s own value (weakly) more than it affects the value of any other player. This last assumption is strong, but in some sense necessary, as it is possible to construct interdependent valuation functions (that violate single-crossing) for which no truthful mechanism can be efficient (i.e. allocate the item to the player with the highest value).

Consider the following CM mechanism for this problem: each player reports her signal, and the player with the highest *value* (note that this may be different than the player with the highest signal) receives the object. In order to determine her payment, define the “threshold signal” $T_i(s_{-i})$ of any player i to be the minimal signal that will enable her to win (given the signals of the other players), i.e. $T_i(s_{-i}) = \min\{\tilde{s}_i \in S_i \mid v_i(\tilde{s}_i, s_{-i}) \geq \max_{j \neq i} v_j(\tilde{s}_i, s_{-i})\}$. The payment of the winner, i , is her value if her signal was $T_i(s_{-i})$, i.e. $P_i(s_{-i}) = v_i(T_i(s_{-i}), s_{-i})$. Clearly, if all players report their true signals, then the player with the highest value receives the item. Truthful reporting is also an ex-post Nash equilibrium, which means the following: if all other players report the true signal (no matter what that is) then it is a best response for i to report her true signal as well.

To verify that truthfulness is indeed an ex-post Nash equilibrium, notice first that each player has a price for winning which does not depend on her declaration. Now, truthful reporting will ensure winning (given that the others are truthful as well) if and only if the true value of the player is higher than her price (i.e. iff winning will yield a positive utility). Thus, when a player “wants to win”, truthful reporting will do that, and when a player “wants to lose”, truthful reporting will do that as well, and so truthfulness will always maximize the player’s utility.

The notion of an ex-post equilibrium is stronger than Bayesian-Nash equilibrium, since, here, even after the signals are revealed no player regrets her declaration (while in Bayesian-Nash equilibrium, since only the *expected* utility is maximized, there are some realizations for which a player can deviate and gain). On the other hand, ex-post equilibrium is weaker than dominant strategies, in which truthfulness is the best strategy no matter what the others choose to declare, while here

truthfulness is a best response only if the others are truthful as well.

As seen above, both for the VCG mechanism as well as for the BB mechanism, adding a “constant” to the prices (i.e. setting $\tilde{P}_i(s_{-i}) = P_i + h_i(s_{-i})$) maintains the strategic properties of the mechanism, since the function $h_i(\cdot)$ does not depend on the declaration of player i . The correlation in the values can help the mechanism designer to extract more payments from the players, as follows. Consider the matrix that describes the conditional probability for a specific tuple of signals of the other players, given i 's own signal. There is a row for every signal s_i of i , a column for every tuple of signals s_{-i} of the other players, and the cell (s_i, s_{-i}) contains the conditional probability $Pr(s_{-i}|s_i)$. In the private value case, the signals of the players are not correlated, hence the matrix has rank one (all rows are identical). As the correlation between the signals “increases”, the rank increases, and we consider here the case when the matrix has full row rank. Let $q_i(s_i, s_{-i})$ be an indicator to the event that i is the winner when the signals are (s_i, s_{-i}) . The expected surplus of player i in the CM mechanism is $U_i^*(s_i) = \sum_{s_{-i}} Pr(s_{-i}|s_i) \cdot (q_i(s_i, s_{-i}) \cdot v_i(s_i, s_{-i})) - P_i(s_{-i})$. ($P_i(s_{-i})$ is defined to be zero whenever i is not a winner). Now find “constants” $h_i(s_{-i})$ such that, for every s_i , $\sum_{s_{-i}} h_i(s_{-i}) \cdot Pr(s_{-i}|s_i) = U_i^*(s_i)$. Note that such an $h_i(\cdot)$ function exists: we have a system of linear equations, where the variables are the function values $h_i(s_{-i})$ for all possible tuples s_{-i} , and the qualifiers are the probabilities and the expected surplus. Since the matrix of qualifiers has full row rank, a solution exists. It is now not hard to verify that, with prices $\tilde{P}_i(\cdot)$, the expected utility of a truthful player is zero.

As mentioned above, truthfulness is still an ex-post equilibrium of this mechanism. It is not ex-post individually rational, though, but rather only ex-ante, since a player pays her expected surplus even if the actual signals cause her to lose. Thus this mechanism can be considered a fair lottery. Also note that the crucial property was the correlation between the values, the interdependence assumption was not important.

10 Future Directions

As surveyed here, the last three decades have seen the theory of mechanism design being developed in many different directions. The common thread of all settings is the requirement to implement some social goal in the presence of incomplete information – the social designer does not know the players’ preferences for the different outcomes. We have seen several alternative assumptions about the structure of players’ preferences, the different equilibria solution concepts that are suitable for the different cases, and several positive examples for elegant solutions. We have also discussed some impossibilities, demonstrating that some attractive definitions may turn out almost powerless. One relatively new research direction in mechanism design is the analysis of new models for the emerging Internet economy, and the development of new alternative solution concepts that better suit this setting. A very recent example is the new model of “dynamic mechanism design”, where the parameters of the problem (e.g. the number of players, or their types) vary over time. Such settings become more and more important as the economic environment becomes more dynamic, for example due to the growing importance of the electronic markets. Examples for such models include e.g. the works by Lavi and Nisan [18] in the context of computer science models, and by Athey and Segal [4] in a more classical economic context, among many other works that study such dynamic settings.

The Internet environment also strengthens the question marks posed on the solution concept of Bayesian incentive compatibility, which was the most common solution concept in mechanism design

literature in the 80's and throughout the 90's, due to the accompanying assumption of a common prior. Such an assumption seems problematic in general, and in particular in an environment like the Internet, that brings together players from many different parts of the world. It seems that the research community agrees more and more that alternative, detail-free solution concepts should be sought. The description of more recently new solution concepts is beyond the scope of this entry, and the interested reader is referred e.g. to the papers by [11, 6, 5] for some recent examples.

Another aspect of mechanism design that is largely ignored in the classic research is the computational feasibility of the mechanisms being suggested. This question is not just a technicality – some classic mechanisms imply heavy computational and communicational requirements that scale exponentially as the number of players increase, making them completely infeasible for even moderate numbers of players. The computer science community has begun looking at the design of computationally-efficient mechanisms, and the recent book by Nisan et. al. [24] contains several surveys on the subject.

References

- [1] A. Archer and E. Tardos. Truthful mechanisms for one-parameter agents. In *Proc. of the 42st Annual Symposium on Foundations of Computer Science(FOCS'01)*, 2001.
- [2] K. Arrow. The property rights doctrine and demand revelation under incomplete information. In M. Boskin, editor, *Economies and Human Welfare*. Academic Press : NY, 1979.
- [3] Kenneth Arrow. *Social Choice and Individual Values*. Wiley, 1951.
- [4] S. Athey and I. Segal. Designing dynamic mechanisms. *American Economic Review*, 97(2), 2007.
- [5] M. Babaioff, R. Lavi, and E. Pavlov. Single-value combinatorial auctions and implementation in undominated strategies. In *Proc. of the 17th Symposium on Discrete Algorithms (SODA)*, 2006.
- [6] D. Bergemann and S. Morris. Robust mechanism design. *Econometrica*, 73:1771 – 1813, 2005.
- [7] S. Bikhchandani, S. Chatterjee, R. Lavi, A. Mu'alem, N. Nisan, and A. Sen. Weak monotonicity characterizes deterministic dominant-strategy implementation. *Econometrica*, 74(4):1109–1132, 2006.
- [8] E. Clarke. Multipart pricing of public goods. *Public Choice*, 8:17–33, 1971.
- [9] J. Cremer and R. McLean. Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica*, 53:345 – 361, 1985.
- [10] C. d'Aspremont and L. Ge'rrard-Varet. Incentives and incomplete information. *Journal of Public Economy*, 11:25–45, 1979.
- [11] E. Dekel and A. Wolinsky. Rationalizable outcomes of large private-value first-price discrete auctions. *Games and Economic Behavior*, 43(2):175–188, 2003.
- [12] A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

- [13] Theodore Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.
- [14] H. Gui, R. Muller, and R. V. Vohra. Characterizing dominant strategy mechanisms with multi-dimensional types, 2004. Working paper.
- [15] P. Jehiel, M. Meyer ter Vehn, B. Moldovanu, and W. R. Zame. The limits of ex-post implementation. *Econometrica*, 74(3):585–610, 2006.
- [16] R. Lavi, A. Mu’alem, and N. Nisan. Towards a characterization of truthful combinatorial auctions. In *Proc. of the 44rd Annual Symposium on Foundations of Computer Science (FOCS’03)*, 2003.
- [17] R. Lavi and C. Swamy. Truthful mechanism design for multidimensional scheduling. In *The Proc. of the 8th ACM Conference on Electronic Commerce (EC’07)*, 2007.
- [18] Ron Lavi and Noam Nisan. Competitive analysis of incentive compatible on-line auctions. *Theoretical Computer Science*, 310:159–180, 2004.
- [19] D. Monderer. Monotonicity and implementability, 2007. Working paper.
- [20] H. Moulin. Incremental cost sharing: Characterization by coalition strategy-proofness. *Social Choice and Welfare*, 16:279 – 320, 1999.
- [21] R. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6:58–73, 1981.
- [22] R. Myerson and M. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29:265–281, 1983.
- [23] N. Nisan and A. Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35:166–196, 2001.
- [24] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge University Press, 2007.
- [25] K. Roberts. The characterization of implementable choice rules. In Jean-Jacques Laffont, editor, *Aggregation and Revelation of Preferences*, pages 321–349. North-Holland, 1979.
- [26] J. C. Rochet. A necessary and sufficient condition for rationalizability in a quasilinear context. *Journal of Mathematical Economics*, 16:191–200, 1987.
- [27] M. Satterthwaite. Strategy-proofness and arrow’s conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, 1975.
- [28] W. Vickrey. Counterspeculations, auctions, and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.
- [29] R. Wilson. Game-theoretic analyses of trading processes. In Truman Bewley, editor, *Advances in Economic Theory: Fifth World Congress*, pages 33–70. Cambridge University Press, 1987.