Social Welfare Maximization in Dynamic Strategic Decision Problems

A dissertation presented
by

Ruggiero Cavallo

to
The School of Engineering and Applied Sciences
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Computer Science

Harvard University
Cambridge, Massachusetts
May 2008

Formatting note

This document differs from the official version of the dissertation submitted to the Harvard University Registrar on May 23, 2008 only in the following way: the inter-line formatting has been changed from double-spaced to single-spaced. Thus the pagination differs from the official version, but all content is identical.

June 3, 2008

Thesis advisor
David C. Parkes

Author
Ruggiero Cavallo

Social Welfare Maximization in Dynamic Strategic Decision Problems

Abstract

Deriving effective group decision-making procedures for complex environments is hard but of fundamental importance, and the challenges grow significantly more daunting when individuals are self-interested. There is an inherent tension in striving to achieve *social* goals in decisions that will impact individuals who are only concerned with *selfish* objectives. Innumerable scenarios fit this mold, from resource allocation to coordinating behavior in the presence of global constraints. The field of mechanism design addresses such problems via specific payment schemes that disarm agent self-interest. This thesis attacks two fundamental issues in this area.

First: How can one implement a decision-making mechanism that maximizes the net welfare of a group of self-interested agents? Classic solutions typically require agents to make large payments to a central coordinator which, from the agents' perspective, purely detracts from social welfare. This thesis provides a mechanism applicable to arbitrary group decision-making problems that yields drastically higher group welfare in important settings, including resource allocation. The *redistribution mechanism* uses structure inherent in the domain to give payments required in the classic solution *back* to the agents in a way that does not yield a budget deficit or distort their incentives to participate truthfully.

Second: How can social welfare maximizing outcomes be reached with selfish agents in a setting that is *dynamic and uncertain*? In the real world, decisions do not exist in isolation, but rather are situated in a temporal context with other decisions. Individuals will act to maximize their *utility over time*, and decisions in the present influence how the world will look in the future, but rarely in completely predictable ways. This thesis addresses the problem of *dynamic mechanism design* for such settings and provides key results including: a characterization of the social welfare maximizing dynamic mechanisms that can be implemented in an ex post equilibrium; an extension to handle dynamically changing agent populations; an application to coordinating research preceding allocation of a resource. Finally, a *dynamic redistribution mechanism* unifies the two main focuses of the thesis, providing a solution with near-optimal social welfare properties for an array of important dynamic problems.

Contents

Acknowledgments

Considering this thesis and the process that led to its completion, it's obvious I owe so much to so many. But these are the best kind of debts – the kind you feel happy and excited about paying off because they're held by such good people. My gratitude is overflowing. Thanks:

To David Parkes, who I want to acknowledge as mentor, collaborator, and friend. First, it's impossible to imagine a more patient or caring advisor; in addition to everything he's taught me about research and all that, whenever I get a little cynical about things before too long I end up turned around by his graciousness, integrity, and humility. Second, so much of this thesis owes to his ideas and the direct collaboration between us—and I'm afraid David often gets the short end of the stick in working with me. Third, on a personal level I feel extremely fortunate to have shared these years here with such a great guy.

To the rest of my dissertation committee: Jerry Green, Barbara Grosz, Leslie Valiant, and Satinder Singh; I very much appreciate the care they took in helping me complete this, and thanks also to Satinder for his collaboration on work that appears here. To other teachers who have helped and inspired me along the way, especially Bart Selman who provided key early guidance, and also Tim Kelly and Patrick Henry Winston.

To my Dad, my ultimate hero, who pointed me to a path leading this way a long time ago and lovingly packed my bag with everything I'd need, and who will always be the source of answers to my most important questions. To my Mom, whose love carries me forward always and everywhere, and whose selflessness I aspire to as much as anything. To Lydia & Jay, Carlo, Corrina, and Cristiana, my siblings and closest compatriots in life. To my extended family, especially those in the Boston area, and more especially my godmother Franny, for helping make this a home. To my Li family, especially my mother-in-law, for her love and support these years.

To my dearest friends and confidants, Jacomo Corbo and Leif Easley: the idea of doing this without doing it together with them is unthinkable – they were with me all the way through and made it life (and good life) rather than just a series of events. Also to good friends Nicolas Janssen, Luis Martos, Alex Sheh, Ben Lee, and Cynthia Jensen, who, among other things, helped make residing here the joy that it has been. To Mike Pittarelli, Calin Trenkov-Wermuth, Sergei Sychov, Rama Chakravarthi, and Ruwan Ratnayake. To friends that have also contributed to a great research environment: Sébastien Lahaie, Laura Kang, Florin Constantin, Loizos Michael, Adam Juda, Ben Lubin, Sven Seuken, Jeff Shneidman, Marco Carbone, Kobi Gal, and others.

To LIN Zheng He, for patiently guiding me out of the woods.

Finally, to my wife Wen. In a very lucky life, the luckiest thing of all was finding her as a teenager in upstate New York so many years ago. Getting to spend our lives together is my greatest blessing. She supports me in all that I do, inspires me with her example, and is proof of the goodness that is possible. Thank you.

Chapter 1

Introduction

Among the most fundamental and important problems associated with social interaction is that of coordinating the behavior of individuals in a group, each with his or her own interests and concerns, towards maximization of social welfare. A self-interested individual will act to maximize his own utility, which will often come at the the expense of utility for the group. There is often an inherent tension between the optimization goals of individuals and those of a coordinator concerned with social welfare.

The field of *mechanism design* (MD) exists to address this tension, using payment schemes to transform a solution of inherently conflicting goals to one in which all interests are aligned with that of the coordinator. For instance, consider an allocation problem in which individuals are competing for a resource. An auction protocol in which the resource is allocated to the highest bidder for a price equal to the second highest bid transforms the scenario from one in which each agent wants to receive the item to one in which an individual wants the item only if his value for it is highest.

The mechanism design approach has seen great success in decision-making problems such as one-shot resource allocation, yet classic solutions still frequently fall short in two very important respects:

First, from a social welfare perspective mechanisms such as the second-price auction are often only desirable if we explicitly consider the value the auctioneer obtains, via the payment made by the highest bidder. If we are instead concerned only with the utility obtained by the bidders, this auction is highly undesirable because a large portion of the value will usually be transferred outside the group to the auctioneer. Imagine, for instance, a group of friends who jointly own a car and must decide who gets to use it on a particular evening. The friends could allocate the car to the one who wants it most via a second-price auction, but they would likely find it unacceptable to make large payments to an outside auctioneer. I will propose a *redistribution mechanism* that, in many environments, allows individuals to maintain the vast majority of the value obtained from decisions that are made.

Second, mechanism design has largely disregarded the context in which decisions are made; in particular, it has not accounted for scenarios in which a *sequence of decisions* is to be made over time with new information obtained incrementally. In fact it is typical that decisions are made in the context of other decisions, and the context of time and future situations is essential to determining both what is best for the group and what is best for each individual. In such settings, as we will see, static solutions that consider each decision in isolation will not work. I will present *dynamic mechanism design* (DMD), which extends MD by explicitly reasoning about sequences of decisions and how individuals value utility obtained now relative to utility obtained in the future.

These two main contributions of the thesis—improving social welfare via redistribution, and addressing dynamics in decision-making scenarios—bring the mechanism design enterprise closer to the real world. In the case of redistribution, individuals will frequently not be satisfied by "solutions" that do not principally benefit the group of interested parties. In the case of moving from a static to a dynamic analysis, when a context of future decisions bears on consideration of a current decision (as it often will), static "solutions" will typically not be solutions at all.

The rest of this chapter gives an intuitive introduction to the challenge of group decision making, the style of solution that mechanism design offers, the shortcomings of previous approaches, and the main contribution areas of the thesis. Finally I'll preview the main results in more detail.

1.1 The setting: group decision-making

This thesis is concerned with scenarios in which a decision must be made that will impact a group of individuals (or "agents"). Each decision will bear a certain amount of *value* for each agent, and will be made by a coordinator or "center". There is a set of outcomes from which the center must choose; each agent holds some information that is pertinent to the decision-making process and, in particular, determines the agent's value for every possible outcome. Agents are rational and self-interested, acting in ways that maximize their own individual utilities. Though it is possible to consider other goals, in this thesis I take the goal of the center to be maximization of social welfare (the sum of agent values), i.e., to choose the outcome that is most preferred by the group as a whole. The center's pursuing this goal can be justified by considering it an actor that intrinsically feels invested in the welfare of the group (imagine a government, for instance), or one who is hired by the agents simply to organize the decision process. In either case social welfare maximization as an objective follows naturally.

Observe, trivially, that if there is no particular behavior on the part of the agents required for determination and implementation of the socially optimal outcome, then there is no problem posed by self-interest. If the center has the power and authority to choose and enforce any outcome and also knows *a priori* the valuations of each agent for each outcome, he can simply choose and implement it. The challenge arises when we assume agents must participate in the mechanism in some way for it to be effective; we will almost always assume that this challenge arises due to *privacy of the agents' information*. If the value of an outcome to a group depends on information that only they know, it will be impossible for the center to determine the social-welfare maximizing outcome without their willing participation.

Consider the example of allocating a single indivisible item among a group of 3 agents. There are 3 possible outcomes, one for allocation to each member of the group. Assuming an agent obtains no value when he does not receive the item, relevant private information can be expressed in a single number: the value the agent would obtain if he were to be allocated the item. A social-welfare maximizing decision procedure simply chooses the outcome in which the agent who values the item most receives it, but in order to determine which agent that is the coordinator must somehow get the agents to reveal their true private information.

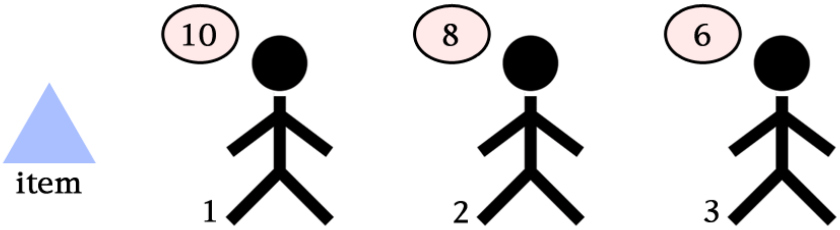


Figure 1.1: Illustration of a typical group decision-making problem: there is a single indivisible resource to be allocated to one person in a group of 3. Each person has his or her own value for the item, and these values are *private*—unknown to the rest of the group. To determine the socially optimal outcome (allocation to person 1, here) the participants must be persuaded to truthfully share their private valuation information. But how?

In this simple single-item allocation problem determining the optimal solution is easy once the agents have revealed their private information, but in many settings this will not be the case. When the outcome space is very large (for instance in *combinatorial* allocation scenarios) even merely doing a comparison of the social value for each outcome according to agent reported information may be impossible. In dynamic environments determining optimal decisions involves solving a complex stochastic optimization problem, which is often intractable. Thus there is an important computational aspect to problems in this area. These considerations are especially pertinent because, we will see, being able to form strong expectations that agents will truthfully reveal private information often requires choosing decisions that are optimal given their reports. While there will always be environments where this is computationally infeasible, an important direction (pursued in two chapters of this thesis and elsewhere) is identification and exploitation of underlying *structure* in decision environments that makes optimal policies tractable.

1.2 Mechanism design

Mechanism design is the enterprise of engineering socially desirable outcomes in the context of self-interested agents with private information; the tool used to achieve this is *monetary transfer payments*. Typically, agents are asked to report their private information to the center, who then chooses an outcome and potentially makes payments to (or demands payments *from*) the agents.

MD works in a context of multiple rational agents, and so it applies a game-theoretic analysis and seeks to implement an efficient choice function *in equilibrium*. We will get to formal equilibrium definitions in Chapter 2, but for now it is enough to say that an equilibrium is a specification of a profile of behaviors or "strategies", one for each agent, such that each agent maximizes his or her own utility by playing the specified strategy.

Consider a simple mechanism for single-item allocation problems in which agents are asked to report their value for receiving the item, and then the item is allocated to the agent who reports the highest value. There is only one equilibrium in this mechanism, and it is for all agents to report the highest value they can. This is a *bad* equilibrium because it does not reliably lead to a socially-welfare maximizing outcome. If all agents report value ∞, how can the center determine whose value is actually highest?

The idea of mechanism design is to use transfer payments to transform a scenario in which agents act only in their own interests—and possibly at the expense of the social interest—into one in which the interests of each individual and that of the group are *both* best served by the same behavior (usually, truthful reporting of private information). Consider now the mechanism, illustrated in Figure 1.2, that allocates the item to the agent that announces the highest value, but then *charges* that agent an amount equal to the second highest reported value. This is called a "second-price" or "Vickrey" auction and, we will see, it has the following nice property: every agent maximizes his utility by reporting his true value for the item, *regardless of what the other agents announce.*

In Chapter 2 we will see a generalization of this mechanism—the Vickrey-Clarke-Groves (VCG) mechanism—that elicits truthful reporting of information in arbitrary decision-making scenarios. Note that once transfer payments are involved, in addition to incentive properties we need to also consider the budgetary properties of mechanisms. Crucially, the "no-deficit" property will typically be required: the mechanism should not, in aggregate, pay out more money to the agents than it takes in from

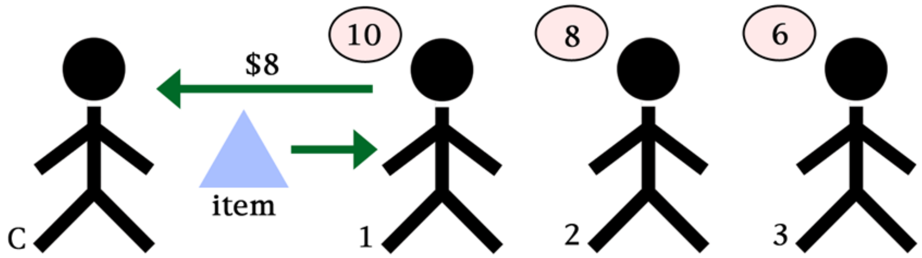


Figure 1.2: Illustration of a Vickrey auction. The item is allocated to the highest bidder for a price equal to the second highest bid, paid to the center.

them. The VCG mechanism has strong incentive properties and at the same time satisfies no-deficit; in fact, in many settings it yields net payments to the center that amount to a very large percentage of the total value yielded from the chosen outcome. In other words, it yields high revenue.

1.3 Contribution 1: Redistribution mechanisms

Whether revenue is a scourge or a blessing depends on the setting, and also sometimes on whom you ask. For instance, in an allocation problem in which the center is the initial holder of the item to be auctioned (and thus also the "auctioneer"), from the center's perspective high revenue means he will personally derive a large portion of the utility yielded from reallocating the resource. Of course the bidders would prefer a mechanism with less or, ideally, no revenue.

In other settings there will be no perspective from which revenue is considered desirable. Consider government allocation of a public good such as wireless spectrum, city green space, or usage rights for an expensive technology like a space telescope or supercomputer. The goal in allocating here is often solely to maximize the welfare of the agents competing for the resource, and large payments made to the government undermine that very goal. One can imagine even more extreme cases. Consider again the example of a group of friends that jointly own a car and must decide who gets to use it on a particular evening. Here there is no inherent center or auctioneer, and the role of a center (if the friends should appoint one) would merely be to facilitate the group's discovery of how best to extract value from the resource they jointly own. Any payments the group must make are clearly waste.

One of the two main contribution areas of this thesis is the study of "redistribution mechanisms", in which payments required under VCG are returned to the agents in a careful way that retains both the incentive and the no-deficit properties of that mechanism. I will demonstrate that in some settings no improvement over VCG is possible, but in others it is. I will propose a specific alternative to VCG that is applicable to any setting and redistributes revenue when doing so does not lead to a deficit or distort incentives. I will demonstrate empirically that in allocation problems the mechanism maintains almost all value within the group of agents while VCG maintains almost nothing.

Consider again the example portrayed in Figure 1.2: the VCG mechanism (the Vickrey auction here, since this is a single-item allocation problem) achieves the efficient outcome in equilibrium, but 80% of the value ($8) is transferred to the center; only $2 remains in the hands of the agents (agent 1, here). In the redistribution mechanism I propose, 87% of the value is maintained within the group of agents in this scenario; the center ends up with a total payment of only 1.33 (see Figure 1.3). Redistribution has not resulted in a deficit and the incentive properties are the same as in the Vickrey case: no agent can every benefit from reporting anything other than his true value for the item.

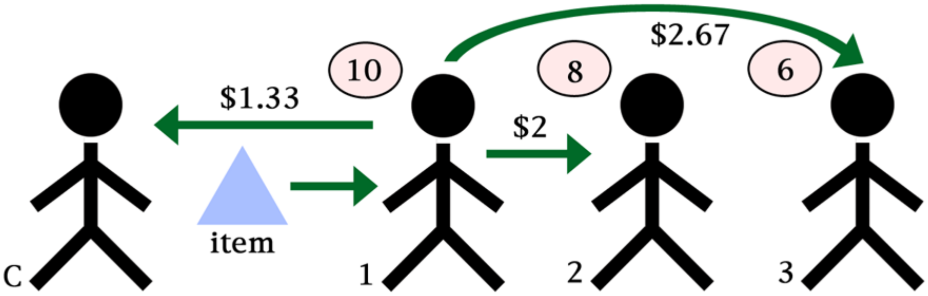


Figure 1.3: Illustration of the *redistribution mechanism* for the single-item allocation problem depicted in Figures 1.1 and 1.2. As in the Vickrey auction, the item is allocated to the highest bidder, but now the majority of the value yielded by the allocation is retained by the agents (8.77, or 87%, here).

1.4 Contribution 2: Dynamic mechanism design

So far the examples of decision-making scenarios I've introduced have all been *static*. A single decision is to be made, and analysis is limited to the isolated consideration of that one decision. We considered the problem of allocating an item, but we did not consider what related decision-problems might arise *after* that allocation decision. Essentially, the static mechanism design approach works under the assumption that nothing that will happen in the future is relevant to analysis of the decision being faced now. Obviously this does not accurately model many important scenarios faced in the real world.

Consider the simple extension of single-item allocation to a case in which the item will be allocated twice: one agent will get it, use it for a time, and return it to the center, at which point it will be allocated again. This makes the problem more complex in several ways. First, the outcome space is now of size $2 \cdot n$ (where n is the number of agents); each outcome is an ordered pair (i, j) indicating allocating to agent i the first time and agent j the second time. Also, an agent's value for receiving the item in the second period may depend on what happens in the first period—imagine the resource is a DVD movie; the agent that gets to watch it in the first period will likely have very low (or 0) value for obtaining it to watch again in the second period. It may even be that an agent's value for the item in the second period depends *unpredictably* on what happens in the first period. If the movie is really great but the viewer was distracted and ended up confused about some of the subtle plot points, he may indeed have high value for seeing it again; the expected value of the second watching depends on what happens during the first watching.

Essentially, the hallmark of dynamic decision-making problems that makes them more complex than static ones is that *new private information arrives over time.* Thus a successful *dynamic mechanism* will specify an optimal decision policy that chooses an outcome in every time period given the information reported by agents in that period, and a transfer payment policy that incentivizes agents to report their true private information *in every period.*

In the arena of dynamic mechanism design main contributions of this thesis include: elaboration of the dynamic mechanism design framework; specification of dynamic mechanisms with desirable equilibrium properties; a characterization of what social-welfare maximizing mechanisms can be implemented in strong equilibrium; a specification of a dynamic redistribution mechanism for an important subclass of settings; dynamic mechanisms for settings where the population of agents is changing (agents "arrive and depart"); and an application of the theory to coordinating research prior to allocation of a resource.

1.5 Outline of chapters and main results

The following is a chapter-by-chapter outline of the thesis. Chapters 2 and 4 are background on static mechanism design and sequential decision making; the related work is provided in the chapter containing results to which it is most relevant. Discussion of future work and detailed discussion of what's left "undone" is, with a few exceptions, left to the concluding chapter.

Chapter 2

Chapter 2 is background on static mechanism design, which provides the foundation for practically the entire thesis. I introduce game theory and game theoretic equilibria, which provide the framework for analysis of mechanism designs solutions. I define the mechanism design paradigm and formalize its goals; I present strong negative results and the assumptions required to move beyond them. I present the hallmark positive result of MD: the Groves class of mechanisms. I define the VCG mechanism and demonstrate its many desirable properties. I finally turn to budget balance: I motivate it as an important concern, and prove that among social-welfare maximizing mechanisms VCG is revenue-maximizing. I present the AGV mechanism, which sacrifices certain desirable properties of VCG in order to always achieve a balanced budget.

Chapter 3

In Chapter 3 I present the redistribution mechanism for static settings, a main contribution of the thesis. Significant results of this chapter have been published in [Cavallo, 2006b; 2006a]. I motivate this work by discussing settings in which payments made to the center by agents can be very undesirable. I demonstrate that without any knowledge about the structure of the decision-making environment, unfortunately, among mechanisms that don't ever run a deficit yet have the strongest incentive properties, the VCG mechanism is in fact *unique* and so these payments are necessary. But I then move to observe that when we do take into account structure—for instance, the fact that in a single-item allocation problem only one agent obtains non-zero value—VCG is not unique and we can do much better in terms of minimizing payments. I propose a general redistribution mechanism that is applicable to *any* static domain (with or without structure), but that deviates from VCG only when the structure of the domain allows it. I demonstrate numerically that in single-item allocation problems the vast majority of value can consistently be retained within the group of agents when the group size is greater than a few.

Chapter 4

Chapter 4 provides background on optimal decision making in dynamic environments. I introduce the dynamic decision-making setting and the Markov decision process (MDP) formalism I will use to model it. I describe handling *multi-agent* decision problems as a basic extension of the framework. I present leading approaches for computing optimal policies ("solving" MDPs), both exact and approximate. I give significant attention to the case of multi-armed bandit (MAB) problems, a subclass of MDPs in which optimal policies can always be computed in time linear in the number of agents, a result due to Gittins. I present Gittins's index policy results and important algorithms for computing such policies.

Chapter 5

Chapter 5 provides an introduction to dynamic mechanism design and presents main results. I introduce the problem and the equilibrium concepts that will constitute solutions. I provide an extension of the Groves class of mechanisms for the static setting to the dynamic case, and prove that (with a reasonable restriction) this "dynamic-Groves" mechanism class corresponds exactly to the mechanisms that implement social-welfare maximizing policies in a strong truthtelling equilibrium. I then present the dynamic-VCG mechanism of Bergemann & Välimäki [2006] and the dynamic-balanced mechanism of Athey & Segal [2007], important developments in the field. I provide new simple proofs of dynamic-VCG's incentives properties, and a novel result that it is revenue-maximizing among all mechanisms that implement social-welfare maximizing decisions. This chapter includes results that have been published in [Cavallo *et al.*, 2006; 2007; Cavallo, 2008], and includes collaborative work with David C. Parkes and Satinder Singh.

Chapter 6

Chapter 6 presents a dynamic redistribution mechanism for multi-armed bandit problems, an important novel result that unifies the two main contribution areas of the thesis (published in [Cavallo, 2008]). I demonstrate that for problems that can be represented as a multi-armed bandit (e.g., repeated allocation of a single item), dynamic-VCG typically yields very high revenue. As in the static setting, in many important environments this revenue is actually waste. I propose a dynamic redistribution mechanism that has a simple and elegant form. It is, conceptually, a natural analogue of the static redistribution mechanism for single-item allocation settings, though more complex. The mechanism redistributes large portions of the revenue back to the agents. I demonstrate empirically that in a group of at least several agents, the group typically retain almost all of the value yielded by decisions ($\sim$97%), while in dynamic-VCG the value retained is small ($\sim$10%).

Chapter 7

Chapter 7 addresses settings in which either the population of agents is changing over time or, more generally, agents periodically go out of communication with the center. I extend positive results in dynamic mechanism design to these settings, focusing on extensions of dynamic-VCG. In addition, I provide an analysis of dynamic mechanism design for settings in which agent valuations are *interdependent* (e.g., when one agent's expected value for an outcome is high if and only if another agent's is high). Results from this chapter appear in [Cavallo *et al.*, 2007], and are collaborative with David C. Parkes and Satinder Singh.

Chapter 8

In Chapter 8 I provide an application of the new theory of dynamic mechanism design to scenarios in which a single resource is to be allocated just once, but where agent valuations for the resource can potentially be changed via a deliberation process. This models settings where, for instance, a new technology becomes available and firms' valuations for obtaining it are subject to improvement given future research. Perhaps research will yield knowledge of a *new way of using* the resource for greater profit, which would increase the firm's willingness to pay for the item. Such settings bear a certain resemblance to multi-armed bandits, though they don't quite fit the model. I provide a new proof that these problems can be *reduced* to multi-armed bandits in a lossless way, and thus that they admit tractable optimal solutions. I also derive a new dynamic mechanism (a variant on dynamic-VCG) to handle scenarios in which one agent can reason about the value of the resource to *another* agent. The results of this chapter are collaborative with David C. Parkes and have been published in [Cavallo and Parkes, 2008].

Chapter 9

Chapter 9 concludes the thesis. I provide an informal summary of the results and describe important directions for future research.

Throughout the thesis when I present results that are due to others, I cite the responsible parties in the name or beginning of the result.

Chapter 2

Static mechanism design

Synopsis

This chapter provides background on mechanism design (MD), the study of engineering desirable equilibrium outcomes in strategic environments via payment schemes. I start with a discussion of what constitutes a solution in the MD framework, and then present both negative and positive results providing indicators of what MD can and cannot achieve. The marquee possibility results are the Groves class of mechanisms in general, and the VCG mechanism in particular, which provide the right incentives for truthful behavior of self-interested agents in dominant strategies.[1]

2.1 Game theory

The group decision making problem can be formalized as follows: an outcome will be chosen from a set O, yielding various amounts of *utility* for each member of a group of agents $I = \{1, \ldots, n\}$. Each agent $i \in I$ has a private type θ_i that encapsulates all information private to i that is relevant to the decision-making process. The outcome space O may depend on the set of agents that participates in the mechanism, but it *does not* depend on any information that is private to the agents.[2] The space of possible private information that agent i could possible have—i.e., i's *type space*—is denoted Θ_i, and the joint type space is $\Theta = \Theta_1 \times \ldots \times \Theta_n$. I assume that agents are rational and self-interested, in that each will act to maximize his own utility.

[1] I call the chapter "*static* mechanism design" to distinguish it from the expanded theory of *dynamic* mechanism design, the presentation of which is a primary topic of this thesis. Static mechanism design is in fact a very important special case of dynamic mechanism design in which there is a single decision to be made in a single time-period. For simplicity I will often just use "mechanism design" to refer to the theory for static (or "one-shot") settings.

[2] For instance, in an allocation decision problem an agent i may bring an item X to the mechanism for allocation, so the outcome in which some agent j receives X is dependent on i's presence; but I assume that the set of items an agent brings *if* he participates is known exogenously.

Game theory[3] models scenarios of strategic interaction between self-interested agents, and will provide the basis for defining what should be considered a "solution" to a mechanism design problem. In the game theoretic approach, each agent i has an action space Λ_i and chooses an action according to a "strategy" σ_i based on his private information; that is, a strategy σ_i is a mapping $\Theta_i \rightarrow \Lambda_i$. Given strategies $\sigma = (\sigma_1, \ldots, \sigma_n)$ played by the agents, an outcome $g(\sigma(\theta))$ results. Each agent i obtains a utility $u_i(\theta_i, g(\sigma(\theta)))$ when his type is θ_i and actions $\sigma(\theta) = (\sigma_1(\theta_1), \ldots, \sigma_n(\theta_n))$ are played. A particular instantiation of a complete type profile and action set for each agent is sometimes called a "game". Such scenarios can be represented concisely in matrix or "normal" form when there are two players with an equal number of actions, an example of which is given in Table 2.1.

	a_2	b_2
a_1	2,2	0,0
b_1	0,0	2,3

Table 2.1: Normal-form representation of a two-agent decision making scenario in which each agent has two actions: a and b. Each cell of the matrix contains the utilities received by agents 1 and 2, respectively, when agent 1 plays the corresponding row action and agent 2 plays the corresponding column action. For instance if both agents play their b action, here, agent 1 obtains utility 2 and agent 2 obtains utility 3.

2.1.1 Game theoretic equilibria

Given that we are considering self-interested agents that can explicitly reason about what *other* agents will do, it is useful to examine what outcomes will occur *in equilibrium* in a given game. The first equilibrium concept we consider is Nash equilibrium, in which each agent acts to maximize utility given the knowledge that other agents are also acting to maximize their own utilities. Here and in many places to come it will be useful to consider the profile of types *excluding* that of some agent i, which I denote θ_{-i}, i.e., $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$. Likewise I write σ_{-i} to denote the profile of strategies excluding i's strategy.

[3] Though there were a few earlier works broaching this area, game theory as we know it was first developed by John von Neumann [von Neumann and Morgenstern, 1944], and was rocketed further forward by the equilibrium analysis of Nash [1950]. See [Osborne and Rubinstein, 1994] or [Mas-Colell *et al.*, 1995] for good introductions to the field.

Definition 2.1 (Nash equilibrium). *Given a type profile θ, strategy profile σ constitutes a Nash equilibrium if and only if:*

$$\forall i \in I, \sigma_i' \neq \sigma_i, \quad u_i(\theta_i, g(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i}))) \geq u_i(\theta_i, g(\sigma_i'(\theta_i), \sigma_{-i}(\theta_{-i}))) \tag{2.1}$$

In a Nash equilibrium no agent can benefit from unilaterally deviating from the specified strategy profile. So, for instance, the example in Table 2.1 has 2 Nash equilibria: (a_1, a_2) and (b_1, b_2).

An even stronger equilibrium-like notion describes scenarios in which each agent has some strategy that is guaranteed to yield him maximum utility, *regardless of what other agents do.* Such a strategy is called a dominant strategy.

Definition 2.2 (dominant strategy). *Given $\theta_i \in \Theta_i$, a strategy σ_i is a dominant strategy for i if and only if:*

$$\forall \theta_{-i} \in \Theta_{-i}, \sigma_{-i}, \sigma_i' \neq \sigma_i, \quad u_i(\theta_i, g(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i}))) \geq u_i(\theta_i, g(\sigma_i'(\theta_i), \sigma_{-i}(\theta_{-i}))) \tag{2.2}$$

This is technically a *weakly* dominant strategy; a *strictly* dominant strategy is one in which the inequality is strict. An example in which an agent has a dominant strategy is illustrated in Table 2.2. In fact in this example agent 1 has 2 weakly dominant strategies. Clearly if there is a strictly dominant strategy it is unique.

	a_2	b_2	c_2
a_1	2,2	4,0	0,3
b_1	2,0	4,1	0,1
c_1	1,0	3,3	0,4

Table 2.2: A two-player game in which player 1 has two dominant strategies, a_1 and b_1, and agent 2 has dominant strategy c_2. The Nash equilibria are thus (a_1, c_2) and (b_1, c_2).

The dominant strategy concept is very strong in that when there is a strictly dominant strategy it allows us to know *with certainty* what action a *rational* agent would take. When all agents are rational and have such a strategy we know exactly what payoffs will be realized.

The final equilibrium notion I describe here applies to settings in which agents have beliefs about the types of other agents. In a *Bayes-Nash equilibrium*, no agent can expect to gain from deviating from the equilibrium strategy, given that other agents don't. Let $b_i(\theta_{-i})$ denote a distribution over the types of agents other than i, representing i's beliefs about them, and let $\tilde{\theta}_{-i}$ be a random variable denoting (from i's perspective) the realization of θ_{-i}.

Definition 2.3 (Bayes-Nash equilibrium). *Given a type profile θ and agent beliefs $b_1(\theta_{-1}), \ldots, b_n(\theta_{-n})$ about other agents' types $(\tilde{\theta}_{-i})$ that are common knowledge,*[4] *strategy profile σ constitutes a Bayes-Nash equilibrium if and only if:*

$$\forall i \in I, \sigma'_i \neq \sigma_i, \quad \mathbb{E}_{b_i(\theta_{-i})}[u_i(\theta_i, g(\sigma_i(\theta_i), \sigma_{-i}(\tilde{\theta}_{-i})))] \geq \mathbb{E}_{b_i(\theta_{-i})}[u_i(\theta_i, g(\sigma'_i(\theta_i), \sigma_{-i}(\tilde{\theta}_{-i})))] \tag{2.3}$$

2.2 The mechanism design framework

Mechanism design[5] addresses group interaction scenarios in which there is a goal in the interaction, usually maximization of the total utility realized, i.e., social welfare. It is typically assumed that there is an entity—"the center"—that has the ability to set the choice function, i.e., specify the mapping from agent actions to outcomes.

Informally, a mechanism defines rules for interaction between the center and the agents that lead to a decision. The "rules" come in the form of specifying ways that agents can communicate with the center, ways in which the center will choose an outcome given actions that agents take, and (usually) ways in which the center will provide *payments* to the agents in order to incentivize the agents into behaving and communicating with the center in ways that the center prefers. Formally, a mechanism is defined as follows:[6]

Definition 2.4 (mechanism). *A tuple (Λ, g, T), where:*

- *$\Lambda = \Lambda_1 \times \ldots \times \Lambda_n$ is a joint action space.*
- *$g : \Lambda \rightarrow O$ is an action-choice function.*
- *$T = (T_1, \ldots, T_n)$, where for each $i \in I$,*
  $T_i : \Theta \rightarrow \Re$ is a transfer payment function (with payments made **to** *agent i).*

Note that in the context of a mechanism an "outcome" is actually bifaceted: there is a choice selected, and also a set of transfer payments defined, one for each agent. For clarity I express these two facets of the outcome separately, so each agent i's utility $u_i : \Theta_i \times O \times \Re \rightarrow \Re$. For instance, $u(\theta_i, o, 5)$ is the utility i obtains if his type is θ_i, outcome o is selected, and he receives transfer payment 5. I will use notation $v(\theta_i, o)$ to denote i's value when no transfer is made (i.e, $u(\theta_i, o, 0)$).

[4] I.e., each agent knows that each i's beliefs are $b_i(\theta_{-i})$, and each agent knows that each agent knows this, etc.

[5] The field began with the work of Hurwicz [1960; 1972]. See [Parkes, 2001] (Chapter 2) or [Jackson, 2000] for other introductory presentations.

[6] The definition I provide here is somewhat narrow, in that it restricts to a setting in which agents simultaneously take a single action. One can imagine natural generalizations.

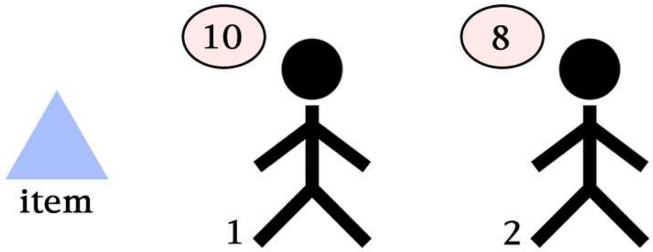


Figure 2.1: Resource allocation is a typical game theoretic scenario that embodies conflicting goals: each individual seeks to obtain the good, while at the same time the allocator may seek to allocate the good to the individual who *wants it most*.

A reasonable way to think about the intuitive motivation for mechanism design is this: the center may only be able to specify action-choice rules that induce games with (perhaps *only*) bad equilibria; the payments allow the center to essentially transform such a game to a new game with better equilibria. For instance consider again the example in Figure 2.1, which portrays a scenario in which a decision must be made regarding allocation of a single item. There are two agents, each with his or her own (privately known) value for the item. Imagine a mechanism in which the actions an agent can take are defined to be the announcing of a value for the item, either 0, 2, or 10.[7] The action-choice rule is defined to allocate the item to the agent who announces the highest value; if there is a tie in announced values it is allocated arbitrarily amongst the agents that announce the highest value. In this mechanism there is only one equilibrium, and that is for both agents to report the highest value possible (10) regardless of their values.

But now consider a mechanism in which we add a transfer payment scheme to this framework, where the winning agent must pay the center the value that the other agent announces for the item. The utilities (or "payoffs")[8] under each mechanism are portrayed in Table 2.3. This mechanism, known as a second-price or "Vickrey" auction, has the property that agents always maximize their payoffs by announcing their true values. This allows the center to determine and subsequently choose the outcome that is social welfare maximizing (allocation to agent 1, here). We will see later in this chapter that the Vickrey auction is a special case of a more general mechanism that has the same property and can be applied not just to single-item allocation scenarios, but to arbitrary decision-making problems.

[7] I restrict possible value reports to this set here only to simplify presentation; more often agents will be able to announce any value.

[8] Assume just for this example that agents have quasilinear utility (described in detail later), with each agent's payoff equal to his value for the outcome plus/minus any payment he receives/makes.

	0	2	10
0	5,1	0,2	0,2
2	10,0	5,1	0,2
10	10,0	10,0	**5,1**

	0	2	10
0	5,1	0,2	0,2
2	10,0	4,0	0,0
10	10,0	**8,0**	0,-4

Table 2.3: Normal-form representation of a two-agent single-item allocation scenario under two different mechanisms. Agent actions are value announcements. In the first mechanism the item is allocated to the agent who announces the highest value; both agents have announcing 10 as a dominant strategy. In the second mechanism payments are added: the agent allocated the item must pay the center the value announced by the other agent; truthful reporting is a dominant strategy. Equilibrium outcomes are in bold; quasilinear utility is assumed (see Definition 2.20).

2.2.1 Implementation

The goal in mechanism design is to achieve a particular outcome in equilibrium. If we redescribe this goal without considering any particular type profile, mechanism design seeks to implement[9] a particular *choice function* f in equilibrium, where a choice function defines a mapping from agent types to outcomes. A choice function is distinct from the more general action-choice function that a mechanism can specify, which is a mapping from *actions* the agents take to outcomes. Very frequently (and in all cases in this thesis) the goal in mechanism design will be to achieve implementation of a social-welfare maximizing or *efficient* choice function.

Definition 2.5 (efficient social choice function). *A choice function $f^* : \Theta \rightarrow O$ is efficient if and only if*[10]

$$\forall \theta \in \Theta,\ f^*(\theta) = \arg\max_{o \in O} \sum_{i \in I} v_i(\theta_i, o) \tag{2.4}$$

If the center seeks to maximize social welfare, the strongest "solution" we can hope for is to design a mechanism in which the agents all have a dominant strategy, and when they play it and action-choice function g is applied, the outcome selected is equivalent to what an efficient choice function f^* would select.

[9] I use the term "implement" in a way distinct from its usual meaning in *implementation theory* (see, e.g., [Jackson, 2001]), which requires achieving the function in *every* equilibrium.

[10] Note that there may be multiple efficient choice functions. I will use f^* to arbitrarily refer to any one of them, since the arbitrary way that ties are broken will not be important for my analysis.

Definition 2.6 (dominant strategy implementation). *A mechanism* (Λ, g, T) *implements choice function* f *in dominant strategies if,* $\forall \theta \in \Theta$*, there exists a strategy profile* σ *such that for each* $i \in I$*,* σ_i *is dominant strategy and* $g(\sigma(\theta)) = f(\theta)$*.*

Solutions based on the other equilibrium concepts we've seen, though somewhat less desirable, are still satisfactory in some environments.

Definition 2.7 (Nash equilibrium implementation). *A mechanism* (Λ, g, T) *implements choice function* f *in Nash equilibrium if,* $\forall \theta \in \Theta$*, there exists a strategy profile* σ *that is a Nash equilibrium and* $g(\sigma(\theta)) = f(\theta)$*.*

Note that any mechanism that implements a given choice function in dominant strategies also implements it in Nash equilibrium. Dominant strategy implementation is also preferable to Bayes-Nash implementation, since agents must have certain beliefs in order for the latter to occur:

Definition 2.8 (Bayes-Nash equilibrium implementation). *Given agent beliefs* $b_1(\theta_{-1}), \ldots, b_n(\theta_{-n})$ *that are common knowledge, a mechanism* (Λ, g, T) *implements choice function* f *in Bayes-Nash equilibrium if,* $\forall \theta \in \Theta$*, there exists a strategy profile* σ *that is a Bayes-Nash equilibrium and* $g(\sigma(\theta)) = f(\theta)$*.*

The implementation goal that I will focus on throughout this thesis is that of an efficient social choice function. I use the term *efficient* to describe a mechanism that achieves this.

Definition 2.9 (efficient). *A mechanism is efficient if it implements a social-welfare maximizing (efficient) choice function.*

The definition may seem slightly underspecified, but we will make clear the intended equilibrium concept whenever we use the term. For instance, stating that a mechanism is "efficient in dominant strategies" is to say that it implements an efficient choice function in dominant strategies.

2.2.2 Direct mechanisms and the revelation principle

The space of possible mechanisms is difficult to fully conceptualize as, technically, there is a distinct mechanism for each distinct set of actions or behaviors that one can imagine allowing the agents to perform. But a particularly appealing subset of this huge space is that consisting of *direct mechanisms*, in which the only action each agent is allowed to perform is communication of a claim about his private type (which determines his preferences). A direct mechanism is a mechanism in which the action space is implicitly defined this way, and so the function that selects outcomes is in fact a choice function (a mapping of agent type profiles to outcomes), a more narrowly defined version of the general action-choice function concept.

Definition 2.10 (direct mechanism). *A tuple (f, T), where:*

- $f : \Theta \rightarrow O$ *is a choice function.*
- $T = (T_1, \ldots, T_n)$*, where for each* $i \in I$*,*
  $T_i : \Theta \rightarrow \Re$ *is a transfer payment function (with payments made* **to** *agent* i*).*

Recall that each agent i's utility is a function of the outcome selected and the transfer payments specified. In the context of a direct mechanism (f, T) we can write $u_i(\theta_i, f(\sigma(\theta)), T(\sigma(\theta)))$ to denote the utility i obtains when the agents have type profile θ and play strategy profile σ. Direct mechanisms are very appealing because of their simplicity. Fortunately, it turns out that if we are solely concerned with implementation of a particular choice function, it is without loss of generality to *only* consider this narrow subspace of all mechanisms.[11]

Theorem 2.1 (the revelation principle). *If there exists a mechanism that implements choice function f in dominant strategy, Nash, or Bayes-Nash equilibrium, then there exists a direct mechanism that implements f in the same equilibrium concept, where the equilibrium strategy for each agent is to report his type truthfully.*

The veracity of the revelation principle can be seen intuitively by imagining, for any indirect mechanism in which agents perform some arbitrary action leading up to an outcome choice, a direct mechanism analogue in which all such actions are "simulated" by the center after the agents communicate their types. The possibility of this is implicit in the definition of the *mechanism* concept, as the center can choose an arbitrary action-choice function g that maps agent actions to outcomes.

The revelation principle is an extremely important result in mechanism design, as it allows a mechanism designer to restrict attention to direct mechanisms in seeking to implement a particular choice function. As we will see throughout this chapter, this principle has proved central to the discovery of key negative and positive results in mechanism design.

In thinking about the ramifications of the revelation principle, though, we must be careful not to conclude that direct mechanisms are the only mechanisms *ever* worth implementing. While direct mechanisms have very compelling attributes, in certain settings there may be *indirect mechanisms* that also implement a desired choice function, but with preferable computational and/or privacy properties. Direct mechanisms require agents to *completely* reveal private types (preferences). In some cases it may be computationally very difficult for an agent to even figure out exactly what his preferences are, or there may be issues of trust that would make agents hesitant to completely share such preferences even when they can be ascertained. Nonetheless in many or most scenarios we will discuss in this thesis, the context of direct mechanisms is very useful for conceptual clarity, and the benefits often appear to outweigh these potential costs.

[11] The revelation principle was first noted by [Gibbard, 1973].

2.2.3 Mechanism desiderata

There are three primary criteria by which a mechanism designer will evaluate potential mechanisms, which can be summarized as follows:

- Incentives: will agents behave in ways that predictably lead to desirable outcomes (i.e., in a direct mechanism, will they be truthful)?
- Participation: will agents want to participate in the mechanism at all?
- Budget: will net transfer payments flow towards the center, towards the agents, or completely cancel out?

Each of these areas of concern is very important, and we will discuss each in turn.

Incentive Properties

Given the revelation principle, it is natural to seek mechanisms which have simple truth-revealing equilibria. Intuitively, if agents are best-off reporting types truthfully (i.e., if an equilibrium strategy profile σ has $\sigma_i(\theta_i) = \theta_i$ for every $i \in I$), the mechanism designer can achieve the desired choice function by simply applying it to agent reports, and also the burden of computing complex utility-maximizing strategies is lifted from the agents.

Definition 2.11 (strategyproofness). *A direct mechanism (f, T) is strategyproof if and only if truthfulness is a dominant strategy for every agent, i.e.,*

$$\begin{aligned}\forall \theta \in \Theta, i \in I, \sigma_i, \sigma_{-i}, \quad & u_i(\theta_i, f(\theta_i, \sigma_{-i}(\theta_{-i})), T(\theta_i, \sigma_{-i}(\theta_{-i}))) \geq \\ & u_i(\theta_i, f(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})), T(\sigma_i(\theta_i), \sigma_{-i}(\theta_{-i})))\end{aligned} \tag{2.5}$$

In a strategyproof mechanism no agent can *ever* benefit from deviating from the simple truth-revealing strategy. The term used when truthfulness holds in Nash rather than dominant strategy equilibrium, for *any* profile of true agent types θ, is *ex post incentive compatibility*:

Definition 2.12 (ex post incentive compatibility). *A direct mechanism (f, T) is ex post incentive compatible if and only if truthfulness is a Nash equilibrium for every possible agent type, i.e.,*

$$\forall \theta \in \Theta, i \in I, \sigma_i, \quad u_i(\theta_i, f(\theta), T(\theta)) \geq u_i(\theta_i, f(\sigma_i(\theta_i), \theta_{-i}), T(\sigma_i(\theta_i), \theta_{-i})) \tag{2.6}$$

Observe that ex post incentive compatibility is in fact equivalent to strategyproofness in our formulation, which is characterized by "private values": each agent's utility depends only on the outcome and *his own* type (not the other agents'). If truthfulness is a utility-maximizing strategy when other agents are truthful *no matter what their* *types*, then it is utility-maximizing however they choose to report types (truthfully or not).[12]

A clear weakening of the truth-revealing property occurs when we move from dominant strategy or Nash equilibrium to Bayes-Nash equilibrium. In an *incentive compatible* mechanism no agent can *expect* to benefit from deviating from truthfulness—so long as other agents don't—given his beliefs about the types of other agents.

Definition 2.13 (incentive compatibility). *A direct mechanism (f, T) is incentive compatible if and only if truthfulness is a Bayes-Nash equilibrium for every agent, i.e., if given common knowledge agent beliefs $b_1(\theta_{-1}), \ldots, b_n(\theta_{-n})$,*

$$\begin{aligned}\forall \theta \in \Theta, i \in I, \sigma_i, \quad & \mathbb{E}_{b_i(\theta_{-i})}[u_i(\theta_i, f(\theta_i, \tilde{\theta}_{-i}), T(\theta_i, \tilde{\theta}_{-i})) \geq \\ & \mathbb{E}_{b_i(\theta_{-i})}[u_i(\theta_i, f(\sigma_i(\theta_i), \tilde{\theta}_{-i}), T(\sigma_i(\theta_i), \tilde{\theta}_{-i}))] \qquad (2.7)\end{aligned}$$

I will frequently use the term **truthful** to describe a mechanism that is strategyproof or incentive compatible, with the intended equilibrium concept clear from context. For instance, to say that a mechanism is "truthful in dominant strategies" is equivalent to saying it is strategyproof.

Participation properties

Perhaps as important as the incentives a mechanism provides to agents that have decided to participate is providing the incentives that make the agents *want* to participate in the first place. Of course in settings in which agents have no choice but to participate this is not a concern, but often agents will have a choice, and it is important to design mechanisms in which agents will generally be better off when they make the choice to take part than they would be if they sat out. The term used for this concept is individual rationality (IR).

In fact this property is often essential to obtaining implementation of an efficient outcome. Consider a single-item allocation domain. The efficient choice function allocates the item to the agent with highest value, but it is impossible to achieve this if the agent with highest value has not even decided to participate in the mechanism.

For simplicity I will assume here that agents that don't participate in a mechanism obtain utility 0, and thus in order for a mechanism to meet the "participation constraint" it must yield non-negative payoff to each agent.[13] We will consider both

[12] In so-called "common values" environments where one agent's value depends on the type (or "signal") of another agent, this will not be the case (see, e.g., [Milgrom and Weber, 1982]). Besides in Chapter 7, Section 7.3, this thesis considers only private values environments.

[13] Technically, if an agent would receive utility x from not participating, the individual rationality constraint demands that the agent obtains utility at least x from participating. But my formulation is without loss of generality, as u_i can be considered utility normalized to each agent's outside option, i.e., the utility the mechanism yields for i minus the utility i would obtain if he did not participate.

"guaranteed" and "in-expectation" versions of this property.

Definition 2.14 (ex post individual rationality). *A direct mechanism (f, T) is ex post individual rational if and only if*

$$\forall \theta \in \Theta, i \in I, \;\; u_i(\theta_i, f(\theta), T(\theta)) \geq 0 \tag{2.8}$$

In an ex post individual rational mechanism each agent has a strategy (truthfulness) that is guaranteed to yield him non-negative utility, regardless of the other agents' types or strategies they play.

A weaker notion is interim individual rationality, in which an agent does not *expect* to be worse off from having participated, given knowledge of his own type and beliefs about others' types. A still weaker notion is ex ante individual rationality, in which each agent i—prior to realization of *all* agent types, including his own—*expects* that his utility will be non-negative from participating truthfully; in actuality his utility may still end up being negative.

Definition 2.15 (interim individual rationality). *A direct mechanism (f, T) is interim individual rational if and only if, given common knowledge agent beliefs $b_1(\theta_{-1}), \ldots, b_n(\theta_{-n})$ about other agents' types ($\tilde{\theta}_{-i}$),*

$$\forall i \in I, \theta_i \in \Theta_i, \;\; \mathbb{E}_{b_i(\theta_{-i})}[u_i(\theta_i, f(\theta_i, \tilde{\theta}_{-i}), T(\theta_i, \tilde{\theta}_{-i}))] \geq 0 \tag{2.9}$$

Definition 2.16 (ex ante individual rationality). *A direct mechanism (f, T) is ex ante individual rational if and only if, prior to realization of agent types and given common knowledge agent beliefs $b_1(\theta), \ldots, b_n(\theta)$ about the complete type profile that will be realized ($\tilde{\theta}$),*

$$\forall i \in I, \;\; \mathbb{E}_{b_i(\theta)}[u_i(\tilde{\theta}_i, f(\tilde{\theta}), T(\tilde{\theta}))] \geq 0 \tag{2.10}$$

Budget properties

As we will soon see, successful mechanisms achieve equilibrium implementation of desirable social choice functions by aligning the interests of each agent with implementation of the choice function specified by the mechanism; this alignment is achieved via the mechanism's transfer payment scheme. But in searching for an effective transfer function, we will often have to consider the *budgetary* properties it yields in addition to the incentive properties. Most importantly, it will frequently be the case that a mechanism designer does not have access to an external source of funds that can be used to bankroll the mechanism, and thus net payments made to the agents must be non-positive. In other words the payment scheme cannot run a deficit:

Definition 2.17 (no-deficit). *A direct mechanism* (f, T) *has the no-deficit property if and only if*

$$\forall \theta \in \Theta, \ \sum_{i \in I} T_i(\theta) \leq 0 \tag{2.11}$$

The no-deficit property is sometimes called *weak budget-balance*. If both individual rationality and no-deficit constraints are satisfied, then all agents and the center will want to participate in the mechanism; no one will be hurt from participating and there is the chance of a utility gain. Then the remaining question is what portion of the value (or "surplus") yielded from a selected outcome is kept by the agents, and what portion is transferred to the center. The value transferred to the center is called the *revenue*:

Definition 2.18 (revenue). *In a mechanism with transfer function* T*, when the reported agent type profile is* $\hat{\theta}$ *the revenue is* $-\sum_{i \in I} T_i(\hat{\theta})$*.*

In some scenarios it will be desirable that revenue equals 0, i.e., that the mechanism is *strongly budget-balanced*:

Definition 2.19 (strong budget-balance). *A mechanism with transfer function* T *has the strong budget-balance property if and only if*

$$\forall \theta \in \Theta, \ \sum_{i \in I} T_i(\theta) = 0 \tag{2.12}$$

A mechanism with this property is no-deficit, and at the same time leaves the entire utility from the outcome in the hands of the agents. Transfer payments effectively move money around between the agents, but do not involve any net transfer from or to the center.

2.2.4 The Gibbard-Satterthwaite theorem and quasilinear utility

Given this framework for analyzing mechanisms, natural first questions to ask are: what kind of choice functions can be implemented in the strong dominant strategy equilibrium? Can we implement social-welfare maximizing choice functions? Or choice functions that maximize the welfare of an arbitrarily selected subset of agents?

Surprisingly, the answers are quite negative. When agent utility functions are unrestricted (i.e., each u_i can be an arbitrary mapping from a type, outcome, and transfer payment to a real-valued utility), without further assumptions there are essentially no interesting choice functions that can be implemented in dominant strategies. Specifically, the only implementable choice functions are *dictatorial*: there is an agent i such that the outcomes chosen are *always* those most preferred by i.

Theorem 2.2 ([Gibbard, 1973; Satterthwaite, 1975]). *Consider an arbitrary social choice function f and assume that: 1) agent utilities are unrestricted, 2) there are at least 3 outcomes ($|O| \geq 3$), and 3) for each $o \in O$ there is a $\theta \in \Theta$ such that $f(\theta) = o$. If f is implementable in dominant strategies then f is dictatorial.*

But this is not the end of the story for mechanism design. In the face of this very negative result there are a few ways we can imagine proceeding. First, one could imagine looking to implementation concepts (e.g., Nash or Bayes-Nash) that are weaker than dominant strategy; I will not give that direction much attention here because I want to focus on the strongest implementation concept. The other alternative is to weaken the conditions of the theorem. Condition (2) is difficult to weaken because that would allow us to consider only very simple domains where an either/or decision between two choices is to be made; (3) is also difficult to attack, as it essentially says that no outcome is excluded from consideration independent of agent preferences. That leaves (1), and weakening this condition will provide a reasonable escape hatch from the Gibbard-Satterthwaite theorem.

We will consider a restricted set of utility functions, those in which each agent's utility can be represented as the sum of a value for the outcome selected by the mechanism and the transfer payment the mechanism makes to that agent. Such a utility function is called *quasilinear*.

Definition 2.20 (quasilinear utility function). *A utility function $u_i : \Theta_i \times O \times \Re \rightarrow \Re$ is quasilinear if and only if there exists a function $v_i : \Theta_i \times O \rightarrow \Re$ such that:*

$$\forall \theta_i \in \Theta, o \in O, x \in \Re, \;\; u_i(\theta_i, o, x) = v_i(\theta_i, o) + x \tag{2.13}$$

This type of utility function is quite natural: each agent obtains a particular amount of utility depending on what outcome is realized—which I will refer to throughout as his *value* v_i—and his utility will increase by however much money you give him (or will decrease by however much you *charge* him if his payment is negative).[14] Note that the utility obtained by an agent with a quasilinear utility function does not depend on the transfer payments that *other* agents receive. Thus we can write $u_i(\theta_i, f(\hat{\theta}), T_i(\hat{\theta}))$ to denote the utility agent i obtains when his true type is θ_i and type profile $\hat{\theta}$ is reported.

Primarily motivated by the Gibbard-Satterthwaite theorem, and because it is judged to be a reasonable model of agent utilities in many settings, the quasilinear assumption is practically omnipresent in mechanism design work. For this entire thesis I will assume agents have quasilinear utility; I will refrain from restating this assumption when presenting results, but it's always there.

[14] The value $v_i(\theta_i, o)$ of an agent i with type θ_i for an outcome o can be considered his "net gain" (independent of transfers) relative to the value he would obtain from not participating in the mechanism when there are outside options (see footnote 13).

2.3 The Groves class of mechanisms

We are now ready to look at specific mechanisms that achieve some of the desirable properties defined in the previous section. I will use the following notational shorthands (some of which will not come into play until later in the section):

- $\boldsymbol{v_{-i}(\theta_{-i}, o)}$: the value obtained by agents other than i for outcome $o \in O$ given type profile θ_{-i}, i.e., $\sum_{j \in I \setminus \{i\}} v_j(\theta_j, o)$.
- $\boldsymbol{v(\theta, o)}$: the social value obtained by the agents for outcome $o \in O$ given type profile θ, i.e., $\sum_{i \in I} v_i(\theta_i, o)$.
- $\boldsymbol{f^*(\theta_{-i}) = \arg\max_{o \in O} v_{-i}(\theta_{-i}, o)}$: the outcome in O chosen by a social choice function that maximizes the welfare of the group of agents excluding i. i is still presumed to be "in the system" (so the outcome space does not change), but his valuation function is disregarded.

The *Groves* class of mechanisms—proposed by Vickrey [1961], Clarke [1971], and Groves [1973]—constitutes the foundation on which practically all of dominant strategy implementation mechanism design rests.

Definition 2.21 (Groves class of mechanisms). *A direct mechanism* (f, T) *is a Groves mechanism if and only if:*[15]

- $\forall \theta \in \Theta,\ f(\theta) \in \arg\max_{o \in O} v(\theta, o)$ *(i.e., it executes* f^**)*
- $\forall i \in I$*, there is a function* $h_i : \Theta_{-i} \rightarrow \Re$ *such that* $\forall \theta \in \Theta$*,*

$$T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - h_i(\theta_{-i}) \tag{2.14}$$

A Groves mechanism chooses a social-welfare maximizing outcome (according to agent reported types), and *pays* each agent i the value that other agents report obtaining for the selected outcome (which we will call the "Groves payment"), minus some quantity that is completely independent of i's report (the "charge"). Note that this definition characterizes a *class* of mechanisms rather than just one, as there are many (in fact an infinite number of) ways of defining the charge function h_i for each agent i.

Every mechanism in the Groves class is *strategyproof.* When this is considered in light of the fact that Groves mechanisms choose an efficient outcome according

[15] I give the definition for the restricted context of *efficient* mechanism design. In fact there are variants in which the choice function maximizes a weighted sum of agent valuations, with transfer functions modified accordingly.

to agent reports, one can see that every Groves mechanism implements an efficient social choice function in dominant strategies.

Theorem 2.3. *Every Groves mechanism is truthful and efficient in dominant strategies.*

Proof. Assume for contradiction the existence of a Groves mechanism (f^*, T) that is *not* truthful in dominant strategies. Since (f^*, T) is a Groves mechanism, for each $i \in I$, $T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - h_i(\theta_{-i})$ for some $h_i : \Theta_{-i} \rightarrow \Re$ (h_i is the only part of the mechanism left unspecified). If (f^*, T) is not truthful in dominant strategies, then there exists an $i \in I$, $\theta_{-i} \in \Theta_{-i}$, and $\theta_i, \theta_i' \in \Theta_i$ such that if i's true type is θ_i he is better off reporting θ_i' when other agents report θ_{-i}, i.e.:

$$v_i(\theta_i, f^*(\theta_i, \theta_{-i})) + v_{-i}(\theta_{-i}, f^*(\theta_i, \theta_{-i})) - h_i(\theta_{-i}) \tag{2.15}$$

$$< v_i(\theta_i, f^*(\theta_i', \theta_{-i})) + v_{-i}(\theta_{-i}, f^*(\theta_i', \theta_{-i})) - h_i(\theta_{-i}), \tag{2.16}$$

which implies that:

$$v(\theta, f^*(\theta)) < v(\theta, f^*(\theta_i', \theta_{-i})) \tag{2.17}$$

But then consider a choice function f' such that $f'(\theta_i, \theta_{-i}) = f^*(\theta_i', \theta_{-i})$ and, $\forall \theta \in \Theta \setminus \{\theta_i'\}$, $f'(\theta) = f^*(\theta)$. By equation (2.17),

$$v(\theta, f^*(\theta)) < v(\theta, f'(\theta)), \tag{2.18}$$

which contradicts the fact that f^* is an efficient choice function. Thus each such (f^*, T) is truthful in dominant strategies. Then, since f^* is the efficient choice function, each is also efficient in dominant strategies. □

In a Groves mechanism each agent obtains some value intrinsically for realization of the chosen outcome, and also is payed the total value the other agents claim to obtain; the sum of these two quantities is *exactly* what the center is optimizing (when the agent is truthful) via an efficient choice function. The center then also *charges* each agent some quantity (via the h_i function), but this is completely beyond the agent's control and thus does not influence his incentives. The payments effectively align each agent's interests with the center's interest, i.e., maximization of social welfare.

Consider a single-item allocation setting in which there are 4 agents—1, 2, 3, and 4—with values 10, 8, 6, and 4, respectively, for the good. The agent valuation functions[16] can conveniently be represented in matrix format,[17] where each row corresponds to an outcome (allocation of the item to one particular agent) and each column corresponds to a different agent's valuation. Let o_i denote the outcome in which agent i is allocated the resource.

	v_1	v_2	v_3	v_4
o_1	10	0	0	0
o_2	0	8	0	0
o_3	0	0	6	0
o_4	0	0	0	4

Table 2.4: Tabular representation of 4 agent's valuation functions for a single-item allocation problem with 4 outcomes. Agents have 0 value for outcomes in which they are not allocated the good.

Consider the "**basic-Groves mechanism**" in which the $h_i(\theta_{-i})$ charge term is defined to be 0 for every i and for every θ. This mechanism pays each agent the value other agents report obtaining, but charges them nothing. Applying basic-Groves to the example in Table 2.4, the item will be allocated to agent 1, and then agents 2, 3, and 4 will each receive a payment of 10. Each agent's total utility—in dominant strategy equilibrium—equals the total value to the group from allocation of the item, i.e., 10. Note also that the basic-Groves mechanism is trivially *ex post individual rational* whenever the social value of each outcome is non-negative. When agent i truthfully reports his type θ_i,

$$u_i(\theta_i, f^*(\theta), T_i(\theta)) = v_i(\theta_i, f^*(\theta)) + v_{-i}(\theta_{-i}, f^*(\theta)) = v(\theta, f^*(\theta)) \tag{2.19}$$

While the basic-Groves mechanism is successful from an incentives and individual rationality perspective, note that it fails severely in its budgetary properties. In the example above, the center must pay out a total of 30, which is 3 times the total value obtained from realization of the outcome. Most often a mechanism that runs this kind of deficit will be completely infeasible. Ideally a mechanism should provide the incentive and IR properties of basic-Groves and also have the *no-deficit* property; that way it can be implemented without requiring some external source of funding to bankroll its execution. Later in this section we will see the difficulty in accomplishing

[16] Note the important distinction between a "valuation function" and a "utility function". For an agent i, the former corresponds to v_i, while the latter is u_i (and equals v_i plus the transfer payment).

[17] Note that this matrix format is completely distinct from the "normal form" matrix representation I used to represent 2-player games.

this. First, and in order to guide our search, we note an extraordinary fact discovered by Green & Laffont in 1977.

2.3.1 Uniqueness of the Groves class

We've seen that Groves mechanisms are all truthful and efficient in dominant strategies. One might wonder whether there are other payments schemes that lead to this same property. The answer is no; the Groves mechanisms are in fact unique among strategyproof mechanisms that choose outcomes to maximize social welfare.

Theorem 2.4 (Groves uniqueness [Green and Laffont, 1977]). *For an unrestricted type space,*[18] *a direct mechanism (f, T) is truthful and efficient in dominant strategies if and only if it is a Groves mechanism.*

A couple years after publication of this result, Holmström followed with a significant strengthening of the theorem. He discovered the result still holds even if we consider significant restrictions on agent valuation functions. One may have thought that the uniqueness of the Groves class rested on the possibility of some bizarre utility function that would never be encountered in the real world. Holmström showed this is not the case.

Definition 2.22 (smoothly connected type space). *A type space Θ_i is smoothly connected if and only if, for any two valuation functions v_i and v_i' admitted by Θ_i, one can be differentiably deformed into the other.*

Theorem 2.5 ([Holmstrom, 1979]). *For any smoothly connected type space, a direct mechanism (f, T) is truthful and efficient in dominant strategies if and only if it is a Groves mechanism.*

These remarkable results are especially important because—like the revelation principle—they allow us to significantly narrow our focus in searching for mechanisms that meet certain criteria, e.g., individual rationality and budget properties. Taking dominant strategy implementation of efficient outcomes as a hard constraint, our freedom is thus limited to defining the agent-independent charge function h_i for each agent i.

It seems that essentially every type space encountered in practice is smoothly connected. Smoothly connected spaces include, for instance, those found in typical allocation problems where agents can have any real-number valuation for any bundle of goods. As Holmström [1979] notes, this result shows that "for all practical purposes" one must be content with the Groves class.

We will now see that there is a difficulty in simultaneously achieving the efficiency, individual rationality, and budget properties we want. But it will be a temporary

[18] I.e., where each agent's value function is an arbitrary mapping from outcomes to real numbers.

roadblock, as in this case there will be another reasonable assumption under which positive results become possible.

2.3.2 Tension between efficiency, IR, and budget-balance

Myerson & Satterthwaite [1983] showed that even if we weaken our solution concept to Bayes-Nash equilibrium, it is impossible to specify a mechanism that meets the typically desired efficiency, participation, and budget properties.[19]

Theorem 2.6 (Myerson-Satterthwaite). *For an unrestricted type space, there exists no mechanism that is truthful and efficient in Bayes-Nash equilibrium, interim individual rational, and no-deficit.*

As in the case of the Gibbard-Satterthwaite theorem, just when things seem to be getting off the ground we are faced with a very negative result. But, again, we will find that the theorem's sweep can be evaded by specifying (or "observing") a restriction on agent valuation functions that applies in many important domains. The following condition states that each agent's value for the outcome that would be optimal if his interests were deemed irrelevant is non-negative. For any $\theta \in \Theta$ and $i \in I$, recall that $f^*(\theta_{-i}) = \arg\max_{o \in O} v_{-i}(\theta_{-i}, o)$, i.e., $f^*(\theta_{-i})$ denotes the outcome (from complete outcome set O) selected by a choice function that is efficient when i is ignored.

Definition 2.23 (no negative externalities). *The no negative externalities property holds when, $\forall i \in I, \theta \in \Theta$, $v_i(\theta_i, f^*(\theta_{-i})) \geq 0$.*

A sufficient (but not necessary) condition for this property to hold is that each agent's value for *every* outcome is non-negative (no outcomes "hurt"). Note that this condition does *not* necessarily hold in, e.g., exchange settings in which an agent i brings goods that he owns to the mechanism; it may (and in fact probably will) be the case that allocating i's goods to other agents is optimal for those other agents, though it certainly leaves i worse off.

2.3.3 The VCG mechanism

The VCG mechanism (named for Vickrey [1961], Clarke [1971], and Groves [1973]) is the most famous mechanism in the Groves class, and for good reason. We will see that with only the above assumption it achieves strategyproofness, efficiency, ex post individual rationality, and no-deficit simultaneously, and that there is a sense in which it is *unique* in this achievement. VCG selects an efficient outcome according to agent

[19] Hurwicz [1975] earlier showed the same thing for dominant strategy implementation; the Myerson-Satterthwaite theorem thus strengthens Hurwicz' theorem.

reports, pays each agent the Groves payment, and charges each agent i the value the other agents *could have achieved if i's interests were disregarded.*[20]

Definition 2.24 (VCG mechanism). *The VCG mechanism is a direct mechanism (f^*, T) where, $\forall i \in I$ and $\theta \in \Theta$:*

$$T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) \tag{2.20}$$

Theorem 2.7. *The VCG mechanism is truthful and efficient in dominant strategies.*

Proof. The theorem follows immediately from the fact that Groves mechanisms are truthful and efficient in dominant strategies, combined with the observation that VCG is a Groves mechanism. VCG is in the Groves class since each agent i's charge term, $v_{-i}(\theta_{-i}, f^*(\theta_{-i}))$, is independent of his reported type. □

Agent utilities under VCG have an intuitively appealing property: each agent obtains (in equilibrium) utility equal to his *marginal contribution to social welfare.* So, for instance, in a problem in which the center seeks to allocate a single item efficiently, if agent i has the highest value, 10, and the second-highest value is 8, agent i will have contributed value 2 to social welfare and this will be his net utility under VCG.

Theorem 2.8. *The VCG mechanism is ex post individual rational when the no negative externalities property holds.*

Proof. Consider any agent i that reports his true type θ_i, and let θ_{-i} denote any reported type profile for agents other than i. Under VCG i's utility will equal:

$$v_i(\theta_i, f^*(\theta)) + v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) \tag{2.21}$$

By the no negative externalities property $v_i(\theta_i, f^*(\theta_{-i})) \geq 0$, so we have that the above is

$$\begin{aligned}
&\geq v_i(\theta_i, f^*(\theta_i)) + v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) - v_i(\theta_i, f^*(\theta_{-i})) && (2.22)\\
&= v(\theta, f^*(\theta)) - v(\theta, f^*(\theta_{-i})) && (2.23)\\
&\geq 0 && (2.24)
\end{aligned}$$

[20] The reader should note that this definition is somewhat non-standard. The typical definition would charge i the value other agents could have achieved if *i were not present in the system.* The distinction is that certain outcomes in O are not available without i (for instance, if i has brought goods to sell to other agents). Defining the mechanism this way will allow us to make a minimal number of assumptions (just 1) in order to obtain the budget and participation properties we seek. When the space of possible outcomes is independent of the agent population, the two definitions coincide.

The final inequality holds because f^* is efficient. This shows that i's expected utility is non-negative, which is exactly the condition for individual rationality. Since we selected θ arbitrarily, the property holds *ex post*. □

Theorem 2.9. *The VCG mechanism is no-deficit.*

Proof. Consider arbitrary reported type profile θ. By definition of $f^*(\theta_{-i})$:

$$\forall i \in I,\ v_{-i}(\theta_{-i}, f^*(\theta_{-i})) \geq v_{-i}(\theta_{-i}, f^*(\theta)) \tag{2.25}$$

From this it follows that:

$$\sum_{i \in I} \Big(v_{-i}(\theta_{-i}, f^*(\theta_{-i})) - v_{-i}(\theta_{-i}, f^*(\theta)) \Big) \geq 0 \tag{2.26}$$

This expression is exactly the revenue under VCG. □

So VCG is truthful, efficient, and ex post IR, and also meets the basic criterion of no-deficit, which is absolutely essential to feasibility of a mechanism when there is no external budget. But in fact we can say more about its budgetary properties: I will now provide a proof that the VCG mechanism is *revenue maximizing*:[21]

Definition 2.25 (revenue maximizing). *Given a specified type space* Θ*, a mechanism* (f, T) *is revenue maximizing in mechanism space* M *if and only if* $(f, T) \in M$ *and,* $\forall \theta \in \Theta$*, there is no mechanism* $(f', T') \in M$ *such that* $T'(\theta) < T(\theta)$*.*

I will use the term "0-value admitting" to refer to a type space in which $\forall i \in I, \exists \theta_i \in \Theta_i$ s.t. $v_i(\theta_i, o) = 0, \forall o \in O$. For instance, a single-item allocation domain is 0-value admitting if it is *not* known a priori that all agents have strictly positive value for the item.

Theorem 2.10. *For any smoothly connected 0-value admitting type space, the VCG mechanism is revenue maximizing among all mechanisms that are truthful and efficient in dominant strategies and ex post individual rational.*

Proof. By Theorem 2.4, we know the revenue maximizing mechanism with these properties is a Groves mechanism. Consider any Groves mechanism (f^*, T) such that, for some $\theta \in \Theta$, revenue is greater than under VCG. Then for some $i \in I$:

$$\begin{aligned} T_i(\theta) &= v_{-i}(\theta_{-i}, f^*(\theta)) - h_i(\theta_{-i}) && (2.27)\\ &< v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) && (2.28) \end{aligned}$$

[21] Krishna & Perry [1998] demonstrate that VCG maximizes *expected* revenue in an incomplete information context among all efficient, IC, and IR mechanisms; the result bears some relationship to the "optimal auction" results of [Myerson, 1981], though Krishna & Perry's restriction to efficient mechanisms allows them to achieve results for *multi-dimensional* types. I show here that VCG maximizes revenue *for any type profile* when dominant strategy truthfulness, efficiency, and ex post IR are required.

This implies that:

$$h_i(\theta_{-i}) > v_{-i}(\theta_{-i}, f^*(\theta_{-i})) \tag{2.29}$$

But consider a $\overline{\theta}_i$ such that $f^*(\overline{\theta}_i, \theta_{-i}) = f^*(\theta_{-i})$ and $v_i(\overline{\theta}_i, f^*(\overline{\theta}_i, \theta_{-i})) = 0$ (for instance, if $v_i(\overline{\theta}_i, o) = 0$ for all $o \in O$ this will hold). Then when the true type profile is $(\overline{\theta}_i, \theta_{-i})$,

$$\begin{aligned}
&u_i(\overline{\theta}_i, f^*(\overline{\theta}_i, \theta_{-i}), T_i(\overline{\theta}_i, \theta_{-i})) && (2.30)\\
&\qquad = v_i(\overline{\theta}_i, f^*(\overline{\theta}_i, \theta_{-i})) + v_{-i}(\theta_{-i}, f^*(\overline{\theta}_i, \theta_{-i})) - h_i(\overline{\theta}_i, \theta_{-i}) && (2.31)\\
&\qquad = v_{-i}(\theta_{-i}, f^*(\overline{\theta}_i, \theta_{-i})) - h_i(\overline{\theta}_i, \theta_{-i}) && (2.32)\\
&\qquad = v_{-i}(\theta_{-i}, f^*(\theta_{-i})) - h_i(\overline{\theta}_i, \theta_{-i}) && (2.33)\\
&\qquad < v_{-i}(\theta_{-i}, f^*(\theta_{-i})) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) = 0 && (2.34)
\end{aligned}$$

Thus the mechanism is not ex post individual rational, and the theorem follows. □

Note that since VCG is revenue maximizing for *any* smoothly connected valuations domain (as long as it admits a valuation of 0 by each agent for each outcome), it is revenue maximizing regardless of whether the domain has the no negative externalities property or some other property that yields ex post IR. So when revenue is the goal, within the context of an efficient and IR mechanism VCG is the best one can do. But revenue is not always a good thing; in particular, from the perspective of the agents payments made to the center simply detract from their welfare. Trying to *minimize* revenue rather than maximize it is the topic of the next chapter. But here in this section on VCG I'll just give mention of one of the results we'll see there: if we don't make further assumptions about agent valuation functions, VCG is also *revenue minimizing* among dominant strategy truthful, efficient, ex post individual rational, and no-deficit mechanisms.

2.4 Strong budget balance

2.4.1 Tension between efficiency and budget-balance

Recall that the primary role of payments in mechanism design is often not to shuffle money to or from the center; rather it is to align incentives. In fact in certain cases any net transfer of money from the center to the agents will be considered unacceptable (no-deficit is required), as will transfers from agents to the center. That is, *strong budget balance* is required. Unfortunately if we want dominant strategy implementation of an efficient choice function, then strong budget balance is not possible in general.

Theorem 2.11 ([Green and Laffont, 1979]). *For an unrestricted type space, there exists no strongly budget balanced mechanism that implements an efficient choice function in dominant strategies.*

2.4.2 The AGV mechanism

In light of Theorem 2.11, in a scenario in which strong budget balance is required we must weaken the implementation solution concept. If we consider Bayes-Nash implementation there is a strongly budget balanced mechanism, though it is only *ex ante* individual rational and requires the existence of a *common prior over agent types.* The AGV mechanism of Arrow [1979] and d'Aspremont & Gerard-Varet [1979] chooses an efficient outcome according to agent reports, and pays each agent the "expected externality" he imposes on other agents, minus a constant. To describe the mechanism we will use notation "$\text{ESW}_{-i}(\theta_i)$" for the expected welfare that results for the other agents when agent i reports type θ_i, given the common prior beliefs about other agents' types $b(\theta_{-i})$.

Definition 2.26 (AGV mechanism). *The AGV mechanism is a direct mechanism (f^*, T) where, $\forall i \in I$ and $\theta \in \Theta$, given common prior beliefs $b(\theta_{-i})$ about the types of agents other than i:*

$$T_i(\theta) = ESW_{-i}(\theta_i) - \frac{1}{n-1} \sum_{j \in I \setminus \{i\}} ESW_{-j}(\theta_j), \text{ where} \tag{2.35}$$

$$\forall i \in I, \theta_i \in \Theta_i, \quad ESW_{-i}(\theta_i) = \mathbb{E}_{b(\theta_{-i})}[v_{-i}(\tilde{\theta}_{-i}, f^*(\theta_i, \tilde{\theta}_{-i}))] \tag{2.36}$$

The "payment" term in each agent's transfer plays the same role as the Groves payment in a Groves mechanism; the difference is that here it is the *expected* welfare the other agents will obtain, given a type report for a particular agent. Then the "charge" term is defined in a way that exactly balances the budget.

Theorem 2.12. *The AGV mechanism is truthful and efficient in Bayes-Nash equilibrium, ex ante individual rational, and strongly budget balanced.*

I will now go through the workings of AGV on a simple example. Imagine a single-item allocation scenario in which there are 3 agents: 1, 2, and 3, and there are three possible values that each agent may have for the item: 0, 10, and 20. The prior distribution over values for each agent is represented in Table 2.5.

Now assume that types are ultimately realized in the following fashion: agent 1's value for the item is 10, agent 2's is also 10, and agent 3's is 20. The AGV mechanism will allocate the item to agent 3, and then make payments. In order to construct those payments we must compute $\text{ESW}_{-i}(\theta_i)$ for each i. For simplicity here, assume that

	0	10	20
$P(v_1 = \cdot)$	0	.2	.8
$P(v_2 = \cdot)$	0	.8	.2
$P(v_3 = \cdot)$	.8	.1	.1

Table 2.5: Tabular representation of the distribution over possible values for the item, for each agent. Agent 1's value for the item is 10 with probability .2 and 20 with probability .8, etc.

the choice function breaks ties in favor of agent 1 over agents 2 and 3, and agent 2 over agent 3. In the truthful equilibrium we have:

$$\text{ESW}_{-1}(\theta_1) = (0.2 \cdot 1 \cdot 20) + (0.1 \cdot 0.8 \cdot 20) = 5.6 \tag{2.37}$$

$$\text{ESW}_{-2}(\theta_2) = (0.2 \cdot 0.9 \cdot 10 + 0.8 \cdot 1 \cdot 20) + (0.1 \cdot 0.2 \cdot 20) = 18.2 \tag{2.38}$$

$$\text{ESW}_{-3}(\theta_3) = (0.8 \cdot 1 \cdot 20) + (0.2 \cdot 0.2 \cdot 20) = 16.8 \tag{2.39}$$

We can then compute payments. Each agent's value for the outcome, transfer payment, and ultimate utility (the sum of these two things) is presented in Table 2.6. Note that the transfers sum exactly to 0 (and would regardless of what types the agents reported).

	v_i	T_i	u_i
Agent 1	0	$5.6 - \frac{1}{2} \cdot (18.2 + 16.8) = -11.9$	-11.9
Agent 2	0	$18.2 - \frac{1}{2} \cdot (5.6 + 16.8) = 7$	7
Agent 3	20	$16.8 - \frac{1}{2} \cdot (5.6 + 18.2) = 4.9$	24.9

Table 2.6: Tabular representation of each agent's value for the outcome, transfer payment received, and total utility in the truthtelling Bayes-Nash equilibrium.

This example illustrates how things can go wrong in the AGV mechanism. The payment scheme is set up such that, in expectation, each agent will be better (or at least as well) off playing than not playing. But in practice agents can end up quite worse off, as is the case with agent 1 in the example (so the mechanism is *not* ex post IR). Agent 1 is charged a large quantity because, in expectation, he will receive the item and derive significant value from it. This charge must be the same regardless of the ultimate realization of his true type (for incentive compatibility) and thus can't be adjusted when things vary from what was expected. Note also that if agent 1 had prior beliefs that did not correspond to the “common prior” he could potentially gain. For instance if he was confident that agent 3’s value would be 20, he could have reported value 0 and received a larger payment (so the mechanism is *not* strategyproof).

Chapter 3

Redistribution mechanisms

Synopsis[*]

This thesis is motivated by the goal of implementing decision-making procedures that have good social welfare properties. The starting point of this chapter is the observation that in many scenarios payments agents are required to make in a mechanism *detract from social welfare* by transferring value out of the agents' hands. I propose a new mechanism that, in allocation environments and others, allows the agents to retain the vast majority of value obtained from the chosen outcome. This is in stark contrast to the ubiquitous VCG mechanism, which in the same environments often requires that the majority of value is payed to the center.

3.1 Motivation and background

In the previous chapter we saw that mechanism design obtains social-welfare maximizing equilibrium outcomes via the execution of specific transfer payments. The right transfer payments align the interests of agents towards social welfare maximization, so they participate truthfully in order to enable the center to choose the efficient outcome. We saw that the Groves class of mechanisms fully characterizes the set of mechanisms that yield efficient outcomes in dominant strategies (the strongest of our solution concepts).

The VCG mechanism—an instance of the Groves class—has particularly nice properties in addition to efficiency; in many domains it is both ex post individual rational (IR) (guaranteeing agents won't be worse off from participating) and no-deficit (never requiring an external budget for implementation). On top of that it yields *the most revenue* possible for a mechanism that is ex post IR and efficient in dominant strate-

[*] Many of the results in this chapter first appeared in a paper titled "Optimal Decision-making with Minimal Waste: Strategyproof Redistribution of VCG Payments", which appeared in the proceedings of the AAMAS 2006 conference [Cavallo, 2006b].

gies. For these reasons VCG is by far the most famous and well-regarded mechanism there is.

The impetus for the work in this chapter is that the revenue properties of VCG, though in some cases sought after, are in many other circumstances a very bad thing. How can this be? Well, first we must think about exactly what it means for a mechanism to have high revenue. Sometimes it's clear: if I own a house and decide to auction it off via the VCG mechanism (a Vickrey auction in this simple setting), the payments that agents (bidders for the house) must make will come to me in exchange for the house. Revenue, for me, would be quite a good thing—the more the merrier.

But consider a somewhat different scenario: what if it there are 3 brothers whose parents have moved south to Florida from New York City and decided to leave the family home to their sons? Each son is given an equal "share" by the parents, but the house is only big enough for one son and his immediate family. How are the brothers to decide who should get to live in the house? Well, there are many factors that could come in to play, but for our purposes let's just assume they decide (or their parents decide) that the son who would derive the most value from the house should get it. We know that the VCG mechanism would accomplish this goal, but it would require the son with the highest value for the house to pay the second-highest value for it. And to whom? Not to one of the brothers, but to a "center" that is external to the group. If the top two values are close to each other, practically none of the value from obtaining the house will be kept within the group of brothers. VCG doesn't look so good here.

One can easily think of other examples: government allocation of usage time on a publicly owned and high-priced piece of technology like a space telescope or supercomputer—the mandate is to "increase the public welfare", not to extract that welfare out of the hands of citizens; a municipality's choice of which neighborhood to build a public park in; a group of housemates or friends that jointly own an automobile and must decide who gets to use it on a given Friday night. This last example is portrayed in Figure 3.1, with the 4 friends' values for the car equal to 10, 8, 6, and 4. Scenarios like this are common—the center is present merely for organizational purposes, or is not present at all. The revenue that VCG generates is not desirable from anyone's perspective; it is merely a "cost of implementation" or, put another way, *waste*.

The ideal mechanism for this kind of scenario would have the desirable properties of VCG (truthfulness and efficiency in dominant strategies, ex post IR), and at the same time run neither a budget surplus nor deficit. As we will see, this *exact* budget-balance is not attainable. However, while it has previously been claimed that no improvement over VCG is possible (see, e.g., [Ephrati and Rosenschein, 1991]), I demonstrate here that this *is not the case* in a broad class of domains (e.g., allocation problems) where valuations have some basic structure. Short of strong budget balance we will take as our goal finding a mechanism that *minimizes* the payments that agents must make to the center. I cast this task as "redistribution" of revenue under VCG,

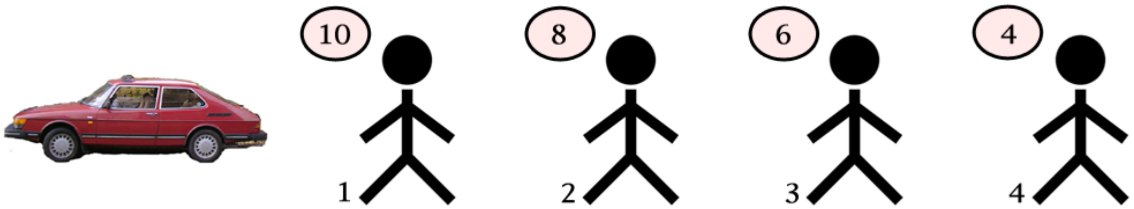


Figure 3.1: Which friend gets to use the jointly-owned car on Friday night?

which is a testament to the central place that VCG enjoys in the mechanism design pantheon.

To preview the results: I will show that in the case of unrestricted type spaces (i.e., when an agent's valuation v_i is an arbitrary mapping from outcomes to real values) *none* of the VCG revenue can be redistributed, i.e., VCG is simultaneously revenue maximizing and minimizing among efficient mechanisms. But that's just the beginning of the story. In fact it is quite rare in the real world that agent valuations have *no* structure. For instance consider the example in which researchers are competing for a time-slot to make observations with a government-owned space telescope—it is reasonable to assume that the researcher that is allocated usage of the telescope may obtain some value, but that the losers obtain 0 value. It turns out that this assumption alone will allows us to implement a mechanism that does vastly better than VCG in terms of net utility to the agents; in fact, as the number of agents participating in the mechanism grows, we will be able to come arbitrarily close to perfect budget balance regardless of the agents' valuations.

I will provide an exact characterization of when it is possible to improve on VCG (i.e., I will specify the property that agent type spaces must have), and will propose redistribution mechanism **RM**, which is applicable to arbitrary decision problems. Intuitively, the mechanism implements VCG and then returns a "redistribution payment" that is proportional to the *guarantee on revenue that would result under VCG independent of the agent's type report.* This mechanism maintains all of VCG's good properties and increases social welfare (decreases payments agents must make). Moreover, we will see that there is a sense in which the mechanism is *optimal.* When a strong fairness constraint is imposed, it redistributes the most revenue back to the agents that is possible—for *every* set of agent valuations—without violating the strong efficiency, IR, and no-deficit properties.

After showing that no improvement over VCG is possible in the unrestricted values setting, I will start by presenting the redistribution mechanism in the special setting of "all-or-nothing" (AON) domains, those in which each outcome yields non-zero reward for just a single agent (every agent gets either all the reward or none). This class of domains is a generalization of the single-item allocation domain, and all the examples

I've mentioned so far are AON examples. Here the redistribution mechanism has a particularly simple and elegant form, and the most redistribution is possible. I will then specify the mechanism for the general case (i.e., in a form applicable to arbitrary type spaces). I'll present the results of some numerical simulations demonstrating the success of the mechanism, and will discuss implementation of the mechanism when there is no center present who can coordinate decisions and payments.

3.1.1 Related work

Work in the area of designing budget-balanced mechanisms is relatively scarce; one contributing explanation for that fact is the negative result I'll present in the next section: it is impossible to get closer to budget balance than VCG in unrestricted values settings. The positive results in this chapter are due to observations about domains in which valuations naturally have some structure; the fact that structure makes redistribution possible is by no means an obvious insight.

Bailey [1997] is one who partially pursued this path, and should be credited (as far as I know) as the first to propose a dominant strategy efficient revenue redistributing mechanism. His approach is, essentially, to consider the revenue that would result if each agent were not present in the mechanism. This yields success in single-item allocation settings—where his mechanism and the one I propose in fact coincide—but is not generally applicable when no-deficit is taken as a hard constraint. Different from my approach is that Bailey focuses his analysis on achieving zero *expected* revenue, forfeiting the no-deficit guarantee. Porter et al. [2004] later also provide this mechanism for the basic single-unit allocation case, but cast it from a cost minimization rather than value maximization perspective, imagining the "imposition" of an undesirable task on subordinate agents (e.g., employees) that have private knowledge of their costs for completing the task.

Other work in striving for budget balance without exploiting structure in type spaces has—necessarily—sacrificed at least one of the following: dominant strategy implementation, ex post individual rationality, or no-deficit. In the previous chapter we saw one proposal for achieving strong budget-balance: the AGV mechanism [Arrow, 1979; D'Aspermont and Gerard-Varet, 1979]. AGV selects the efficient outcome according to reported types, but determines transfer payments based on a model of agent valuations that the center maintains. This mechanism is interesting in that it always leads to strong budget-balance; however, it is implementable only in Bayes-Nash (rather than dominant strategy) equilibrium, and if the center's model is produced via iterative execution, serious problems regarding incentive compatibility could arise. Moreover, AGV is only ex ante IR—instances in which an agent is worse off for participating are possible.

Parkes et al. [2001] describe a payment rule that approximates VCG and achieves strong budget-balance in exchange settings where VCG runs a deficit, though truthfulness (and thus efficiency) is sacrificed; the mechanism seeks to minimize the incentive to deviate subject to strong budget balance. Faltings [2004] formulates the problem more closely to the way I do here, though his approach also attains strong budget-balance at the expense of efficiency. His mechanism chooses the outcome that is social-welfare maximizing among a subset of agents, and distributes the VCG revenue among agents that are not part of that subset. In a similar vein, Feigenbaum et al. [2001] analyze the Shapley-value mechanism for sharing multicast transmission costs, which comes closest to the efficient outcome among all budget-balanced mechanisms for that domain, though their results suggest its implementation is computationally intractable.

My approach is significantly different than these—I characterize the extent to which budget-balance can be approximated in dominant strategies, without sacrificing ex post individual rationality, efficiency, or no-deficit guarantees at all. Since publication of the main results of this chapter ([Cavallo, 2006b]), Guo & Conitzer have extended the theory for the specific setting of multi-unit auction settings in a series of papers [2007; 2008a; 2008b; 2008c]. Notably, [Guo and Conitzer, 2007] is a worst-case analysis (without imposing the fairness constraint introduced in this chapter); the mechanism derived is essentially identical to one independently derived by Moulin [2007]. [Guo and Conitzer, 2008a] (related to [Faltings, 2004]) looks to make gains for agent welfare by considering inefficient outcomes. [Guo and Conitzer, 2008b] considers redistribution that leverages a prior distribution over agent valuations. [Guo and Conitzer, 2008c] provides a technique for taking a redistribution mechanism such as the one I propose here, and through an iterative process squeezing out more redistribution from the residual revenue left by the original mechanism. Finally, Hartline & Roughgarden [2008] do mechanism design for settings in which payments are impossible, but where money can be *burned*; there too the goal is to minimize value not retained within the group of agents.

3.2 Uniqueness of VCG

In the last chapter we saw that the VCG mechanism is truthful and efficient in dominant strategies, is ex post individual rational when the no negative externalities condition holds, and is no-deficit. Moreover, it is *revenue maximizing* among all mechanisms with these properties. Thus from the perspective of social-welfare maximization that I take in this chapter and more generally in this thesis, the VCG mechanism is the *worst* among all mechanisms with these good properties. It is natural to ask if we can do better, and the first answer is negative. Consider the following property a type space might satisfy:

Definition 3.1 (potential for universal relevance nullification (PURN)). *A type space Θ has the PURN property if and only if, $\forall i \in I$ and $\theta_{-i} \in \Theta_{-i}$, there exists a $\theta_i \in \Theta_i$ such that $\forall j \in I$ (including i), $f^*(\theta) = f^*(\theta_{-j})$.*

This is a "broadness" condition on a type space, the opposite of a restriction. In a domain that satisfies the PURN property, for any agent i and any profile of reports for the other agents, i could always potentially report a value that renders null the influence on the outcome of any single agent's report alone, even his own. This is a property that often *will not* hold, and as we will see, it is exactly in those settings where redistribution is possible. But in this section I present a strong negative result for type spaces in which there is not enough structure to preclude the PURN property.

Definition 3.2 (revenue minimizing). *Given a specified type space Θ, a mechanism (f, T) is revenue minimizing in mechanism space M if and only if $(f, T) \in M$ and, $\forall \theta \in \Theta$, there is no mechanism $(f', T') \in M$ such that $T'(\theta) > T(\theta)$.*

Theorem 3.1. *For any smoothly connected 0-value admitting type space that satisfies no negative externalities and PURN, the VCG mechanism is revenue minimizing among all mechanisms that are truthful and efficient in dominant strategies, ex post individual rational, and no-deficit.*

Proof. By Theorem 2.4, we know that the revenue minimizing mechanism with these properties is a Groves mechanism. Assume for contradiction that there is a no-deficit Groves mechanism (f^*, T) such that, for some $\theta \in \Theta$, revenue is *less* than under VCG, i.e., $\exists h_1, \ldots, h_n$ s.t.:

$$0 \geq \sum_{i \in I} T_i(\theta) = \sum_{i \in I} \Big(v_{-i}(\theta_{-i}, f^*(\theta)) - h_i(\theta_{-i}) \Big) \tag{3.1}$$

$$> \sum_{i \in I} \Big(v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) \Big) \tag{3.2}$$

Then for some $i \in I$,

$$T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - h_i(\theta_{-i}) \tag{3.3}$$

$$> v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})), \tag{3.4}$$

which implies that:

$$h_i(\theta_{-i}) < v_{-i}(\theta_{-i}, f^*(\theta_{-i})) \tag{3.5}$$

But consider a $\overline{\theta}_i$ such that $f^*(\overline{\theta}_i, \theta_{-i}) = f^*(\theta_{-i})$ and, $\forall j \in I \setminus \{i\}$, $f^*(\overline{\theta}_i, \theta_{-i}) = f^*(\overline{\theta}_i, \theta_{-i,j})$. Since Θ satisfies PURN we know such a $\overline{\theta}_i$ exists. This would hold, for instance, letting $j \in \arg\max_{k \in I \setminus \{i\}} v_k(\theta_k, f^*(\theta_{-i}))$, if $v_i(\overline{\theta}_i, f^*(\theta_{-j})) = v_j(\theta_j, f^*(\theta_{-i}))$ and $v_i(\overline{\theta}_i, o) = 0$ for all $o \neq f^*(\theta_{-j})$. We know (by Theorem 2.10) that an upper bound on the revenue generated by $f^*, T)$ is the revenue generated by VCG. For type profile $(\overline{\theta}_i, \theta_{-i})$ this revenue equals:

$$v_{-i}(\theta_{-i}, f^*(\overline{\theta}_i, \theta_{-i})) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) + \tag{3.6}$$

$$\sum_{j \in I \setminus \{i\}} \Big(v_{-j}(\theta_{-j}, f^*(\overline{\theta}_i, \theta_{-i})) - v_{-j}(\theta_{-j}, f^*(\overline{\theta}_i, \theta_{-i,j})) \Big) \; = \; 0 \tag{3.7}$$

Thus if $h_i(\theta_{-i})$ were $< v_{-i}(\theta_{-i}, f^*(\theta_{-i}))$ a deficit would result when i's type is $\overline{\theta}_i$. Then since we picked i and θ arbitrarily, h_i can never be defined as less than the charge that VCG prescribes, for any reported type profile. □

Theorems 2.10 and 3.1 together yield the following:

Corollary 3.1. *For any smoothly connected, 0-value admitting type space that meets the no negative externalities and PURN conditions, VCG is the only mechanism that is truthful and efficient in dominant strategies, ex post individual rational, and no-deficit.*

Note again that saying a type space satisfies PURN and is 0-value admitting is a statement about the broadness of the type space. Corollary 3.1 is essentially a negative result about what is possible in mechanism design, and stating the result in a way that applies to domains that are *not* completely unrestricted makes it *stronger*. In the next section we will see that by restricting valuation spaces by adding assumptions, we can sometimes design mechanisms with stronger properties than would hold for the unrestricted case. Of course an unrestricted type space (the extreme in broadness) satisfies PURN and is 0-value admitting, so we automatically get the weaker result:

Corollary 3.2. *For a type space that meets the no negative externalities condition but is otherwise unrestricted, VCG is the only mechanism that is truthful and efficient in dominant strategies, ex post individual rational, and no-deficit.*

3.3 Restricted type spaces

The VCG mechanism has great IR and no-deficit properties and always chooses an efficient outcome, but often requires agents to transfer much of the utility they gain to the center. In the face of Corollary 3.1, how do we proceed? Is there anything we can do? The answer turns out to be yes; in many real-world settings of great import and interest, agent valuations have structure that precludes the PURN property and can be exploited to design better mechanisms.

In general when we restrict the type space that we consider—i.e., assume agents have *fewer* possible types they could report—we can get *more* positive results. Intuitively, a restriction on type spaces amounts to restrictions on the ways that agents could potentially manipulate the system, and at the same time may exclude certain

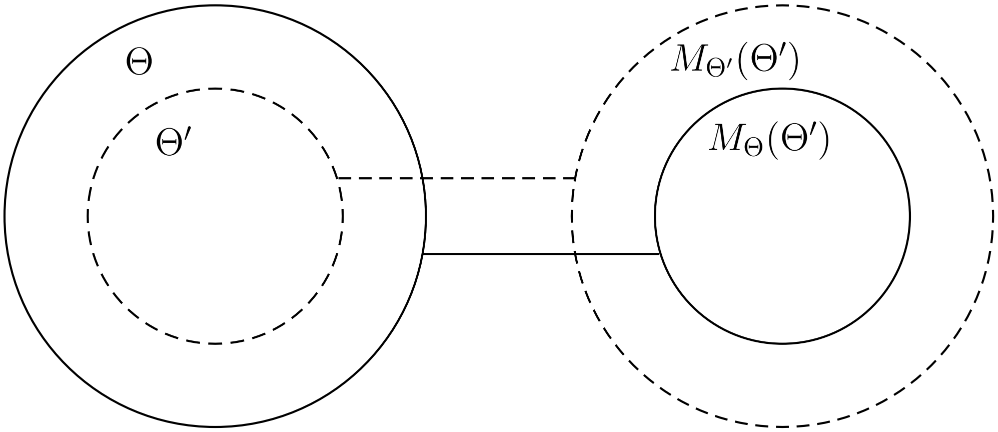


Figure 3.2: On the left, the solid circle represents one type space and the dashed circle represents a restriction on that space. Intuitively, the circles on the right can be considered the corresponding spaces of dominant strategy efficient no-deficit mechanisms. Crucially, there may be mechanisms that *become* dominant strategy efficient and no-deficit once a restriction is placed on the type space.

undesirable outcomes that could otherwise occur. Considering no-deficit and putting things more formally, for any two type spaces Θ and Θ' with $\Theta' \subseteq \Theta$, let $M_{\Theta}(\Theta')$ be the space of all dominant strategy efficient "mechanisms on Θ'" that satisfy no-deficit, i.e., all mappings from Θ' to outcomes and transfers that (combined with some mapping for $\Theta \setminus \Theta'$) are efficient in dominant strategies and never run a deficit given that the true type space is Θ. $M_{\Theta'}(\Theta')$ is weakly bigger than $M_{\Theta}(\Theta')$.

Theorem 3.2. *For all Θ, for all $\Theta' \subseteq \Theta$, $M_{\Theta'}(\Theta') \supseteq M_{\Theta}(\Theta')$.*

Proof. Assume for contradiction existence of $\Theta' \subseteq \Theta$ with $M_{\Theta'}(\Theta') \subset M_{\Theta}(\Theta')$. Consider arbitrary $m \in M_{\Theta}(\Theta')$ with $m \notin M_{\Theta'}(\Theta')$. $m \in M_{\Theta}(\Theta')$ entails that $\forall \theta \in \Theta$, given m, all agents $i \in I$ will truthfully report θ_i when that is their type and a deficit will not result. Then since $\Theta' \subseteq \Theta$, $\forall \theta' \in \Theta'$, given m, each agent $i \in I$ will truthfully report θ'_i when that is his type and no deficit will result. Thus $m \in M_{\Theta'}(\Theta')$, a contradiction. □

Figure 3.2 gives a graphical portrayal of the theorem. Importantly, the mechanism space set inclusion is *strict* in some cases—i.e., for some choices of Θ and Θ', there will be ways of computing transfers on types in Θ' that would not be strategyproof if the type space were $\Theta \supset \Theta'$. One such type of domain[1] we will see in this chapter is that

[1] I use the term "domain" throughout to refer to a specification of a type space. E.g., a smoothly connected valuations domain is one in which the type space is defined such that agent valuation spaces are smoothly connected.

	v_1	v_2	v_3	v_4
o_1	10	0	0	0
o_2	0	8	0	0
o_3	0	0	6	0
o_4	0	0	0	4

Table 3.1: Example of valuations in an AON domain. All values off the main diagonal are a priori known to be 0. Each agent may have a non-zero value only for the outcome associated with him (e.g., in which he is allocated the item in a single-item allocation problem).

of *single-item allocation*, where the decision to be made is in who gets to acquire what goods. In these domains (and in fact in a generalization of them) there are no-deficit mechanisms with better social-welfare properties than VCG, and these mechanisms only *become* strategyproof when the type space is restricted.

3.4 Redistribution in AON Domains

We will first look at the special case of *all-or-nothing (AON)* domains, in which the redistribution mechanism has a simple and elegant form.

Definition 3.3 (AON domain). *A type space Θ constitutes an AON domain if and only if, for every $\theta \in \Theta$ and $o \in O$, a maximum of one agent i has a value $v_i(\theta_i, o)$ for o that is non-zero.*

AON domains generalize single-item allocation problems in which an agent obtains non-zero value only if he is allocated the item. Consider the example portrayed in Figure 3.1, where 4 friends are deciding who gets to use the car. The friends' valuations can be represented in tabular form as in Table 3.1.

I identify o_i with the unique outcome that may yield positive value to agent i, and use the following short-hand notation:

- Let a_i denote the agent with the i^{th} highest reported value. (So in a social welfare-maximizing mechanism a_1 is the "winning" agent—the one that receives positive value from the selected outcome.)
- Let $\mathbb{V}_{a_i}$ denote the *true* value of a_i for the outcome he favors (i.e., $v_{a_i}(\theta_{a_i}, o_{a_i})$).
- Let $\hat{\mathbb{V}}_{a_i}$ denote the *reported* value of a_i for the outcome he favors.

Using this terminology, in an AON domain VCG chooses the outcome favored by agent a_1, and a_1 pays the center $\hat{\mathbb{V}}_{a_2}$. It is a dominant strategy for each agent a_i to report $\mathbb{V}_{a_i} = \hat{\mathbb{V}}_{a_i}$. The final distribution of value obtained from the outcome is as follows:

- $u_{a_1} = \mathbb{V}_{a_1} - \hat{\mathbb{V}}_{a_2}$.
- $u_{a_i} = 0$, $\forall i \neq 1$.
- The center obtains $\hat{\mathbb{V}}_{a_2}$.

For instance, in the car allocating example of Figure 3.1 and Table 3.1, the person with highest value for the car obtains utility 2, the center obtains 8, and all other agents obtain 0. This is a highly undesirable situation for the housemates who own the car—why should they want to pay someone 80% of the value they derive from using it?

The redistribution mechanism takes the second-highest value payment, made by the winning agent to the center, and redistributes a large portion of it back to the agents. This process is done in a careful way such that each agent's "redistribution payment" is independent of his type report. The mechanism has the following very simple form in AON domains:[2]

Definition 3.4 (Redistribution mechanism RM for AON domains). *The efficient outcome (i.e., the outcome preferred by agent a_1) is chosen, and the following transfers are executed:*

1. *The winning agent a_1 pays the center an amount equal to the second highest bid ($\hat{\mathbb{V}}_{a_2}$).*
2. *The center pays the winner and the second highest bidder an amount equal to the third highest bid divided by the number of agents ($\hat{\mathbb{V}}_{a_3}/n$), and pays all other agents the second highest bid divided by the number of agents ($\hat{\mathbb{V}}_{a_2}/n$).*

I will use notation Z_i to denote the redistribution payment made to agent i under **RM**.

Theorem 3.3. RM *for AON domains is truthful and efficient in dominant strategies.*

[2] Again, this mechanism coincides with Bailey's [1997] in this domain; in other domains it does not, but I present the AON case first to build intuition since it is simplest.

Proof. By Theorem 2.3, to prove the theorem it is sufficient to demonstrate that **RM** is a Groves mechanism. **RM** is identical to VCG (which is a Groves mechanism) modified by the addition of redistribution payments (step 2, above); so to show that **RM** is a Groves mechanism it is sufficient to show that each agent's redistribution payment is independent of his type report.

But observe that each agent i's redistribution payment Z_i can be described as *the second highest reported value amongst the other agents, divided by n.* In the case of the first and second highest bidders this will equal the third highest bid divided by n, and for all other bidders this will equal the second highest divided by n. □

No agent can influence his redistribution payment because it is defined as a function of only the *other* agents' reported values. To possibly reach more clarity on this, we can consider each agent's scenario in turn. Letting a_k below refer to the agent with the k^{th} highest bid under truthful reporting, we have:

Agent a_1 receives redistribution payment $\hat{\mathbb{V}}_{a_3}/n$. Over-reporting $\hat{\mathbb{V}}_{a_1}$ changes nothing. Under-reporting could put a_1 in the second or third position (beyond the third it's obvious that nothing could change). In the second position, he would still receive $\hat{\mathbb{V}}_{a_3}/n$. In the third he would receive $\hat{\mathbb{V}}_{a_3}/n$ as well, since the second position would then be held by the actual a_3.

Agent a_2 receives redistribution payment $\hat{\mathbb{V}}_{a_3}/n$. Over-reporting could move him to the first position, in which case his payment would be the same. Under-reporting could put him in the third position or beyond, but then the second position would be held by the actual a_3, so a_2's payoff would be the same.

Agent a_3 receives redistribution payment $\hat{\mathbb{V}}_{a_2}/n$. Under-reporting changes nothing. Over-reporting could put him in the first or second position. In both cases he receives the same Z payment, since the third position would then be held by the actual a_2. The same holds for all $a_{j>3}$.

Theorem 3.4. **RM** *for AON domains is ex post individual rational and no-deficit.*

Proof. Ex post individual rationality follows trivially from the fact that VCG has that property (in an AON domain the no negative externalities condition is satisfied), since **RM** modifies VCG only by *paying* the agents an additional sum. No-deficit is also almost immediate: total revenue obtained by the center equals:

$$\hat{\mathbb{V}}_{a_2} - \sum_{i \in I} Z_i \tag{3.8}$$

$$= \hat{\mathbb{V}}_{a_2} - \left(\frac{n-2}{n} \cdot \hat{\mathbb{V}}_{a_2} + \frac{2}{n} \cdot \hat{\mathbb{V}}_{a_3}\right) \tag{3.9}$$

$$\geq \hat{\mathbb{V}}_{a_2} - \hat{\mathbb{V}}_{a_2} = 0 \tag{3.10}$$

□

Theorem 3.5. *Assuming a finite bound on the value of each agent for each outcome, as the number of participating agents n goes to ∞ the amount of extracted wealth that cannot be redistributed among the agents under* **RM** *goes to 0. That is,* **RM** *is asymptotically strongly budget-balanced for AON domains.*

Proof. Again we can look at the total revenue in a scenario with n agents:

$$\begin{aligned} & \hat{\mathbb{V}}_{a_2} - \sum_{i \in I} Z_i \\ = \; & \hat{\mathbb{V}}_{a_2} - \frac{n-2}{n} \cdot \hat{\mathbb{V}}_{a_2} - \frac{2}{n} \cdot \hat{\mathbb{V}}_{a_3} \\ = \; & \frac{2}{n} \cdot (\hat{\mathbb{V}}_{a_2} - \hat{\mathbb{V}}_{a_3}) \end{aligned}$$

As n goes to ∞ this quantity is arbitrarily close to 0, regardless of the value of $\mathbb{V}_{a_2}$. □

As n increases, we may expect the payments to the center to be pushed down by a convergence of $\mathbb{V}_{a_2}$ and $\mathbb{V}_{a_3}$, but regardless of this **RM** achieves perfect budget balance in the limit. VCG will always lose $\hat{\mathbb{V}}_{a_2}$, no matter the number of agents.

Consider once more the example illustrated in Table 1. The following payoffs are obtained under **RM**:

$$\begin{aligned} & u_1 = 10 - 8 + \frac{5}{4} = \frac{13}{4}, \quad u_2 = \frac{5}{4}, \quad u_3 = \frac{8}{4}, \quad u_4 = \frac{8}{4} \\ & \text{payments to center} = 8 - \left(\frac{5}{4} + \frac{5}{4} + \frac{8}{4} + \frac{8}{4}\right) = \frac{3}{2} \end{aligned}$$

Even in this example with just 4 agents, the vast majority (81%) of the VCG revenue has been redistributed. If $\mathbb{V}_{a_3}$ were 8 rather than 5, 100% would be redistributed.

3.5 Redistribution in the general case

In this section I will present the redistribution mechanism in its general form. It is applicable to *any* domain and is, in a sense, "parameterized" by the type space that defines the domain. It redistributes a portion of the VCG revenue for any instance in which doing so is possible without distorting the incentives of the agents. The intuition is as follows: for each agent we can compute a guarantee—independent of that agent's report—on the revenue that would result under VCG. This quantity represents an upper bound on the amount that can be redistributed to that agent without violating no-deficit or strategyproofness, and, as we will see, also an indication that we can *definitely* redistribute some revenue (again without violating no-deficit or strategyproofness). In other words: *our ability to redistribute VCG revenue to an agent is directly tied to the extent to which we can know revenue will exist independent of that agent's reported type.*

Definition 3.5 (revenue-guarantee $\mathcal{G}_i(\Theta_i, \theta_{-i})$). *The lower-bound on VCG revenue that would result, computed over all possible reported types $\theta_i \in \Theta_i$ for agent i, given the type profile θ_{-i} reported by the other agents, i.e.,*

$$\mathcal{G}_i(\Theta_i, \theta_{-i}) = \min_{\theta_i \in \Theta_i} \sum_{j \in I} \Big[v_{-j}(\theta_{-j}, f^*(\theta_{-j})) - v_{-j}(\theta_{-j}, f^*(\theta)) \Big] \tag{3.11}$$

Breaking down equation (3.11), within brackets is an expression representing agent j's payment to the center under VCG, given reported type profile θ. Computing this quantity for each j (including i) and summing them all together, we get the complete revenue under VCG. In the equation θ_i is chosen to *minimize* this quantity; thus $\mathcal{G}_i$ represents the minimum level of revenue we can guarantee will occur independent of what type agent i ultimately reports.

It will also be useful in what follows to consider the revenue-minimizing report that an agent i could make, given the reports of other agents, i.e.,

$$\underline{\theta}_i = \arg\min_{\theta_i \in \Theta_i} \sum_{j \in I} \Big[v_{-j}(\theta_{-j}, f^*(\theta_{-j})) - v_{-j}(\theta_{-j}, f^*(\theta)) \Big] \tag{3.12}$$

$\underline{\theta}_i$ is determined by the context of reported type profile θ_{-i} and type space Θ_i, but those will be clear whenever I refer to $\underline{\theta}_i$.

I will use the term "redistribution mechanism" to refer to any mechanism that executes the payments of VCG modified by some redistribution term. Let $Z_i^T : \Theta \to \Re$ denote a redistribution payment function for i (which is a part of the overall transfer function T_i). I now show that in any redistribution mechanism (f^*, T) an agent's revenue-guarantee is an upper bound on his redistribution payment Z_i^T.

Lemma 3.1. *For any smoothly connected type space Θ, in any redistribution mechanism (f^*, T) that is truthful and efficient in dominant strategies and no-deficit: $\forall i \in I$ and $\theta \in \Theta$, $Z_i^T(\theta) \leq \mathcal{G}_i(\Theta_i, \theta_{-i})$.*

Proof. Assume otherwise, i.e., define T_i for some $i \in I$ such that:

$$T_i(\theta) > v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) + \mathcal{G}_i(\Theta_i, \theta_{-i}), \tag{3.13}$$

Then if i reports $\underline{\theta}_i$ and other agents report θ_{-i} a deficit will result, by definition of $\mathcal{G}_i$. □

But $\mathcal{G}_i$ provides both a constraint on the amount we can redistribute and a pointer towards creating a mechanism that *does* redistribute significant portions of revenue.

Lemma 3.2. *There exists a mechanism (f^*, T) that is truthful and efficient in dominant strategies, ex post individual rational when the no negative externalities property holds, no-deficit, and—for any type space Θ and reported type profile $\theta \in \Theta$—yields social-welfare at least as great as VCG plus* $\min_{i \in I} \mathcal{G}_i(\Theta_i, \theta_{-i})$.

Proof. Consider a mechanism (f^*, T) that picks an agent i arbitrarily (but independent of reported types) and defines:

$$T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) + \mathcal{G}_i(\Theta_i, \theta_{-i}), \text{ and} \tag{3.14}$$

$$T_j(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})),\ \forall j \in I \setminus \{i\} \tag{3.15}$$

This mechanism is a Groves mechanism: it is equivalent to VCG for all agents but i, and for i it is VCG plus an extra redistribution payment that is independent of his report; thus it is truthful and efficient in dominant strategies. It is ex post individual rational since VCG is (every agent is weakly better off under this mechanism, for every type profile), and it is no-deficit since $\mathcal{G}_i(\Theta_i, \theta_{-i})$ is by definition less than the revenue that VCG generates on θ. Finally it yields social welfare at least $\min_{i \in I} \mathcal{G}_i(\Theta_i, \theta_{-i})$ better than VCG since, no matter what agent i is selected randomly, $\mathcal{G}_i(\Theta_i, \theta_{-i}) \geq \min_{j \in I} \mathcal{G}_j(\Theta_j, \theta_{-j})$. □

The mechanism specified in the proof of Lemma 3.2 succeeds in redistributing some revenue, but note that it fails to meet an *anonymity* property: agents that have the same type space and report the same value may obtain different utilities, since only one will be randomly chosen to receive the redistribution payment. Anonymity properties are motivated by fairness concerns—the idea is that agents may not be satisfied if they are treated "unequally" without compelling reasons.

Here I introduce an anonymity notion that derives relevance from the context of redistribution mechanisms and the fact that they are possible based on computations of revenue-guarantee quantities for each agent. It is formulated to exclude situations in which agents have the same revenue-guarantee but receive different redistribution payments.

Definition 3.6 (redistribution-anonymity). *A redistribution mechanism (f, T) is redistribution-anonymous if and only if it maps agent-specific revenue-guarantees ($\mathcal{G}_i$) to redistribution payments (Z_i^T) according to a single deterministic function that is invariant to domain information that does not apply identically to every agent, i.e., if for any $i, j \in I$, for any Θ_i and Θ_j, $\forall \theta_i \in \Theta_i$ and $\theta_j \in \Theta_j$:*

$$\mathcal{G}_i(\Theta_i, \theta_{-i}) = \mathcal{G}_j(\Theta_j, \theta_{-j}) \Rightarrow Z_i^T(\theta) = Z_j^T(\theta) \tag{3.16}$$

In a *redistribution-anonymous* mechanism, whenever two agents have the same revenue-guarantee they receive the same redistribution payment. In fact, two agents participating in two different instances of the mechanism run with different agent type spaces must receive the same redistribution payment if their revenue guarantees are the same. It turns out we can do better than the bound of the mechanism in Lemma 3.2 *and* satisfy redistribution-anonymity. I now present the main results of the chapter: I define the full version of the redistribution mechanism; I show that it has all the good properties of VCG, yields significantly greater social-welfare in many domains, and is redistribution-anonymous; I show that among all redistribution-anonymous mechanisms it is *optimal* in the strongest sense.

Definition 3.7 (Redistribution mechanism RM). RM *is a direct mechanism* (f^*, T) *where,* $\forall i \in I$ *and* $\theta \in \Theta$*:*

$$T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) + \frac{\mathcal{G}_i(\Theta_i, \theta_{-i})}{n} \tag{3.17}$$

Theorem 3.6. RM *is truthful and efficient in dominant strategies, ex post individual rational if the no negative externalities condition holds, no-deficit, and redistribution-anonymous.*

Proof. Truthfulness and efficiency in dominant strategies follows immediately from dominant strategy truthfulness and efficiency of VCG plus the fact that $\mathcal{G}_i$ is independent of i's reported type (i.e., since **RM** is a Groves mechanism). Likewise ex post individual rationality for no negative externalities domains is immediate since VCG has the property and **RM** only increases agent utilities over VCG.

VCG is no-deficit (see Theorem 2.9), so to prove **RM** is no-deficit it is sufficient to show that the sum of the redistribution payments is always less than the VCG revenue. Abstracting away from notation a bit, we have that, for any reported type profile θ, for every i,

$$\mathcal{G}_i(\Theta_i, \theta_{-i}) = \min_{\theta'_i \in \Theta_i} \text{VCG-revenue}(\theta'_i, \theta_{-i}) \tag{3.18}$$

$$\leq \text{VCG-revenue}(\theta) \tag{3.19}$$

Thus $\sum_{i \in I} \mathcal{G}_i(\Theta_i, \theta_{-i})/n \leq \text{VCG-revenue}(\theta)$, and so **RM** is no-deficit.

Finally, redistribution-anonymity of **RM** holds since the mechanism does, in fact, map all $\mathcal{G}_i$ to a redistribution payment according to a single deterministic function: division by n. □

Lemma 3.1 and Theorem 3.6 taken together provide an exact characterization of when redistribution is possible: when there is an agent i with a positive revenue-guarantee $\mathcal{G}_i$.

Theorem 3.7. *For any smoothly connected type space* Θ *in which the no negative externalities property holds, there exists a redistribution mechanism that is truthful and efficient in dominant strategies, ex post individual rational, no-deficit, and redistributes a positive amount of VCG revenue on some type profile if and only if* $\exists \theta \in \Theta$ *and* $i \in I$ *such that* $\mathcal{G}_i(\Theta_i, \theta_{-i}) > 0$.

Proof. First, that $\exists \theta \in \Theta$ and $i \in I$ with $\mathcal{G}_i(\Theta_i, \theta_{-i}) > 0$ is sufficient follows immediately from Theorem 3.6. On the other hand when $\forall \theta \in \Theta, i \in I$, $\mathcal{G}_i(\Theta_i, \theta_{-i}) = 0$, by Lemma 3.1 any redistribution sacrifices no-deficit. □

3.5.1 A sense in which RM is optimal

We will now see that if redistribution-anonymity is taken as a hard constraint, then there is no mechanism that *ever* redistributes more than **RM** while retaining the good (and essential) properties of VCG.

Definition 3.8 (optimal redistribution mechanism). *Given a specified type space Θ, (f^*, T) is an optimal redistribution mechanism in redistribution mechanism space M if and only if $(f^*, T) \in M$ and, $\forall \theta \in \Theta$, no mechanism in M redistributes more VCG revenue on θ.*

This is a very strong optimality condition, requiring that no other mechanism in the space ever redistribute more revenue ever.

Theorem 3.8. *For any smoothly connected domain,* **RM** *is an optimal redistribution mechanism among all redistribution mechanisms that are truthful and efficient in dominant strategies, no-deficit, and redistribution-anonymous.*

Proof. Let $\mathcal{G}^*(\theta)$ denote the VCG revenue that results given a reported type profile θ. Each agent i's revenue-guarantee $\mathcal{G}_i(\Theta_i, \theta_i) = \min_{\theta_i \in \Theta_i} \mathcal{G}^*(\theta_i, \theta_{-i})$, and $\forall i \in I$,

$$\underline{\theta}_i \in \arg\min_{\theta_i \in \Theta_i} \mathcal{G}^*(\theta_i, \theta_{-i}) \tag{3.20}$$

For any profile of true agent types θ, note that there is a set of type spaces such that $\underline{\theta}_i = \theta_i$, for all $i \in I$ (for instance, consider the case where each agent i's type space consists of just a single type, θ_i). For such a type space Θ',

$$\mathcal{G}(\Theta'_1, \theta_{-1}) = \mathcal{G}(\Theta'_2, \theta_{-2}) = \ldots = \mathcal{G}(\Theta'_n, \theta_{-n}) = \mathcal{G}(\theta)^* \tag{3.21}$$

In a redistribution-anonymous mechanism, a single deterministic function $z : \Re \Rightarrow \Re$ maps each agent i's revenue-guarantee $\mathcal{G}_i$ to a redistribution payment. So by redistribution-anonymity of **RM**,

$$z(\mathcal{G}(\Theta'_1, \theta_{-1})) = z(\mathcal{G}(\Theta'_2, \theta_{-2})) = \ldots = z(\mathcal{G}(\Theta'_n, \theta_{-n})) = z(\mathcal{G}^*(\theta)) \tag{3.22}$$

But then in order to satisfy the no-deficit property,

$$z(\mathcal{G}^*(\theta)) \leq \frac{\mathcal{G}^*(\theta)}{n} \tag{3.23}$$

Since this holds for any type profile θ (and thus any possible VCG revenue $\mathcal{G}^*(\theta)$), mechanism **RM** is an optimal redistribution mechanism given constraints of truthfulness and efficiency in dominant strategies, ex post individual rationality, and redistribution-anonymity. □

A few words about anonymity

Note that **RM** is *not* optimally balanced when the anonymity constraint is significantly reduced, for instance by allowing redistributions to vary with both the revenue-guarantee *and* the type space explicitly. Imagine a redistribution function that is derived from both the revenue-guarantee and a type space. Redistribution-anonymity requires that:

$$\forall \mathcal{G}_i,\, \Theta,\, \Theta',\;\; z(\mathcal{G}_i, \Theta) = z(\mathcal{G}_i, \Theta') \tag{3.24}$$

If this requirement is dropped, greater redistribution is possible. For example, consider an AON allocation problem with type space such that:

$$\mathbb{V}_1 \in [0,1],\; \mathbb{V}_2 \in [1,2],\; \mathbb{V}_3 \in [1,2]$$

A mechanism that redistributes the entire VCG revenue to agent 1 would not violate strategyproofness, no-deficit, or ex post IR. However, redistribution-anonymity *would* be violated since, for instance, if every agent i's type space were such that $\mathbb{V}_i \in [0,1]$, the mechanism could not be implemented in dominant strategy equilibrium.

It is also worth describing the relationship of redistribution-anonymity to a distinct fairness/anonymity constraint that is sometimes imposed in mechanism design, which I will refer to as *valuation-anonymity.* Informally, valuation-anonymity holds when the outcome and transfer functions are completely invariant to agent identity, including individual type spaces.[3] One of the things valuation-anonymity implies is that if two agents have the same type and report truthfully they will obtain the same expected utility. The VCG mechanism is valuation-anonymous: two agents with the same type make the same "marginal contribution" to social welfare, regardless of identity. In a redistribution-anonymous mechanism in the case of *symmetric* domains[4] this holds true, but the agents must also obtain the same expected utility in some cases in which they report *different* types—i.e., when they contribute equally to social welfare and the computed redistribution-guarantees are the same. In asymmetric domains redistribution-anonymous mechanisms permit different utilities for identical types (but different type *spaces*), unlike valuation-anonymous mechanisms.

The relationship between the two anonymity concepts is illustrated in Figure 3.3. There are mechanisms that are redistribution-anonymous but not valuation-anonymous, others that are valuation-anonymous but not redistribution-anonymous,

[3] Technically, to handle tie-breaking while achieving valuation-anonymity we would need to allow the choice function to select an outcome randomly from a subset of outcomes that are social welfare maximizing (should ties occur), and then let the transfer function depend on the selected outcome. The notion of redistribution-anonymity extends naturally in the same way.

[4] A *symmetric domain* is one in which the type space is identical for each agent.

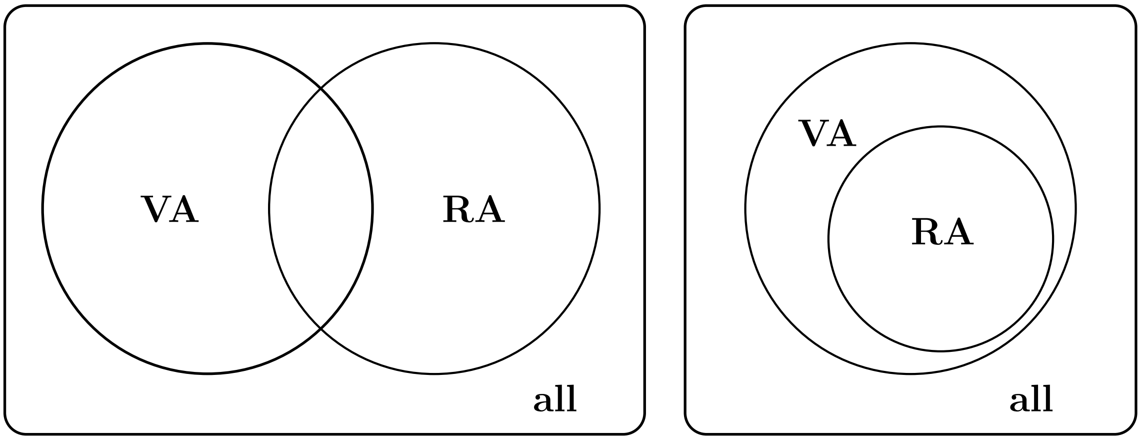


Figure 3.3: In many type spaces the set of all redistribution-anonymous mechanisms and the set of all valuation-anonymous mechanisms are overlapping subsets of the set of all mechanisms; this is represented in the left box. In *symmetric* type spaces, the set of all redistribution-anonymous mechanisms is a subset of the set of all valuation-anonymous mechanisms; this is represented in the box on the right.

and still others that have both properties. In symmetric domains redistribution-anonymity implies valuation-anonymity; thus **RM** is *redistribution-anonymous and valuation-anonymous* when applied to symmetric domains.

There is no optimal redistribution mechanism (among those that are truthful and efficient in dominant strategies and no-deficit) when anonymity considerations are completely dropped. I don't provide a proof here, but note, for instance, that a mechanism that redistributes $\mathcal{G}_i$ to a randomly selected agent i and none to other agents will in some cases redistribute more revenue than **RM**; however, such a mechanism is clearly not optimal since it may pick i with $\mathcal{G}_i < \max_j \mathcal{G}_j$.

3.5.2 Redistribution in combinatorial allocation problems

We've been discussing redistribution in the context of general social choice or decision-making problems. An important subclass of decision problems consists of those in which a decision must be made about how to allocate goods amongst a group of competing agents. The most basic allocation problem is that in which a single item is available; this limited case (usually) falls within the AON class.[5] More generally, an allocation problem consists of a number of agents with preferences over goods that are potentially combinatorial in nature.

[5] The exception is when there are externalities, e.g., when agent i is happy if his friend agent j is allocated the item.

Sometimes in such domains allocation mechanisms are selected to maximize revenue to the seller (the center), but there are many significant cases in which it is preferable to keep as much wealth as possible in the hands of the agents. For example, consider the allocation of job time on a publicly owned super-computer among a community of researchers. Resources like this are often established with the mandate of maximizing benefit to the public good (social-welfare); if a VCG-based allocation were implemented a revenue would result, redistribution of which would go further toward satisfying this mandate.

In the vast majority of allocation problems, the following are generally accepted to hold: agents that aren't allocated anything receive no value (normalization); agents' values monotonically increase as they receive more goods (free disposal); and agents have no preferences over allocations to other agents (no externalities). These intrinsic elements of the allocation domain map directly to constraints on the type space that can allow for significant strategyproof redistribution of VCG revenue. Formally, let G be the set of goods to be allocated, and for any bundle of goods $B \subseteq G$ let $v_i(B)$ be agent i's value for obtaining B. Each agent i's valuation conforms to the following:

$$\begin{aligned}&v_i(\emptyset) = 0\\ &v_i(B) \leq v_i(B \cup \{g\}), \;\; \forall_{B \subseteq G,\, g \in G}\end{aligned}$$

Consider the following valuations of three agents in an allocation problem with two goods, c and d, where $\{X_1, X_2, X_3\}$ represents the outcome in which bundles X_1, X_2, and X_3 are allocated to agents 1, 2, and 3 respectively.

	v_1	v_2	v_3
$\{cd, \emptyset, \emptyset\}$	12	0	0
$\{\emptyset, cd, \emptyset\}$	0	10	0
$\{\emptyset, \emptyset, cd\}$	0	0	11
$\{c, d, \emptyset\}$	4	6	0
$\{c, \emptyset, d\}$	4	0	5
$\{d, c, \emptyset\}$	5	7	0
$\{d, \emptyset, c\}$	5	0	7
$\{\emptyset, c, d\}$	0	7	5
$\{\emptyset, d, c\}$	0	6	7

Table 2: 2-good, 3-agent allocation problem.

The efficient outcome is to allocate c to agent 3 and d to agent 2. Applying **RM**, agent utilities are as follows:

$$
\begin{aligned}
u_i &= v_i(\theta_i, f^*(\theta)) + v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i})) + \frac{\mathcal{G}_i(\Theta_i, \theta_{-i})}{n} \\
u_1 &= 0 + 13 - 13 + \frac{8}{3} = \frac{8}{3} \\
u_2 &= 6 + 7 - 12 + \frac{9}{3} = 4 \\
u_3 &= 7 + 6 - 12 + \frac{9}{3} = 4
\end{aligned}
$$

Using VCG with no redistribution, total social utility is 2 and payment to the center is 11. 8.67 of this (79% of revenue) can be redistributed to the agents under **RM**, yielding a social utility of 10.67, a nearly 5-fold improvement.

3.6 Simulations

3.6.1 Computing redistribution payments

Determining redistribution payments under **RM** amounts to computing revenue-guarantee $\mathcal{G}_i$ for each agent i, and then merely dividing by n. In some domains, a simple algorithm for computing $\mathcal{G}_i$ exists. For example, in AON and some other allocation settings, $\mathcal{G}_i$ can be computed by determining the revenue that would result if i were just not present. Unfortunately this simple algorithm does not hold in general (i.e., not for all possible sets of value constraints); notably, it does not hold for combinatorial allocation domains. However, we can *always* compute $\mathcal{G}_i$ through a mixed-integer programming (MIP) specification of the revenue-guarantee equation (3.11). I outline the formulation here.

Equation (3.11) can be rewritten in an expanded form as follows:

$$
\begin{aligned}
\mathcal{G}_i(\Theta_i, \theta_{-i}) &= \min_{\theta_i} \Big[\sum_{j\in I} \Big(\max_{o'\in O} \sum_{k\in I\setminus\{j\}} v_k(\theta_k, o') - \sum_{k\in I\setminus\{j\}} v_k(\theta_k, o^\dagger) \Big) \Big] \\
&= \min_{\theta_i} \Big[\sum_{j\in I} \Big(\max_{o'\in O} \sum_{k\in I\setminus\{j\}} v_k(\theta_k, o') \Big) - (n-1) \sum_{j\in I} v_j(\theta_j, o^\dagger) \Big] \qquad (3.25)
\end{aligned}
$$

subject to the constraint that the $o^\dagger$ and θ_i in the above minimization must satisfy the following:

$$
o^\dagger = \arg\max_{o\in O} \sum_{j\in I} v_j(\theta_j, o) \qquad (3.26)
$$

The objective in the MIP for agent i's payment is to minimize VCG revenue, and the primary variables are $v_i(\theta_i, o)$ for each $o \in O$. Representing the program constraints is relatively straightforward, but some care must be taken in handling equation (3.26) and the inner maximization in (3.25). In both cases there are non-linearities. In (3.26), for instance, we must specify $c_o \cdot v_i(o)$ for each outcome o, where c_o is a boolean variable representing whether or not o is chosen as the outcome that maximizes social welfare. We can get around this issue by representing $c_o \cdot v_i(\theta_i, o)$ with a new variable $v_i'(\theta_i, o)$, and including the following constraints:

$$\begin{aligned} v_i'(\theta_i, o) &\leq v_i(\theta_i, o) \\ v_i'(\theta_i, o) &\leq c_o \cdot M, \end{aligned}$$

where M is a value larger than the maximum possible value an agent could have for any outcome. $v_i'(\theta_i, o)$ will then be $v_i(\theta_i, o)$ if $o = f^*(\theta)$ and 0 otherwise, as desired.

While solving a mixed-integer program has exponential worst-case running time, in practice I was able to quickly find solutions to very large problems. Determining redistribution payments in a 100 agent, 100 outcome problem took 24 seconds for each agent.[6] Note that the MIPs for calculating agent payments (one for each agent) are independent of each other, and thus all can be solved in parallel.

3.6.2 Empirical results

In order to understand how much redistribution can be achieved for valuations under different levels of mutual constraint, I performed an empirical analysis on large sets of randomly generated problem instances (sets of valuations), each with the same number of outcomes as agents. I generated valuations according to the following process, where e is an "exclusivity" (between agent valuations) parameter representing the extent to which a domain has "all-or-nothing properties." I chose a maximum value $maxval_i$ for each agent i's valuation function uniformly at random between 0 and 100. I then chose each $v_i(\theta_i, o_i)$ uniformly at random between 0 and $maxval_i$, and $v_i(\theta_i, o_{j\neq i})$ uniformly at random between 0 and $(1-e)\cdot maxval_i$. So when $e = 1$, we have a completely AON domain; when $e = 0$ it is an unrestricted type space.

The graph in Figure 3.4 plots the percentage of VCG revenue redistributed by **RM** as a function of the parameter e, for problems with various numbers of agents. For each number of agents, 100 samples were computed for each value of e between 0 and 1 in increments of 0.05, and I took the average. Notably, redistribution remains nearly constant for values of e between 0 and 0.5, and then increases roughly linearly with e from 0.5 to 1. As expected, the possibility for redistribution grows with the number of agents.

It is also instructive to compare directly the percentage of total value retained by the agents under **RM** compared to that under VCG. Figure 3.5 plots exactly this for population sizes of 4, 8, and 20. The left column of graphs have agent valuations drawn uniformly at random between 0 and 100 for one outcome and from between 0

[6] Solutions were obtained using the commercial solver CPLEX, run on a 1.6 GHz Pentium 4 PC.

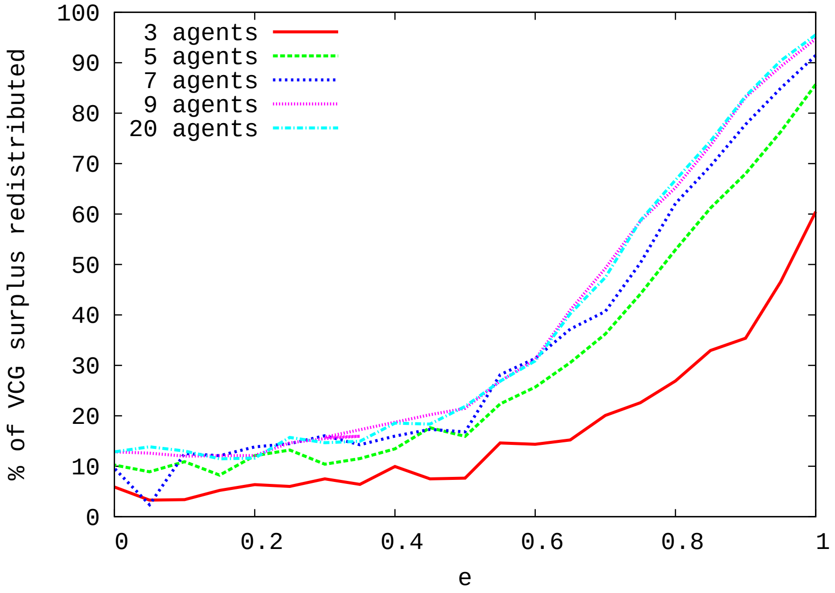


Figure 3.4: Percent of revenue redistributed as a function of mutual "exclusivity" between agent valuations.

and $(1-e)\cdot 100$ for the rest. The right column does the same except with a normal distribution with mean 50 (or $(1-e)\cdot 50$) and standard deviation 12.5 (or $(1-e)\cdot 12.5$). The simulations demonstrate that **RM** maintains a more or less constant amount of value within the population of agents for different values of e, except as e goes to 1 (i.e., as we trend towards AON domains), where value maintained increases somewhat sharply. On the other hand VCG falls off quite sharply as **RM** starts to rise. Again, on the left side of the axis are represented domains in which no redistribution is possible (the type spaces are unrestricted), and there **RM** coincides with and thus yields the same value as VCG. On the right hand side are more purely "competitive" environments, where an outcome that is good for one agent is necessarily not good for others—this is where the differences between VCG and **RM** become stark.

3.7 A center-less implementation

Consider the problem of building a decision-making mechanism that doesn't require a center. This is intuitively desirable, as it's easy to imagine scenarios in which a group of agents must determine an allocation or decide on an action to take without the benefit of a trusted central authority.[7]

Theorem 3.9. **RM** *can be implemented in Nash equilibrium without the facilitation of a center when the following conditions hold:*

i) no agent acting alone has the power to obstruct realization of a specified outcome or transfer payments, while $n-1$ agents acting together do.

ii) agents can simultaneously "broadcast" valuations to all other agents, and can compute and perform transfers.

iii) agents have the capacity to publicly destroy money.

Proof sketch. The first condition brings abiding by the mechanism into Nash equilibrium—unilaterally deviating cannot be beneficial. The second and third conditions ensure that agents will be able to execute what the mechanism prescribes. With (ii), all agents can recognize the efficient outcome and deliver the appropriate redistribution payment to each "receiving" agent. (iii) allows the "paying" agents (just a_1 in AON domains) to demonstrably receive the appropriate payoff. The quantity that cannot stay within the group of agents would normally be transfered to a center, but its destruction is a satisfactory substitute. □

In AON domains, at least, where communicating a valuation amounts merely to a public announcement of a single (value, outcome) pair, satisfaction of these conditions is plausible. Then using **RM**, it is possible in equilibrium for a large group of self-interested agents to independently reach the socially optimal outcome, and jointly reap nearly all fruits of the chosen action via redistribution.

3.8 Discussion

In this chapter I argued for the desirability of redistributing VCG payments back among participants in a mechanism, and showed it is feasible in a broad range of settings, including allocation problems. I presented **RM**: a mechanism that is truthful and efficient in dominant strategies, ex post individual rational, no-deficit, and satisfies a fairness constraint pertinent to the setting (redistribution-anonymity). **RM** improves the social welfare properties of VCG drastically in, e.g., allocation settings, and is optimal in a strong sense when all the properties just mentioned are required as hard constraints. In "all-or-nothing" domains, a class which encompasses typical single-item allocation problems, the mechanism is strongly budget-balanced in the limit. So for decision problems in which the large transfers VCG requires from the agents are considered waste, the mechanism specified in this chapter is a superior solution in terms of social welfare.

[7] See [Shneidman and Parkes, 2004] or [Petcu *et al.*, 2006] for further discussion of distributed implementations that seek to minimize the role of a center.

In a single-item allocation domain it so happens that an agent's revenue-minimizing report is 0, which is equivalent to the agent simply not being present in the system at all. Bailey [1997] recognized this fact. But the approach of simply giving each agent a share of the revenue that would result if he were absent does not extend. In the case of combinatorial allocations, for instance, an agent's revenue-minimizing report may *not* be simply a 0-valuation report; if we computed redistribution payments based on that hypothetical, deficits would result. **RM**, on the other hand, will never run a deficit regardless of the domain.

Returning to the question of optimality, there are a few points worth making. First, the redistribution-anonymity fairness constraint may not, in fact, be important in practice. If (or when) that is the case, in the absence of an optimality result for **RM** what effort, if any, should be expended on trying to improve on it? One of the remarkable attributes of **RM** is that it is applicable to any and every domain (type space). When no redistribution is possible, it coincides with VCG. Whenever redistribution *is* possible, **RM** will achieve some redistribution. In some cases it will be possible to design mechanisms that are restricted to only a specific setting, e.g., single-item allocation problems, and these mechanisms may make some gains. But, to use that domain as an example, the empirical evidence shows that the room for improvement over **RM** is in fact quite limited when there are more than a few agents. Considering the graphs in Figure 3.5 for 8 agents, $\sim$96–98% of the value is retained within the group in AON domains (the far right side of the axis). 100% is the theoretical limit. Perhaps there are other domains in which **RM** will not be as successful where another mechanism could be.

But also deserving of consideration is the fact that **RM** can be described intuitively and concisely: run VCG, compute an agent-independent "guarantee on revenue" for each agent, and give each agent this quantity divided by the number of agents. In the real world this may matter; in considering possible alternatives that achieve an extra percentage or two of value over **RM**, this benefit must be weighed against factors like simplicity and understandability.

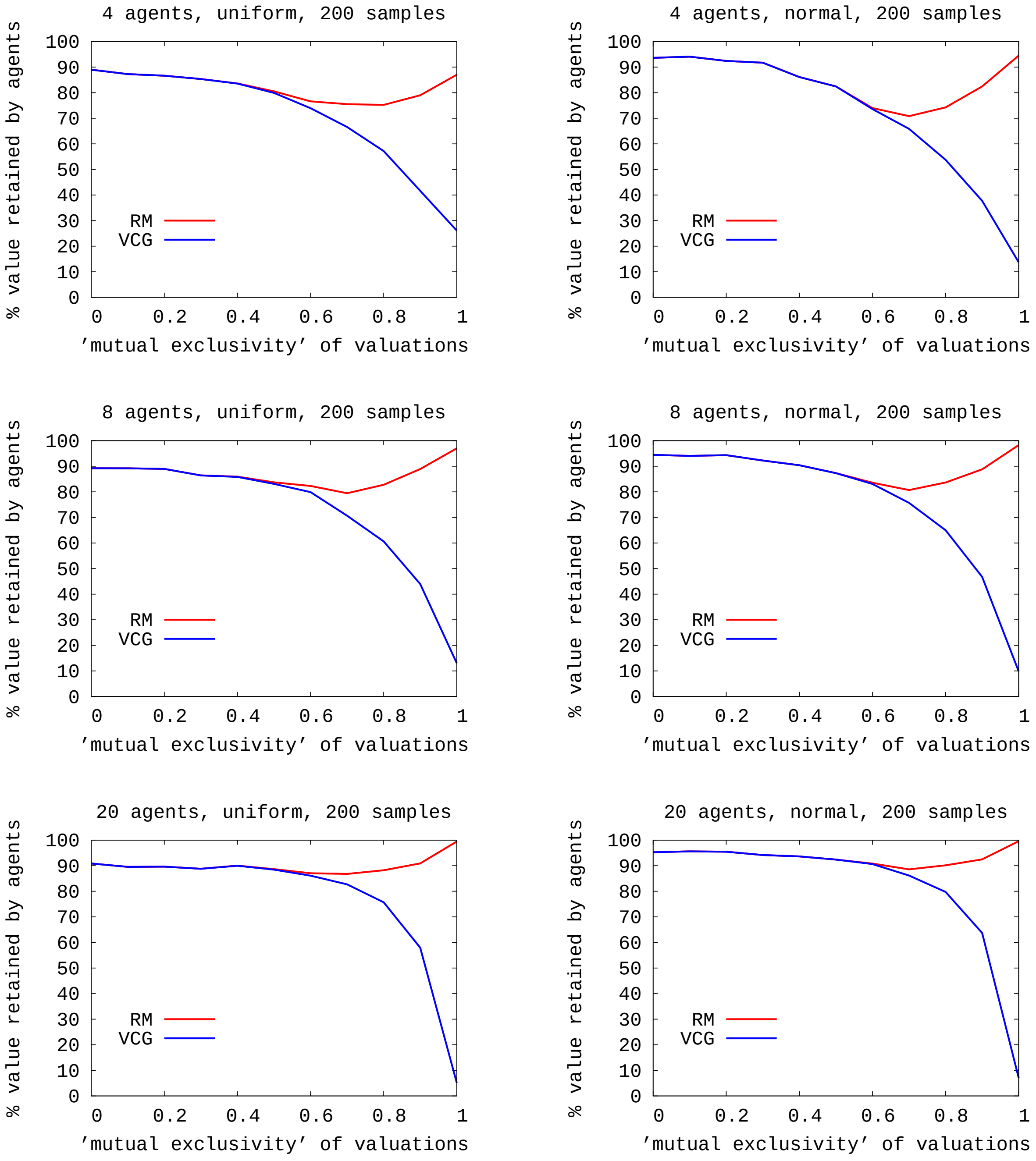


Figure 3.5: Comparison of the percentage of value from the socially optimal decision retained by the agents (i.e., social value minus payments to the center, divided by social value) under RM and VCG. Average over 200 samples for agent population sizes of 4, 8, and 20. Plots for valuations drawn from both uniform (left) and normal (right) distributions.

Chapter 4

Optimal decision-making in dynamic settings

Synopsis

In this chapter I introduce the problem of multi-agent sequential decision-making, where the goal is to maximize total social welfare accumulated over time. I introduce Markov decision processes (MDPs) as a representation for such problems and present leading exact and approximate solution techniques. I then give special focus to the subset of decision problems that can be modeled as multi-armed bandits, since this class is exceptional in that it admits computationally tractable optimal solutions and at the same time, to a reasonable approximation, effectively models many important real-world problems.

4.1 Introduction

Thus far we have considered decision-making problems that are static or *one-shot*, without a context of time and other decisions that could possibly be of relevance. The results for this setting, as we have seen, are vast, and they have found fairly successful application in a number of real-world scenarios. But a static model can only go so far.

Consider, for instance, the simple problem we discussed in the previous chapter (illustrated in Figure 3.1) in which 4 housemates must decide who gets to use their jointly owned car on a Friday night. They agree to implement a mechanism that achieves the efficient allocation (the one who wants or needs the car most will get to use it). We saw that both the VCG mechanism (the Vickrey auction in this case) and redistribution mechanism **RM** achieve the efficient outcome in dominant strategy equilibrium even when each agent acts only to maximize his own welfare. But what if the decision of who gets to use the car is *not* an isolated, one-time event? What if the same scenario arises every day? Well, one can see the added complexity that arises even from looking at the relatively simpler case in which the decision-making problem arises only twice, once on Friday and then again on Saturday.

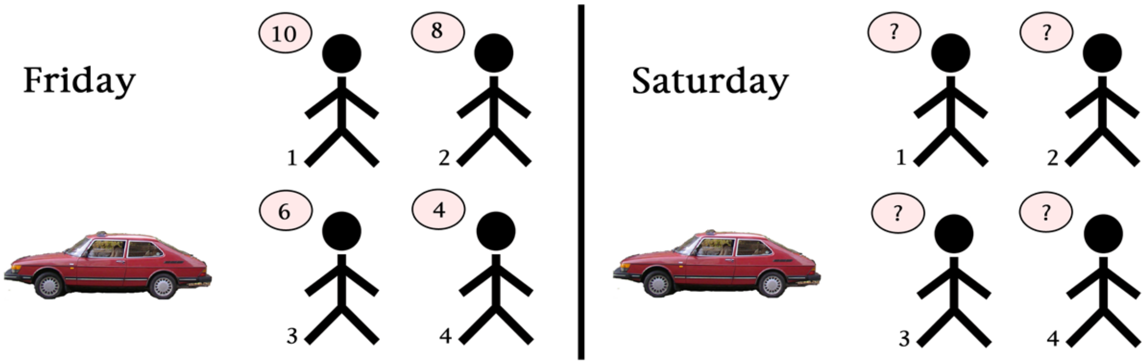


Figure 4.1: Illustration of a 4-agent 2-period decision problem. A decision about whom to allocate the car to is made on Friday and again on Saturday.

Each agent would obtain some value on Friday night should he get to use the car. One might plan to go buy groceries, and would obtain value 6; another might drive to the movies with a date, obtaining value 10; still another would use it to help his brother move into a new apartment, etc. Each agent can calculate this value and report it (e.g., to a center) so that it can be determined who has the highest value for the car on Friday. But this information is *not* sufficient to determine what decision is optimal, even just on Friday; the reason is that the decision on Friday will influence the values that agents could obtain on Saturday.

For instance, consider the agent (call him agent 3) who would use the car to help his brother move. What if he can't complete the job in a single day, but if he had the car for both Friday and Saturday he could? His value for the car on Saturday will depend on whether or not he is allocated the car on Friday. Perhaps his value for using the car for either (but not both) Friday or Saturday is 4, but if he gets the car both days he will obtain value 20 once the job is finished.

To determine efficient decisions, the center needs to know each agent's value for *every set of decisions*. Let $o_{(i,j)}$ denote the outcome-pair in which the car is allocated to agent i on Friday and agent j on Saturday. With 4 agents there are 16 different outcome-pairs. Agent 3's total value for $o_{3,1}$, $o_{3,2}$, $o_{3,4}$, $o_{1,3}$, $o_{2,3}$, or $o_{4,3}$ is 4; his value for $o_{3,3}$ is 20, and his value for all other outcomes is 0. Notice that a problem like this can actually be treated as a static problem, in which each outcome is an outcome *pair* from the 2-step version; there is nothing to keep the center from computing the efficient choice-pair in the first time-step. And as far as incentives, the payments prescribed by a Groves mechanism will yield truthful revelation of values for outcome-pairs as a dominant strategy for each agent at the beginning of the mechanism.

But this is not the end of the story. In many dynamic scenarios agents' expectations of the value they will receive in the future change as decisions are made. That is, there is uncertainty in agent valuations, either due to ignorance on the part of agents or true randomness in the realization of outcomes. For instance, consider agent 3 in the car example again. What if there is some chance that he can't complete the moving job even if he is granted the car for both days, and what if he will find out whether or not he'll be able to finish after using the car for the first day. Before Friday agent 3 will have an *expectation* about what will happen during the first day, but after Friday he will *know for sure* what has happened, and thus will know exactly what his value is for getting the car for another day.

Crucially, the center cannot implement optimal decisions at a given time t unless agents have shared all relevant information they have accumulated *through* t. In a setting like this where agents are obtaining new private information, one can consider agent types as "evolving" or gradually revealed by nature over time. An agent can't possibly report all private information that will be relevant in the beginning, since he hasn't yet obtained all that information himself. The incentive issues get significantly more complicated when we need agents to be truthful every time-period; dealing with this problem is one of the two main focuses of this dissertation, and aspects of the problem will be addressed in Chapters 5, 6, 7, and 8. But first in this chapter I will provide an overview of computing efficient decision policies for dynamic settings *assuming all information held by the agents is known to the center* (i.e., ignoring incentive issues). As we will see, this problem is very hard in its own right, and represents a significant challenge to implementation of efficient mechanisms for dynamic settings.

4.2 Markov decision processes as a representation of type in dynamic settings

In order to make decisions at every point in time that *maximize expected value*, we must have a clear and effective way of representing information—about both the immediate value a decision will bring and the impact it will have on the results of future decisions. A convenient formalism for this task is a *Markov Decision Process (MDP).*[1]

In an MDP, the notion of a *state* encompasses or "contains" all information relevant to: 1) the immediate value that will be obtained from decisions, and 2) the (expected) bearing decisions will have on *future* decisions (or *actions* in MDP parlance). A *transition function* describes how, upon taking a given action in a given state, a new state is reached. Finally a *reward function* describes the value that taking

[1] See, e.g., [Bellman, 1957] for an early appearance, and [Sutton and Barto, 1998] for a very good reference.

a given action in a given state will yield. Formally, an MDP $M = \{S, A, \tau, r\}$, where S is a state space and A is an action space; $\tau : S \times A \times S \rightarrow \Re$ is a transition function, with $\tau(s, a, s')$ denoting the probability that taking action a in state s will yield new state s'; finally, $r : S \times A \times S \rightarrow \Re$ is a reward function, with $r(s, a, s')$ denoting the value obtained when a is taken in s and successor state s' results. It will be convenient at times to refer to the random variable representing the state reached after executing an action a in a state s; I use notation $\tau(s, a)$ for this purpose. I will use $r(s, a)$ to denote the expected reward when a is taken in s, i.e., $\sum_{s' \in S} \tau(s, a, s')\, r(s, a, s')$. Note that implicit in this framework is the *Markov assumption*: that the effect of taking any given action in any given state depends only on that state and, in particular, not on the previous history of states that have been realized.[2]

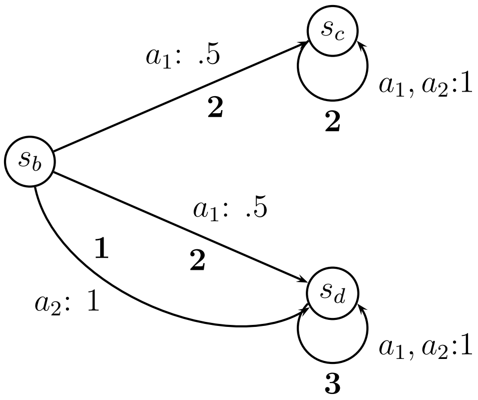


Figure 4.2: Simple 3-state, 2-action MDP example. Actions and their respective transition probabilities from one state to another are labeled in italics on the arc between the two states. Rewards are labeled in bold.

For instance, consider the simple example in Figure 4.2. There are two actions: $A = \{a_1, a_2\}$, and three states: $S = \{s_b, s_c, s_d\}$. Taking action a_1 in state s_b will yield new state s_c or s_d with equal probability (i.e., $\tau(s_b, a_1, s_c) = \tau(s_b, a_1, s_d) = 0.5$) and yields value 2 (i.e., $r(s_b, a_1) = 2$); taking action a_2 will definitely lead to state s_d and yield value 1. From states s_c or s_d, taking either action leads to the same state; value 2 is yielded from s_c and 3 from s_d. Thus the connections between states in an MDP represent the ways in which decisions (actions) taken now influence decision-making scenarios in the future—deterministically (e.g., taking action a_2 in state s_b above) or non-deterministically (e.g., taking a_1 in s_b). The timing of events in a sequential decision-making procedure is illustrated in Figure 4.3. An action being taken leads to the start of each period, then a state transition occurs and value is obtained.

[2] But this does not constrain the representational power of MDPs, since the state space can always

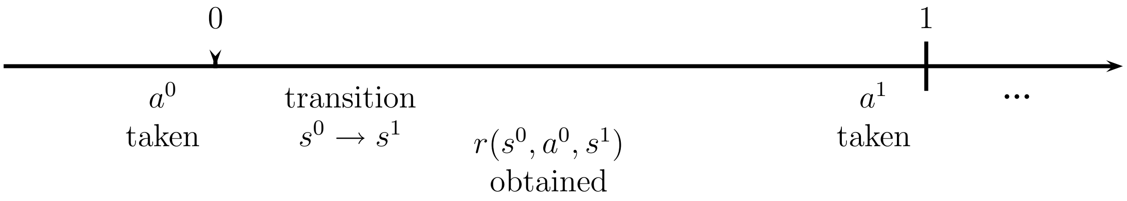


Figure 4.3: A graphical representation of the timing in a sequential decision-making procedure. The time-step "ticks forward" when an action is taken. Value is obtained, a state transition occurs, and finally the next action is taken leading into the next time-step tick.

In dynamic decision problems we must also consider the possibility that agents are *impatient*, i.e., that they value reward received sooner more than they do the same reward received later. This is typically modeled via an *exponential discount factor* γ, where a reward of x received t time-steps in the future is evaluated to yield utility $\gamma^t x$. Of course whether or not this is the right model of how agents perceive value relative to time is a domain-specific question; the exponential discounting model seems to be a relatively satisfying approximation for many settings (see, e.g., [Keller and Strazzera, 2002]). But perhaps more relevant to its ubiquity is that for problems with an infinite *time-horizon* (i.e., no limit to the number of decisions to be taken), the analysis becomes clean and tractable with an exponential discount factor. Observe that in the infinite horizon case in which reward in all time-steps is evaluated equally, all sequences of decisions yield utility ∞ over time.

Decision policies and MDP values

In a dynamic setting with sequences of decisions made over time, we must specify decision *policies*, i.e., ways of mapping information to decisions. A decision policy $\pi : S \to A$ specifies an action $\pi(s)$ to be taken in every state s.

Given any decision policy π, we can compute the expected total discounted utility $V^\pi(s)$ that will be achieved from any state s; i.e., letting s^t be a random variable denoting the state reached at a time $t > 0$:

$$\forall s \in S,\ V^\pi(s) = \mathbb{E}\Big[\sum_{t=0} \gamma^t r(s^t, \pi(s^t)) \,\Big|\, s^0 = s\Big] \tag{4.1}$$

I will use notation π^* for an *optimal* policy, i.e., one which maximizes expected discounted value going forward from any state. Letting Π denote the space of all decision

be "blown up" such there is a unique state for each history, when history matters.

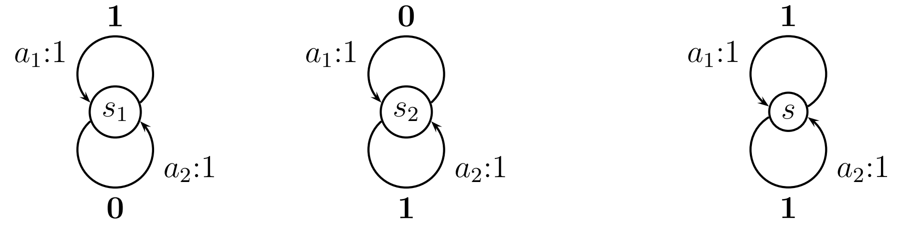
policies,

$$\pi^* \in \arg\max_{\pi \in \Pi} V^\pi(s),\ \forall s \in S \tag{4.2}$$

I will use V^* as shorthand for V^{π^*}, the expected value that is obtained under the optimal policy. V^* is also sometimes called the "MDP value".

4.2.1 Multi-agent MDPs

In multi-agent decision making scenarios it is conceptually convenient to consider the problem distinctly from each agent's perspective, with a different MDP associated with each agent. For instance, consider the following extremely simple problem: imagine that there is a decision to be made repeatedly with two possible outcomes and two agents; the scenario is the same at each time-step, no matter how many decisions have already occurred. Each outcome yields value 1 for one agent and 0 for the other agent. This scenario can be represented graphically via two local agent MDPs as in the first frame of Figure 4.4.



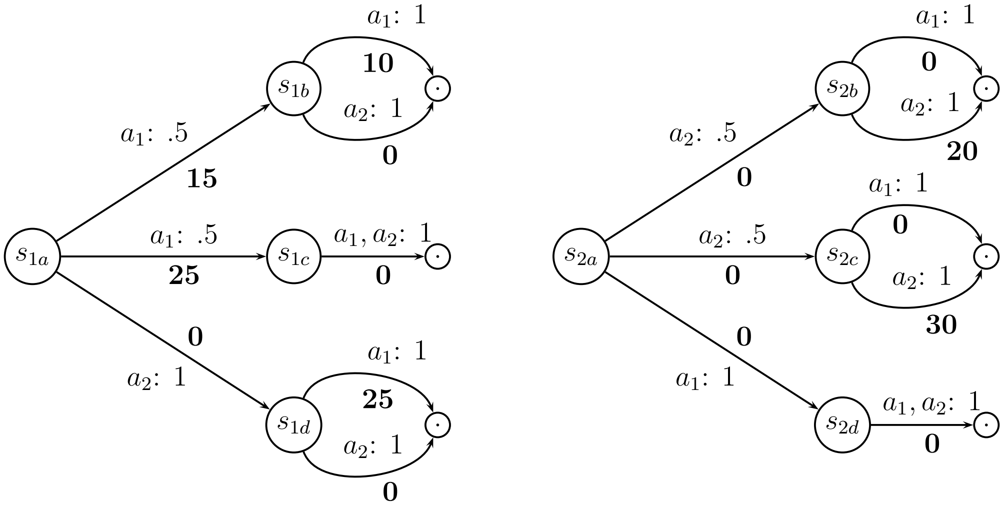
(a) Two agent MDPs.

(b) The joint MDP.

Figure 4.4: A simple two-agent decision-making scenario represented via two agent MDPs (left) and the joint MDP representing the social problem they induce.

The center's problem can be represented as a *joint MDP* formed by combining the various agent *local* MDPs. Let $M_i = \{S_i, A, r_i, \tau_i\}$ denote any agent i's (local) MDP model of the problem. The joint MDP is $M = \{S, A, r, \tau\}$, where: $S = \times_{i \in I} S_i$ is a joint state space, with $s \in S$ specifying a local state s_i for every $i \in I$; A is the decision space; $\forall s \in S, a \in A$, $r(s, a) = \sum_{i \in I} r_i(s_i, a)$, and $\tau(s, a, s')$ is the probability that joint state s' will result if action a is taken in joint state s (with $\tau(s, a)$ the random variable notation as before).

Assuming a common discount factor γ across all agents, given a set of local MDPs $\{M_1, \ldots, M_n\}$, "solving" the corresponding joint MDP M will yield a *socially* optimal policy π^*. However, in the worst case there is an exponential blow-up in the number of states, so solving the joint MDP may not be easy. Consider again the example in which a car is to be allocated two days in a row. Assume now that there are just two agents, whose valuations for the two decisions are represented in Figure 4.5. a_i represents allocating the car to agent i.



(a) Agent 1

(b) Agent 2

Figure 4.5: Two local agent MDPs: each represents a different agent's value structure (with uncertainty) for allocation of a car over a two-day period, once on Friday and once on Saturday.

Agent 1 will get total value 25 if allocated the car for both days or just the second day, and if allocated for just Friday will get either value 15 or the full 25 (depending on chance). Agent 2 will get value either 25 or 30 if he obtains the car both days; otherwise he gets no value. While each local MDP in this example has only 4 states, the joint MDP will have 11: $S = \{(s_{1a}, s_{2a}), (s_{1b}, s_{2b}), (s_{1b}, s_{2c}), (s_{1b}, s_{2d}), (s_{1c}, s_{2b}), \ldots\}$. The action space $A = \{a_1, a_2\}$ and both actions are available from every state. For γ close to 1, the optimal policy specifies action a_2 on Friday (i.e., in state (s_{1a}, s_{2a})), a_2 on Saturday if state (s_{1a}, s_{2c}) results, and a_1 on Saturday otherwise (i.e., if state (s_{1a}, s_{2b}) results). If γ is low enough it will instead be optimal to allocate only to agent 1 since he will obtain reward on Friday while the payoff for agent 2 comes only after Saturday and thus requires some "patience".

Though I will restrict attention in this thesis to scenarios in which actions are only taken by the center, in some scenarios "decisions" will in fact specify *local* actions for each agent to take (see, e.g., [Boutilier, 1996]). Reaching efficient outcomes will require two things: 1) agents to truthfully report private information so that the center can compute an optimal policy, and 2) agents to take local actions according to the computed optimal policy and as communicated by the center. Note that in such cases each agent i will have a distinct action space, and joint action space $A = A_1 \times \ldots \times A_n$. This generalizes the type of centralized setting we've seen in cases like the car allocation example, which can be modeled under this framework by assuming each local action space is identical; but the agents—rather than autonomously executing actions—are simply subject to the effects of actions executed by the center.

4.3 Computing optimal policies

Given a joint MDP for a multi-agent decision making problem, determining an efficient decision policy amounts to "solving" the MDP. Though this task can be computationally difficult (depending on the size of the MDP), in theory a solution can always be achieved via dynamic programming based methods. There are a few notable algorithms, all of which leverage the fact that the value function V^* associated with any optimal policy π^* satisfies the so-called "Bellman equations":

$$V^*(s) = \max_{a \in A} \Big[r(s,a) + \gamma \sum_{s' \in S} \tau(s,a,s') \cdot V^*(s') \Big], \ \forall s \in S \tag{4.3}$$

This fact implies that in order to find an optimal policy π^* it is sufficient to compute V^*. See [Sutton and Barto, 1998] for a very nice introduction and detailed discussion of the theory and algorithms touched upon in this section.

Policy iteration

Policy iteration is one algorithm for finding an optimal policy. The idea is to start with an arbitrary policy π, compute the expected value it will yield, and then check if there is any state s for which switching $\pi(s)$ with some other action yields higher expected value. If so, then make that switch, go back to the evaluation step, and loop. These greedy improvements ultimately lead to an optimal policy—once the looping stops, by definition the action specified for each state is the best.

Value iteration

In value iteration the evaluation and improvement steps of policy iteration are combined into one sweep through the set of states. Progressively more accurate "estimates" of V^* are computed by using the Bellman equation as an update rule. Specifically, value iteration proceeds in a number of "rounds", where in each round k an estimate $V^k(s)$ of $V^*(s)$ is formed for each $s \in S$. Starting with an arbitrary value function estimate V^0, we let:

$$V^k(s) = \max_{a \in A} \Big[r(s,a) + \gamma \sum_{s' \in S} \tau(s,a,s') \cdot V^{k-1}(s') \Big], \ \ \forall k > 0, s \in S \qquad (4.4)$$

When $V^k(s) = V^{k-1}(s), \forall s \in S$, then V^k must equal V^* and determining π^* is a simple matter of reading off the optimal action according to the Bellman equation for each state.[3]

4.3.1 Approximate and online algorithms

The mechanisms we will see in the chapters that follow provide incentives for agents to participate truthfully in a dynamic decision-making environment, given that the center is executing an *optimal* decision policy. But due to the exponential blow-up in the size of the state space that occurs in multi-agent problems, often it will not be computationally tractable to compute such policies. In such cases approximate algorithms are often employed. Also, sometimes *online algorithms* are used, wherein actions are taken in the world and policies are improved step-by-step, in some cases ultimately converging to (but not starting with) an optimal policy. Here, I briefly mention some the most notable of these techniques.

Linear programming

There is a natural exact linear programming (LP) formulation of the problem of computing an optimal policy in sequential environments (see, e.g., [Puterman, 1994]). Specifically, one can formulate an objective function of *minimizing* the sum of $V^*(s)$ over all $s \in S$, and add a constraint for each $s \in S$ that essentially specifies that the right hand side of the Bellman equation for s is greater than or equal to $V^*(s)$. An LP formulation can then leverage high-powered commercial solvers such as CPLEX (see www.ilog.com).

But in non-trivial practical cases (especially involving multiple agents) the state-space will be massive, leading to a prohibitive number of state variables and constraints in the LP. De Farias & Roy [2003] have recently made progress in linear programming for approximate dynamic programming problems like this (see [Schweitzer and Seidmann, 1985] for an initial proposal in this space). Their approach fits a linear combination of pre-selected basis functions to the objective function, and they obtain analytical bounds on the error of the approximate solution given the choice of basis functions.

[3] In practice the process would be stopped when $V^k(s) - V^{k-1}(s)$ is less than some tolerance level ε.

Sparse sampling

Kearns, Mansour, & Ng [1999] provide a "near-optimal" algorithm for sequential decision making that works with a "generative model" of an MDP—i.e., a black box that, given any state-action pair, returns a successor state in accordance with the transition function of the MDP. Given any "current state", their algorithm uses such a generative model to produce a relatively small number of potential paths forward (sparse sampling), and then uses these samples as the basis for determining an action to take. This process results in a near-optimal policy. Though the run-time of the algorithm is exponential in the time-horizon, it is completely independent of the size of the state space and thus evades the "curse of dimensionality" inherent in most multi-agent problems.

Q-learning

Q-learning, though still subject to critical challenges if the state space is very large, is an online method of simultaneously taking actions and learning that eventually converges to an optimal policy. The algorithm maintains a table of values, one for every state-action pair (s, a), representing an estimate of the total discounted value that will be obtained (starting from state s) if action a is taken in state s and an optimal policy is followed in all future time-steps. Let $Q(s, a)$ denote the entry in the table for pair (s, a). Q-learning specifies that at every time-step, from any current state s, an action $a^* = \arg\max_{a \in A} Q(s, a)$ is taken; when new state s' is observed, the (s, a^*) entry in the table is updated as follows:

$$Q(s, a^*) = Q(s, a^*) + \alpha[r(s, a^*) + \gamma \max_{a \in A} Q(s', a) - Q(s, a^*)] \tag{4.5}$$

α is a "learning rate" determining how new information is weighted against old estimates. As α approaches 1, the "old" estimate is completely updated with the new estimate every time-period. As α approaches 0, values are never updated at all. Typically α is gradually decreased over time. Notably, letting α_t denote the choice of α at time t, if $\sum_{t=1}^{\infty} \alpha_t = \infty$ and $\sum_{t=1}^{\infty} \alpha_t^2 = \infty$ then Q-learning is guaranteed to converge to the true optimal values V^* so long as all state-action pairs are visited infinitely often [Watkins, 1989]. For instance if α^t is set to $1/t$ these conditions are satisfied from any t forward.

Interestingly, Q-learning will converge to the optimal values V^* even if one does not choose actions according to $\arg\max_{a \in A} Q(s, a)$. In fact actions can be chosen arbitrarily and convergence will still occur. Note that this also implies that Q-learning is "model-free", i.e., it will converge to an optimal policy even when the decision-maker has no knowledge of the reward or transition functions that define the underlying MDP—these are implicitly learned and encoded over time in the table of Q values that is constructed.

E^3

The “Explicit Explore or Exploit” (E^3) algorithm of Kearns & Singh [1998] provides a mechanism for reinforcement learning (or “solving” an MDP) that is guaranteed to reach a near-optimal policy in a *finite* amount of time. The algorithm is designed to perform a mix of targeted exploration of the state space and offline optimization. The algorithm thus explicitly addresses the exploration-exploitation tradeoff; there is a polynomial bound (in the time-horizon K and number of states S) on the number of actions taken and computation performed in order to achieve near-optimal performance.

4.4 Multi-armed bandits: an important special case

In general multi-agent decision-making gets computationally hard very quickly. Even for relatively basic problems in which each agent’s local MDP model has only 10 states, if there are 10 agents then there are 10^{10} joint states. But there is an important subclass of decision-making problems that does not suffer from this blow-up in state space, consisting of those that can be modeled as *multi-armed bandits*.

In a multi-armed bandit (MAB) problem[4] one must choose to activate, at every time-period, one among a set of *independent Markov chains*. A Markov chain (MC) is an MDP in which there is only one “action” available from each state. So in a single Markov chain MDP there is thus no problem of which action to take; but in a MAB there *is* a choice: since only one MC can be activated at a time, which should be chosen when in order to maximize total discounted expected social reward over time?

In a MAB the Markov chains are independent in that when one is activated, the states of the others remain unchanged. So, letting $s^t = (s_1^t, \ldots, s_n^t)$ denote the joint state of n Markov chains at time t and letting $\tau(s_i)$ be the random variable representing the local state of process i after being activated once from state s_i, if we execute a policy π with $\pi(s^t) = i$ then the resultant joint state at time $t+1$ will be s^{t+1} with:

$$\begin{aligned} s_i^{t+1} &= \tau(s_i^{t+1}) \\ s_j^{t+1} &= s_j^t, \qquad \forall j \neq i \end{aligned}$$

[4] See [Bellman, 1956] for an early study, and [Berry and Fristedt, 1985] for an in-depth analysis of various bandit-style problems.

Optimal Bayesian-learning

Among the many classes of problems that can be modeled as MABs, one stands out as particularly intriguing: that of optimally *learning* how to choose among a set of reward generating processes to obtain highest reward. Given some prior beliefs about the reward each process will generate, the goal is to optimize in a Bayesian sense: to activate processes in a way that is optimal at every time-step given current beliefs.

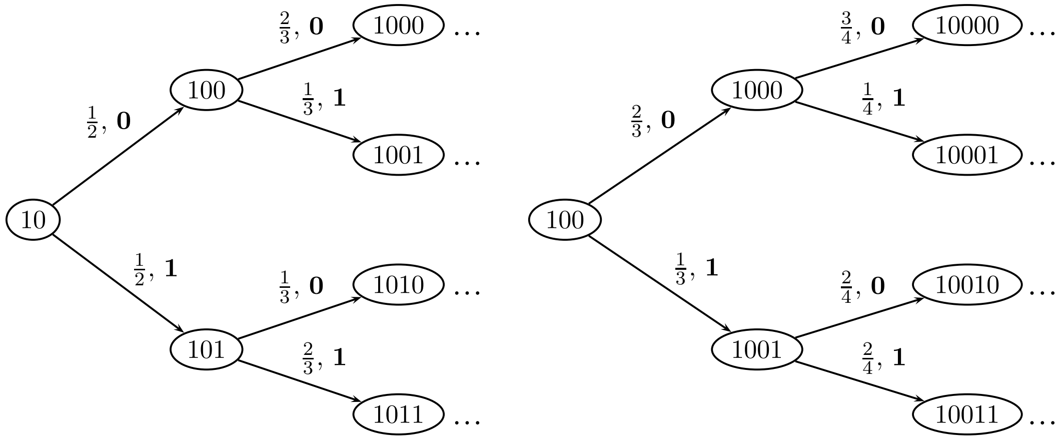
For instance, consider a scenario in which an oil company is deciding in which among a set of regions to drill for oil. Say the oil company only has resources to drill in one region at a time. There is uncertainty as to the amount of oil that each region contains. Every time drilling is performed in a given region, there is feedback regarding the oil in that region (i.e., in the Markov chain representing beliefs about how rich in oil that region is, there is a state change); beliefs about regions that were not drilled in do not change. In such a setting an optimal policy will trade off the benefits of learning about the potential yield from regions that are not well-known (exploration) with capitalizing on the information already obtained about regions that are known to have high yield (exploitation).

It is from analogy to another setting of this kind that multi-armed bandit problems get their name. Imagine a gambler choosing among a set of slot machines ("one-armed bandits") to play at each point in time, where each slot machine has a distinct payout rate. The gambler will want to make the series of selections that maximizes his expected discounted reward over time, which will involve both using information he already has and obtaining more information about relatively unknown slot machines (arms) in order to avoid missing a potential goldmine.

Figure 4.6 depicts two "Bernoulli bandit" process, where each process yields reward either 1 or 0 (i.e., it either "succeeds" or "fails") every time it is activated and has some unknown rate of success p. A state in a Bernoulli bandit can be represented as a beta distribution (α, β) with parameter α corresponding to the number of successes observed plus 1, and β corresponding to the number of failures observed plus 1. So process 1 in state 10 (state "one zero", not "ten") represents the uniform beta distribution prior of $(\alpha, \beta) = (1, 1)$. Process 2 in state 100 denotes that a single failure (0) has been observed, and so $(\alpha, \beta) = (1, 2)$. The expected success rate given a beta distribution with parameters (α, β) is $\frac{\alpha}{\alpha+\beta}$, so the estimated success rate of process 2 in state 100 is 1/3. Thus with probability 2/3 the successor state will be 1000 and with probability 1/3 it will be 1001.

A multi-agent interpretation

It is natural to consider a multi-agent interpretation of the multi-armed bandit problem, in which each process is associated with an agent, and when an agent's process is activated it generates some private reward for that agent. This is exactly



(a) Process 1

(b) Process 2

Figure 4.6: A two-armed Bernoulli bandit problem. Each process yields reward either 0 or 1 upon every activation, where the probability of reward 1 is stationary over time but unknown. A state representation (roughly) corresponds to the number of 0 and 1 rewards the process has yielded so far.

the multi-agent MDP model we discussed in Section 4.2.1, but with specific limitations placed on the types of MDPs that agent world models correspond to. We can model these limitation specifically as follows: for each agent i, for each local state s_i, there is exactly one action a_i (taken by either the center or the agent) that yields a state transition and non-zero reward; all other actions self-loop back to the current state and yield no reward. Moreover, the only joint actions that can be taken specify a_i for exactly one $i \in I$.

A very natural problem domain that fits this model is *repeated allocation of a single item* in an infinite horizon environment. An item is allocated to one agent for one time-step, and in the next time-step it is reallocated, and so on. The agent that receives the item in any given time-step may obtain non-zero reward and possible also learn something about his value, but all other agents receive zero reward and their states do not change.

The infinite-horizon version of the problem is better modeled as a MAB than the finite-horizon version since in the finite case every time an agent is not allocated the item something does change—the amount of possible future times it could receive the item. For instance in Figure 4.5 agent 2 obtains value only if he gets to use the car for 2 days; so in a 2-day setting if he is not allocated the car on the first day there is no way he can obtain value—his state changes.

4.4.1 Gittins's solution

In the late 1970s John C. Gittins achieved one of the seminal results in the area of sequential decision-making: he proved that solving multi-armed bandits problems is computationally tractable. Specifically, he demonstrated that in any MAB, an optimal policy can be constructed by: 1) computing an *index* (henceforth called the "Gittins index") for each process that depends only on the current state of that process (and *not* on any other processes), and 2) activating the process with the highest index.

Theorem 4.1. [Gittins and Jones, 1974; Gittins, 1989] *Given Markov processes* $\{1, \ldots, n\}$*, joint state space* $S = S_1 \times \ldots \times S_n$*, discount factor* $0 \leq \gamma < 1$*, and an infinite time-horizon, there exists a function* $\nu : S_1 \cup \ldots \cup S_n \rightarrow \Re$ *such that the optimal policy* $\pi^*(s)$ *specifies activation of* $\arg\max_i \nu(s_i)$*,* $\forall s \in S$*.*

Since the index for each process does not depend on the other processes, as the number of processes increases there is only a linear increase in computation required—one extra process requires only computing and comparing one extra Gittins index. So in the multi-agent interpretation we have:

Corollary 4.1. *In an infinite-horizon dynamic multi-agent decision-making scenario in which each agent's local world model is a Markov chain and only one can be activated per time-step, the computation required to compute an efficient decision policy grows only linearly in the number of agents.*

Gittins defined what is now known as the Gittins index (which he originally called the "dynamic allocation index") as follows.

Definition 4.1 (Gittins index). *The Gittins index of a process* i *in state* s_i *is:*

$$\nu(s_i) = \sup_{\rho > 0} \frac{\mathbb{E}\Big[\sum_{t=0}^{\rho-1} \gamma^t r(s_i^t, a_i) \mid s_i^0 = s_i\Big]}{\mathbb{E}\Big[\sum_{t=0}^{\rho-1} \gamma^t \mid s_i^0 = s_i\Big]} \tag{4.6}$$

where ρ *is a stopping time,* s_i^t *is a random variable denoting the state of process* i *after being activated* t *times starting from state* s_i*, and* $r(s_i^t, a_i)$ *is the reward yielded when process* i *is activated in state* s_i^t*.*

Intuitively, a stopping time ρ specifies "stop" or "go" for every possible state that could be reached. So the Gittins index is the expected total reward divided by discounted time that goes by when a stopping policy is selected to maximize that ratio.

Gittins first proved Theorem 4.1 via a so-called "interchange argument", explicitly comparing the expected value from any alternative sequence of activations against the sequence that results from always choosing the highest Gittins-indexed process. In the time that has passed since his proof other proofs have been found that do not rely on the analytical form of the Gittins index in Equation 4.6, notably [Weber, 1992], [Whittle, 1980], and [Bertsimas and Nino-Mora, 1996]. Frostig & Weiss [1999] provide an excellent presentation of all 4 of these proofs.

Conceptualizing the Gittins index as an indifference point

Whittle's [1980] proof of Gittins's theorem provides an important insight into the essence of what the Gittins index captures. Imagine a scenario in which there is a single process i, and that at every point in time the decision-maker has the option to permanently "retire" from playing i and receive one-time lump sum μ (constant over time). The decision-maker must trade-off the possibility that continuing activation of i will yield high reward in the future with the chance to get a big reward μ *now.* The expected value obtained from playing optimally in such a scenario given state s_i can be expressed as follows:

$$V(s_i, \mu) = \max \left[r(s_i, i) + \gamma \sum_{s'_i \in S_i} \tau(s_i, a, s'_i) \cdot V(s'_i, \mu),\ \mu \right] \tag{4.7}$$

When μ is less than the lowest reward i could possibly generate it is obviously optimal to always keep playing; when μ is very large it is optimal to stop. There is some point in between at which both stopping and continuing are optimal choices. It is at this value that μ equals the Gittins index (scaled by a function of the discount factor). Let $\mu^*(s_i)$ be the value of μ at which $r(s_i, i) + \gamma \sum_{s'_i \in S_i} \tau(s_i, a, s'_i) \cdot V(s'_i, \mu) = \mu$. Then:

$$\nu(s_i) = (1 - \gamma)\mu^*(s_i) \tag{4.8}$$

So the Gittins index can alternatively be described as ($(1-\gamma)$ times) the point at which one would be indifferent between playing and retiring, given that the option for retirement is always available in the future. For instance in the extremely simple example of Figure 4.7, if $\gamma = 0.75$, the indifference point is μ such that $\mu = 1 + 0.5 \cdot 0.75\mu + 0.5 \cdot \frac{0.75}{1-0.75}$. Solving for μ yields 4, so the Gittins index equals $(1-0.75) \cdot 4 = 1$.[5]

Computing Gittins indices

Though Gittins's result demonstrates that the required computation in determining optimal MAB policies scales well with the number of agents, there is still the

[5] See also chapter 3 of Michael Duff's dissertation [Duff, 2002] for a nice presentation of the "indifference point" interpretation of the Gittins index and related issues.

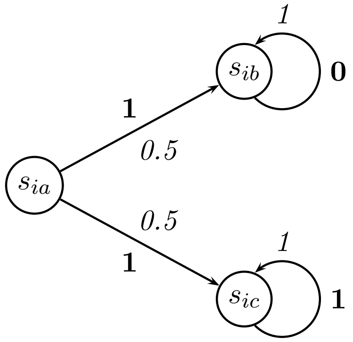
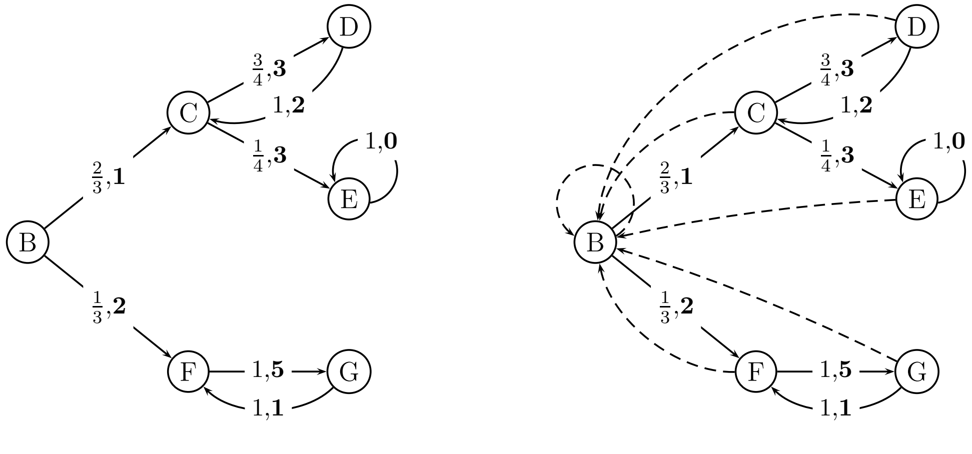


Figure 4.7: A very simple Markov chain process; after the first non-deterministic transition all future rewards will either be 0 or 1. The lump-sum reward μ that makes a decision-maker indifferent between playing this process or retiring for μ (where the option to retire for μ persists in the future) is 4.

matter of actually computing the Gittins index for a given process in a given state. In some domains there is structure that makes the computation easier. For instance, in a Bernoulli bandit scenario where beliefs are represented via a beta distribution one can feasibly compute a very close approximation of the index numerically via explicitly expanding out Equation 4.7 for a "guessed" μ value, where the error in the approximation caused by expanding out only a finite number of levels is bounded via the discount factor (any reward received after t time steps will yield value $\leq \frac{\gamma^t}{1-\gamma}$). One can then use binary search to ultimately find the μ value that leads to indifference, thus obtaining the Gittins index. Gittins [1989] discusses several methods for estimating Gittins indices in other structured environments.

Katehakis & Veinott [1987] provide a method for computing Gittins indices for arbitrary finite-state Markov chains (i.e., making no assumptions about structure) which allows appeal to well studied algorithms for solving MDPs to compute the Gittins index for a process in a given state. Consider a set of "restart-in-i" MDPs, defined for each process at each time-step as follows: There is a state in the restart-in-s_i MDP for process i at time t corresponding to each possible sequence of states that could be visited from s_i forward. From any state, two actions are possible: *restart* and *continue*. If *restart* is selected, the process transitions instantaneously (no delay) to state s_i with probability 1 and then also makes a transition and receives a reward as though activated in s_i. If *continue* is selected, the MDP proceeds (makes a transition and receives a reward) as though activated in the associated state. Figure 1 depicts the restart-in-s_i MDP for a 6-state bandit process currently in state B.

Katehakis and Veinott showed that the value of state s_i in the restart-in-s_i MDP gives the Gittins index of a process in s_i. Thus, at every time-step, given joint state s, the optimal process to activate can be determined by solving n MDPs: the restart-in-s_i process for each i.



(a) Markov process

(b) Restart-in-B MDP

Figure 4.8: A Markov chain and the corresponding Katehakis & Veinott "restart-in-B" MDP, the optimal value of which gives the Gittins index. The instantaneous *restart* transitions are represented by dashed lines. *Continue* transitions and the transition that occurs from B immediately after a *restart* action are represented by solid lines.

4.5 Summary

In this chapter I presented the computational aspects of reaching efficient decisions in dynamic, multi-agent environments; I put the mechanism design (incentive & budget) issues to one side—they'll be addressed in the following chapters. We saw that Markov decision processes (MDPs) provide an appropriate representational formalism for dynamic decision problems, and that it is natural to model the decision problem from each agent's perspective as a distinct "local" MDP. We saw how local MDPs can be combined into a joint MDP, and that solving this MDP leads to a decision policy that maximizes social welfare. We saw various algorithms for solving MDPs: some that lead to optimal decisions at every point in time (policy-iteration, value-iteration), and some approximate or online algorithms (e.g., sparse-sampling, Q-learning) that either provide bounds on the suboptimality they achieve, or reach an optimal policy in infinite-horizon settings after executing for long enough. Finally, I presented the very important subclass of dynamic decision problems that can be represented as multi-armed bandits—a class that includes repeated allocation of a single good. In these settings determining an optimal policy is computationally tractable (via "Gittins indices"); as the number of agents in the system grows the necessary compute time likewise grows only linearly.

Chapter 5

Dynamic mechanism design

Synopsis[*]

In the last chapter we saw how preferences can be represented and efficient decision policies can be computed in dynamic decision problems. We now turn to incentive issues: for the center to implement an efficient policy the participation of the agents is typically required. In Chapter 2 we saw foundational results in eliciting truthful participation of agents in one-shot settings (static mechanism design); in this chapter I present *dynamic mechanism design*, an extension of the mechanism design paradigm to scenarios in which a *sequence* of decisions is to be made over time, with new private information potentially arriving at every time-step.

We will see that many of the foundational results of static mechanism design have rather direct, though significantly more complex, analogues in the dynamic setting. Perhaps most centrally:

- The Groves class of mechanisms has a natural analogue in dynamic settings, and this class completely characterizes the set of efficient mechanisms that can be implemented in a strong dynamic truthtelling equilibrium (with a reasonable restriction).

We will also see important extensions of results regarding particular static mechanisms to the dynamic case:

- The VCG mechanism has a natural analogue in dynamic settings—dynamic-VCG, due to Bergemann & Välimäki [2006]—and this mechanism yields equilibrium utility for each agent equal to his marginal contribution to social welfare.

- The dynamic-VCG mechanism is revenue-maximizing among all mechanisms that are efficient in a strong truthtelling dynamic equilibrium.

[*] Many results from this chapter have appeared in the following papers: [Cavallo, 2008], [Cavallo *et al.*, 2006], and [Cavallo *et al.*, 2007], the latter two of which are collaborative with David C. Parkes and Satinder Singh.

- The AGV mechanism has a natural analogue in dynamic settings—the dynamic-balanced mechanism, due to Athey & Segal [2007]—and this mechanism is strongly budget balanced but achieves weaker incentive and participation properties.

In Chapter 6 (as in Chapter 3) I will observe that revenue is waste in many environments, and propose a *dynamic redistribution mechanism* that goes a long way towards minimizing this waste in some important settings. Before getting to these results I will properly introduce dynamic mechanism design, including the notion of a dynamic type and equilibrium concepts.

5.1 Dynamic mechanism design defined

Dynamic mechanism design addresses settings in which there is a group of agents and a sequence of decisions or "actions" to be taken, one per time-step, that bear utility for the agents. At each time-step, each agent has some private information that determines the expected value he would obtain for every possible action that could be taken in the current time-step, and also determines a probability distribution over future private information, given any future sequence of decisions. When an action is taken, new private information potentially arrives for each agent.

To formalize these notions I use the Markov decision process (MDP) framework. Each agent's type in a dynamic setting corresponds to an MDP along with an indication of the current state, and only the current state changes over time—the MDP model stays the same. Formally, there is a set of agents $I = \{1, \ldots, n\}$, an exogenously defined action space[1] A, with actions to be chosen from A for K successive time periods (where K is potentially infinite), and a space of agent *types* $\Theta_1 \times \ldots \times \Theta_n$. Each agent i's type at time t, $\theta_i^t \in \Theta_i$, induces a tuple $(s_{\theta_i^t}, r_{\theta_i^t}, \tau_{\theta_i^t})$ that represents i's private information at t. There is a local state space S_i defined by Θ_i, and for type θ_i^t, $s_{\theta_i^t}$ is the "current" local state. $r_{\theta_i^t} : S_i \times A \times S_i \rightarrow \Re$ is a value (reward) function, with $r_{\theta_i^t}(s_i, a, s_i')$ denoting the immediate value that i obtains if action a is taken when i is in local state s_i and new state s_i' results. $\tau_{\theta_i^t} : S_i \times A \times S_i \rightarrow \Re$ is a probability function, with $\tau_{\theta_i^t}(s_i, a, s_i')$ denoting the probability that taking action a while i is in local state s_i will yield new local state s_i' for i in the next period.

Given any $\theta_i^t \in \Theta_i$, in this way A, S_i, $r_{\theta_i^t}$, and $\tau_{\theta_i^t}$ define a local MDP for agent i. In dynamic mechanism design the *center* elicits reports from each agent regarding private types in every period, and then takes an action. I let $\theta_c^t \in \Theta_c$ denote the "type" of the center at time t. This represents any information known to the center, and can for instance include the history of reported types, the index of the current time-period, etc. I then denote the joint type space $\Theta = \Theta_c \times \Theta_1 \times \ldots \Theta_n$. As

[1] This is analogous to there being a fixed outcome space O in a static setting, independent of the behavior of the participants.

a notational simplification, for any $\theta \in \Theta$ I write $\tau(\theta, a)$ for the random variable representing the joint type in Θ that results when action a is taken in $s_\theta \in S$. I write $r_i(\theta_i, a)$ for the expected immediate value to i when a is taken and i's type is θ_i, i.e., $\sum_{s'_i \in S} \tau(s_{\theta_i}, a, s'_i)\, r_{\theta_i}(s_{\theta_i}, a, s'_i)$. I then write $r(\theta, a)$ for $\sum_{i \in I} r_i(\theta_i, a)$ (the immediate expected *social* value of taking action a in the joint state s_θ), and $r_{-i}(\theta_{-i}, a)$ for $\sum_{j \in I \setminus \{i\}} r_j(\theta_j, a)$ (the immediate value to agents other than i). I assume agents exponentially discount future reward at rate $\gamma \in [0, 1)$, so a reward of x received t steps in the future is valued at $\gamma^t x$.

Note that this set-up technically places us in a private values setting, excluding scenarios where an agent's value for an action depends on the private information of other agents.[2] But I still allow for *serial correlation* of types, where, e.g., the fact that an agent i has transitioned from some type θ_i to θ'_i allows us to know with certainty that if j's type in the last period were θ_j his current state would be some θ''_j. This can be modeled by explicitly considering a stochastic process φ representing the (random) events of nature; then for each i, θ_i, and a, we have $\tau(\theta_i, a, \varphi)$ representing the transition—this allows a coupling (only through the realization of random events) of agent type transitions. For simplicity of exposition I will omit φ from the notation going forward.

There is a dynamic version of the revelation principle due to Myerson [1986][3], so I restrict attention to *direct dynamic mechanisms*:

Definition 5.1 (direct dynamic mechanism). *A tuple* (π, T)*, where:*

- $\pi : \Theta \rightarrow A$ *is a decision policy.*
- $T = (T_1, \ldots, T_n)$*, where for each* $i \in I$*,*
  $T_i : \Theta \rightarrow \Re$ *is a transfer payment function.*

In a direct dynamic mechanism, in every time-step each agent makes a claim about his current type, an action is taken, and transfer payments are executed. Decision policy π maps a reported joint type to an action,[4] and T_i, for each $i \in I$, maps a reported joint type to a monetary payment delivered *from* the center *to* agent i. We will see that, as in static mechanism design, certain transfer payment schemes will succeed in aligning the interests of all agents towards execution of certain decision policies, such that each agent will be best off participating truthfully given that the

[2] At the end of chapter 7 I will consider relaxing this assumption, allowing one agent's expected type transitions to depend on *other* agents' current states.

[3] See also [Segal and Toikka, 2007] for a description.

[4] I will consider mechanisms in which truthtelling is an equilibrium, and thus it will only be necessary for agents to report local state s_i^t at each time t (as r_i and τ_i are constant), but formally a dynamic mechanism will allow agents to report their entire type at every period (allowing for the possibility of an agent i being truthful *in the future* from a time in which i has misreported r_i or τ_i).

center chooses actions according to the policy. I will focus on the socially optimal or *efficient* decision policy π^*, which maximizes expected discounted social value (reward) according to agent reported types.

Agents report types according to *strategies*. Let $\sigma_i : \Theta_i \rightarrow \Theta_i$ denote a reporting strategy for agent i, where $\sigma_i(\theta_i) \in \Theta_i$ is the type that i reports when his true type is θ_i. Although an agent is potentially aware of his entire *history* of types (not just the current type), this formulation is without loss of generality as whenever histories may play a role in an agent's strategy we can consider that type spaces are defined with each type θ_i^t containing a representation of i's entire type history through time t.[5] I let $\sigma = (\sigma_1, \ldots, \sigma_n)$, and for any $\theta \in \Theta$, $\sigma(\theta) = (\theta_c, \sigma(s_1), \ldots, \sigma(\theta_n))$.

An illustration of the timing of events in a dynamic mechanism is given in Figure 5.1.[6] Importantly, value (r_i) and transfers (T_i) obtained in the same period are discounted at the same rate. Time "ticks forward" when an action is taken.

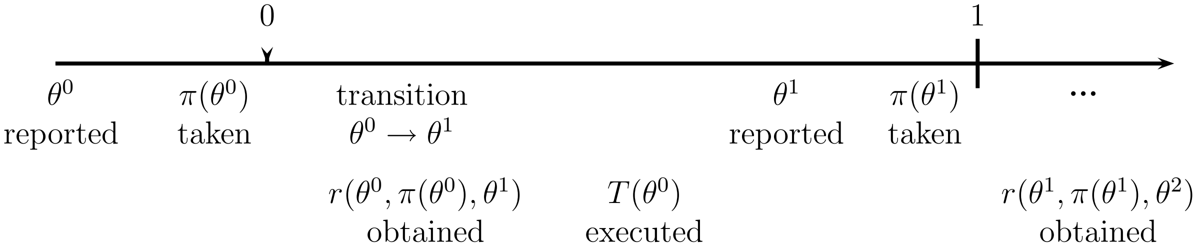

0
1
θ^0 reported
$\pi(\theta^0)$ taken
transition $\theta^0 \rightarrow \theta^1$
θ^1 reported
$\pi(\theta^1)$ taken
...
$r(\theta^0, \pi(\theta^0), \theta^1)$ obtained
$T(\theta^0)$ executed
$r(\theta^1, \pi(\theta^1), \theta^2)$ obtained

Figure 5.1: An illustration of the timing of a dynamic mechanism (π, T). The pre-time-step 0 "period" is unique in that there is no value obtained and no transfers are executed. All subsequent time-steps follow the outline of period 0 as portrayed, except at the end of time-step K (for finite-horizon problems) no action is taken.

Note that in a direct dynamic mechanism the report history (potentially tracked in θ_c) is not relevant to determining the optimal decision policy (when agents are truthful). Moreover, we will pursue an equilibrium solution concept in which agents are best off playing according to the equilibrium *no matter their previous history of reports*, so it is natural to focus on dynamic mechanisms in which the transfer functions also do not depend on histories:

Definition 5.2 (history-independent dynamic mechanism). *A history independent dynamic mechanism is a* (π, T) *such that,* $\forall \theta, \theta' \in \Theta$ *with* θ *and* θ' *identical*

[5] For instance, in a tree-structured MDP there is a unique history for every possible state.

[6] At the end of this chapter and in Chapter 7 we will see a variant in which transfers are made just *after* type reports; this will allow for stronger mechanisms when an assumption is imposed about experienced values being discernible from individual types.

except for the history of reports tracked in θ_c and θ'_c, respectively,

$$\pi(\theta) = \pi(\theta') \quad \textit{and} \quad T(\theta) = T(\theta') \tag{5.1}$$

So if the only difference between two type profiles is the history of reports, a history-independent dynamic mechanism makes the same decisions on each type profile. Important mechanisms such as dynamic-VCG and the dynamic redistribution mechanism I present in the next chapter fit this model.

Note that a one-shot setting is a special case of the dynamic framework (the number of decisions K to be made equals 1), and thus any negative results from static mechanism design will immediately apply to dynamic mechanism design if we seek mechanisms that achieve certain properties for any K. So as in the static setting in order to get traction we will assume quasilinear utility functions throughout.

Notation

I use the following notational shorthand:

- $V_i(\theta_i^t, \theta_{-i}^t, \pi, \sigma_i, \sigma_{-i})$ (or $V_i(\theta^t, \pi, \sigma)$, more concisely) is the expected discounted sum of value to be obtained by agent i in the future given that his true type is θ_i^t, the other agents' true joint type is θ_{-i}^t, the center executes decision policy π, i follows reporting strategy σ_i, and other agents follow reporting strategy profile σ_{-i}. Algebraically,

  $$V_i(\theta^t, \pi, \sigma) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} r_i(\theta_i^k, \pi(\sigma(\theta^k))) \,\Big|\, \theta^t, \pi, \sigma\Big] \tag{5.2}$$

  Here and in all other places going forward, the expectation is taken over future true types of the agents given the decision policy, and is based on current *true type* θ (not the reported type).

  When I omit σ_i or σ_{-i}, I intend that the truthful strategy is followed. When I omit π, I intend that the expectation is based on execution of π^*. So, for example,

  $$V_i(\theta) = V_i(\theta_i, \theta_{-i}, \pi^*, \sigma_i, \sigma_{-i}), \tag{5.3}$$

  where σ is the truthful strategy profile. $V_i(\theta)$ is the expected utility to agent i given joint type θ, truthful reporting by other agents, and execution of the socially optimal policy π^*. Letting Π be the set of all possible decision policies,

  $$\forall \theta^t \in \Theta,\ \pi^* \in \arg\max_{\pi \in \Pi} \sum_{i \in I} V_i(\theta^t, \pi) \tag{5.4}$$

- V_{-i} is defined analogous to V_i, but is the expected value to agents *other than* i (i.e., $V_{-i}(\,\cdot\,) = \sum_{j \in I \setminus \{i\}} V_j(\,\cdot\,)$). We will at times consider the value to agents other than i of a policy π^*_{-i} that is optimal for them, given state θ^t (i.e., $\pi^*_{-i} \in \arg\max_\pi \sum_{j \in I \setminus \{i\}} V_j(\theta^t, \pi)$). I use $V_{-i}(\theta^t_{-i})$ to denote this value when agents other than i are truthful, as it is completely independent of i's state or strategy. For any σ_i,

$$V_{-i}(\theta^t_{-i}) = V_{-i}(\theta^t_i, \theta^t_{-i}, \pi^*_{-i}, \sigma_i) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} r_{-i}(\theta^k_{-i}, \pi^*_{-i}(\theta^k_{-i}))\,\Big|\, \theta^t, \pi^*_{-i}\Big] \tag{5.5}$$

- V is defined analogous to V_i and V_{-i}, but is the expected value to all agents (i.e., $V(\,\cdot\,) = \sum_{i \in I} V_i(\,\cdot\,)$).

- $\mathcal{T}_i(\theta^t_i, \theta^t_{-i}, \pi, \sigma_i, \sigma_{-i})$ (more concisely, $\mathcal{T}_i(\theta^t, \pi, \sigma)$) is the expected discounted sum of transfer payments received by agent i under a dynamic mechanism (π, T).

$$\mathcal{T}_i(\theta^t,\pi, \sigma) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} T_i(\sigma(\theta^k))\,\Big|\, \theta^t, \pi, \sigma\Big] \tag{5.6}$$

  Variants for expected transfers received by agents other than i, under truthful reporting, and under policy π^* are defined analogous to the V notation.

Given this notation, the restriction to quasilinear utility functions can be expressed as an assumption that each agent i's total expected discounted utility given mechanism (π, T) executed forward from a joint state θ^t given that agents play strategy profile σ is:

$$V_i(\theta^t, \pi, \sigma) + \mathcal{T}_i(\theta^t, \pi, \sigma) \tag{5.7}$$

5.1.1 Dynamic equilibrium concepts and properties

The goal in dynamic mechanism design is to achieve implementation of desirable decision policies—typically, the efficient policy π^*—in a game theoretic equilibrium. In the static setting we saw that the Groves class of mechanisms achieves truthfulness and efficiency in dominant strategies, i.e., truthtelling is utility-maximizing for each agent "no matter what". The strongest notion one could imagine in a dynamic setting would be for truthfulness to always maximize the utility an agent receives, no matter what other agents do, now or in the future. But there cannot be an equilibrium notion that is quite "no matter what" in dynamic environments with uncertainty. One can maximize *expected* utility given expectations about how future types will be realized, but without knowing every random realization that would occur for every hypothetical action, one cannot act in a way that would never lead to hindsight regret.

So we take as our goal to achieve an equilibrium in which each agent is best off playing the equilibrium strategy when others do, in expectation "knowing everything that is knowable"[7] (i.e., knowing the private types of other agents, whatever they are, but not future state transitions). This is exactly what the *within-period ex post Nash equilibrium* concept describes, where there is a strategy profile in which each agent maximizes his payoff (expected discounted utility) by playing the equilibrium strategy, given that the other agents do, for *every possible joint type.*

Definition 5.3 (within-period ex post Nash equilibrium). *Given dynamic mechanism (π, T), a strategy profile σ constitutes a within-period ex post Nash equilibrium if and only if, at all times t, for all agents $i \in I$, for all possible true types $\theta^t \in \Theta$, and for all σ'_i,*

$$V_i(\theta^t, \pi, \sigma_i, \sigma_{-i}) + \mathcal{T}_i(\theta^t, \pi, \sigma_i, \sigma_{-i}) \geq \tag{5.8}$$

$$V_i(\theta^t, \pi, \sigma'_i, \sigma_{-i}) + \mathcal{T}_i(\theta^t, \pi, \sigma'_i, \sigma_{-i}) \tag{5.9}$$

A mechanism is *incentive compatible (IC)* in this equilibrium if each agent maximizes his payoff by reporting truthfully when others do, for every possible joint type. A mechanism is *individual rational (IR)* in this equilibrium if each agent's payoff is non-negative in expectation from any possible joint type, given that agents play equilibrium strategies.

Definition 5.4 (within-period ex post incentive compatible). *A dynamic mechanism (π, T) is within-period ex post incentive compatible if and only if, at all times t, for all agents $i \in I$, for all possible true types $\theta^t \in \Theta$, and for all σ_i,*

$$V_i(\theta^t, \pi) + \mathcal{T}_i(\theta^t, \pi) \geq V_i(\theta^t, \pi, \sigma_i) + \mathcal{T}_i(\theta^t, \pi, \sigma_i) \tag{5.10}$$

One can (and I will) alternatively refer to a within-period ex post incentive compatible mechanism as "truthful in within-period ex post Nash equilibrium".

Note that in a within-period ex post incentive compatible mechanism agents will have the incentive to be truthful going forward (when others are) from any state no matter what has happened in the past, and in particular even if they have previously deviated from truth. It may initially surprise some that we are in a private values setting, yet ex post incentive compatibility does not generally yield truthtelling as a best-response to *non-truthful* strategies on the part of other agents. To see why, consider a dynamic setting in which an agent i will misreport type information in the current period (and will be truthful subsequently), leading to an action that restricts the possibility for high social value in future periods. Assume payments have aligned all agents' incentives towards maximizing social welfare. An agent $j \neq i$ may benefit

[7] Thanks to Susan Athey and David Miller for this nicely descriptive phrasing; see also [Athey and Segal, 2007].

from reporting a false type to mitigate or counterbalance i's misreport—the two misreports combined may restore the efficient decision. Thus within-period ex post incentive compatibility is really a gold standard for dynamic settings.[8]

Definition 5.5 (within-period ex post individual rational). *A dynamic mechanism (π, T) is within-period ex post individual rational if and only if there exists a within-period ex post Nash equilibrium strategy profile σ such that at all times t, for all agents $i \in I$, for all possible true types $\theta^t \in \Theta$,*

$$V_i(\theta^t, \pi, \sigma) + \mathcal{T}_i(\theta^t, \pi, \sigma) \geq 0 \tag{5.11}$$

As in the static case, it will often be important that the mechanism is not required to make payments to the agents in aggregate, as this would mean implementing the mechanism requires some external source of funding.

Definition 5.6 (no-deficit). *A dynamic mechanism (π, T) has the no-deficit property if and only if $\forall \theta^t \in \Theta$,*

$$\sum_{i \in I} T_i(\theta^t) \leq 0 \tag{5.12}$$

Also as in the static case, the strong ex post equilibrium concept can be weakened—there may be mechanisms that fail to achieve truthfulness, efficiency, etc. in the within-period ex post Nash equilibrium solution concept yet meet weaker criteria. The following definitions extend naturally from the static setting, so I maintain the same terminology. In a dynamic setting a *Bayes-Nash equilibrium* will entail that agents playing according to the equilibrium maximize expected payoff *given beliefs about other agents' types*, and given that they play according to the equilibrium. Let $b_i^t(\theta_{-i}^t)$ denote a distribution over the types of agents other than i at t, representing i's beliefs about them at t, and let $\tilde{\theta}_{-i}^t$ be a random variable denoting (from i's perspective) the realization of θ_{-i}^t.

Definition 5.7 (Bayes-Nash equilibrium). *Given dynamic mechanism (π, T) a strategy profile σ constitutes a Bayes-Nash equilibrium if and only if, for all agents $i \in I$, for all possible true types $\theta^t \in \Theta$, for common-knowledge beliefs b_i^t for each i regarding the type profile θ_{-i}^t for the other agents at every time t that are formed consistent with Bayesian-updating, and for all σ_i',*

$$\mathbb{E}_{b_i^t(\theta_{-i}^t)}[V_i(\theta_i^t, \tilde{\theta}_{-i}^t, \pi, \sigma_i, \sigma_{-i})] + \mathbb{E}_{b_i^t(\theta_{-i}^t)}[\mathcal{T}_i(\theta_i^t, \tilde{\theta}_{-i}^t, \pi, \sigma_i, \sigma_{-i})] \tag{5.13}$$

$$\geq \mathbb{E}_{b_i^t(\theta_{-i}^t)}[V_i(\theta^t, \tilde{\theta}_{-i}^t, \pi, \sigma_i', \sigma_{-i})] + \mathbb{E}_{b_i^t(\theta_{-i}^t)}[\mathcal{T}_i(\theta_i^t, \tilde{\theta}_{-i}^t, \pi, \sigma_i', \sigma_{-i})] \tag{5.14}$$

[8] It is interesting to note that an ex post Nash (rather than dominant strategy) equilibrium concept is adopted in analysis of iterative combinatorial auctions for similar reasons (see [Parkes, 2006]).

We can consider a weakening of the individual rationality property in which each agent can expect, a priori, to gain from participating in the mechanism (whatever his type and the types of other agents), yet may in future time-steps expect to lose from participating further. In an ex ante individual rational mechanism agents will "sign up" for the mechanism in the beginning (and could potentially even be contractually bound) but may end up wanting to opt-out depending on how things evolve. Since the property holds for *any* true type profile θ^0, it does not depend whatsoever on agent beliefs about other agents' types; the uncertainty is regarding the realization of random type transitions in the future.

Definition 5.8 (ex ante individual rational). *A dynamic mechanism (π, T) is ex ante individual rational if and only if there exists a strategy profile σ such that, for all agents $i \in I$, for all possible initial true types $\theta^0 \in \Theta$, σ is a Bayes-Nash equilibrium strategy profile and*

$$V_i(\theta^0, \pi, \sigma) + \mathcal{T}_i(\theta^0, \pi, \sigma) \geq 0 \tag{5.15}$$

There is an analogous weakening of the no-deficit property to "in expectation from the beginning of the mechanism":

Definition 5.9 (ex ante no-deficit). *A dynamic mechanism (π, T) ex ante no-deficit if and only if there exists a within-period ex post Nash equilibrium strategy profile σ such that, for all possible initial true types $\theta^0 \in \Theta$,*

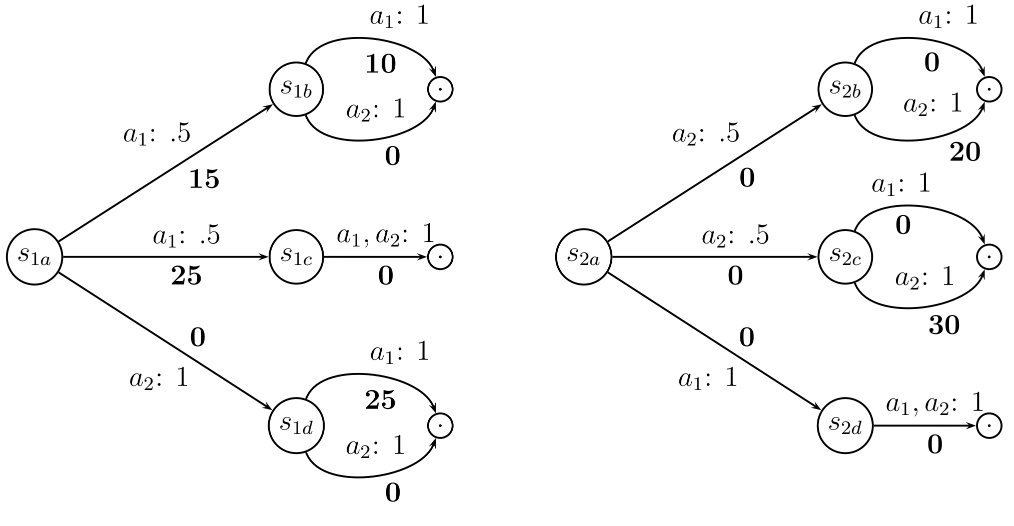
$$\sum_{i \in I} \mathcal{T}_i(\theta^0, \sigma) \leq 0 \tag{5.16}$$

5.2 A simple efficient mechanism

Given this framework for evaluating mechanisms via dynamic equilibrium properties, we are ready to begin looking for satisfactory instantiations. Given the results of static mechanism design, a natural place to start is the Groves class of mechanisms, in which each agent's utility equals total social utility minus a constant.

Extending this basic idea to the dynamic setting is straightforward. Consider the following *dynamic-basic-Groves* mechanism,[9] which takes efficient decisions and pays each agent the reported expected value other agents receive that period:

[9] This mechanism first appeared in [Cavallo *et al.*, 2006], where we termed it the "sequential-Groves" mechanism.



(a) Agent 1

(b) Agent 2

Figure 5.2: Two local agent MDPs: each represents a different agent's value structure (with uncertainty) for allocation of a car over a two-day period, once on Friday and once on Saturday.

Definition 5.10 (dynamic-basic-Groves). *The dynamic-basic-Groves mechanism executes decision policy π^* and, $\forall i \in I$ and $\theta^t \in \Theta$, transfers:*

$$T_i(\theta^t) = r_{-i}(\theta^t_{-i}, \pi^*(\theta^t)) \tag{5.17}$$

The two time-step car allocation example portrayed in Figure 4.5, which I represent here in Figure 5.2, provides a convenient setting for illustrating the mechanism. Take $\gamma = 0.9$. Here the efficient policy specifies action $\pi^*(s_{1a}, s_{2a}) = a_2$ (i.e., allocate to agent 2 on Friday), $\pi^*(s_{1a}, s_{2c}) = a_2$ (i.e., allocate to agent 2 on Saturday if his state is s_{2c}), and $\pi^*(s_{1a}, s_{2b}) = a_1$ (allocate to agent 1 on Saturday otherwise). No other joint states occur in equilibrium. So under the dynamic-basic-Groves mechanism, if agents report truthfully: in the first time-step the car will be allocated to agent 2 (action a_2), agent 1 will be payed $r_2(s_{2a}, a_2) = 0$, and agent 2 will be payed $r_1(s_{1a}, a_2) = 0$. Assume that random state transition $(s_{1a}, s_{2a}) \rightarrow (s_{1d}, s_{2b})$ is realized. Then in the second time-step the car will be allocated to agent 1 (action a_1), agent 1 will be payed $r_2(s_{2b}, a_1) = 0$, and agent 2 will be payed $r_1(s_{1d}, a_2) = 25$.

This mechanism is truthful and efficient in within-period ex post Nash equilibrium. I won't provide a full proof here because this fact is entailed by a broader result proved in the following section, but the intuition is simply that each agent's (expected) utility from truthful participation equals the expected *social* utility; then since the center is acting to maximize this quantity (by implementing policy π^*) deviating from truthfulness can never be beneficial. In the last step agent 2 *wants* agent 1 to receive the car, since agent 1's value of 25 is higher than agent 1's value of 20. But, like the basic-Groves static mechanism, this mechanism's budget properties severely undermine its applicability. Typically we will require that mechanisms meet the no-deficit property, i.e., that total payments made by the center are non-positive so that no external budget is required for the mechanism's implementation.

Fortunately, we will see in the next section that, as was the case with the static-setting basic-Groves mechanism, dynamic-basic-Groves is just the simplest special case of a broad class of mechanisms that are truthful and efficient in a within-period ex post Nash equilibrium: the *dynamic-Groves* mechanisms. We will see, again analogously to the static setting, that this class completely characterizes the mechanisms with these properties.

5.3 A characterization of efficient incentive compatible mechanisms

The results of this section build on and follow closely the analysis of Groves [1973] and Green & Laffont [1977]—in fact the result we will see is a rather direct analogue of Green & Laffont's characterization of the Groves mechanism as the set of efficient and strategyproof static mechanisms (see Theorem 2.4). Consider the following class of "dynamic-Groves" mechanisms, which I name thus because they are the natural extension of the static Groves class, in which each agent's transfer payment equals the reported value of the other agents for the chosen outcomes, minus some constant.

Definition 5.11 (dynamic-Groves mechanism class). *A dynamic-Groves mechanism executes efficient decision policy π^* and a transfer function T such that at every time t, $\forall \theta^t \in \Theta$, $\forall i \in I$, $\forall \sigma_i$, there is a function $C_i : \Theta \rightarrow \Re$ such that, letting $\mathcal{C}_i(\theta^t, \sigma_i) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} C_i(\sigma_i(\theta_i^k), \theta_{-i}^k)) \mid \theta^t, \pi^*, \sigma_i]$:*

$$\mathcal{T}_i(\theta^t, \sigma_i) = V_{-i}(\theta^t, \sigma_i) - \mathcal{C}_i(\theta^t, \sigma_i), \tag{5.18}$$

and for any two strategies σ_i' and σ_i'' for agent i,

$$\mathcal{C}_i(\theta^t, \sigma_i') = \mathcal{C}_i(\theta^t, \sigma_i'') \tag{5.19}$$

So in any given period the payment to an agent i in a dynamic-Groves mechanism *does not* need to equal the reward received by other agents minus a constant, but summed over time the expected *total* payments must. Observing that this class of mechanisms can be described as follows will be useful for my analysis. Lemma 5.1 specifies that the difference in expected total discounted transfer payments for two different reporting strategies, given any true type, is the expected difference in value the other agents obtain (given that they're truthful) from decisions based on those reports. This is the essential attribute of a dynamic-Groves mechanism.

Lemma 5.1. *A dynamic mechanism (π^*, T) is a dynamic-Groves mechanism if and only if:*

$$\forall i \in I,\, \theta^t \in \Theta,\, \sigma_i', \sigma_i'',\;\; \mathcal{T}_i(\theta^t, \sigma_i') - \mathcal{T}_i(\theta^t, \sigma_i'') = V_{-i}(\theta^t, \sigma_i') - V_{-i}(\theta^t, \sigma_i'') \tag{5.20}$$

Proof. First, it is obvious that any dynamic-Groves mechanism satisfies (5.20). Now, for any mechanism (π^*, T) there is some $C_i : \Theta \to \Re$ such that $\mathcal{T}_i(\theta^t, \sigma_i) = V_{-i}(\theta^t, \sigma_i) - \mathcal{C}(\theta^t, \sigma_i)$, for every σ_i; in particular, we can let $C_i(\theta^t) = r_{-i}(\theta^t, \pi^*(\theta^t)) - T_i(\theta^t), \forall \theta^t$. Assume (π^*, T) satisfies (5.20). Then, substituting in (5.20) for $\mathcal{T}$ with V_{-i} and $\mathcal{C}$ corresponding to the expected discounted sum of such a C, we have:

$$\begin{aligned} &(V_{-i}(\theta^t, \sigma_i') - \mathcal{C}(\theta^t, \sigma_i')) - (V_{-i}(\theta^t, \sigma_i'') - \mathcal{C}(\theta^t, \sigma_i'')) \\ &= V_{-i}(\theta^t, \sigma_i') - V_{-i}(\theta^t, \sigma_i'') \end{aligned} \tag{5.21}$$

This implies $\mathcal{C}(\theta^t, \sigma_i') = \mathcal{C}(\theta^t, \sigma_i'')$, and thus (π^*, T) is a dynamic-Groves mechanism. □

This fact allows us to establish the sufficiency direction of the characterization:

Theorem 5.1. *All dynamic-Groves mechanisms are truthful and efficient in within-period ex post Nash equilibrium.*

Proof. By Lemma 5.1, for any dynamic-Groves mechanism (π^*, T), for any θ^t, i, and σ_i, if agents other than i report truthfully:

$$\begin{aligned} & (V_i(\theta^t) + \mathcal{T}_i(\theta^t)) - (V_i(\theta^t, \sigma_i) + \mathcal{T}_i(\theta^t, \sigma_i)) && (5.22)\\ =\; & (V_i(\theta^t) + V_{-i}(\theta^t)) - (V_i(\theta^t, \sigma_i) + V_{-i}(\theta^t, \sigma_i)) && (5.23)\\ =\; & V(\theta^t) - V(\theta^t, \sigma_i) && (5.24)\\ \geq\; & 0 && (5.25) \end{aligned}$$

The final inequality follows from the definition (optimality) of π^*. If it did not hold, then one could construct a socially superior policy π such that $\forall \theta \in \Theta$, $\pi(\theta) = \pi^*(\sigma_i(\theta_i), \theta_{-i})$. □

I will now demonstrate that every history-independent dynamic mechanism that is truthful and efficient in within-period ex post Nash equilibrium is a dynamic-Groves mechanism. This direction is significantly more involved, and the proof follows the broad strokes of the Green & Laffont [1977] proof, though things become more complex in the dynamic setting. We will see that if the difference in expected transfers from two reporting strategies does not equal the expected difference in value obtained by the other agents, then one can construct a hypothetical *true* type for an agent such that he would gain by executing the reporting strategy that yields greater transfers.

I will use notation $\mathcal{A}(\theta^t, \sigma)$ to reason about the future sequence of actions that will occur given true type θ^t, reporting strategy profile σ, and decision policy π^*. The distribution over actions that might be taken at time $k > t$ is determined by the realization of *random* events through time $k-1$, and so I let $\mathcal{A}(\theta^t, \sigma)$ be an "action sequence mapping" from the space of possible (given θ^t, σ, and π^*) random event realizations to a sequence of actions. Given $\dot{\theta}^t, \overline{\theta}^t \in \Theta$, σ', and σ'', then, $\mathcal{A}(\dot{\theta}^t, \sigma') = \mathcal{A}(\overline{\theta}^t, \sigma'')$ means that $\pi^*(\sigma'(\dot{\theta}^t)) = \pi^*(\sigma''(\overline{\theta}^t))$, and moreover (given the decision at t) for every possible realization of random events at t the decision taken at time $t+1$ will be the same whether the joint type and strategy at t was $(\dot{\theta}^t, \sigma)$ or $(\overline{\theta}^t, \sigma')$, and so on for times $t+2, \ldots, K$.

The proof of the characterization result for dynamic-Groves is simplified by the following lemma, which says that in any within-period ex post efficient and IC history-independent mechanism, given the reported types at time t of agents other than some i, if two reports by i would cause the center to take the same action at t, i's transfer at t is the same regardless of which of the two types he reports.

Lemma 5.2. *For an unrestricted type space, if a history-independent dynamic mechanism (π^*, T) is truthful and efficient in within-period ex post Nash equilibrium, then $\forall i \in I$, $\theta^t_{-i} \in \Theta_{-i}$, and $\dot{\theta}^t_i, \overline{\theta}^t_i \in \Theta_i$,*

$$\pi^*(\dot{\theta}^t_i, \theta^t_{-i}) = \pi^*(\overline{\theta}^t_i, \theta^t_{-i}) \Rightarrow T_i(\dot{\theta}^t_i, \theta^t_{-i}) = T_i(\overline{\theta}^t_i, \theta^t_{-i}) \tag{5.26}$$

Proof. Consider an arbitrary history-independent dynamic mechanism (π^*, T) for which there exists an agent i and types $\dot{\theta}^t_i$, $\overline{\theta}^t_i$, and θ^t_{-i} such that $\pi^*(\dot{\theta}^t_i, \theta^t_{-i}) = \pi^*(\overline{\theta}^t_i, \theta^t_{-i})$ and $T_i(\overline{\theta}^t_i, \theta^t_{-i}) > T_i(\dot{\theta}^t_i, \theta^t_{-i})$. Consider an agent whose true type at t is $\dot{\theta}^t_i$. If i reports truthfully in all time periods following t, the value and transfers he obtains after t will be the same regardless of whether he reports $\dot{\theta}^t_i$ or $\overline{\theta}^t_i$ at t since the same action will be taken at t. We have:

$$\mathbb{E}\Big[V_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\dot{\theta}^t_i, \theta^t_{-i}))) + \mathcal{T}_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\dot{\theta}^t_i, \theta^t_{-i})))\Big] \tag{5.27}$$

$$= \mathbb{E}\Big[V_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\overline{\theta}^t_i, \theta^t_{-i}))) + \mathcal{T}_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\overline{\theta}^t_i, \theta^t_{-i})))\Big] \tag{5.28}$$

Given this equality, and since $r_i(\dot{\theta}^t_i, \pi^*(\dot{\theta}^t_i, \theta^t_{-i})) = r_i(\dot{\theta}^t_i, \pi^*(\overline{\theta}^t_i, \theta^t_{-i}))$ and $T_i(\overline{\theta}^t_i, \theta^t_{-i}) > T_i(\dot{\theta}^t_i, \theta^t_{-i})$, i is better off reporting $\overline{\theta}^t_i$ rather than true type $\dot{\theta}^t_i$ at t, and thus the mechanism is not truthful in within-period ex post Nash equilibrium. □

Theorem 5.2. *For an unrestricted type space, if a history-independent dynamic mechanism (π^*, T) is truthful and efficient in within-period ex post Nash equilibrium, it is a dynamic-Groves mechanism.*

Proof. Assume for contradiction existence of a history-independent dynamic mechanism (π^*, T) that is *not* a member of the dynamic-Groves class yet is truthful and efficient in within-period ex post Nash equilibrium. There is an $i \in I$, joint type $(\overline{\theta}^t_i, \theta^t_{-i})$, strategies σ'_i and σ''_i for agent i, and $\epsilon > 0$ such that:

$$\mathcal{T}_i(\overline{\theta}^t_i, \theta^t_{-i}, \sigma'_i) - \mathcal{T}_i(\overline{\theta}^t_i, \theta^t_{-i}, \sigma''_i) = V_{-i}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma'_i) - V_{-i}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma''_i) + \epsilon \qquad (5.29)$$

Consider a type $\dot{\theta}^t_i$ *correlated* with $\overline{\theta}^t_i$ such that any path of state transitions forward from initial state $\dot{\theta}^t_i$ would indicate exactly what state transitions *would have* occurred if the initial state were instead $\overline{\theta}^t_i$. Then, there are strategies $\sigma'_{\overline{\theta}^t_i}$ and $\sigma''_{\overline{\theta}^t_i}$ such that $\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}) = \mathcal{A}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma'_i)$ and $\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\overline{\theta}^t_i}) = \mathcal{A}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma''_i)$. Consider that $\dot{\theta}^t_i$ is also such that $\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}) = \mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\overline{\theta}^t_i})$ and, for some $0 < \delta < \epsilon$,

$$\begin{aligned} V_i(\dot{\theta}^t_i, \theta^t_{-i}) &= -V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}) + \delta && (5.30)\\ &= V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\overline{\theta}^t_i}) = -V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\overline{\theta}^t_i}) + \delta, && (5.31)\\ V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}) &= -V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}), && (5.32) \end{aligned}$$

and i's expected value $V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma_i)$ for any strategy σ_i that yields any action sequence mapping that is not equal to $\mathcal{A}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma''_i)$ or $\mathcal{A}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma'_i)$ is -1 times the other agents' combined expected value (V_{-i}) for that mapping.[10] The valuation implied by type $\dot{\theta}^t_i$ is valid, as the expected social value of π^* executed on truthful reports is δ better than the expected social value of any policy that yields any alternate action sequence mapping.

$\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}) = \mathcal{A}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma''_i)$ combined with Lemma 5.2 entails that the expected transfers to i are the same if i's type at t is $\dot{\theta}^t_i$ and i is truthful, or if it is $\overline{\theta}^t_i$ and i

[10] The proof can still go through if we require that values are always non-negative; one can just take c to be some constant greater than $V_{-i}(\theta^t_{-i})$, and specify V_i as above but with c added to the value for each action-sequence mapping.

follows reporting strategy σ''_i. We have:

$$\begin{aligned}
&\mathcal{T}_i(\overline{\theta}^t_i, \theta^t_{-i}, \sigma'_i) - \mathcal{T}_i(\overline{\theta}^t_i, \theta^t_{-i}, \sigma''_i) && (5.33)\\
&= \mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}) - \mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}) && (5.34)\\
&= V_{-i}(\overline{\theta}^t_{-i}, \theta^t_{-i}, \sigma'_i) - V_{-i}(\overline{\theta}^t_i, \theta^t_{-i}, \sigma''_i) + \epsilon && (5.35)\\
&= V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}) - V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\overline{\theta}^t_i}) + \epsilon && (5.36)\\
&= V_i(\dot{\theta}^t_i, \theta^t_{-i}) - \delta - V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}) + \epsilon, && (5.37)
\end{aligned}$$

from which we can see that:

$$\mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}) + V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\overline{\theta}^t_i}) > \mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}) + V_i(\dot{\theta}^t_i, \theta^t_{-i}) \tag{5.38}$$

When i's type is $\dot{\theta}^t_i$ he is better off reporting according to $\sigma'_{\overline{\theta}^t_i}$ rather than truthfully, and so the mechanism is not truthful in within-period ex post Nash equilibrium. □

Theorem 5.3. *For an unrestricted type space, a history-independent dynamic mechanism is efficient and within-period ex post incentive compatible if and only if it is a dynamic-Groves mechanism.*

Proof. Follows immediately from Theorems 5.1 and 5.2. □

5.4 Ex ante charge dynamic mechanisms

The results of the previous section provide a complete mapping of the space of possible (history-independent) mechanisms we can consider if we require efficiency and incentive compatibility in within-period ex post Nash equilibrium. But, as discussed earlier, there are additional criteria that will typically be applied to the design of a mechanism. Individual rationality is central; one could legitimately argue that a mechanism that is not IR has no hope of being truly efficient, because reaching efficient outcomes requires the participation of agents, and self-interested agents who may be worse off from participating may not do so. No-deficit is often essential for feasibility of the mechanism.

The dynamic-basic-Groves mechanism satisfies the within-period ex post IC and IR properties (IR requires an assumption that expected social value is positive, which is natural) but is not no-deficit. But since we know that *any* dynamic-Groves mechanism is within-period ex post IC, we can try to define the "charge" term C_i for agent i in a dynamic-Groves mechanism such that it recovers just enough of the transfer payments to (weakly) balance the budget, but not so much that it makes participating in the mechanism undesirable (i.e., breaks the IR property).

The dynamic-basic-Groves mechanism pays each agent the (reported) value obtained by other agents, and it is thus a dynamic-Groves mechanism where the constant C charge functions are null. But we can see that any mechanism that modifies dynamic-basic-Groves by imposing a charge on each agent at each time period that is independent of anything that agent has ever reported will also be a dynamic-Groves mechanism. We can specify the charge for each agent in a way such that, in expectation from the beginning of the mechanism, a deficit will not result and at the same time agent payoffs will be non-negative. I will refer to this as the dynamic-EAC (ex ante charge) mechanism:

Definition 5.12 (dynamic-EAC). *The dynamic-EAC mechanism executes decision policy π^* and, $\forall i \in I$ and $\theta^t \in \Theta$, transfers:*

$$T_i(\theta^t) = r_{-i}(\theta^t_{-i}, \pi^*(\theta^t)) - (1-\gamma)V_{-i}(\theta^0_{-i}) \tag{5.39}$$

Theorem 5.4. *The dynamic-EAC mechanism is truthful and efficient in within-period ex post Nash equilibrium, ex ante individual rational, and ex ante no-deficit.*

Proof. First observe that for any $i \in I$, $\theta^0 \in \Theta$, and any two strategies σ'_i and σ''_i, $(1-\gamma)V_{-i}(\theta^0_{-i}, \pi^*_{-i}, \sigma'_i) = (1-\gamma)V_{-i}(\theta^0_{-i}, \pi^*_{-i}, \sigma''_i)$, since $V_{-i}(\theta^0_{-i})$ depends only on the states of other agents at the beginning of the mechanism, which i cannot possibly impact. Thus dynamic-EAC is a dynamic-Groves mechanism with $C_i(\theta^t) = (1-\gamma)V_{-i}(\theta^0_{-i})$, $\forall \theta^t$. Therefore, by Theorem 5.3, dynamic-EAC is truthful and efficient in within-period ex post Nash equilibrium.

The mechanism is also ex ante IR. Consider the truthful reporting strategy profile. In expectation from the beginning of the mechanism the reward obtained intrinsically by any agent plus the payment he receives is greater than the charge he must pay:

$$\mathbb{E}\Big[\sum_{k=0}^{K} \gamma^k \Big(r_i(\theta^k_i, \pi^*(\theta^k)) + r_{-i}(\theta^k_{-i}, \pi^*(\theta^k)) - (1-\gamma)V_{-i}(\theta^0_{-i})\Big) \,\Big|\, \theta^0, \pi^*\Big] \tag{5.40}$$

$$= V_i(\theta^0) + V_{-i}(\theta^0) - V_{-i}(\theta^0_{-i}) \tag{5.41}$$

$$= V(\theta^0) - V_{-i}(\theta^0_{-i}) \tag{5.42}$$

$$\geq 0, \tag{5.43}$$

where the final inequality holds by optimality of π^*.

Finally, note that in expectation from the beginning of the mechanism, in the truthful equilibrium the payments made by the mechanism to any agent i equal $V_{-i}(\theta^0) - V_{-i}(\theta^0_{-i})$. $V_{-i}(\theta^0_{-i}) \geq V_{-i}(\theta^0)$ by optimality (for agents other than i) of π^*_{-i}, and thus the mechanism is ex ante no-deficit □

There are a couple additional interesting things to note about this mechanism. First, the properties (incentives and budget) do not depend at all on when the charge terms imposed on the agents are executed, as long as they are scaled appropriately according to the discount factor. For instance, the mechanism could just as well have been defined to have each agent i pay $V_{-i}(\theta^0_{-i})$ at time 0 with no further charges in the periods to follow.

Also, if the initial state θ^0 is common knowledge (and only the realized state transitions are private), each i can be charged $(1-\gamma)V_{-i}(\theta^0)$ rather than $(1-\gamma)V_{-i}(\theta^0_{-i})$ each period; Theorem 5.4 continues to hold, and in fact the expected revenue from the beginning of the mechanism equals 0, since the expected aggregate value of payments made to any agent are 0.

5.5 The dynamic-VCG mechanism

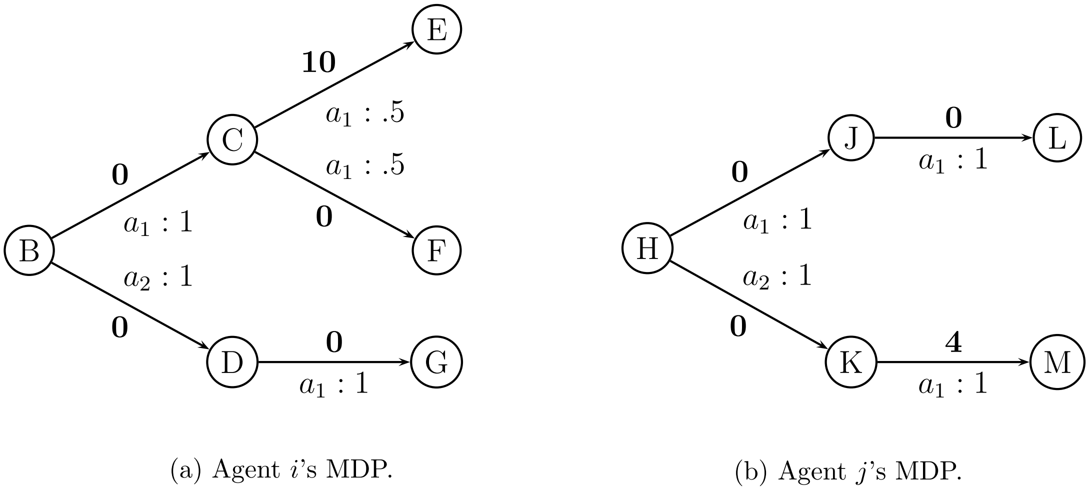
The results of the previous section are positive, but leave room for improvement. In many scenarios the distinction between ex ante individual rationality or no-deficit and within-period ex post individual rationality or *guaranteed* no-deficit will be significant. The mechanism I present in this chapter strengthens the dynamic-EAC mechanism in just these ways. Bergemann & Välimäki's [2006] dynamic-VCG mechanism, we will see, is efficient, IC, and IR in within-period ex post Nash equilibrium, and is also no-deficit. I will demonstrate efficiency and IC, again, by showing that dynamic-VCG is a dynamic-Groves mechanism and then referring to Theorem 5.3. The nature of the proof will at the same time demonstrate the IR and no-deficit properties of the mechanism.

Finally, the revenue a mechanism generates—or, how much of the value from a sequence of decisions is acquired by the center rather than kept by the agents—is also an important evaluation metric. Of course in many business settings a mechanism designer would seek to implement a mechanism in which revenue is high, extracting as much value as possible; I will show that dynamic-VCG is optimal here (if efficiency is required). In the next chapter I follow the approach of Chapter 3 and try to *minimize* rather than maximize revenue.

Definition 5.13 (dynamic-VCG). [Bergemann and Valimaki, 2006] *The dynamic-VCG mechanism executes policy π^* and, $\forall i \in I$ and $\theta^t \in \Theta$, transfers:*

$$T_i(\theta^t) = r_{-i}(\theta^t_{-i}, \pi^*(\theta^t)) + \gamma \mathbb{E}[V_{-i}(\tau(\theta^t_{-i}, \pi^*(\theta^t)))] - V_{-i}(\theta^t_{-i}) \tag{5.44}$$

Recall that $V_{-i}(\theta^t_{-i})$ denotes $V_{-i}(\theta^t_i, \theta^t_{-i}, \pi^*_{-i}, \sigma_i)$, and thus $\mathbb{E}[V_{-i}(\tau(\theta^t_{-i}, \pi^*(\theta^t)))]$ is the expected value that agents other than i would obtain from a policy that is optimized for them forward from the joint type that results when the socially optimal policy π^* is followed for one time-step. Intuitively, at each time-step each agent must pay the center a quantity equal to the extent to which his current type report inhibits other agents from obtaining value in the present and in the future.[11]



(a) Agent i's MDP.

(b) Agent j's MDP.

Figure 5.3: A simple 2-agent, 2 time-step environment. Action a_2 is omitted in the second time-step, which indicates that it yields 0 value to both agents (the state-transition it yields in the final period is irrelevant).

Consider the example illustrated in Figure 5.3. The optimal policy is to take action a_1 in initial joint state B, H. Dynamic-VCG specifies that agent 1's payment $T_1(B, H)$ in this first period equal -4, since this is the long-term cost to the other agent from taking a_1 rather than a_2 in the first time-step; $T_2(B, H) = 0$ since there is no "cost" to agent 1. In the second period both $T_1(C, J)$ and $T_2(C, J)$ are 0, since the system-optimal decision a_1 is also optimal for both agents individually (trivially here, as it is the only choice yielding non-zero value).

I now present a new, simple proof of the incentive compatibility of dynamic-VCG (originally proved by Bergemann & Välimäki [2006]), which becomes possible because my analysis of dynamic-Groves mechanisms allows me to cast the question of whether or not dynamic-VCG is efficient in a truthtelling ex post Nash equilibrium as a question of whether or not it is a dynamic-Groves mechanism. I will show that it is, by observing that when other agents are truthful the expected sum, over time, of the first term in (5.44) equals $V_{-i}(\theta^t, \sigma_i)$, and then the expected sum of the rest of the payment can be represented as a function independent of anything i reports.

Theorem 5.5. *The dynamic-VCG mechanism is a dynamic-Groves mechanism.*

[11] Bergemann & Välimäki [2007] have more recently referred to their mechanism as the "dynamic marginal contribution mechanism".

Proof. Pick any agent i and joint type θ^t, assume all other agents report truthfully, and consider any strategy σ_i for i. Let θ^k be the random variable denoting the true joint type at time $k > t$ given that other agents are truthful, i follows σ_i, and π^* is executed from θ^t. We have:

$$\mathcal{T}_i(\theta^t, \sigma_i) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}(r_{-i}(\theta_{-i}, \pi^*(\sigma_i(\theta_i^k), \theta_{-i}^k)) + \gamma\, V_{-i}(\theta_{-i}^{k+1}) - V_{-i}(\theta_{-i}^k)) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] \tag{5.45}$$

Extracting out the sum over the first term and reversing the second and third terms, we see this:

$$= \; V_{-i}(\theta^t, \sigma_i) - \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}(V_{-i}(\theta_{-i}^k) - \gamma V_{-i}(\theta_{-i}^{k+1})) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] \tag{5.46}$$

Expanding out the summation, then extracting $V_{-i}(\theta_{-i}^t)$ out from the first summation and canceling out, we see this:

$$= V_{-i}(\theta^t, \sigma_i) - \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} V_{-i}(\theta_{-i}^k) - \gamma \sum_{k=t}^{K} \gamma^{k-t} V_{-i}(\theta_{-i}^{k+1}) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] \tag{5.47}$$

$$= V_{-i}(\theta^t, \sigma_i) - V_{-i}(\theta_{-i}^t) - \mathbb{E}\Big[\gamma \sum_{k=t}^{K-1} \gamma^{k-t} V_{-i}(\theta_{-i}^{k+1}) - \gamma \sum_{k=t}^{K} \gamma^{k-t} V_{-i}(\theta_{-i}^{k+1}) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] \tag{5.48}$$

$$= V_{-i}(\theta^t, \sigma_i) - V_{-i}(\theta_{-i}^t) \tag{5.49}$$

The move from (5.48) to (5.49) is valid because for any θ_{-i}^{K+1}, $V_{-i}(\theta_{-i}^{K+1}) = 0$ (there are no more decisions to be made after time K). Thus dynamic-VCG is a dynamic-Groves mechanism, since I have shown that, letting $C_i(\theta^t) = V_{-i}(\theta_{-i}^t) - \gamma\mathbb{E}[V_{-i}(\tau(\theta_{-i}^t, \pi^*(\theta^t)))]$, $\mathcal{C}_i(\theta^t, \sigma_i) = V_{-i}(\theta_{-i}^t)$ for any σ_i. □

Theorems 5.3 and 5.5 together yield the following:

Corollary 5.1. [Bergemann and Valimaki, 2006] *The dynamic-VCG mechanism is truthful and efficient in within-period ex post Nash equilibrium.*

The following statements about the expected equilibrium payoffs follow immediately from the proof of Theorem 5.5:

Corollary 5.2. *Utility to any agent i in the truthful equilibrium under dynamic-VCG, in expectation forward from any any type θ^t, is:*

$$V(\theta^t) - V_{-i}(\theta_{-i}^t) \tag{5.50}$$

By optimality of π^*, we have:

Corollary 5.3. *The dynamic-VCG mechanism is within-period ex post individual rational.*

Corollary 5.4. *Social utility in the truthful equilibrium under dynamic-VCG, in expectation forward from any θ^t, is:*

$$n \cdot V(\theta^t) - \sum_{i \in I} V_{-i}(\theta^t_{-i}) \tag{5.51}$$

Corollary 5.5. *Expected revenue in the truthful equilibrium under dynamic-VCG, in expectation forward from any θ^t, is:*

$$\sum_{i \in I} V_{-i}(\theta^t_{-i}) - (n-1)V(\theta^t) \tag{5.52}$$

Dynamic-VCG never runs a deficit in any period, regardless of what types agents report; thus the no-deficit property is robust to “off-equilibrium” play:

Theorem 5.6. *The dynamic-VCG mechanism is no-deficit.*

Proof. By optimality (for agents other than i) of π^*_{-i}, for any type θ^t and any strategy profile σ, $r_{-i}(\sigma_{-i}(\theta^t_{-i}), \pi^*(\sigma(\theta^t))) + \gamma \mathbb{E}[V_{-i}(\tau(\sigma_{-i}(\theta^t_{-i}), \pi^*(\sigma(\theta^t)))] \leq V_{-i}(\sigma_{-i}(\theta^t_{-i}))$. Thus the net payment to each agent *in every time period* is at most 0 and no deficit can ever result. □

It is worth noting that agents can potentially end up worse off from participating in dynamic-VCG (so the mechanism is not ex post IR in as strong a sense as it is no-deficit), though in expectation from any state they will not. Consider again the example in Figure 5.3 with truthful reporting. The optimal policy is to take action a_1 in the first period. In this time-step agent 1 will obtain no value and receives payment -4 (he pays 4 to the center). However, there is only a 0.5 probability that he will transition from state C to E in the second time-step, obtaining value 10. If instead he transitions to state F he will obtain no value and his total payoff will be -4. Thus non-negative payoff is only achieved in expectation forward from every state.

5.5.1 Revenue maximization

I now show that if individual rationality is required in addition to efficiency and incentive compatibility, no history-independent dynamic mechanism yields more expected revenue than dynamic-VCG in within-period ex post Nash equilibrium, forward from *any* joint type θ^t.

Theorem 5.7. *For an unrestricted type space, among all history-independent mechanisms that are efficient, incentive compatible, and individual rational in within-period ex post Nash equilibrium, dynamic-VCG yields the most expected revenue in the truthful equilibrium going forward from every θ^t.*

Proof. The expected equilibrium revenue under dynamic-VCG given any joint type θ^t is:

$$\sum_{i \in I} [V_{-i}(\theta^t_{-i}) - V_{-i}(\theta^t)] \tag{5.53}$$

Consider any history-independent dynamic-Groves mechanism (π^*, T) that yields more revenue (this is without loss of generality by Theorem 5.3). This mechanism must define $C_1, \ldots, C_n$ such that

$$\sum_{i \in I} \mathcal{C}_i(\theta^t) > \sum_{i \in I} V_{-i}(\theta^t_{-i}), \tag{5.54}$$

since revenue under a dynamic-Groves mechanism is $\sum_{i \in I} [\mathcal{C}(\theta^t) - V_{-i}(\theta^t)]$. This in turn implies there is an $i \in I$ such that:

$$\mathcal{C}_i(\theta^t) > V_{-i}(\theta^t_{-i}) \tag{5.55}$$

Recall that $\mathcal{C}_i(\theta^t)$ must be independent of i's type reports, and thus independent of his actual realized types. So consider the case in which $V(\theta^t) = V_{-i}(\theta^t_{-i})$ (for instance, this holds when i's value is always 0). Then agent i's expected payoff is:

$$\begin{aligned} V(\theta^t) - \mathcal{C}_i(\theta^t) &= V_{-i}(\theta^t_{-i}) - \mathcal{C}_i(\theta^t) && (5.56) \\ &< V_{-i}(\theta^t_{-i}) - V_{-i}(\theta^t_{-i}) && (5.57) \\ &= 0, && (5.58) \end{aligned}$$

and thus the mechanism is not within-period ex post individual rational. □

Given that dynamic-VCG is revenue maximizing, it is natural to ask whether there are other dynamic-Groves mechanisms with the same desirable efficiency, IC, IR, and no-deficit properties that yield *less* revenue. I will pursue that question in the following chapter. But first I briefly review a positive result in which the budget is perfectly balanced (revenue is always 0) at the expense of weaker equilibrium incentive and individual rationality properties.

5.6 The dynamic-balanced mechanism

The dynamic-balanced mechanism of Athey & Segal [2007] provides an analogue of the AGV mechanism for dynamic settings. It is a *strongly* budget-balanced mechanism that implements the efficient decision policy, but in a weaker equilibrium concept than the dynamic-Groves mechanisms, and with a weaker individual rationality property than dynamic-VCG. Intuitively, in every time period each agent receives a payment equal to the extent to which his most recent type report changes the expected value that the other agents will receive in the future; each agent is then also charged a "rebalancing" quantity.

Definition 5.14 (dynamic-balanced mechanism). [Athey and Segal, 2007] *Letting θ^{t-1} denote the type reported in time $t-1$ (tracked in the center's type θ_c^t), and letting $\tilde{\theta}_i^t$ be a random variable representing i's type at t, the dynamic-balanced mechanism executes decision policy π^* and, $\forall i \in I$ and $\theta^t \in \Theta$, transfers:*

$$T_i(\theta^t) = \Delta_i(\theta_i^t, \theta^{t-1}) - \frac{1}{n-1} \sum_{j \in I \setminus \{i\}} \Delta_j(\theta_j^t, \theta^{t-1}), \textit{ where} \tag{5.59}$$

$$\Delta_i(\theta_i^t, \theta^{t-1}) = \mathbb{E}[V_{-i}(\theta_i^t, \tilde{\theta}_{-i}^t) \mid \theta_{-i}^{t-1}] - \mathbb{E}[V_{-i}(\tilde{\theta}_i^t, \tilde{\theta}_{-i}^t) \mid \theta^{t-1}] \tag{5.60}$$

It is important to note that the payments at time t are actually based on reported type profile θ^{t+1} (i.e., $T_i(\theta^t)$ actually occurs in time-step $t-1$). This timing is distinct from that employed in the other mechanisms presented in this chapter, and goes as follows: time-step t starts when the center takes an action; during t each agent will experience some value as he undergoes a type transition; still in the same period he can report his next type to the center and then receive a transfer payment; when the center takes the next action the next period begins. But the very beginning of execution is special, with initial types reported and transfers executed prior to the first action and thus the commencing of what I am calling time-step 0. So there is a transfer payment immediately preceding 0 based on reported initial types *and prior beliefs about agent types* (which could be considered θ^{-1}), and then there is a transfer made in time-period 0 based on type θ^1, which is reported in time 0 after the first action has occurred.[12]

Theorem 5.8. [Athey and Segal, 2007][13] *The dynamic-balanced mechanism is truthful and efficient in Bayes-Nash equilibrium, strongly budget-balanced, and ex ante individual rational.*

[12] In Chapter 7 I describe timing issues in detail; the curious reader can look ahead to Figure 7.3 for an illustration of how this timing works.

[13] These incentive and budget properties are entailed by Proposition 2 in [Athey and Segal, 2007], which is a somewhat broader result indicating that any mechanism that is truthful and efficient in Bayes-Nash equilibrium can be transformed into a strongly budget-balanced mechanism with the same equilibrium incentives.

Consider, for a moment, only the first term in each agent's transfer payment at some period in which type θ^t has been reported; the second term is a rebalancing term and for now we will imagine it was not there. It can be shown that the expected discounted sum of transfers going forward for each agent i would equal $\mathbb{E}[V_{-i}(\theta_i^t, \tilde{\theta}_{-i}^t, \sigma_i) \mid \theta_{-i}^{t-1}]$; everything else cancels out in the expectation. This aligns agent i's interests with truthtelling since the center is following an efficient policy π^*. The equilibrium is Bayes-Nash and not within-period ex post, though, because if i *knew* θ_{-i}^t, this could change his expected payoff (and his incentives) from the expectation of his payoff based on θ_{-i}^{t-1}. Now considering the rebalancing term, it can be shown that this does not distort incentives (in the context of a Bayes-Nash equilibrium).

The mechanism, or a simple variant of it, is ex ante IR. We can assume that the total expected social value from running the mechanism is non-negative—otherwise why even run it? Then since all value is maintained within the group of agents, only the question of who gets how much remains. One can then make ex ante payments prior to any types being reported that distribute the expected social value in a way such that each agent's expected utility is non-negative.[14]

5.7 Summary

In this chapter I addressed the problem area of *dynamic mechanism design*, which is intuitively the study of how good outcomes can be reached among a group of self-interested agents in sequential decision problems.

I specified the dynamic-Groves class of mechanisms, and proved that it completely characterizes the set of dynamic mechanisms that are efficient and incentive compatible in within-period ex post Nash equilibrium given a context of history-independent transfers. I presented an instance of the Groves class (dynamic-EAC) that imposes a charge on each agent that depends only on the *initial* types reported by *other* agents; this allows the mechanism to achieve ex ante IR and ex ante no-deficit without distorting incentives. I presented Bergemann & Välimäki's dynamic-VCG mechanism, which has stronger IR and no-deficit properties, and I used the dynamic-Groves characterization result to produce a simple proof that the mechanism is truthful and efficient in within-period ex post Nash equilibrium. I proved that dynamic-VCG is revenue *maximizing* (payoff minimizing for the agents) among all within-period ex post IR and no-deficit mechanisms in the dynamic-Groves class. Finally, I presented Athey & Segal's [2007] dynamic-balanced mechanism, which is *strongly budget-balanced*, but is truthful and efficient in only Bayes-Nash equilibrium and satisfies only ex ante

[14] I am grateful to Susan Athey for a personal explication of this point and others regarding the mechanism; ex ante IR is not discussed in [Athey and Segal, 2007], but special conditions are identified under which a stronger interim individual rationality property will hold.

individual rationality. There are many important open problems and directions for future research; I defer a discussion of these until Chapter 9.

Chapter 6

A dynamic redistribution mechanism

Synopsis*

Imagine a city that has invested in an expensive mobile health clinic to serve the medical needs of the poor and uninsured. There are five separate neighborhoods in the city that would like to use the clinic, and so the city government decides to allocate it repeatedly to a single neighborhood for one week periods, reevaluating every week. The government wants the clinic to go to the neighborhood that needs it most and can use it most effectively each week. For the government to determine which choice is best, neighborhood leaders must make weekly claims about their estimated value for the clinic. When the clinic is allocated to a particular neighborhood in one week, in the *next* week that neighborhood's value for it is likely to change—perhaps a significant portion of the needs have been filled, or perhaps the local population has learned about its presence and is thus better able to exploit it. The government does not want to extract large payments from the communities that use the clinic—it is a public resource. The goal is simply to maximize the aggregate welfare of the city's communities. With settings like this in mind, in this chapter I provide a mechanism for multi-armed bandit problems that seeks to minimize the payments made by agents to the center in a dynamic context.

In the previous chapter we saw that the dynamic-VCG mechanism provides the proper incentives to achieve efficient decisions in sequential problems, and is also within-period ex post individual rational and never runs a deficit. Moreover we saw that it is *revenue maximizing* among all mechanisms with these properties. In the static setting for unrestricted valuations we found that it is impossible to provide greater payoff to the agents while maintaining the essential properties of VCG; VCG

[*] Many of the results in this chapter appear in a paper titled "Efficiency and Redistribution in Dynamic Mechanism Design", which will be published in the proceedings of the ACM-EC 2008 conference [Cavallo, 2008].

is simultaneously revenue maximizing and revenue minimizing, i.e., it is the *only* mechanism with these desirable properties (see Corollary 3.1).

Redistribution is possible in the static setting only by using *domain information about agent type spaces*. For instance, in single-item allocation problems it is typically known, independent of any agent's report, that all agents that don't receive the item obtain value 0. I follow the same approach here, considering a class of problems that includes scenarios in which a single item is to be allocated *repeatedly*. This domain and others fall in the category of multi-armed bandit settings, the inherent structure of which will allow for significant redistribution of VCG revenue. The dynamic redistribution mechanism I propose generalizes the core ideas underlying its static counterpart, but the extension is not straightforward because of dependencies between *future* types and current type reports.

6.1 Multi-armed bandits and dynamic-VCG

Recall from Chapter 4 (Section 4.4) that multi-armed bandit problems are sequential decision-making problems with a strong factorization of the state space. Specifically, in a MAB there are n Markov chains and exactly one can be activated per time-step. When a process is activated at time t, a reward is obtained that depends only on the local state of that process, at which point the process's state changes (and all other processes' states remain unchanged).

Among the many good reasons to consider multi-armed bandit problems are: the range of real-world problems that, to a reasonable approximation, fit the restrictions of the model (including the health clinic scenario described at the beginning of this chapter); the elegance of the solutions we can achieve; and perhaps most importantly, the computational tractability of actually determining efficient decision policies. In a seminal result, Gittins showed that the optimal decision policy in a MAB setting can be computed in time linear in the number of processes (see Theorem 4.1).

There is a natural multi-agent interpretation of multi-armed bandits: a Markov process is associated with each agent, and the state of that process is the local state (type) of the agent. Note that the MAB setting is simply a specialization of the general MDP model I used in Chapter 5, in which MDPs are restricted to be Markov chains and only one can be activated per time-step. The most natural class of real-world multi-agent MAB problems is probably that of repeated single-item allocation, e.g., of an expensive public good such as a supercomputer, space telescope, wireless bandwidth, etc. Gittins's result is remarkable in that it implies that all problems of this nature have a computationally scalable solution, as the complexity grows only linearly in the number of agents. This is in stark contrast to the general MDP case, in which the computation required to determine efficient policies effectively grows exponentially with the number of agents in the worst case.

6.1.1 Dynamic-VCG in MAB settings

In multi-agent MAB domains, the dynamic-VCG payment structure reduces to a very simple form. Because the policy that would be optimal if we disregarded an agent that is not "activated" (e.g., allocated the resource) is optimal if we *do* consider him, the immediate externality any such agent imposes is 0, and thus his payment is 0. For the agent i that *is* activated, the externality he imposes on the other agents is simply the cost of them having to wait one period since their types do not change. I will write $\pi(\theta) = i$ to indicate that i is selected by π when type θ is reported.

Definition 6.1 (Dynamic-VCG in MAB worlds). [Bergemann and Valimaki, 2006] *Executes decision policy π^* and, $\forall i \in I$ and $\theta^t \in \Theta$, transfers:*

$$T_i(\theta^t) = \begin{cases} -(1-\gamma)V_{-i}(\theta^t_{-i}) & \textit{if } \pi^*(\theta^t) = i \\ 0 & \textit{otherwise} \end{cases}$$

To understand that this is in fact the form dynamic-VCG takes in MAB settings, consider that if $\pi^*(\theta^t) = i$, then:

$$\begin{aligned} & r_{-i}(\theta^t_{-i}, \pi^*(\theta^t)) + \gamma \mathbb{E}[V_{-i}(\tau(\theta^t_{-i}, \pi^*(\theta^t))] - V_{-i}(\theta^t_{-i}) && (6.1)\\ &= 0 + \gamma V_{-i}(\theta^t_{-i}) - V_{-i}(\theta^t_{-i}) && (6.2)\\ &= -(1-\gamma)V_{-i}(\theta^t_{-i}), && (6.3) \end{aligned}$$

since $\tau(\theta^t_{-i}, \pi^*(\theta^t)) = \theta^t_{-i}$ when $\pi^*(\theta^t) = i$ in a MAB. Then for all $j \in I \setminus \{i\}$, note that the immediate value obtained by agents other than j is the immediate value obtained by i, and $\pi^*(\theta^t) = \pi^*(\theta^t_{-j}) = i$, so:

$$\begin{aligned} & r_{-j}(\theta^t_{-j}, \pi^*(\theta^t)) + \gamma \mathbb{E}[V_{-j}(\tau(\theta^t_{-j}, \pi^*(\theta^t))] - V_{-j}(\theta^t_{-j}) && (6.4)\\ &= r_i(\theta^t_i, \pi^*(\theta^t)) + \gamma \mathbb{E}[V_{-j}(\tau(\theta^t_{-j}, \pi^*(\theta^t))] - r_i(\theta^t_i, \pi^*(\theta^t)) - \gamma \mathbb{E}[V_{-j}(\tau(\theta^t_{-j}, \pi^*(\theta^t))] && (6.5)\\ &= 0 && (6.6) \end{aligned}$$

The expected revenue generated by dynamic-VCG in a MAB setting is quite large. At the end of this chapter I present results of an empirical analysis that demonstrates, among other things, that on average over a uniform distribution of agent valuations, only about 10–20% of the value of a decision policy is enjoyed by the agents (the rest is payed to the center). This motivates the substance of this chapter, the introduction of a *dynamic redistribution mechanism.*[1]

[1] This is the first time, to my knowledge, that the idea of redistribution has been applied to a dynamic setting.

6.2 Dynamic-RM

The dynamic redistribution I define extends the intuition underlying the static mechanism **RM**: if we can compute a quantity each period such that the expected sum of these values over time in the future is both independent of i's reporting strategy and guaranteed to be lower than the revenue that will result, we can give back a share to i without distorting his incentives or running a deficit.

For all time-periods t and all possible reported types θ^t, let $w(\theta^t, \pi^*)$ denote the revenue that would result in period t under dynamic-VCG (i.e., $(1-\gamma)V_{-j}(\theta^t_{-j})$ for $\pi^*(\theta^t) = j$). Similarly, for any $i \in I$, let $w(\theta^t_{-i}, \pi^*_{-i})$ be the revenue that would result at t if dynamic-VCG were executed and agent i was not present in the system (i.e., $(1-\gamma)V_{-i,j}(\theta^t_{-i,j})$ for $\pi^*_{-i}(\theta^t) = j$).

Now let $W(\theta^t, \pi^*)$ denote the *total* expected discounted future revenue that results under dynamic-VCG, given that agents report truthfully; i.e., $W(\theta^t, \pi^*) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} w(\theta^k, \pi^*) \,|\, \theta^t, \pi^*]$. Likewise, let $W(\theta^t_{-i}, \pi^*_{-i}) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} w(\theta^k_{-i}, \pi^*_{-i}) \,|\, \theta_{-i}, \pi^*_{-i}]$. So $W(\theta^t_{-i}, \pi^*_{-i})$ is the expected revenue that would result going forward from θ^t if agent i were not present in the system. I now use these concepts to define *dynamic-RM*:

Definition 6.2 (Dynamic-RM). *Executes decision policy π^* and, $\forall i \in I$ and $\theta^t \in \Theta$, transfers:*

$$T_i(\theta^t) = \begin{cases} -(1-\gamma)V_{-i}(\theta^t_{-i}) + \frac{1}{n}\,(1-\gamma)W(\theta^t_{-i}, \pi^*_{-i}) & \textit{if } \pi^*(\theta^t) = i \\ \frac{1}{n}\, w(\theta^t_{-i}, \pi^*_{-i}) & \textit{otherwise} \end{cases}$$

The mechanism is dynamic-VCG plus a revenue "redistribution payment". Let $Z_i : \Theta \rightarrow \Re$ denote this redistribution payment function. We have, $\forall i \in I$ and $\theta^t \in \Theta$,

$$Z_i(\theta^t) = \begin{cases} \frac{1}{n}\,(1-\gamma)W(\theta^t_{-i}, \pi^*_{-i}) & \text{if } \pi^*(\theta^t) = i \\ \frac{1}{n}\, w(\theta^t_{-i}, \pi^*_{-i}) & \text{otherwise} \end{cases}$$

As we are about to see, this payment is defined such that the expected sum of redistribution over time to each agent i is a constant fraction of the expected revenue that would have resulted if i were not present in the system.

Theorem 6.1. *Dynamic-RM is truthful and efficient in within-period ex post Nash equilibrium.*

Proof. Since dynamic-VCG is a dynamic-Groves mechanism, by Theorem 5.1 it is sufficient to show that for every agent i, at all times t, for all $\theta^t \in \Theta$ and all σ'_i, σ''_i, letting $\mathcal{Z}(\theta^t, \sigma_i) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} Z_i(\sigma_i(\theta_i^k), \theta_{-i}^k) \,|\, \theta^t, \pi^*, \sigma_i]$,

$$\mathcal{Z}(\theta^t, \sigma_i') = \mathcal{Z}(\theta^t, \sigma_i'') \tag{6.7}$$

This would imply that dynamic-RM is a dynamic-Groves mechanism. Consider an arbitrary indicator function $h : \mathbb{N} \rightarrow \{0, 1\}$, and define $Y_h : \Theta_{-i} \times \mathbb{N} \rightarrow \Re$ as follows:

$$Y_h(\theta_{-i}, t) = \begin{cases} 0 & \text{if } t > K \text{, else} \\ (1-\gamma)W(\theta_{-i}, \pi_{-i}^*) + \gamma Y_h(\theta_{-i}, t+1) & \text{if } h(t) = 0 \\ w(\theta_{-i}, \pi_{-i}^*) + \gamma \sum_{\theta'_{-i} \in \Theta_{-i}} \tau(\theta_{-i}, \pi_{-i}^*, \theta'_{-i}) Y_h(\theta'_{-i}, t+1) & \text{if } h(t) = 1, \end{cases}$$

where $\tau(\theta_{-i}, \pi_{-i}^*, \theta'_{-i})$ denotes the probability that $\theta'_{-i} \in \Theta_{-i}$ will result when $\pi_{-i}^*(\theta_{-i})$ is taken with current type $\theta_{-i} \in \Theta_{-i}$.

Observe that $\frac{1}{n} Y_h(\theta_{-i}^t, t)$ corresponds exactly to the expected discounted value of total future redistribution payments to i given θ^t and truthful reporting by all $j \neq i$ under a policy that chooses i exactly when $h(k) = 1$, for all times $k \geq t$. This is because, crucially, in MAB settings $\forall \theta \in \Theta$ s.t. $\pi^*(\theta) \neq i$, $\pi^*(\theta) = \pi_{-i}^*(\theta_{-i})$. Let h^1 denote the indicator function with $h^1(k) = 1$, $\forall k \geq 0$. By definition, for all t, θ^t, and i, $Y_{h^1}(\theta_{-i}^t, t) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} w(\theta_{-i}^k, \pi_{-i}^*) \,|\, \theta_{-i}^t, \pi_{-i}^*] = W(\theta_{-i}^t, \pi_{-i}^*)$. I will now show that for all t, θ^t, and i, for any indicator function h,

$$Y_h(\theta_{-i}^t, t) = Y_{h^1}(\theta_{-i}^t, t) = W(\theta_{-i}^t, \pi_{-i}^*) \tag{6.8}$$

Take arbitrary t, θ^t, i, and $h \neq h^1$, and assume for contradiction that $\exists \varepsilon > 0$ s.t. $|Y_h(\theta_{-i}^t, t) - Y_{h^1}(\theta_{-i}^t, t)| \geq \varepsilon$. Now consider the greatest $k \leq K$ such that $h(k) = 0$; call this k_h. Assume first that k_h exists (it may not if $K = \infty$). Define h' to be identical to h except with $h'(k_h) = 1$. Consider any type $\dot{\theta}_{-i}^{k_h}$ associated with a $(k_h - t)^{th}$ expansion of Y_h, given θ_{-i}^t and h. We have that:

$$\begin{aligned} & Y_h(\dot{\theta}_{-i}^{k_h}, k_h) - Y_{h'}(\dot{\theta}_{-i}^{k_h}, k_h) && (6.9) \\ &= (1-\gamma)W(\dot{\theta}_{-i}^{k_h}, \pi_{-i}^*) + \gamma \mathbb{E}\Big[\sum_{k=0}^{K} \gamma^k w(\dot{\theta}_{-i}^{k_h+k}, \pi_{-i}^*)\Big] - \mathbb{E}\Big[\sum_{k=0}^{K} \gamma^k w(\dot{\theta}_{-i}^{k_h+k}, \pi_{-i}^*)\Big] && (6.10) \\ &= (1-\gamma)W(\dot{\theta}_{-i}^{k_h}, \pi_{-i}^*) + \gamma W(\dot{\theta}_{-i}^{k_h}, \pi_{-i}^*) - W(\dot{\theta}_{-i}^{k_h}, \pi_{-i}^*) && (6.11) \\ &= 0 && (6.12) \end{aligned}$$

Note that for an indicator $h^{1'}$ that has $h^{1'}(k) = 1$ for all $k \geq k_h$, $Y_{h^{1'}}(\dot{\theta}_{-i}^{k_h}, k_h) = \mathbb{E}[\gamma^k \sum_{k=0}^{K} w(\theta_{-i}^{k_h+k}, \pi_{-i}^*) \,|\, \dot{\theta}_{-i}^{k_h}, \pi_{-i}]$. This allows the move to (6.10). The move from (6.10) to (6.11) is just by definition of $W(\theta_{-i}, \pi_{-i}^*)$ for any θ_{-i}.

Since $Y_h(\theta^t_{-i}, t)$ and $Y_{h'}(\theta^t_{-i}, t)$ differ only from the $(k_h - t)^{th}$ expansion onwards, and since I showed $Y_h(\dot{\theta}^{k_h}_{-i}, k_h) - Y_{h'}(\dot{\theta}^{k_h}_{-i}, k_h) = 0$ for arbitrary type $\dot{\theta}^{k_h}_{-i}$, this proves that $Y_h(\theta^t_{-i}, t) - Y_{h'}(\theta^t_{-i}, t) = 0$. So for an arbitrary h, switching the last "0-bit" ($\leq K$) to a "1-bit" does not change $Y_h(\theta^t_{-i}, t)$. We can imagine repeating this process, applying it to the resulting function h' yielding h'', and then to h'' yielding h''', and so on. This chain can be continued until we reach h^1, establishing that $Y_h(\theta^t_{-i}, t) - Y_{h^1}(\theta^t_{-i}, t) = 0$.

Now for the case in which there is no finite k_h, consider the indicator function $\hat{h}$ identical to h except with $\hat{h}(k) = 1$ for all $k \geq$ some k_h. We can choose k_h arbitrarily high enough such that $\gamma^{k_h} |Y_{\hat{h}}(\theta^{k_h}_{-i}, k_h) - Y_h(\theta^{k_h}_{-i}, t)| < \varepsilon$ for any $\theta^{k_h}_{-i}$ (since we assume the maximum immediate value any action can yield for any agent is finite). Then since $Y_{\hat{h}}(\theta^{k_h}_{-i}, k_h) = Y_{h^1}(\theta^t_{-i}, t)$ (by the first part of the proof), we have that $|Y_h(\theta^t_{-i}, t) - Y_{h^1}(\theta^t_{-i}, t)| < \varepsilon$. This contradicts our assumption that $|Y_h(\theta^t_{-i}, t) - Y_{h^1}(\theta^t_{-i}, t)| \geq \varepsilon$. Since ε was chosen arbitrarily, this proves the validity of (6.8).

Note again that any agent i's only influence on his redistribution payments is via the policy that is implemented. Then, if we imagine $h(t), h(t+1), \ldots$ as the sequence corresponding to execution of one sequence of actions, and $h'(t), h'(t+1), \ldots$ as that corresponding to any other, we can see that the total expected discounted redistribution payments for i are the same. This combined with equation (6.8) implies that for any reporting strategies σ_i' and σ_i'',

$$\mathcal{Z}(\theta^t, \sigma_i') = \mathcal{Z}(\theta^t, \sigma_i'') = \frac{1}{n} W(\theta^t_{-i}, \pi^*_{-i}) \tag{6.13}$$

So dynamic-RM is a dynamic-Groves mechanism, and the theorem follows by appeal to Theorem 5.1. □

Theorem 6.2. *Dynamic-RM is within-period ex post individual rational.*

Proof. Since dynamic-VCG is within-period ex post IR, it is sufficient to show that $\forall i \in I$ and $\theta^t \in \Theta$, $\mathcal{Z}_i(\theta^t) \geq 0$. This holds trivially from the definition of $Z_i, \forall i \in I$, as the hypothetical revenue that would result for any subset of agents in I is always greater than or equal to 0. This can be seen directly from the dynamic-VCG payment rule, from which revenue expectations are derived. □

Theorem 6.3. *Dynamic-RM is no-deficit.*

Proof. Since dynamic-VCG is no-deficit and yields revenue $w(\theta, \pi^*)$ in any period in which joint state θ was reported, it is sufficient to show that, for every $\theta^t \in \Theta$, letting $i^* = \pi^*(\theta^t)$:

$$\sum_{i \in I} Z_i(\theta^t) \leq w(\theta^t, \pi^*) = (1-\gamma)\, V_{-i^*}(\theta^t_{-i^*}) \tag{6.14}$$

This, in turn, follows if for all $i \in I$ and $\theta^t \in \Theta$, $n \cdot Z_i(\theta^t) \leq w(\theta^t, \pi^*)$. First note that $\forall i \neq i^*$:

$$n \cdot Z_i(\theta^t) = w(\theta^t_{-i}, \pi^*_{-i}) \leq w(\theta^t, \pi^*), \tag{6.15}$$

where the inequality holds simply by observation that $V_{-i,j}(\theta_{-i,j}) \leq V_{-i}(\theta_{-i})$, $\forall \theta \in \Theta$, $i, j \in I$ (by optimality of a policy $\pi^*_{-i,j}$ for the group of agents excluding i and j). To finish the proof we must show that $n \cdot Z_{i^*}(\theta^t) \leq w(\theta^t, \pi^*)$, i.e., that $(1 - \gamma)W(\theta^t_{-i^*}, \pi^*_{-i^*}) \leq (1 - \gamma)V_{-i^*}(\theta^t_{-i^*})$, or,

$$W(\theta^t_{-i^*}, \pi^*_{-i^*}) \leq V_{-i^*}(\theta^t_{-i^*}) \tag{6.16}$$

But this holds immediately by within period ex post individual rationality of dynamic-VCG (Corollary 5.3)—if in a world without some agent i^* the expected discounted payments made to the center were more than the expected discounted value obtained by the agents, some agent would necessarily expect to pay more than the value he obtains from the decision policy. The theorem follows. □

Theorem 6.4. *Utility to each agent i in the truthful equilibrium under dynamic-RM, in expectation forward from any θ^t, is:*

$$V(\theta^t) - V_{-i}(\theta^t_{-i}) + \frac{1}{n} \sum_{j \in I \setminus \{i\}} \Big[V_{-i,j}(\theta^t_{-i,j}) - V_{-i,j}(\theta^t_{-i}) \Big] \tag{6.17}$$

Proof. From Corollary 5.5, under dynamic-VCG given any θ^t expected revenue in the truthful equilibrium equals:

$$\sum_{j \in I} V_{-j}(\theta^t_{-j}) - (n-1)V(\theta^t) = \sum_{j \in I} \Big[V_{-j}(\theta^t_{-j}) - V_{-j}(\theta^t) \Big] \tag{6.18}$$

From equation (6.13), in dynamic-RM the expected payoff to i is increased by $\frac{1}{n}$ times $W(\theta^t_{-i}, \pi^*_{-i})$, the expected revenue that would result under dynamic-VCG from θ^t forward if agent i were not in the system. $W(\theta^t_{-i}, \pi^*_{-i})$ can be written:

$$\sum_{j \in I \setminus \{i\}} \Big[V_{-i,j}(\theta^t_{-i,j}) - V_{-i,j}(\theta^t_{-i}) \Big] \tag{6.19}$$

Adding the payoff under dynamic-VCG and $\frac{1}{n}$ times (6.19) yields (6.17). □

Corollary 6.1. *Social utility in the truthful equilibrium under dynamic-RM, in expectation forward from any θ^t, is:*

$$n \cdot V(\theta^t) - \frac{1}{n} \sum_{i \in I} \Big[(2n-2)V_{-i}(\theta^t_{-i}) + \sum_{j \in I \setminus \{i\}} V_{-i,j}(\theta^t_{-i,j}) \Big] \tag{6.20}$$

Corollary 6.2. *The social utility gain from redistribution in the truthful equilibrium, in expectation forward from any θ^t, is:*

$$\frac{1}{n}\sum_{i\in I}\sum_{j\in I\setminus\{i\}}\Big[V_{-i,j}(\theta^t_{-i,j}) - V_{-i,j}(\theta^t_{-i})\Big] \tag{6.21}$$

6.2.1 Empirical analysis

I now present results of an empirical analysis I ran to determine what the analytical results for social welfare improvement brought by dynamic-RM map to on problem instances. The punchline is that the vast majority of value yielded from decisions is retained by the agents under dynamic-RM, while very little of it is retained under dynamic-VCG.

I examined settings in which activation of a bandit (allocation of the item in an allocation problem) yields either value 1 ("success") or 0 ("failure"), and I represented agent types as beta distributions. Each agent's private information can thus be fully represented by two parameters, α and β, and the probability of success for the next activation equals $\alpha/(\alpha+\beta)$. When an agent is activated if he observes a success his α parameter is updated to $\alpha+1$, and if he observes a failure his β is updated to $\beta+1$.

I generated agent types by selecting a number x between 2 and 20 at random for the number of "prior observations" $(\alpha+\beta)$, and then selecting α at random from 1 to $x-1$, with $\beta = x-\alpha$. I examined both a uniform distribution and a normal distribution. Essentially this process generates either a uniform or normal distribution over prior knowledge in the agent population, and a uniform or normal distribution over valuation levels. We would expect that dynamic-RM would perform better on distributions with less variance, as "similarity" of agent valuations allows greater redistribution in general. This is borne out in the gain in utility achieved with valuations drawn from a normal distribution compared to from a uniform distribution.

I examined different size populations (n). A complete "sample instance" (i.e., a joint type θ) consists of n types drawn randomly as above. For each instance I computed[2] the expected social value of the optimal policy ($V(\theta)$), the expected percentage of that value that is retained by the agents under dynamic-VCG (see Corollary 5.4), and the expected percentage retained by the agents under dynamic-RM (see Corollary 6.1). I computed results for a few different discount factors (γ), but there were not significant differences. Figure 6.1 plots the expected percentage of value retained under each mechanism for a range of different population sizes, with $\gamma = 0.8$. For each population size I computed 100 samples and plotted the average.

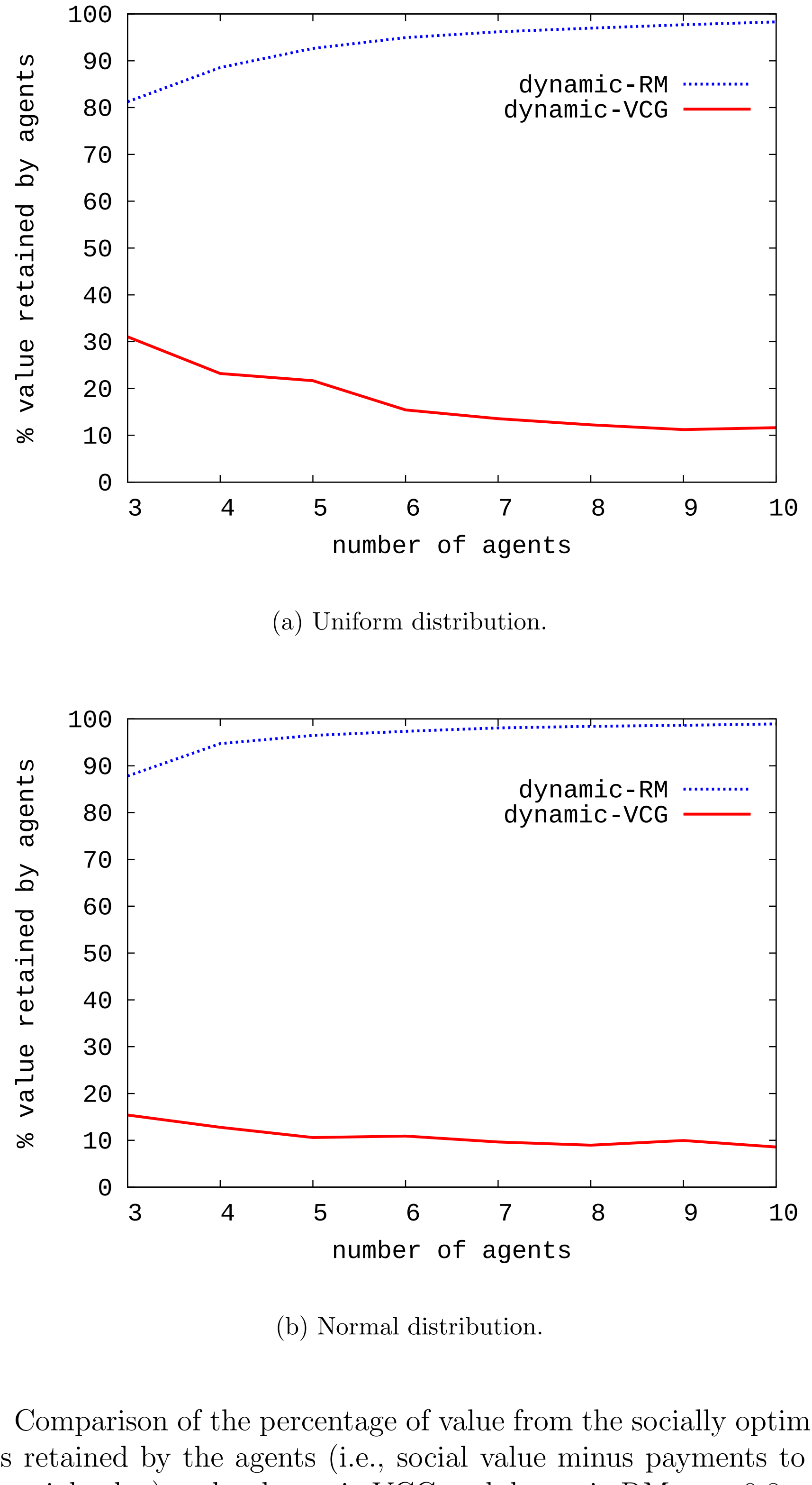
[2] Estimated to within 2–3% accuracy by using the exponential decay of the discount factor.

6.3 Discussion

The motivation for dynamic-RM is that, while dynamic-VCG achieves the very important properties of efficiency, IC, IR, and no-deficit in within-period ex post Nash equilibrium, in many settings the value of decisions is wasted because it is largely not kept within the population of agents. Athey & Segal's [2007] mechanism keeps all value within the group of agents but sacrifices on the equilibrium and, more importantly, on the IR property. A mechanism that is not IR in every time-period (theirs is not) raises significant questions about implementability. For the context of repeated single-item auctions, dynamic-RM does not sacrifice any of these properties while achieving near-perfect budget-balance even with a relatively small number of agents.

Bergemann & Välimäki [2007] observe that dynamic-VCG is unique among mechanisms that satisfy the "efficient exit" condition: agents that will *definitely* no longer have influence on the chosen actions no longer receive or make payments. Clearly dynamic-RM does not satisfy this condition, yet it does not lead to the difficulty that led Bergemann & Välimäki to consider it, namely that agents no longer influencing decisions may leave the mechanism and not make payments owed. In a redistribution mechanism after an agent's exit period he will only *receive* payments.

As in the static setting with mechanism **RM**, the question of optimality naturally arises. Is dynamic-RM "optimal" in the strong sense that I showed of the static version when the analogous fairness constraint is imposed? I suspect the answer is yes, but it may not be extremely consequential in practice since: a) the fairness constraint is probably too strong, and b) dynamic-RM demonstrably performs so well. If it's not optimal and there's a significantly more complex and less scrutable alternative, the ceiling for improvement is low, as we already can maintain almost all value among the agents in bandits settings with more than a few agents. That said, a worst-case analysis could provide some security against any "bad" outcomes, however rare.



(a) Uniform distribution.

(b) Normal distribution.

Figure 6.1: Comparison of the percentage of value from the socially optimal sequence of decisions retained by the agents (i.e., social value minus payments to the center, divided by social value) under dynamic-VCG and dynamic-RM. $\gamma = 0.8$; average over 100 samples for each population size.

Chapter 7

Periodic inaccessibility and dynamically changing populations

Synopsis*

Thus far in the thesis I have considered dynamics in the form of new information arriving for a fixed population of agents; this is the hallmark of *dynamic mechanism design* (DMD). But prior to the DMD model a sequential setting with different dynamics was proposed and studied: in *online mechanism design* (OMD) [Lavi and Nisan, 2000; Parkes and Singh, 2003] the *agent population* changes, though each agent's type does not. The first contribution of this chapter is to extend the online model to one in which agents can be "accessible" or "inaccessible"; an inaccessible agent is defined as one that cannot communicate with the center or make or receive payments. This generalizes arrival/departure dynamics, since in that case agents can be considered inaccessible (prior to arrival), then accessible for some period, then inaccessible again forever (after departure).

So the challenge of DMD is in dealing with new private information held by a static population, while the challenge of OMD is in dealing with a changing population or agent inaccessibility more generally, where each agent has only *static* private information. This chapter's main contribution is in providing a unification of these two models, introducing and addressing one that allows for periodically-inaccessible agents with dynamic types. I present generalizations of the dynamic-VCG mechanism that are effective in the context of periods of inaccessibility together with stochastic local dynamics, which significantly expands the domains to which dynamic mechanisms can be applied.

I first consider a setting in which all agents are persistent (always known to the center, or "identified"), yet each agent has the possibility of becoming inaccessible

*David C. Parkes and Satinder Singh are collaborators on the work presented in this chapter, a version of which appears in a paper titled "Efficient Online Mechanisms for Persistent, Periodically Inaccessible Self-Interested Agents" [Cavallo *et al.*, 2007].

for some period of time (I assume all agents eventually "come back"). The variant on dynamic-VCG I propose here works with a *belief type* about inaccessible agents since their true types can't be communicated, and "logs" the payments that would be executed if they were accessible; when the agents return these payments are executed in a lump sum. I then provide a dynamic-VCG variant for a setting in which agents are not persistent, but rather "arrive" (become accessible and known to the mechanism) and then "depart" (become inaccessible) permanently. Interestingly, the mechanism for this setting is equivalent to the earlier online-VCG mechanism of Parkes and Singh [2003] in the special case in which agents *do not* receive any new private information after arrival (i.e., if we go back to a pure OMD setting). Finally, motivated by an interdependence that results from considering general agent arrival dynamics, I provide an analysis of dynamic mechanism design in settings where the private values assumption does not hold, discussing how the approach of [Mezzetti, 2004] for static environments has a natural application with even stronger results in dynamic settings.

7.1 Persistent agents with periodic inaccessibility

As in Chapters 5 and 6, in this section I consider a fixed set I of persistent agents, but now each agent $i \in I$ may periodically become inaccessible to the center.

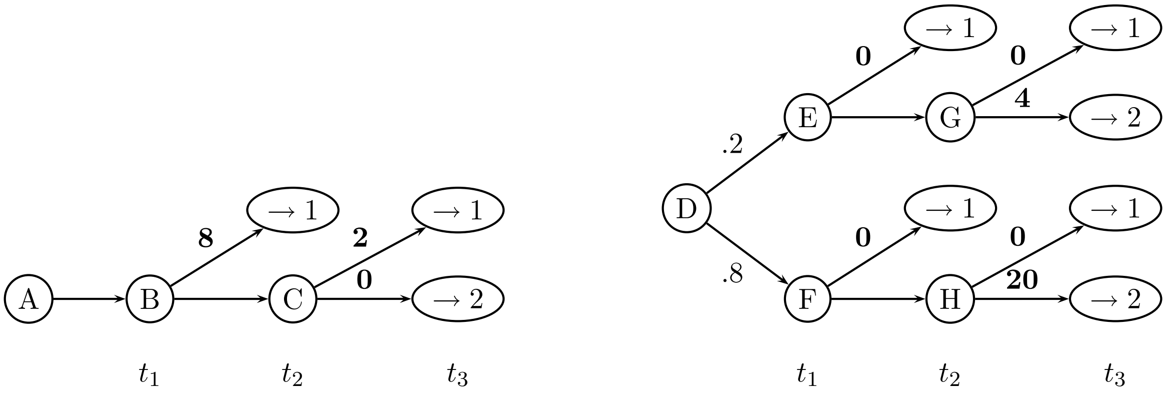
Definition 7.1 (agent inaccessibility). *When an agent is* **inaccessible***, he cannot send messages to the center or receive or make payments. For any agent i, for any inaccessible state θ_i, for any strategy σ_i, $\sigma_i(\theta_i) = \phi$ (the* null *message). An agent cannot claim to be accessible (by making a non-null report) when he is actually inaccessible, but he can pretend to be inaccessible when he is in fact accessible.*

A model accounting for inaccessibility applies, e.g., to environments in which an agent might periodically lose contact with the center due to faulty communication links or choose to temporarily leave the mechanism to do something else for a while. Let $H(\theta^t) \subseteq I$ denote the set of accessible ("here") agents given any true joint state θ^t. For simplicity, I make the following assumption:

Assumption 7.1. *Every agent is accessible and thus able to report a type at time $t = 0$; i.e., for any possible θ^0, $\forall i \in I$, $i \in H(\theta^0)$.*

The results I will describe allow agents to undergo transitions (obtain new types) and receive value while inaccessible in a way that depends on actions taken by the center. Whether an agent transitions to an *inaccessible* state can also depend on the action taken and the agent's current type. The goal is the following: *To design an efficient mechanism in which an agent will truthfully report his type whenever he can, i.e., whenever he is accessible.*

To see the difficulty that inaccessibility poses, consider the simple dynamic-basic-Groves payment scheme with a naive policy that ignores the existence of any inaccessible agents, following a policy that would be optimal if the set of agents were fixed to those accessible in any given period. Each accessible agent is payed the value the other agents obtain in each period (according to the types they report).



(a) Agent 1's MDP.

(b) Agent 2's MDP.

Figure 7.1: Two-agent, 3 time-step single-item allocation problem. Actions ({allocate to 1, allocate to 2, don't allocate}) are implicit in the state transitions. Agent 1 is accessible in each period with probability $1-\varepsilon$ (for negligible ε). Agent 2 is inaccessible at t_1 with probability 1, and with probability $1-\varepsilon$ becomes accessible at t_2.

Example 1. Consider Figure 7.1 with the following accessibility dynamics: with very high probability agent 1 is accessible in all periods; agent 2 is definitely inaccessible[1] in period 0, but will become accessible in period 1 or 2 or, with negligible probability $\varepsilon > 0$, not at all. Under a simple scheme that ignores inaccessible agents and makes Groves payments, if agent 2 is not accessible in period 1, then agent 1 should pretend to be inaccessible to avoid receiving the item. It is likely that agent 2 will become accessible in period 2, be allocated the item, and obtain a higher (expected) value then agent 1 would have. Agent 1 will then be payed that value.

What's happening here? The Groves payments have aligned agent incentives towards maximization of social welfare, but the policy is suboptimal. We see in the example that agent 1 deviates from truth in order to achieve a (socially) superior outcome to the one that would be achieved by the center's naive policy applied to truthful reports.

In an environment with agent inaccessibility, to implement an optimal policy the center must reason about the distribution of possible states for an agent that

[1] Although this appears to violate the assumption that agents are all initially present, one can simply imagine a prior time-step not illustrated in which both agents are accessible and can report their types.

is currently inaccessible. To model this we will consider that the center's type θ_c^t keeps the necessary information to form the proper (Bayesian) beliefs, derived from agent reports. At each period t, from θ_c^t a "belief type" $\tilde{\theta}^t = (\theta_c^t, \tilde{\theta}_1^t, \ldots, \tilde{\theta}_n^t) \in \tilde{\Theta} = \Theta_c \times \tilde{\Theta}_1 \times \ldots \tilde{\Theta}_n$ can be formed, where $\tilde{\theta}_i^t$ defines a probability distribution over agent i's current type at t. For an agent i that is accessible at t ($i \in H(\theta^t)$ for true joint type θ^t) and reports type $\hat{\theta}_i^t$, $\tilde{\theta}_i^t$ assigns probability 1 to $\hat{\theta}_i^t$. If i is inaccessible at t, the distribution $\tilde{\theta}_i^t$ over i's current type is derived from i's last reported type.[2] For instance, considering agent 2 from Example 1 (Figure 7.1), if he does not make a report at t_1, $\tilde{\theta}_2^1$ assigns probabilities 0.2 and 0.8 to states E and F, respectively.[3]

Given belief type $\tilde{\theta}^t$, the expected immediate social value of taking action a is $r(\tilde{\theta}^t, a) = \mathbb{E}[r(\theta^t, a) \,|\, \tilde{\theta}^t]$. Let $V(\tilde{\theta}^t) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} r(\theta^k, \pi(\theta^k)) \,|\, \tilde{\theta}^t, \pi]$, with analogous variants for V_i and V_{-i}. The optimal policy $\pi^* : \tilde{\Theta} \rightarrow A$ maximizes the expected discounted value in every belief type; i.e., letting Π now denote the space of all mappings from joint belief types to actions, for any belief type $\tilde{\theta}^t \in \tilde{\Theta}$, $\pi^* \in \arg\max_{\pi \in \Pi} V(\tilde{\theta}^t, \pi)$, $\forall \tilde{\theta}^t \in \tilde{\Theta}$. The dynamic-VCG mechanism is now defined on belief types:

Definition 7.2 (dynamic-VCG for belief types). *Executes decision policy π^* and, for any $i \in I$ that makes a report at t, for any $\tilde{\theta}^t \in \tilde{\Theta}$, transfers:*

$$T_i(\tilde{\theta}^t) = r_{-i}(\tilde{\theta}_{-i}^t, \pi^*(\tilde{\theta}^t)) + \gamma \mathbb{E}[V_{-i}(\tau(\tilde{\theta}_{-i}^t, \pi^*(\tilde{\theta}^t)))] - V_{-i}(\tilde{\theta}_{-i}^t) \qquad (7.1)$$

$\tilde{\theta}^t$ is the belief type derived from θ_c^t, updated every period based on reported types.

In π^* we have moved from a suboptimal policy to an optimal one given the communication constraints posed by inaccessibility, but this mechanism also fails. This time the payments *do not* succeed in aligning incentives towards social welfare maximization. Dynamic-VCG is designed to make each agent's payoff equal to his contribution to social welfare, but here agents can "evade" paying for the externality they impose on others by faking inaccessibility.

Example 2. Consider again the example in Figure 7.1. If agent 2 is truthful and accessible in period t_1, and in state E, he is better off pretending to be inaccessible. If he were truthful, agent 1 would be allocated the item at t_1 and agent 2's payoff would be zero. By lying, the policy will delay making an allocation until period 2

[2] Recall the assumption that agents are all accessible at time 0, and that an agent's type describes both its current state and a probability distribution over future state transitions contingent on actions. This allows the center to form beliefs on an agent's current type when inaccessible.

[3] The appropriate computational model for this environment is the Partially Observable MDP (POMDP) model. The social problem can then be formulated as a belief-state MDP (see [Kaelbling *et al.*, 1996]).

because $8 < (0.2)4 + (0.8)20 = 16.8$ (ignoring ε). Both agents' payments in period 1 will be zero (agent 2's because he is inaccessible). Agent 2 can then report state G in period 2, receive the item, and make a payment of -2 for net payoff $4 - 2 = 2$. Note the efficiency loss: the center should have allocated to agent 1 in period 1.

The problem is that this formulation of dynamic-VCG is not a dynamic-Groves mechanism. To understand this, define a *true belief type* at time t as the belief type the center would have, given the decision policy, if *every* agent reports his true state whenever he is accessible. A dynamic mechanism (π^*, T) is IC in this environment if in any true belief type, for every agent i, when other agents are truthful i maximizes his payoff by following the truthful strategy. This was not satisfied in dynamic-VCG for belief types; however, a slightly smarter payment scheme *will* work.

It will be useful to formally state the analogue of Theorem 5.1 for this belief-type environment, implicitly extending the dynamic-Groves class to deal with belief types. Here I let $\tilde{\sigma}_i : \Theta_i \rightarrow \tilde{\Theta}_i$ be derived from σ_i to form a belief type in the context of possible null reporting (either by choice when i is accessible or necessity when he is inaccessible). $\tilde{\sigma}_i(\theta_i)$ assigns probability 1 to $\sigma_i(\theta_i^t)$ when $\sigma_i(\theta_i) \neq \phi$, and is otherwise defined by Bayesian updating from the last report. I also extend notation $V(\theta^t)$ in the following natural way (with the analogous extensions of V_{-i}, V, and $\mathcal{T}_i$); it is i's expected value going forward given his true state θ_i^t, true belief state $\tilde{\theta}_{-i}^t$ for and truthful reporting by the other agents, and given that i follows strategy σ_i:

$$V_i(\theta_i^t, \tilde{\theta}_{-i}^t, \tilde{\sigma}_i) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} r_i(\theta_i^k, \pi^*(\tilde{\sigma}_i(\theta_i^k), \tilde{\theta}_{-i}^k)) \,\Big|\, \theta_i^t, \tilde{\theta}_{-i}^t, \pi^*, \tilde{\sigma}_i\Big] \tag{7.2}$$

Lemma 7.1. *A dynamic mechanism (π^*, T) is truthful and efficient in within-period ex post*[4] *Nash equilibrium with persistent, periodically-inaccessible agents if, for all $i \in I$, any true belief type $\tilde{\theta}_{-i}^t \in \tilde{\Theta}_{-i}$ regarding the types of agents other than i, any $\theta_i^t \in \Theta_i$, and any σ_i, there is a function $C_i : \tilde{\Theta} \rightarrow \Re$ such that, letting $\mathcal{C}_i(\theta_i^t, \tilde{\theta}^t, \tilde{\sigma}_i) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} C_i(\tilde{\sigma}_i(\theta_i^k), \tilde{\theta}_{-i}^k)) \,|\, \theta_i^t, \tilde{\theta}^t, \pi^*, \tilde{\sigma}_i]$:*

$$\mathcal{T}_i(\theta_i^t, \tilde{\theta}_{-i}^t, \tilde{\sigma}_i) = V_{-i}(\theta_i^t, \tilde{\theta}_{-i}^t, \tilde{\sigma}_i) - \mathcal{C}_i(\theta_i^t, \tilde{\theta}_{-i}^t, \tilde{\sigma}_i), \tag{7.3}$$

and for any two strategies σ_i' and σ_i'' for agent i, $\mathcal{C}_i(\theta_i^t, \tilde{\theta}_{-i}^t, \tilde{\sigma}_i') = \mathcal{C}_i(\theta_i^t, \tilde{\theta}_{-i}^t, \tilde{\sigma}_i'')$.

Proof. Fix arbitrary agent i and assume all other agents are truthful. Assume the mechanism is *not* within-period ex post IC. Then there must be some θ_i^t, σ_i, and $\tilde{\theta}_{-i}^t$ for which $V(\theta_i^t, \tilde{\theta}^t, \tilde{\sigma}_i) - \mathcal{C}_i(\theta_i^t, \tilde{\theta}^t, \tilde{\sigma}_i) > V(\theta_i^t, \tilde{\theta}_{-i}^t) - \mathcal{C}_i(\theta_i^t, \tilde{\theta}_{-i}^t)$, i.e., $V(\theta_i^t, \tilde{\theta}^t, \tilde{\sigma}_i) > V(\theta_i^t, \tilde{\theta}_{-i}^t)$.

[4] There is a nuance here: the equilibrium property is "ex post" with respect to belief types, which incorporate all information attainable given the communication constraints of the environment, although it is still possible that—if an agent knew an *inaccessible* agent's state (which I presume is impossible)—a deviation could be beneficial.

But we can then construct a policy π where, for every θ_i^k and $\tilde{\theta}_{-i}^k$ yielding true full belief type $\tilde{\theta}^k$, $\pi(\tilde{\theta}^k) = \pi^*(\tilde{\sigma}_i(\theta_i^k), \tilde{\theta}_{-i}^k)$. Now $V(\tilde{\theta}^k, \pi) > V(\tilde{\theta}^k)$, which contradicts optimality of π^*. □

Looking again at the formulation of dynamic-VCG I specified, incentive compatibility is not achieved because payments are not made to an agent in periods during which he is inaccessible, and thus that agent's incentives are not correctly aligned. In Example 2, agent 2 is able to mimic the effect of reporting state F by hiding because he is likely to be in state F anyway (according to the belief type), and by hiding he can avoid making the payment of 6 he would otherwise be forced to make. To isolate the problem, imagine for a moment that payments are always possible and modify dynamic-VCG for belief types so that the payment form specified for accessible agents is also applied to inaccessible agents.

Lemma 7.2. *When payments can be made in every period, dynamic-VCG for belief types is truthful and efficient in within-period ex post Nash equilibrium with persistent, periodically-inaccessible agents.*

Proof. Fix arbitrary agent i and assume all other agents are truthful. Consider any θ_i^t, $\tilde{\theta}_{-i}^t$, and strategy σ_i for i. The total expected discounted payment to agent i forward is:

$$V_{-i}(\theta_i^t, \tilde{\theta}^t, \tilde{\sigma}_i) + \mathbb{E}\Big[\sum_{k=t}^{K} \gamma V_{-i}(\tilde{\theta}_{-i}^{k+1}) - \sum_{k=t}^{K} V_{-i}(\tilde{\theta}_{-i}^k)) \,\Big|\, \tilde{\theta}^t, \pi^*, \tilde{\sigma}_i\Big] \qquad (7.4)$$

From this point the rest of the proof goes through completely analogously to the proof of Theorem 5.5. The mechanism yields expected payoff for i equal to $V(\theta_i^t, \tilde{\theta}^t, \sigma_i) - V_{-i}(\tilde{\theta}_{-i}^t)$, which implies truthfulness and efficiency in within-period ex post Nash equilibrium by Lemma 7.1. □

But in fact agents cannot receive payments in every period, and their incentives are not correctly aligned. The mechanism I now propose addresses this problem by "logging" the payments that an inaccessible agent *should* be making, and then executing them (appropriately scaled for the discount factor) once the agent becomes accessible.

Definition 7.3 (dynamic-VCG#). *A dynamic mechanism $(\pi^*, \hat{T})$, where, for any $i \in I$ that makes a report at t, for any $\tilde{\theta}^t \in \tilde{\Theta}$:*

$$\hat{T}_i(\tilde{\theta}^t) = \sum_{k=t-\delta(t)}^{t} \frac{T_i(\tilde{\theta}^k)}{\gamma^{t-k}}, \text{ and} \tag{7.5}$$

$$T_i(\tilde{\theta}^t) = r_{-i}(\tilde{\theta}^t_{-i}, \pi^*(\tilde{\theta}^t)) + \gamma \mathbb{E}[V_{-i}(\tau(\tilde{\theta}^t_{-i}, \pi^*(\tilde{\theta}^t)))] - V_{-i}(\tilde{\theta}^t_{-i}) \tag{7.6}$$

where $\tilde{\theta}^k$ for every $k \geq 0$ is the belief type derived from θ^k_c (all can be tracked in θ^t_c), and $\delta(t) \geq 0$ is the number of successive periods prior to t that i reported inaccessibility.

The payments here (as before) are defined with respect to the belief type, which incorporates uncertainty about inaccessible agents' types—the only difference in this mechanism is that transfers are now made *cumulatively* for windows of inaccessibility when an agent returns, defined to be (discounting-adjusted) equivalent to those he would have received if payments could have been executed while he was inaccessible.

I now introduce a new assumption that ensures agents cannot *permanently* evade paying their dues. Informally, an agent can run but cannot hide forever (or, "you must pay the piper"). Given this, the expected discounted stream of payments under dynamic-VCG# is the same as in the first dynamic-VCG variant when payments can be made in every period.

Assumption 7.2. *Each agent must eventually make any payments he owes.*

Lemma 7.3. *Given Assumption 7.2, the expected payoff to each agent i forward from any true θ^t_i and true belief type $\tilde{\theta}^t_{-i}$ under dynamic-VCG#, for any strategy σ_i, is equal to that in dynamic-VCG for belief types when payments in that mechanism can be made in every period.*

Proof. The policy is the same and the values intrinsically obtained by each agent for actions taken are the same. Left to show is that the expected discounted stream of payments is the same. We need that for every agent i, time t, θ^t_i, $\tilde{\theta}^t_{-i}$, and σ_i,

$$\mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\, T_i(\tilde{\sigma}_i(\theta^k_i), \tilde{\theta}^k_{-i}) \,\Big|\, \theta^t_i, \tilde{\theta}^t_{-i}, \tilde{\sigma}_i, \pi^*\Big] \tag{7.7}$$

$$= \mathbb{E}\Big[\sum_{\substack{k=t\\H}}^{K} \gamma^{k-t}\, \hat{T}_i(\tilde{\sigma}_i(\theta^k_i), \tilde{\theta}^k_{-i}) \,\Big|\, \theta^t_i, \tilde{\theta}^t_{-i}, \tilde{\sigma}_i, \pi^*\Big], \tag{7.8}$$

where the second summation restricts to time-periods in which agent i reports he is "here" (i.e., $\sigma_i(\theta^k_i) \neq \phi$). To see that this holds, consider any *realization* of i's types

$\boldsymbol{\theta}_i^t, \ldots, \boldsymbol{\theta}_i^K$ and belief types for other agents $\tilde{\boldsymbol{\theta}}_{-i}^t, \ldots, \tilde{\boldsymbol{\theta}}_{-i}^K$. We have:

$$\sum_{k=t}^{K} \gamma^{k-t} T_i(\tilde{\sigma}_i(\boldsymbol{\theta}_i^k), \tilde{\boldsymbol{\theta}}_{-i}^k) \tag{7.9}$$

$$= \sum_{\substack{k=t \\ H\wedge NF}}^{K} \gamma^{k-t} T_i(\tilde{\sigma}_i(\boldsymbol{\theta}_i^k), \tilde{\boldsymbol{\theta}}_{-i}^k) \;+\; \sum_{\substack{k'=t \\ H\wedge F}}^{K} \gamma^{k'-t} \sum_{k=k'-\delta(k)}^{k'} \frac{T_i(\tilde{\sigma}_i(\boldsymbol{\theta}_i^k), \tilde{\boldsymbol{\theta}}_{-i}^k)}{\gamma^{k'-k}}, \tag{7.10}$$

where the first summation restricts to states in which agent i reports his accessibility and this is not the first time $(H \wedge NF)$ after being inaccessible (I also put time-period t here, if accessible), and the second summation is over time-periods that are "first times back" for i after being inaccessible for $\delta(k) > 0$ periods $(H \wedge F)$. Simple algebra completes the proof when combined with Assumption 7.2, which ensures that the final state is not inaccessible. □

Given this, we obtain the desired theorem:

Theorem 7.1. *Dynamic-VCG# is truthful and efficient in within-period ex post Nash equilibrium with persistent agents that are periodically inaccessible when each agent must eventually make payments owed to the center (Assumption 7.2).*

Proof. Follows right away from Lemmas 7.2 and 7.3. □

By introducing the constraint that payments must eventually be made we avoid a manipulation in which an agent does not "re-enter" because he faces a large payment. Returning again to Example 2, the earlier manipulation goes away. Agent 2 can no longer benefit from pretending to be inaccessible when he is in fact accessible and in state E, because he will face a payment of $-6 - 2$ if he makes himself accessible in period 2. But if he *could* avoid payments altogether, a deviation could still be useful—Assumption 7.2 is key.

7.2 Dynamic agent population with arrival process

I now depart from the standard model in which all agents remain "present" in the system, even though they may undergo periods of inaccessibility. I now consider a *dynamically changing population*, with each agent initially inaccessible, then accessible at an *arrival* period, and then becoming inaccessible again at a *departure* period forever.[5] The key difference is that now agents are not "identified" when they are not present in the system—they are unknown to the center, and thus cannot individually

[5] This captures the arrival/departure semantics of Parkes and Singh [2003], generalizing it here to allow for agents with dynamic types.

exert influence on the system or have payments accrue. I will conceptualize the first and last periods in which an agent is accessible as his arrival and departure periods. One can imagine that becoming accessible corresponds to an agent learning his model, or learning of the existence of the mechanism. I assume that an agent obtains no value and undergoes no type transitions while inaccessible (i.e., before arriving or after departing). I continue to allow local dynamics to depend on actions after arrival, with new periodic information arriving as in the DMD model I've considered throughout. Once an agent departs he can participate in no further payments at any time.

Formally, I allow for the set of agents $I = \{1, \ldots, \infty\}$ to be unbounded. The joint type space is now characterized by $\theta = (\theta_c, \{\theta_i\}^{i \in H(\theta_c)}) \in \Theta$, where θ_c keeps sufficient history to model the dynamics of agent arrivals, and $H(\theta) \subseteq I$ is the set of present agents given θ. Type transition probabilities $\tau : \Theta \times A \times \Theta \rightarrow \Re$ are induced by 1) an *arrival model* $\tau_c : \Theta \times A \rightarrow \Theta_c$ that is known to the center (in θ_c) and defines the process by which agents become accessible, and 2) the dynamics $\tau_i : \Theta_i \times A \times \Theta_i \rightarrow \Re$ for each accessible agent, as before. The type space of an agent includes an *absorbing, inaccessible* departure type; once an agent has arrived his own type determines when he will become inaccessible (i.e., depart).

The goal is as before: *to define an efficient mechanism in which each agent will report his true type information in every period in which he is accessible.* Consider a slight modification (or just a reinterpretation) of the dynamic-VCG mechanism to handle agent inaccessibility:

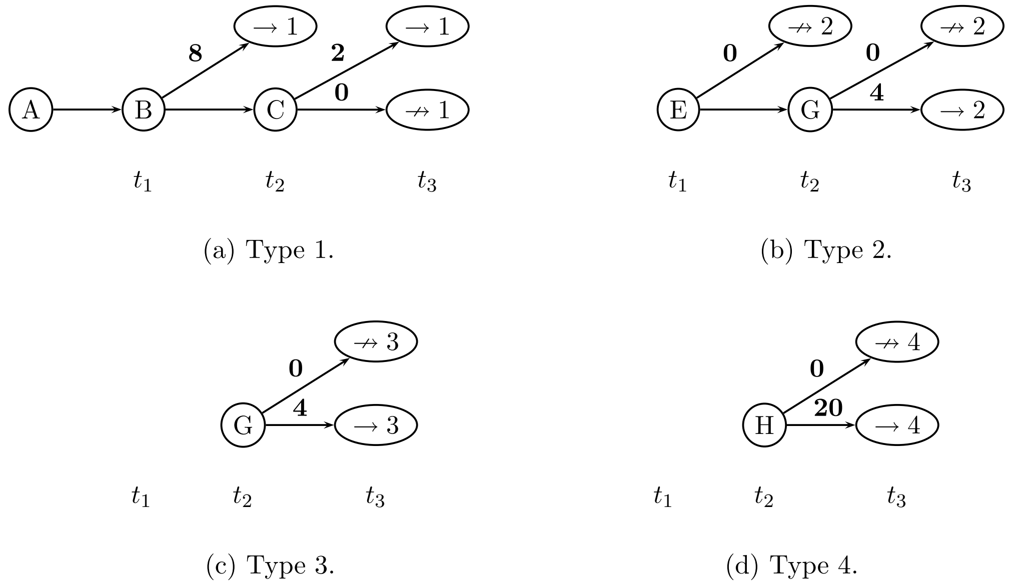
Definition 7.4 (online-dynamic-VCG). *The online-dynamic-VCG mechanism executes decision policy π^* and, given any reported type θ^t, $\forall i \in H(\theta^t)$, transfers:*

$$T_i(\theta^t) = r_{-i}(\theta^t_{-i}, \pi^*(\theta^t)) + \gamma \mathbb{E}[V_{-i}(\tau(\theta^t_{-i}, \pi^*(\theta^t)))] - V_{-i}(\theta^t_{-i}), \tag{7.11}$$

where π^* explicitly incorporates the arrival model and maximizes the expected social utility going forward, including to agents that haven't yet arrived, *and $V_{-i}(\theta^t_{-i})$ is now the expected value going forward including agents that haven't yet arrived but excluding agent i, for the policy that is optimal excluding i.*

Without an additional assumption the online-dynamic-VCG mechanism fails for a subtle reason.

Example 3. Consider an adaptation of Example 2 depicted in Figure 7.2. There are 4 *arrival types*. Define an arrival process so that a single agent of type 1 always arrives at t_0 while at most one agent among types 2, 3, or 4 can arrive (ever), and it is very likely that a type 4 agent will arrive at t_2. If an agent of type 2 arrives at t_1, then he will hide and claim to be inaccessible. The optimal policy will wait to allocate the resource because it likely that a type 4 agent will arrive in the next period. At t_2, the type 2 agent can truthfully report state G (posing as a type 3 agent that just arrived), and will be allocated the item and have to make a payment of 2. This causes an efficiency loss because the item should have been allocated to agent 1 at t_1.



(a) Type 1.

(b) Type 2.

(c) Type 3.

(d) Type 4.

Figure 7.2: Illustration of Example 3. One type 1 agent arrives at t_1, and at most one agent of types 2, 3, or 4 will arrive (at t_2 or t_3, depending on the type). In the absence of Assumption 7.3, even under the online-dynamic-VCG mechanism a type 2 agent arriving at t_1 is better off "posing" as a type 3 agent.

The incentive compatibility of dynamic-Groves mechanisms (Theorem 5.1, but with the optimal policy π^* incorporating the arrival process dynamics) continues to hold in this environment, and the proof that dynamic-VCG is a dynamic-Groves mechanism (Theorem 5.5) also remains valid—except for the last step. The payoff to agent i in any accessible state is, as before, $V_{-i}(\theta^t, \pi^*_{-i})$. But now this can not necessarily be represented as $V_{-i}(\theta^t_{-i}, \pi^*_{-i})$, since now it may be the case that $\pi^*_{-i}(\theta^t) \neq \pi^*_{-i}(\theta^t_{-i})$. That is, π^*_{-i} need not be independent of i's type, and thus, of his strategy σ_i. Although we maintain the private values assumption for the set of accessible agents (with individual values and transitions independent of other agents' types), the probability of future agent arrivals can depend on the *arrival* of agent i: the type reported by i upon his arrival (i.e., his arrival type), or his failure to arrive can influence the center's beliefs about subsequent arrivals.

Fully general arrival dynamics introduce this new interdependence between agents. In Section 7.3 I go into detail regarding conditions under which dynamic mechanisms will work in settings with interdependence more broadly, but for now let's just observe that the following intuitive condition is sufficient to avoid this interdependence:

Assumption 7.3 (conditionally independent arrivals). *The center's arrival model which specifies the distribution over new agent arrival types in period* $t+1$ *is independent of arrivals in times* $0, \ldots, t$ *when conditioned on the sequence of actions taken in times* $0, \ldots, t$.

Theorem 7.2. *The online-dynamic-VCG mechanism is truthful and efficient in within-period ex post Nash equilibrium in this dynamic population, become-accessible-once environment given Assumption 7.3.*

Proof sketch. The proof follows exactly the same lines as the proof for dynamic-VCG in settings without inaccessibility. Each agent's expected payoff is the expected social value (including to agents that haven't yet arrived) minus some constant; then since the center is following an optimal policy (incorporating expectations about arrivals) there can be no gain from deviation. □

7.2.1 A previous mechanism as a special case

The conditionally independent arrivals assumption was implicitly made in the work of Parkes and Singh [2003] (PS) in their online mechanism design framework; I will now unify that earlier framework with the current framework, obtaining the PS setting as a special case with a restrictive assumption. The *online-VCG* mechanism of PS is payoff-equivalent to the online-dynamic-VCG mechanism when coupled with the following assumption:

Assumption 7.4. *Each agent's type is deterministic—i.e., for any* $i \in I$, $\theta_i \in \Theta_i$ *and* $a \in A$, $\tau(\theta_i, a, \theta_i')$ *assigns probability 1 to some* $\theta_i' \in \Theta_i$.

In other words, an agent obtains no *new* private information after it arrives in the mechanism—the local MDP induced by his type is deterministic. Since the only uncertainty is due to the arrival model, agents can report information with a single message (upon arrival). An agent's local problem is now defined by a deterministic finite-state automaton, which plays the role of type in the model of PS. But we can give a dynamic-VCG style interpretation of online-VCG in an environment with discounting (which [Parkes and Singh, 2003] did not account for):

Definition 7.5 (online-VCG-γ). *Each agent can report to the center a single claim θ_i^t about his (deterministic) type. The online-VCG-γ mechanism executes decision policy π^* which is optimal given the arrival model and reported types, and, $\forall \theta^t \in \Theta$, for any $i \in I$ that is accessible at t, transfers:*

$$T_i(\theta^t) = \begin{cases} -r_i(\theta_i^t, \pi^*(\theta^t)) + V(\theta^t) - V_{-i}(\theta_{-i}^t) & \textit{if REPORT}_i \\ -r_i(\theta_i^t, \pi^*(\theta^t)) & \textit{otherwise,} \end{cases} \tag{7.12}$$

where the expected values V and V_{-i} are taken with respect to the arrival model of the center and the agent reports, and REPORT$_i$ indicates that this is the period in which i reports his arrival and type.

The cumulative effect of the payments is that agent i pays to the center the total (reported) value he obtains for the sequence of decisions, and receives a payment of $V(\theta^t) - V_{-i}(\theta_{-i}^t)$ in the period in which he announces his type. This payment is equal to the expected marginal contribution to social utility by i given the arrival model of the center and the reported types of agents.

Theorem 7.3. *Online-VCG-γ is truthful and efficient in within-period ex post Nash equilibrium*[6] *in this dynamic population, become-accessible-once setting given Assumptions 7.3 and 7.4.*

Proof. An application of the dynamic-Groves result holds here: assume an agent i, arriving at time t in which the joint reported type for other agents is θ_{-i}^t and other agents reporting in future periods will do so truthfully, has expected discounted transfers going forward equal to $V_{-i}(\theta^t, \sigma_i) - \mathcal{C}(\theta^t, \sigma_i)$ when he plays strategy σ_i, where $\mathcal{C}(\theta^t, \sigma_i') = \mathcal{C}(\theta^t, \sigma_i'')$ for any σ_i' and σ_i''. Then truthfulness is a payoff maximizing strategy for i.

Fix some agent i, type θ_i^t, strategy σ_i, and joint reported type θ_{-i}^t for other agents at time t in which agent i has arrived, and assume all other agents reporting in the future will be truthful. Let $t' \geq t$ be the time that i reports his arrival given θ^t and

[6] The nuance here is that the mechanism is efficient modulo the 1-report restriction. That is, it may be inefficient (though it is still truthful) if an agent has falsely reported his type (which we will see never happens in equilibrium), since the mechanism does not allow agents to report "corrections" to previous lies.

σ_i. The expected discounted sum of transfers to i forward from θ^t is:

$$\mathbb{E}\Big[\gamma^{t'-t}\Big(-V_i(\theta^{t'},\sigma_i)+V(\theta^{t'},\sigma_i)-V_{-i}(\theta^{t'}_{-i})\Big)\,\Big|\,\theta^t,\pi^*,\sigma_i\Big] \tag{7.13}$$

$$=\mathbb{E}\Big[\gamma^{t'-t}\Big(V_{-i}(\theta^{t'},\sigma_i)-V_{-i}(\theta^{t'}_{-i})\Big)\,\Big|\,\theta^t,\pi^*,\sigma_i\Big] \tag{7.14}$$

$$=V_{-i}(\theta^t,\sigma_i)-\mathbb{E}\Big[\sum_{k=t}^{t'}\gamma^{k-t}r_{-i}(\theta^k_{-i},\pi^*(\theta^k_{-i}))+\gamma^{t'-t}V_{-i}(\theta^{t'}_{-i})\,\Big|\,\theta^t,\pi^*,\sigma_i\Big] \tag{7.15}$$

$$=V_{-i}(\theta^t,\sigma_i)-V_{-i}(\theta^t_{-i}) \tag{7.16}$$

Assumption 7.3 is exhibited in my expression of $V_{-i}(\theta^t,\pi^*_{-i})$ as $V_{-i}(\theta^t_{-i})$ (in the description of the mechanism and in the above equations). Converting equation (7.13) to (7.14) follows by simple addition of the first two terms. The move from (7.14) to (7.15) is by adding $\mathbb{E}[\sum_{k=t}^{t'}\gamma^{k-t}r_{-i}(\theta^k_{-i},\pi^*(\theta^k_{-i}))\,|\,\theta^t,\pi^*,\sigma_i]$ to the first term of (7.14) and then subtracting it again. Finally, combining the terms within the expectation in (7.15) yields (7.16). Since the agent's total expected payment (7.16) has the dynamic-Groves form specified in the first part of the proof, the mechanism is truthful in within-period ex post Nash equilibrium. □

One reason to adopt online-VCG-γ rather than online-dynamic-VCG in the special environment in which Assumption 7.4 holds is that the payments require solving $V_{-i}(\theta^t_{-i})$ only once for each agent arrival, whereas in online-dynamic-VCG this problem is solved in every period in which the agent remains accessible according to his report. But this same savings in computation could be achieved with the online-dynamic-VCG transfer scheme if redundant recomputation is avoided by explicitly considering the implications of Assumption 7.4. That assumption is very restrictive, and the extension from online-VCG to dynamic-VCG# brings a vast expansion of the domains to which a mechanism accounting for population dynamics can be applied.

7.3 Dynamic mechanisms for interdependent settings

Throughout the thesis I have assumed a *private values* environment, where an agent's expected value for a given outcome (or sequence of outcomes) is a function of his own type, but not of the other agents' types. This is fairly typical in mechanism design, but there are in fact important domains that don't fit the mold. For instance, if an oil company is evaluating a field that's up for bid, the values that *other* firms have for the same field may be relevant, in that these values could indicate information about the oil contents of the field that only these other firms know. One firm's expected value for an outcome depends on another's. In static mechanism design there is a rather strong negative result about what can be achieved when agent valuations may be interdependent in this way:

Theorem 7.4 (entailed by [Jehiel and Moldovanu, 2001]). *In static environments where agent valuations may be arbitrarily interdependent,*[7] *there exists no mechanism that is truthful and efficient in an ex post Nash equilibrium.*

But when it is only agents' *expected* (rather than *actual*) values that are interdependent, this negative result only holds if we construe mechanism design as confined to making payments *ex ante* of realization of a selected outcome. The intuition for the problem is that if the payment I receive depends on your expected value for an outcome, and your expected value for an outcome depends on what I report my expected value to be, I can potentially gain from reporting a false expected value. This kind of reasoning breaks down, though, if we imagine executing payments *ex post* of the outcome realization: if your *expected* value for an outcome is influenced by my report but your *actual* value is not, then basing payments on *actual* values (as reported *after* they are realized) nullifies the effect my reported value has on my payment.

This is exactly the idea formalized by Mezzetti [2004], who formulates a two-stage mechanism for one-shot settings: in the first stage agents are asked to report their private types and an outcome is selected and implemented; in the second stage agents are asked to report the value they experienced from the outcome, and these reports form the basis for payments made by the mechanism. Mezzetti argues that such mechanisms *always* (i.e., regardless of any interdependencies in values) allow for implementation of efficient outcomes in equilibrium. It is worth noting, though, that the equilibrium is fragile in that agents are indifferent about reporting their true experienced values in the second stage, since the outcome has already been selected and there are no further implications of an agent's report for his own utility.

Dynamic environments in which sequences of decisions are to be made provide natural settings for a Mezzetti-like approach to interdependence,[8] and the fragility of the equilibrium in the static problem goes away in dynamic settings (except in the last period for finite-horizon problems). Interdependence would take the form of agent transition functions now being mappings from a complete *joint* type to a successor-type. Imagine a repeated allocation scenario in which some agent i was just allocated the resource; in the next period some agent j's expected value for the resource may change if he finds out that i obtained very little value from it—perhaps the resource is faulty in some way that is only observable by an agent to whom it has been allocated.

[7] Where "arbitrarily independent" precludes the kind of structure provided by the single-crossing condition of [Dasgupta and Maskin, 2000]; see also [Krishna, 2002].

[8] This fact is noted by Athey & Segal [2007]. The analysis in this section should be viewed as working out details of this observation and adding further analysis.

We will need to make a basic but important assumption: that agents can accurately quantify and report the immediate value they have experience after execution of a decision.[9] This assumption is also present in [Mezzetti, 2004]. For convenience, I will assume that such a report can be encapsulated in an agent's broader (i.e., more informative) type report:

Assumption 7.5 (privately realized types). *For any agent i with current true type θ_i, the last immediate value $\hat{r}_i(\theta_i)$ actually obtained by i can be discerned from θ_i.*

I will say that a mechanism (π, T) is implemented in the *interdependent-dynamic* framework if it is characterized by the following timing:

Definition 7.6 (interdependent-dynamic mechanism framework). .

- *Each agent i reports to the center a claim θ_i^0 about his initial type.*
- *The center executes action $\pi(\theta^0)$.*
- *At every time step $t = 0, \ldots, K$:*

1. *Each agent i obtains value $r(\theta_i^t, \pi(\theta^t), \theta_i^{t+1})$ as transition $\theta_i^t \rightarrow \theta_i^{t+1}$ is realized.*
2. *Each agent i reports to the center a claim θ_i^{t+1} about his type.*
3. *The center makes a payment $T_i(\theta^{t+1})$ to each agent i.*
4. *The center executes action $\pi(\theta^{t+1})$.*

Note that time "ticks forward" each time an action is taken. In each period an agent obtains value from the action just taken, undergoes a type transition, makes a report, and receives a transfer payment. See Figure 7.3 for an illustration of this timing. A transfer $T_i(\theta^{t+1})$ is actually executed in time-period t, but in the interdependent-dynamic framework it is based on the joint type the agents report transitioning to in t rather than the type they started t with. With the exception of the dynamic-balanced mechanism, previous to now I've defined mechanisms such that an agent's transfer at t is a function of θ^t (the type at the beginning of t) and is thus based on an *expectation* regarding the type transition that will occur at t.[10]

A version of any of the dynamic mechanisms described in this thesis can be implemented in a way that fits the interdependent-dynamic framework; the definition

[9] It should be noted that this assumption holds in many, but not all, of the domains that are typically thought of as interdependent. It will not hold when an agent's *actual* value depends on the type of another agent, for instance in a case where owning a particular item like a car brings "prestige" value when other people admire it.

[10] For instance, I described the dynamic-basic-Groves mechanism in a way such that $T_i(\theta^t) = r_i(\theta_{-i}^t, \pi^*(\theta^t))$; the payment to i equals the *expected* value agents other than i will receive in time-step t.

simply specifies a variation on the order in which actions, reports, and payments occur. This reordering allows for reports of actual “just experienced” values rather than expected values for the period, and it turns out that this can yield new possibilities when type transitions are interdependent and types are privately realized.

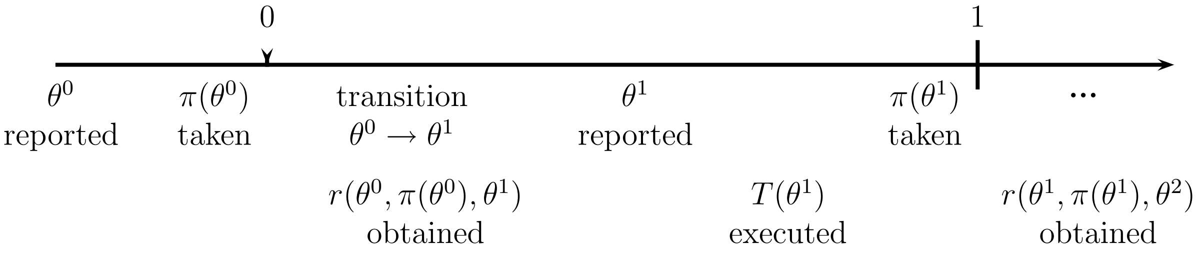

0
1
θ^0 reported
$\pi(\theta^0)$ taken
transition $\theta^0 \to \theta^1$
θ^1 reported
$\pi(\theta^1)$ taken
···
$r(\theta^0, \pi(\theta^0), \theta^1)$ obtained
$T(\theta^1)$ executed
$r(\theta^1, \pi(\theta^1), \theta^2)$ obtained

Figure 7.3: An illustration of the timing of a mechanism implemented in the interdependent-dynamic framework. Note that value and transfers received in the same period are discounted in the same way. Compare to the previous dynamic mechanism timing illustrated in Figure 5.1.

In private values settings, we can express (and I have) the expected actual value that agents other than some i will obtain, $V_{-i}(\theta^t, \sigma_i)$, as $\mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} r_i(\theta_i^k, \pi(\sigma(\theta^k))) \mid \theta^t, \pi^*, \sigma]$. This is because private values entails that:

$$\mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} r_i(\theta_i^k, \pi(\sigma(\theta^k))) \,\Big|\, \theta^t, \pi, \sigma_i\Big] = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k+1-t} \hat{r}_i(\theta_i^k) \,\Big|\, \theta^t, \pi, \sigma_i\Big] \tag{7.17}$$

A version of Theorem 5.1 continues to hold in dynamic environments with interdependent values if we define $V_{-i}(\theta^t, \sigma_i) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k+1-t} \hat{r}_{-i}(\theta_{-i}^k) \,\Big|\, \theta^t, \pi^*, \sigma_i]$. The intuition is similar to before: now if payments are defined such that, in equilibrium, each agent’s *actual* utility going forward equals exactly the *actual* social welfare (minus a constant), no agent can possibly gain from deviating, by optimality of π^*.

Lemma 7.4. *Even with interdependent type transitions, if Assumption 7.5 is satisfied then any mechanism (π^*, T) implemented in the interdependent-dynamic framework is truthful and efficient in within-period ex post Nash equilibrium if, $\forall i \in I$, there exists some $C_i : \Theta \to \Re$ such that $\forall \theta^t \in \Theta$,*

$$\mathcal{T}_i(\theta^t) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k+1-t} \hat{r}_{-i}(\theta_{-i}^k) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] - \mathcal{C}_i(\theta^t), \tag{7.18}$$

where $\mathcal{C}_i(\theta^t, \sigma_i) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} C_i(\sigma_i(\theta_i^k), \theta_{-i}^k) \mid \theta^t, \pi^*, \sigma_i]$ and, for any σ_i' and σ_i'',

$\mathcal{C}_i(\theta^t, \sigma_i') = \mathcal{C}_i(\theta^t, \sigma_i'')$.

Proof. Fix any agent $i \in I$, and assume all other agents are truthful. For any $\theta^t \in \Theta$, in such a mechanism the difference in i's expected utility going forward from being truthful or playing strategy σ_i equals:

$$\mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\hat{r}(\theta^k) - \sum_{k=t}^{K} \gamma^{k-t} C_i(\theta^k) \,\Big|\, \theta^t, \pi^*\Big] - \tag{7.19}$$

$$\mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\hat{r}(\theta^k) - \sum_{k=t}^{K} \gamma^{k-t} C_i(\sigma_i(\theta_i^k), \theta_{-i}^k) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] \tag{7.20}$$

$$= \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\hat{r}(\theta^k) \,\Big|\, \theta^t, \pi^*\Big] - \mathcal{C}_i(\theta^t) - \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\hat{r}(\theta^k) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] + \mathcal{C}_i(\theta^t, \sigma_i) \tag{7.21}$$

$$= \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\hat{r}(\theta^k) \,\Big|\, \theta^t, \pi^*\Big] - \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\hat{r}(\theta^k) \,\Big|\, \theta^t, \pi^*, \sigma_i\Big] \tag{7.22}$$

By optimality of π^* this quantity is greater than or equal to 0, so truthfulness maximizes expected utility going forward. □

Incentive compatibility of an interdependent-dynamic implementation of the dynamic-basic-Groves mechanism follows as an immediate corollary.

Theorem 7.5. *When Assumption 7.5 holds, an interdependent-dynamic version of the dynamic-basic-Groves mechanism that, at every time t, makes payment $\hat{r}_{-i}(\theta_{-i}^{t+1})$ to each each agent i given joint reported type θ^{t+1}, is truthful and efficient in within-period ex post Nash equilibrium even when type transitions are interdependent.*

Recall that the basic-dynamic-Groves mechanism has serious budgetary problems. Of course an ex ante charge of the type dynamic-EAC imposes will yield a mechanism that is ex ante no-deficit, but the cost of that approach is a weakening from ex post to ex ante IR. I will now turn to dynamic-VCG, with an eye towards the strong no-deficit and IR properties it simultaneously achieves.

7.3.1 The independence requirements of dynamic-VCG

In the truthful equilibrium under dynamic-VCG, the payoff to any agent i playing strategy σ_i is $V(\theta^t, \sigma_i) - V_{-i}(\theta^t, \pi_{-i}^*)$. In fact the argument for incentive compatibility in the private values setting was based on this very observation: since i cannot influence $V_{-i}(\theta^t, \pi_{-i}^*)$, he will choose σ_i to maximize $V(\theta^t, \sigma_i)$; this strategy is truth, by optimality of π^*.

Note that this reasoning *requires* that i cannot influence $V_{-i}(\theta^t, \pi_{-i}^*)$; if he could, it would distort his incentives away from maximization of social welfare. In Section

7.2 we saw scenarios in which the policy that would be optimal for the set of agents other than i depends on the current state of i, due to an interdependence in the probabilities of arrival for different agents.

With the proper interpretation, the private values assumption is strong enough to exclude such scenarios. Considering "inaccessible" as a state for an agent, if whether or not every other agent arrives (i.e., transitions to an "accessible" state) is independent of whether every other agent is in an accessible or inaccessible state, then conditionally independent arrivals (Assumption 7.3) is satisfied, and agents cannot benefit from deviating from truth. But in fact the private values assumption may be viewed as (just slightly) stronger than what is required for incentive compatibility of an interdependent-dynamic implementation of dynamic-VCG to hold. A somewhat weaker sufficient condition when agent strategies are inherently limited (e.g., when an inaccessible agent cannot report accessibility), is the following.

Assumption 7.6. $\forall i \in I, \theta^t \in \Theta, \sigma_i', \sigma_i'', \;\; V_{-i}(\theta^t, \pi_{-i}^*, \sigma_i') = V_{-i}(\theta^t, \pi_{-i}^*, \sigma_i'')$.

Recall that for any θ^t, $V_{-i}(\theta^t, \pi_{-i}^*, \sigma_i)$ denotes the expected value to agents other than i when the policy that is optimal for them is executed from joint type θ^t, i plays strategy σ_i, and other agents are truthful.

Theorem 7.6. *The interdependent-dynamic implementation of dynamic-VCG is within-period ex post incentive compatible and efficient when Assumptions 7.5 and 7.6 hold.*

Proof. In an interdependent-dynamic implementation (with Assumption 7.5 in force), the dynamic-VCG payment function for each $i \in I$ can be defined as:

$$\forall \theta^{t+1} \in \Theta, \;\; T_i(\theta^{t+1}) = \hat{r}_{-i}(\theta_{-i}^{t+1}) + \gamma V_{-i}(\theta_{-i}^{t+1}) - V_{-i}(\theta_{-i}^t) \tag{7.23}$$

Then by Lemma 7.4 it is sufficient to show that, given Assumption 7.6 and truthful reporting by all other agents going forward, for any two strategies σ_i' and σ_i'', for any time $t > 0$ and θ^t,

$$\mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\Big(\gamma V_{-i}(\theta^{k+1}, \pi_{-i}^*, \sigma_i') - V_{-i}(\theta^k, \pi_{-i}^*, \sigma_i')\Big) \,\Big|\, \sigma_i'\Big] \tag{7.24}$$

$$= \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\Big(\gamma V_{-i}(\theta^{k+1}, \pi_{-i}^*, \sigma_i'') - V_{-i}(\theta^k, \pi_{-i}^*, \sigma_i'')\Big) \,\Big|\, \sigma_i''\Big] \tag{7.25}$$

Skipping some of the algebra (since it follows exactly the proof of Theorem 5.5), for any σ_i,

$$\mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}\Big(\gamma V_{-i}(\theta^{k+1}, \pi_{-i}^*, \sigma_i) - V_{-i}(\theta^k, \pi_{-i}^*, \sigma_i)\Big) \,\Big|\, \sigma_i\Big] = -V_{-i}(\theta^t, \pi_{-i}^*, \sigma_i) \tag{7.26}$$

Then, by Assumption 7.6 we have that for any σ_i' and σ_i'',

$$\mathbb{E}\Big[\sum_{k=t}^{K}\gamma^{k-t}\Big(\gamma V_{-i}(\theta^{k+1},\pi^*_{-i},\sigma_i') - V_{-i}(\theta^k,\pi^*_{-i},\sigma_i')\Big)\;\Big|\;\sigma_i'\Big] \tag{7.27}$$

$$= \; - V_{-i}(\theta^t,\pi^*_{-i},\sigma_i') = -V_{-i}(\theta^t,\pi^*_{-i},\sigma_i'') \tag{7.28}$$

$$= \mathbb{E}\Big[\sum_{k=t}^{K}\gamma^{k-t}\Big(\gamma V_{-i}(\theta^{k+1},\pi^*_{-i},\sigma_i'') - V_{-i}(\theta^k,\pi^*_{-i},\sigma_i'')\Big)\;\Big|\;\sigma_i''\Big], \tag{7.29}$$

and the theorem follows. □

While I've relaxed the private values condition slightly, the takeaway point from this analysis is primarily negative: Assumption 7.6 was required for the last step of the proof to go through, and so it seems that the natural interdependent-dynamic implementations of dynamic-VCG and its inaccessibility-handling variants will probably not work without making some private-values style assumptions. Though types may be privately realized, Assumption 7.6 is not generally satisfied when type transitions can be interdependent, since one agent's expectation of the types he will realize in the future may be contingent on another agent's reported type, which is a function of the strategy he chooses. Of course this analysis does not in and of itself preclude the possibility for other dynamic-VCG-like mechanisms that could be effective even in interdependent settings.

7.4 Summary

In this chapter we saw dynamic mechanisms for two different environments in which agents may be inaccessible: those in which agents are *persistent* and cannot remain inaccessible forever; and those in which agents are initially "unidentified", arrive at a certain time-step, stay for a while, and then depart permanently. I provided mechanisms for each of these environments that generalize dynamic-VCG, which is defined for the persistent, always-accessible agents case.

Finally, I addressed the issue of dynamic mechanism design for interdependent values settings, where the type transitions of one agent may depend on the local information held by other agents. I found that the intuitions of [Mezzetti, 2004] for static environments have a very natural application in dynamic settings. Moreover, the fragility of the equilibrium achieved in the static setting becomes more robust in the dynamic case, since the reports of agents continually form the basis of both payment and policy decisions. I specified an extension of the dynamic-Groves mechanism class for interdependent dynamic settings. In lieu of a successful direct interdependent-dynamic implementation of dynamic-VCG, future research should explore the possibility of other approaches towards achieving no-deficit without sacrificing IR.

Chapter 8

Efficient metadeliberation auctions

Synopsis[*]

In this chapter I consider a resource allocation scenario in which the interested parties can, at a cost, individually research ways of using the resource to be allocated, potentially increasing the value they would achieve from obtaining it. Each agent has a private model of his research process and obtains a private realization of his improvement in value, if any. From a social perspective it is optimal to coordinate research in a way that strikes the right tradeoff between value and cost, ultimately allocating the resource to one party– thus this is a problem of *multi-agent metadeliberation.* I provide a reduction of computing the optimal deliberation-allocation policy to computing Gittins indices in multi-armed bandit worlds, and apply a modification of the dynamic-VCG mechanism to yield truthful participation in a within-period ex post Nash equilibrium. The mechanism achieves equilibrium implementation of the optimal policy even when agents have the capacity to deliberate about other agents' valuations, and thus addresses the problem of *strategic deliberation.*

8.1 Motivation and background

Imagine a group of firms competing for the allocation of a new technology. Each firm initially has some estimate of how valuable the technology is to its business, and is able to learn new ways of using the technology for greater profit through research. If such research were costless and instantaneous, the socially optimal plan would have all firms research the technology in all ways possible, at which point it would be allocated to the firm with highest value. But in reality performing such research will come at a cost. To maximize expected social welfare an optimal tradeoff should be

[*] David C. Parkes is a collaborator on the work presented in this chapter, a version of which appears in a paper with the same name published in the proceedings of the AAAI 2008 conference [Cavallo and Parkes, 2008].

struck between value and cost, with firms following a coordinated research policy. In addition to gathering information from the outside world, agents may improve their values for the resource by performing some costly computation, for instance finding better business plans involving the resource. I adopt the general term *deliberation* for any such value-improving process, and I consider the social planner's *metadeliberation* problem—deciding when and how to perform deliberation, including when to stop and allocate the resource.

The main contributions of this chapter lie, first, in describing a method of reducing such deliberation-allocation problems to the *multi-armed bandit* problem, thus providing a computationally efficient way of determining optimal policies. This is non-trivial because the local problem of each agent (or firm) includes two actions in each state—deliberation and allocation—and is thus not modeled as a simple Markov chain. The second contribution is in applying dynamic mechanism design to achieve equilibrium implementation in the face of selfish, strategic parties. My solution provides a *metadeliberation auction*, in which agents will choose to reveal private information about their deliberation processes and also to voluntarily perform deliberation as and when specified by the optimal solution.

In an extension, I allow that agents may have deliberation processes for the value of *other* agents for the resource. This borrows from the earlier model of Larson and Sandholm [2005], in which agents have costly deliberation processes and can perform "*strategic deliberation*" about the value of other agents. But whereas they exclude solutions in which the mechanism is actively involved in coordinating the deliberation of agents, I allow for this and obtain positive results where they have impossibility results. In particular, when the optimal policy calls for one agent to perform research on behalf of another, we can achieve this. In the proposed mechanism an agent is paid for increasing (via his deliberation process) the value of the item to another agent, and thus enjoys the beneficial results of the deliberation it performs.

Related work

On the policy computation side, the most important result for our purposes is that of Gittins [1974], who showed that the multi-armed bandit problem has a solution with complexity that grows linearly in the number of agents (see Chapter 4, Section 4.4.1). Glazebrook [1979] extended this result to "stoppable" bandits, where execution of the system can be halted for a final reward. The multi-agent deliberation-allocation problem I consider falls within his framework and my reduction to the bandits problem is a special case of his reduction. This noted, I provide a new proof that elucidates the reduction and leverages the special structure in this environment.

Of previous studies that address related problems, perhaps closest is that of Weitzman [1979], whose foundational result demonstrated that an index policy can be used to optimally search among a set of alternatives, where the exact value of an alternative is revealed for a cost. Weitzman's result, intuitively, has much in common with that of Gittins, as both observe that a "reservation price" can be computed for each alternative independent of the others, with the optimal policy defined by choosing the option with highest reservation price. Weitzman's setting is different from Gittins's (and closer to this chapter's and Glazebrook's), though, since one accumulates options as search proceeds and ultimately stops the procedure. However, it is limited to addressing settings where "searching" an option is a one-step affair. This chapter can be viewed as an extension of Weitzman's work, where the options are more complex and exploring them can be multi- time-stepped with incremental results.[1]

Bergemann and Välimäki [2006] look at the problem of information acquisition by bidders in a single-item auction, and show that when such acquisition is one-shot and simultaneous among the group, the Vickrey auction provides the right *ex ante* incentives. Larson [2006] and Cremer et al. [2007] use Weitzman's result to form an optimal-search auction model with sequential information acquisition, but also assume that a buyer's acquisition process is instantaneous (not multi time-stepped, with incremental information). Parkes [2005] addresses the role of auction design given participants that have costly or limited value refinement capabilities, especially the tradeoff between sealed bid and iterative designs, but does not provide an optimal method. Results presented in previous chapters on dynamic mechanism design—in particular Bergemann and Välimäki's dynamic-VCG mechanism—will find application in the solution I propose here.

8.2 The setting

Members of a set I of n agents ($\{1, 2, \ldots, n\}$) compete for allocation of a resource. Each agent $i \in I$ has an initial value for the resource, and can refine his value repeatedly via costly "deliberation". To keep things simple, until Section 8.5 I will assume that each agent has only one such deliberation process, and moreover that no agent has a deliberation process about the value of any other agent.

Each agent i's type $\theta_i \in \Theta_i$ induces an MDP model $M_i = (S_i, A_i, \tau_i, r_i)$ of his value for the resource and how it will change subject to deliberation. In this chapter I will give significant focus to computational issues and provide results that are of interest in a general MDP context independent of incentive issues, and so I will use the MDP representation notation and language (rather than the type abstraction) explicitly for most of the exposition. S_i is i's local state space. The action space $A_i = \{\alpha_i, \beta_i\}$, where α_i allocates the resource to i and β_i is deliberation by i. States evolve according to a (possibly nondeterministic) transition function. I use $\tau_i(s_i, a_i) \in S_i$ for

[1] But note that Weitzman's result can be applied to undiscounted settings, unlike ours or those of Gittins and Glazebrook. Castañon et al. [1999] analyze a simpler model in which one can "sample" from options without limit in a potentially undiscounted setting, but with each sample drawn independently from the identical distribution, which does not model the economic settings of [Weitzman, 1979] or the current chapter.

the random variable representing the state that results when a_i is taken in state s_i, defined so that $\tau_i(s_i, \alpha_i) = \phi_i$, where $\phi_i \in S_i$ is a special *absorbing* state entered after allocation from which no additional actions are available. Reward function r_i can be described in terms of the value $v_i(s_i)$ i obtains if allocated the resource (performing no further deliberation) while in given state $s_i \in S_i$, and the cost c_i that i incurs from performing deliberation (for simplicity I assume c_i is constant, though my results hold as long as c_i is a bounded function of i's state). If the resource hasn't yet been allocated: if deliberation is performed reward $-c_i$ is obtained, and if allocation is performed value $v(s_i)$ is obtained.

A set of further assumptions placed on this framework defines a *domain.* In the setting as I described it, researching new uses may yield a greater value for the resource, but agents won't forget previously known uses, so the following is natural:

Assumption 8.1 (uncertainly improvable values). *Agent valuations never decrease, i.e.,* $\forall s_i, s_i' \in S_i$ *such that* $Pr(\tau(s_i, \beta_i) = s_i') > 0$, $v_i(s_i) \leq v_i(s_i')$.

Consider the agent MDP represented in Figure 8.1. If the agent deliberates once, with probability 0.33 his valuation for the resource (i.e., the value it would obtain if allocated) will increase from 0 to 3, and with probability 0.67 it will increase only to 1. If a second deliberation action is taken and the current value is 1, with equal probability the valuation will stay the same or increase to 4; if the current value is 3, it will increase to 4 with certainty.

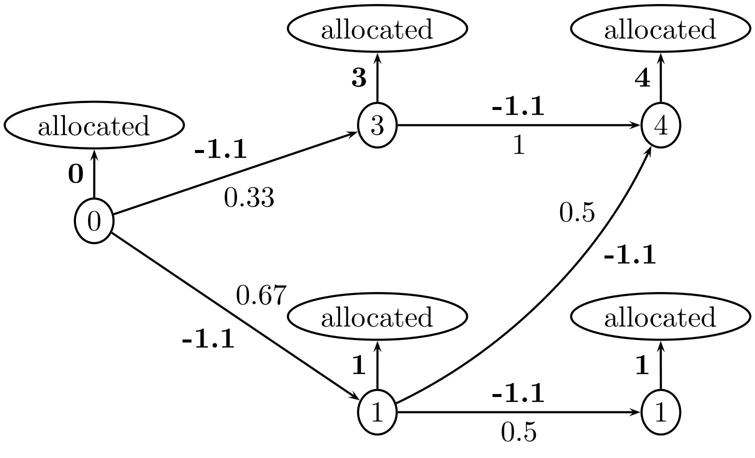


Figure 8.1: Example of an agent's MDP model of how his valuation for the resource would change upon deliberation, labeled with transition probabilities and instantaneous rewards (in bold). The agent's cost of deliberation is 1.1.

I make the following additional assumptions:

Assumption 8.2. *Agent deliberation processes are independent of each other.*

Assumption 8.3. *Agents cannot deliberate prior to the beginning of the mechanism.*

Assumption 8.4. *Only one action can be taken per time-step (i.e., multiple agents cannot deliberate concurrently).*

Assumption 8.2 is already implicit in the setup, with agent transitions and rewards functions of only *local* states, and deliberation actions for one agent causing transitions only in his own MDP. Assumption 8.3 can be motivated by considering that the resource is "revealed" only at the beginning of the mechanism. Finally, though restrictive in some cases, Assumption 8.4 is without loss of generality when the discount factor is high enough because it would be socially optimal to deliberate sequentially in that case anyway.

Combining the agent problems, we have a multi-agent MDP (see Chapter 4, Section 4.2.1) $M = (S, A, \tau, r)$ in which $S = S_1 \times \ldots \times S_n$ and $A = A_1 \cup \ldots \cup A_n$, with, $\forall s \in S$ and $i \in I$, $\tau(s, \beta_i) = (s_1, \ldots, \tau_i(s_i, \beta_i), \ldots, s_n)$ and $\tau(s, \alpha_i) = (\phi_1, \ldots, \phi_n)$, i.e., in this joint action space transitions now occur for an agent if his own deliberation action is taken or if an allocation action is taken for *any* agent, in which case he enters his absorbing state. I assume that each agent has a correct model for his local deliberation process; from this the multi-agent MDP is also correct. Notation v and c denote a valuation profile $(v_1, \ldots, v_n)$ and cost profile $(c_1, \ldots, c_n)$ respectively. Given this, the joint reward function $r(s, a)$ for the multi-agent MDP is defined as $\sum_{i \in I} r_i(s_i, a)$, with, $\forall i \in I, s \in S, a \in A$:

$$r_i(s_i, a) = \begin{cases} 0 & \text{if } s_i = \phi_i \text{ or } a \notin \{\alpha_i, \beta_i\} \\ v_i(s_i) & \text{if } s_i \neq \phi \text{ and } a = \alpha_i \\ -c_i & \text{if } s_i \neq \phi \text{ and } a = \beta_i, \end{cases}$$

This captures the essential aspect of the problem: the process "stops" once the resource has been allocated, and upon allocation the agent that receives the item obtains the value associated with his current state. I will write $v(s)$ for $\max_{i \in I} v_i(s)$ in what follows.

This formulation is quite general and allows, for example, for the local states to represent "information states" in the sense of models of optimal Bayesian learning [Bellman and Kalaba, 1959], as well as performance profile trees of the form proposed by Larson and Sandholm [2001] for normative metadeliberation (with the added restriction that values cannot decrease).

Consider decision policy π, where $\pi(s) \in A$ is the action prescribed in state s. Like in previous chapters but now using MDP state notation rather than the type abstraction, I write $V_i(s^t, \pi) = \mathbb{E}[\sum_{k=t}^{\infty} \gamma^{k-t} r_i(s^k, \pi(s^k)) \mid s^t, \pi], \forall s^t \in S$, where $s^k = \tau(s^{k-1}, \pi(s^{k-1}))$ for $k > t$. I write $V(s, \pi) = \sum_{i \in I} V_i(s, \pi), \forall s \in S$. π^* is a socially optimal policy, i.e., $\pi^* \in \arg\max_{\pi \in \Pi} V(s, \pi), \forall s \in S$, where Π is the space of all policies. We will at times consider a policy π_i^* that is optimal for agent i, i.e., $\pi_i^* \in \arg\max_{\pi \in \Pi} V_i(s, \pi)$, $\forall s \in S$. I use $V(s)$ as shorthand for $V(s, \pi^*)$, and $V_i(s_i)$ for

$V_i(s, \pi_i^*)$. Letting Π_{-i} denote the policies that never choose deliberation or allocation for i (as though i were not present in the system), I write $\pi_{-i}^* \in \arg\max_{\pi \in \Pi_{-i}} V_{-i}(s, \pi)$, and $V_{-i}(s_{-i})$ as shorthand for $V_{-i}(s, \pi_{-i}^*)$. I also define, $\forall s \in S, a \in A$:

$$\begin{aligned}
&Q(s, a) = \sum_{i \in I} r_i(s_i, a) + \gamma \mathbb{E}[V(\tau(s, a), \pi^*)], \\
&Q_i(s_i, a) = r_i(s_i, a) + \gamma \mathbb{E}[V_i(\tau(s_i, a), \pi_i^*)], \text{ and} \\
&Q_{-i}(s_{-i}, a) = \sum_{j \in I \setminus \{i\}} r_j(s_j, a) + \gamma \mathbb{E}[V_{-i}(\tau(s_{-i}, a), \pi_{-i}^*)]
\end{aligned}$$

As in previous chapters, I will consider procedures in which agents report private information to a *center* such as an auctioneer. Now in this setting the center executes a deliberation-allocation ("metadeliberation") policy, in each period either suggesting to some agent that he take a deliberation action or allocating the resource (and ending the process). Self-interested agents may subvert the process by misreporting information *or* by not following a deliberation action suggested by the center.

8.3 Efficient computation

I first present results regarding efficient computation of an optimal metadeliberation policy. As discussed earlier, Gittins [1974] provides a scalable optimal solution to any problem that can be modeled as a multi-armed bandit (MAB). To review: in MAB problems there is a set of n reward-generating Markov processes, $\{1, \ldots, n\}$, and exactly one process can be activated every time-step. The reward that a process i generates if activated at time t is a function only of its state s_i^t at t (and not of any other process's state). If i is chosen at t, a reward $r_i(s_i^t)$ is obtained and successor state s_i^{t+1} is reached (perhaps non-deterministically) according to s_i^t; for all $j \neq i$, $s_j^{t+1} = s_j^t$ and no reward is generated at t. Gittins proved that the complexity of computing an optimal policy is linear in the number of processes.

But the metadeliberation problem is not quite a bandits problem. If our agents are considered the arms of the MAB problem, each arm has *two* local actions—allocate and deliberate—and is not a Markov chain. There is also special structure to the problem: if an allocation action α_i is taken then the whole system stops. Glazebrook [1979] considered a similar setting, in which the local MDP for each arm could be reduced to a Markov chain by pre-solving for the optimal local policy, supposing that the arm was activated in every time-step. This approach also applies here: his "condition (b)" is our uncertainly improvable values (UIV) condition, which will allow us to prune away one action from every state of an agent's local MDP, yielding Markov chains. I thus reduce the problem to a multi-armed bandit, which is then solvable via Gittins indices. I offer an independent proof of Glazebrook's result, exposing additional structure of the problem when this UIV property holds;

Glazebrook's proof is for a more general condition shown to be implied by his condition (b).

Recalling the MDP model for a single agent depicted in Figure 8.1; Figure 8.2 portrays the same MDP after the pruning away of actions that would not be optimal *in a world in which the agent existed alone.*

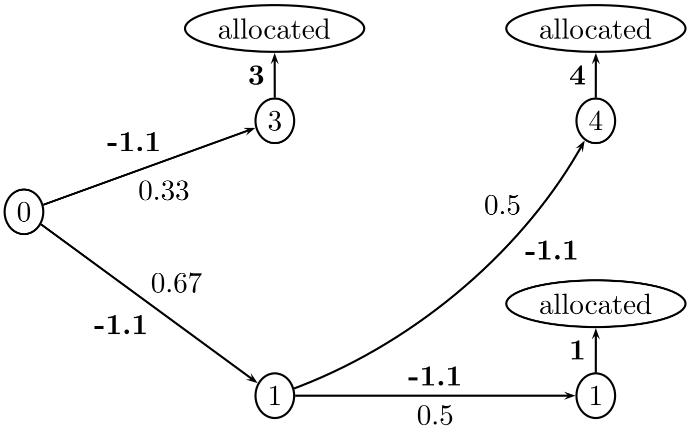


Figure 8.2: MDP world model from Figure 8.1 for a single agent i, after pruning of actions that would not be optimal in a world with no other agents. $\gamma = 0.95$, $c_i = 1.1$.

Definition 8.1 (Efficiently MC-prunable). *A domain is efficiently MC-prunable if and only if, for any agent i, for any agent MDP models, any action that would not be optimal in a world with no agents other than i is not socially-optimal in the multi-agent MDP problem, i.e.,*

$$\forall i \in I, \forall a \in \{\alpha_i, \beta_i\}, \forall s \in S, \;\; a \notin \pi_i^*(s_i) \Rightarrow a \notin \pi^*(s) \tag{8.1}$$

I will establish this property for our setting, which will lead to the validity of the following procedure:

- Convert each agent's MDP model into a Markov chain by determining the policy that would be optimal if no other agents were present.
- Execute the following deliberation-allocation policy: compute an index for each agent MC at every time-period and always activate an MC with highest index.

The following lemma shows that to test for efficient MC-prunability in our domain, we can restrict analysis to the pruning of deliberation actions.

Lemma 8.1. *A domain is efficiently MC-prunable if and only if*

$$\forall i \in I, \; \forall s \in S, \;\; \beta_i \in \pi^*(s) \Rightarrow \beta_i \in \pi_i^*(s_i) \tag{8.2}$$

Proof. Considering the contrapositive of (8.1), efficient MC-prunability requires that (8.2) and the following hold:

$$\forall i \in I,\ \forall s \in S,\quad \alpha_i \in \pi^*(s) \Rightarrow \alpha_i \in \pi_i^*(s_i) \tag{8.3}$$

It turns out that (8.3) holds for any domain. Observe that $Q(s,a) \geq Q_i(s_i,a), \forall a \in A$, as π^* is optimized over policy space Π, and $\pi_i^* \in \Pi$. Assume that $\alpha_i \in \pi^*(s)$ and, for contradiction, that $\alpha_i \notin \pi_i^*(s)$, i.e., that $Q(s,\alpha_i) \geq Q(s,a), \forall a \in A$, and $Q_i(s,\beta_i) > Q_i(s,\alpha_i)$. We have:

$$Q(s,\alpha_i) \geq Q(s,\beta_i) \geq Q_i(s_i,\beta_i) > Q_i(s_i,\alpha_i) = Q(s,\alpha_i),$$

a contradiction. □

I will now show that any domain with uncertainly improvable values satisfies equation (8.2) and is thus *efficiently MC-prunable.* This will allow us to disregard—without loss in terms of computing a socially optimal policy—those parts of each agent's allocate-deliberate MDP that would not be optimal if that agent were alone in the world.[2]

The proof is rather long and detailed, and works by analyzing the "per-period gain", denoted $g(s,a)$, achieved when a deliberation action a is taken in state s; this is the difference between the value of allocating in s and the expected (discounted) value of allocating in the successor state minus the cost of the action. I use the following notational definitions:

- $g(s,a) = \mathbb{E}[\gamma v(\tau(s,a))] - v(s^t) - c_a$, for any state s and deliberation action a, where c_a is shorthand for c_j when $a = \beta_j$.
- ρ is a random variable representing the the stopping time for policy π^* given an initial state s^0 (the identity of which will be clear from context); $\mathbb{E}[\rho \,|\, s^0, \pi^*]$ is the expected number of deliberation steps that will be taken prior to allocation.
- $G(s^0,\pi^*) = \mathbb{E}[\sum_{t=0}^{\rho-1} \gamma^t g(s^t, \pi^*(s^t)) \,|\, s^0]$.
- $h_i(s)$ is a boolean variable equal to 1 if $\pi^*(s) = \beta_i$, and 0 otherwise. Likewise, $h_{-i}(s)$ is a variable that is 1 if $\pi^*(s) = \beta_j$ for some $j \in I \setminus \{i\}$, and 0 otherwise.
- $G_i(s^0,\pi^*) = \mathbb{E}[\sum_{t=0}^{\rho-1} h_i(s^t) \cdot \gamma^t g(s^t, \pi^*(s^t)) \,|\, s^0]$
- $G_{-i}(s^0,\pi^*) = \mathbb{E}[\sum_{t=0}^{\rho-1} h_{-i}(s^t) \cdot \gamma^t g(s^t, \pi^*(s^t)) \,|\, s^0]$

[2] Actually doing the pruning requires "solving" each agent's local MDP, which can be achieved via value iteration or one of the other methods described in Chapter 4, Section 4.3.

Note that the sum of the per-period gains of deliberation under π^* from any state $s^0 \in S$ equals the total difference between the value achieved by π^* and a policy that does no deliberation:[3]

$$G(s^0, \pi^*) = \mathbb{E}\Big[\sum_{t=0}^{\rho-1} \gamma^t \Big(\gamma v(s^{t+1}) - v(s^t) - c_{\pi^*(s^t)}\Big) \,\Big|\, s^0, \pi^*\Big] \tag{8.4}$$

$$= \mathbb{E}\Big[\sum_{t=1}^{\rho} \gamma^t v(s^t) - \sum_{t=0}^{\rho-1} \gamma^t v(s^t) - \sum_{t=0}^{\rho-1} \gamma^t c_{\pi^*(s^t)} \,\Big|\, s^0, \pi^*\Big] \tag{8.5}$$

$$= \mathbb{E}\Big[\gamma^\rho v(s^\rho) - v(s^0) - \sum_{t=0}^{\rho-1} \gamma^t c_{\pi^*(s^t)} \,\Big|\, s^0, \pi^*\Big] \tag{8.6}$$

$$= V(s^0, \pi^*) - v(s^0) \tag{8.7}$$

On the way to proving that uncertainly improvable values domains are all efficiently MC-prunable, I establish in Lemma 8.3 that under the optimal policy π^*, from any state going forward, for every agent i, the total expected discounted per-period gains for the deliberation actions taken by i are non-negative. I will make use of this tiny helper lemma:

Lemma 8.2. *For any a, b, c, and x such that $b \geq c$ and $x > 0$:*

$$\max(a, b) - \max(a, c) \leq \max(a - x, b) - \max(a - x, c) \tag{8.8}$$

Proof. By case analysis:

Case 1: $b \geq c \geq a$. In this case $\max(a, b) = b$, $\max(a, c) = c$, $\max(a - x, b) = b$, and $\max(a - x, c) = c$. So the lemma requires $b - c \leq b - c$, which is obviously true.

Case 2: $b \geq a \geq c$. In this case $\max(a, b) = b$, $\max(a, c) = a$, $\max(a - x, b) = b$, and $\max(a - x, c) = c$. So the lemma requires that $b - c \leq b - \max(a - x, c)$, which is true since $\max(a - x, c) \geq c$.

Case 3: $a \geq b \geq c$. In this case $\max(a, b) = a$ and $\max(a, c) = a$. So the lemma requires that $a - a \leq \max(a - x, b) - \max(a - x, c)$. This is true since $b \geq c$ implies $\max(a - x, b) - \max(a - x, c)$. □

Lemma 8.3. $\forall s \in S, i \in I,\ G_i(s, \pi^*) \geq 0$.

[3] Doing the analysis in terms of $t = 0$ will make the notation less clumsy and there is no loss of generality.

Proof. For contradiction, say $G_i(s^0, \pi^*) < 0$ for some $i \in I$ and $s^0 \in S$. We have:

$$G(s^0, \pi^*) = \mathbb{E}\Big[\sum_{t=0}^{\rho-1} \gamma^t g(s^t, \pi^*(s^t)) \,\Big|\, s^0, \pi^*\Big] \tag{8.9}$$

$$= \mathbb{E}\Big[\sum_{t=0}^{\rho-1} h_i(s^t)\gamma^t g(s^t, \pi^*(s^t)) + \sum_{t=0}^{\rho-1} h_{-i}(s^t)\gamma^t g(s^t, \pi^*(s^t)) \,\Big|\, s^0, \pi^*\Big] \tag{8.10}$$

$$= G_i(s^0, \pi^*) + G_{-i}(s^0, \pi^*) \tag{8.11}$$

$$< G_{-i}(s^0, \pi^*) = \mathbb{E}\Big[\sum_{t=0}^{\rho-1} h_{-i}(s^t)\gamma^t g(s^t, \pi^*(s^t)) \,\Big|\, s^0, \pi^*\Big] \tag{8.12}$$

Now let ρ' be a stopping time that follows ρ (i.e., is derived from policy π^*) except that it potentially "stops earlier", specifically as soon as a state is reached with non-positive G value: $\rho' = \inf\Big\{0 < \hat{\rho} \leq \rho \,\Big|\, G_{-i}(s^{\hat{\rho}}, \pi^*) \leq 0\Big\}$. Such a ρ' must exist, as $G_{-i}(s^\rho, \pi^*)$ can't be positive by optimality of π^*. We have:

$$G_{-i}(s^0, \pi^*) = \mathbb{E}\Big[\sum_{t=0}^{\rho'-1} h_{-i}(s^t) \cdot \gamma^t g(s^t, \pi^*) + \gamma^{\rho'} G_{-i}(s^{\rho'}, \pi^*) \,\Big|\, s^0, \pi^*\Big] \tag{8.13}$$

$$\leq \mathbb{E}\Big[\sum_{t=0}^{\rho'-1} h_{-i}(s^t) \cdot \gamma^t g(s^t, \pi^*) \,\Big|\, s^0, \pi^*\Big] \tag{8.14}$$

Define $G'(s^t, \pi^*) = \mathbb{E}[\sum_{k=t}^{\rho'-1} h_{-i}(s^k) \cdot \gamma^k g(s^k, \pi^*) \mid s^t, \pi^*]$ for any s^t that has positive probability of being the state at t given s^0 and π^*. Then $G'(s^0, \pi^*) =$ (8.14). We will now consider the times $t_0, \ldots, t_{m-1}$ at which a deliberation action other than β_i is taken; m is a random variable representing the number of times this happens, and t_k (for $0 \leq k < m$) is a random variable representing the time at which a deliberation action other than β_i is taken for the k^{th} time, under π^* from s^0. We have that (8.14)

$$= \mathbb{E}\Big[h_{-i}(s^0) \cdot g(s^0, \pi^*) + h_{-i}(s^0) \cdot \gamma g(s^1, \pi^*) + \dots \tag{8.15}$$

$$+ h_{-i}(s^{\rho'-1}) \cdot \gamma^{\rho'-1} g(s^{\rho'-1}, \pi^*) \;\Big|\; s^0, \pi^*\Big] \tag{8.16}$$

$$= \mathbb{E}\Big[\gamma^{t_0} g(s^{t_0}, \pi^*) + \gamma^{t_1} g(s^{t_1}, \pi^*) + \dots + \gamma^{t_{m-1}} g(s^{t_{m-1}}, \pi^*) \;\Big|\; s^0, \pi^*\Big] \tag{8.17}$$

$$= \mathbb{E}\Big[\gamma^{t_0}\Big(\gamma v(s^{t_0+1}) - v(s^{t_0}) - c_{\pi^*(s^{t_0})}\Big) + \tag{8.18}$$

$$\gamma^{t_1}\Big(\gamma v(s^{t_1+1}) - v(s^{t_1}) - c_{\pi^*(s^{t_0})}\Big) + \dots +$$

$$\gamma^{t_{m-1}}\Big(\gamma v(s^{t_{m-1}+1}) - v(s^{t_{m-1}}) - c_{\pi^*(s^{t_{m-1}})}\Big) \;\Big|\; s^0, \pi^*\Big] \tag{8.19}$$

$$= \mathbb{E}\Big[\gamma^{t_0}\Big(\gamma \max\{v_i(s_i^{t_0}), v_{-i}(s_{-i}^{t_0+1})\} - \max\{v_i(s_i^{t_0}), v_{-i}(s_{-i}^{t_0})\} - c_{\pi^*(s^{t_0})}\Big) +$$
$$\gamma^{t_1}\Big(\gamma \max\{v_i(s_i^{t_1}), v_{-i}(s_{-i}^{t_1+1})\} - \max\{v_i(s_i^{t_1}), v_{-i}(s_{-i}^{t_1})\} - c_{\pi^*(s^{t_0})}\Big) + \dots +$$
$$\gamma^{t_{m-1}}\Big(\gamma \max\{v_i(s_i^{t_{m-1}}), v_{-i}(s_{-i}^{t_{m-1}+1}) - \max\{v_i(s_i^{t_{m-1}}), v(s^{t_{m-1}})\} - c_{\pi^*(s^{t_{m-1}})}\Big) \;\Big|\; s^0, \pi^*\Big] \tag{8.20}$$

By Assumption 8.1, for all $i \in I$, $a \in \{\beta_1, \dots, \beta_n\}$, s^t and s^{t+1} such that s^{t+1} could succeed state s^t, $v_i(\tau(s_i^t, a)) \geq v_i(s_i^t)$. Using this fact and Lemma 8.2, we have that (8.20)

$$\leq \mathbb{E}\Big[\gamma^{t_0}\Big(\gamma \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_0+1})\} - \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_0})\} - c_{\pi^*(s^{t_0})}\Big) +$$
$$\gamma^{t_1}\Big(\gamma \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_1+1})\} - \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_1})\} - c_{\pi^*(s^{t_0})}\Big) + \dots +$$
$$\gamma^{t_{m-1}}\Big(\gamma \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_{m-1}+1}) - \max\{v_i(s_i^0), v(s^{t_{m-1}})\} - c_{\pi^*(s^{t_{m-1}})}\Big) \;\Big|\; s^0, \pi^*\Big] \tag{8.21}$$

By the definition of G' (in particular, the fact that $G'(s^0, \pi^*), \dots, G'(s^{\rho-1}, \pi^*)$ are all non-negative), we can see that removing the discounting "gaps" in (8.21) can only increase the value. We have that (8.21)

$$\leq \mathbb{E}\Big[\Big(\gamma \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_1})\} - \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_0})\} - c_{\pi^*(s^{t_0})}\Big) +$$
$$\gamma\Big(\gamma \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_2})\} - \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_1})\} - c_{\pi^*(s^{t_0})}\Big) + \dots +$$
$$\gamma^{m-1}\Big(\gamma \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_{m-1}+1})\} - \max\{v_i(s_i^0), v(s^{t_{m-1}})\} - c_{\pi^*(s^{t_{m-1}})}\Big) \;\Big|\; s^0, \pi^*\Big] \tag{8.22}$$

Noting that $s_{-i}^{t_0} = s_{-i}^0$ and canceling out intermediate terms we have that this

$$= \mathbb{E}\Big[\gamma^m \max\{v_i(s_i^0), v_{-i}(s_{-i}^{t_{m-1}+1})\} - \sum_{k=0}^{m-1} \gamma^k c_{\pi^*(s^{t_k})} \,\Big|\, s^0, \pi^*\Big] - \max\{v_i(s_i^0), v_{-i}(s_{-i}^0)\} \tag{8.23}$$

$$= \mathbb{E}\Big[\gamma^m \max\{v_i(s_i^0), v_{-i}(s_{-i}^\rho)\} - \sum_{k=0}^{m-1} \gamma^k c_{\pi^*(s^{t_k})} \,\Big|\, s^0, \pi^*\Big] - v(s^0) \tag{8.24}$$

$$\leq V(s^0, \pi) - v(s^0), \tag{8.25}$$

where policy π is optimal among all policies that never select deliberation action β_i. We have that $G(s^0, \pi^*) = V(s^0, \pi^*) - v(s^0) < G_{-i}(s^0, \pi^*) = V(s^0, \pi) - v(s^0)$, and thus that $V(s^0, \pi^*) < V(s^0, \pi)$, a contradiction by optimality of π^*. The lemma follows. □

Lemma 8.4. *All uncertainly improvable values domains are efficiently MC-prunable.*

Proof. Consider any uncertainly improvable values domain and any $i \in I$ and $s^0 \in S$ such that $\beta_i \in \pi^*(s^0)$. Consider the sequence of times $t_0, \ldots, t_{m-1}$, where m is a random variable representing the number of times β_i is taken, and t_k (for $0 \leq k < m$) is a random variable representing the time at which β_i is taken for the k^{th} time, under π^* from s^0. Since $\beta_i \in \pi^*(s^0)$, $Pr(m \geq 1) = 1$. Using Lemmas 8.3 and 8.2, we have that:

$$0 \leq G_i(s^0, \pi^*) = \mathbb{E}\Big[\sum_{k=0}^{m-1} \gamma^{t_k}\Big(\gamma \max\{v_i(s_i^{t_{k+1}}), v_{-i}(s_{-i}^0)\} - \max\{v_i(s_i^{t_k}), v_{-i}(s_{-i}^0)\} - c_{\pi^*(s^{t_k})}\Big) \,\Big|\, s^0, \pi^*\Big] \tag{8.26}$$

From Lemma 8.3 we know that for any s^k reached with positive probability from s^0 given π^*, if $\pi^*(s^k)$ specifies deliberation then $G_i(s^k, \pi^*) \geq 0$. Then as in the previous lemma, we can remove the discounting gaps between the terms, so the above:

$$\leq \mathbb{E}\Big[\sum_{k=0}^{m-1} \gamma^{k}\Big(\gamma \max\{v_i(s_i^{t_{k+1}}), v_{-i}(s_{-i}^0)\} - \max\{v_i(s_i^{t_k}), v_{-i}(s_{-i}^0)\} - c_{\pi^*(s^{t_k})}\Big) \,\Big|\, s^0, \pi^*\Big] \tag{8.27}$$

Noting that $s_i^{t_0} = s_i^0$ and canceling out intermediate terms we have that this

$$= \mathbb{E}\Big[\gamma^m \max\{v_i(s_i^{t_{m-1}+1}), v_{-i}(s_{-i}^0)\} - \sum_{k=0}^{m-1} \gamma^k c_{\pi^*(s^{t_k})} \,\Big|\, s^0, \pi^*\Big] - \max\{v_i(s_i^0), v_{-i}(s_{-i}^0)\} \tag{8.28}$$

$$\leq \mathbb{E}\Big[\gamma^m v_i(s_i^{t_{m-1}+1}) - \sum_{k=0}^{m-1} \gamma^k c_{\pi^*(s^{t_k})} \,\Big|\, s^0, \pi^*\Big] - v_i(s_i^0) \tag{8.29}$$

$$\leq Q_i^*(s_i^0, \beta_i) - v_i(s_i^0) \tag{8.30}$$

I have shown that $0 \leq G_i(s^0, \pi^*) \leq Q_i^*(s_i^0, \beta_i) - v_i(s_i^0)$. $Q_i^*(s_i^0, \beta_i) \geq v_i(s_i^0)$ implies $\beta_i \in \pi_i^*(s_i^0)$, and the theorem follows by appeal to Lemma 8.1. □

This enables a "without loss" reduction from local MDPs to local MCs. The remaining challenge is that the Gittins index policy is only optimal for problems with an infinite time-horizon. This issue can be handled when $\gamma < 1$ by replacing the one-time reward of $v_i(s_i)$ in a state s_i in which agent i is allocated the item with a reward of $(1-\gamma)v_i(s_i)$ received per period in perpetuity. It is then a simple matter to show that the optimal MAB policy will always continue to activate agent i's MC after it first does so when i is in an "allocation state". Thus the resulting policy is valid for the original problem with absorbing states. Returning to our example, Figure 8.3 displays the infinite horizon, pruned MC for the problem earlier depicted in Figure 8.2.

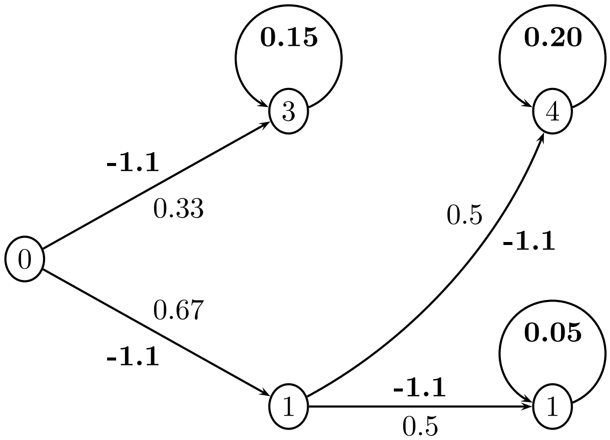


Figure 8.3: Agent-optimal Markov chain from Figure 8.2 after expansion to an infinite-horizon.

Theorem 8.1. *Given Assumptions 8.1–8.4, the deliberation-allocation policy defined by activating, at every time-step t, the pruned, locally-optimal Markov chain of an agent with the highest Gittins index is optimal.*

Proof. The theorem follows immediately from Lemma 8.4 and Theorem 4.1 (Gittins's theorem). □

8.4 Handling selfish agents

Having specified a computationally efficient, socially-optimal decision procedure for uncertainly improvable values domains, the challenge of implementing it in a context of selfish, strategic agents remains. What's best for the system as a whole (the socially optimal policy) will often not be best for every individual. In the absence of an incentive-aligning mechanism, selfish agents will game the system by misreporting private information or disobeying the center's deliberation prescriptions.

I combine the index-policy solution to the multi-agent metadeliberation problem with the dynamic-VCG mechanism to obtain a *metadeliberation auction*, in which the center chooses actions based on private valuation information that agents report. Note that, in the case of a deliberation action, "chooses" means "suggests to the agents"; for an allocation action, the center simply executes it.

Definition 8.2 (metadeliberation auction). .

- *Each agent i computes his locally optimal, infinite-horizon Markov chain, and reports a claim about it to the center along with a claim s_i^0 about his initial local state.*
- *At every time-step t, while the resource has not yet been allocated:*
  1. *The agent i activated in the previous time-step reports a claim s_i^t about his current state (except in $t = 1$ before any action has been taken).*[4]
  2. *The center chooses the action specified by activation of an agent i^* with highest Gittins index.*
  3. *Agent i^* pays the center:*

$$\begin{array}{ll} (1-\gamma)\, V_{-i^*}(s_{-i^*}^t) & \textit{if deliberation was performed} \\ V_{-i^*}(s_{-i^*}^t) & \textit{if the item was allocated} \end{array}$$

[4] Technically (and as in the mechanisms of Chapters 5, 6, and 7), at every time-step t each agent i must have a chance to report his entire "type" θ_i^t, i.e., a new claim about his entire Markov chain model *plus* his current state, but this simpler presentation is consistent with the equilibrium behavior (the same applies to the mechanism specified in Definition 8.3).

Observe that agents are doing more than just reporting types here; in the initial time period the mechanism has them solve their local MDP model and report the optimal *pruned* MC. We could alternatively have agents simply report their entire types to the center who could then do the pruning, but the payment scheme provides the incentives for distributing computational work among the agents in this way.

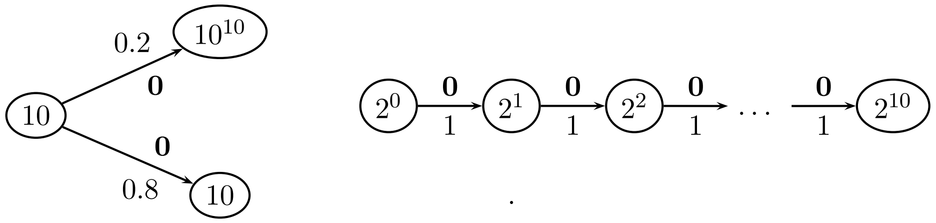
Theorem 8.2. *Given Assumptions 8.1–8.4, the metadeliberation auction is truthful, efficient, and individual rational in within-period ex post Nash equilibrium, and never runs a deficit.*

Proof. The result follows from within-period ex post efficiency, IC, and IR of the dynamic-VCG mechanism, the efficiency of the Gittins index policy for MABs, and the lossless reduction result (Corollary 5.1 and Theorems 4.1 and 8.1). Recall that dynamic-VCG requires that each agent i pay the center an amount equal to the negative externality his presence imposes on the other agents at t. In our setting, for the agent who deliberates at t this is equal to the cost to the other agents of having to wait one time-step to implement the policy that would be optimal for them, i.e., $(1-\gamma)\, V_{-i^*}(\hat{s}^t_{-i^*})$; for all other agents it is 0. When the item is allocated to an agent, that agent imposes an externality equal to the total value agents could get from the current state forward if he were not present. □

This provides the result we want: each agent will first prune away his suboptimal local actions, and then truthfully report his (pruned) MC to the center. From that point forward, the center will suggest deliberation actions according to the optimal deliberation-allocation policy, collecting a payment from the agent that deliberates. Agents will choose to follow these suggestions and truthfully report new local states, and the center will eventually allocate the resource. At that point the agent will consume the resource with no further deliberation, by optimality of the deliberation-allocation policy. I will now demonstrate the workings of the mechanism on an example that illustrates the way the payment scheme causes agents to internalize the *social* costs and benefits of both performing deliberation and being allocated the item.

Example 1

Consider execution of the metadeliberation auction on the example in Figure 8.4 (for simplicity I've switched to a more concise MDP representation, omitting allocation nodes). The optimal policy has agent 1 deliberate first; if his value increases to 10^{10} he is then allocated the item. Otherwise the optimal policy has agent 2 deliberate for 10 time-steps and then allocates to him. The discount factor $\gamma = 0.9$. Under the metadeliberation auction, in the first time-step agent 1 must pay the "immediate externality" imposed on agent 2 assuming the policy optimal for agent 2 would be executed in all following periods, i.e., his cost of waiting one period, or $(1-0.9)\cdot 0.9^{10}\cdot 2^{10}$. If agent 1's deliberation yields the high value (10^{10}) he will be allocated the item and must then pay $0.9^{10}\cdot 2^{10}$.



(a) Agent 1

(b) Agent 2

Figure 8.4: Agent 1 has initial value 10. With small probability his value will increase to 10^{10} if he deliberates once. Agent 2's value is $\min(2^x, 2^{10})$, where x is the number of times he has deliberated. $c_1 = c_2 = 0$ and $\gamma = 0.9$.

If agent 1's deliberation does not yield the improvement, then in every period that follows prior to allocation (with agent 2 deliberating) agent 2 must pay $(1-0.9)\cdot 10 = 1$. In the final allocation step agent 2 pays 10. Would agent 2 rather avoid making these payments by lying? Bear in mind that he *discounts* value (rewards *and* costs) in the future by a factor of 0.9. We can compute agent 2's expected utility (from the first time he is asked to deliberate) for being truthful, and compare it to his expected utility if he misreports his MC such that he is asked to instead deliberate for only $k < 10$ time-steps, and then finishes his deliberation once he receives the resource. If he deliberates k times (for any $k \geq 0$), the total discounted payments he makes will equal:

$$\begin{aligned} & (1-\gamma)10 + \gamma(1-\gamma)10 + \ldots + \gamma^{k-1}(1-\gamma)10 + \gamma^k 10 \\ &= 10 - \gamma 10 + \gamma 10 - \ldots - \gamma^{k-1}10 + \gamma^{k-1}10 - \gamma^k 10 + \gamma^k 10 \\ &= 10 \end{aligned}$$

So his *discounted payments* are the same regardless of how many times he deliberates. Then since it is optimal for agent 2 to deliberate 10 times, whether he does so inside or outside the context of the mechanism, his total discounted utility will always equal $\gamma^{10}2^{10} - 10$, and so truthful participation is a utility maximizing strategy.

8.5 Multiple deliberation processes

So far, in order to simplify analysis I've assumed that each agent has only one way of deliberating. However, the results we've seen also apply when agents have multiple independent deliberation methods. For instance, imagine an agent that has three different research programs it could pursue (potentially with distinct associated costs per time-step)—the agent merely has to report all three models to the center, who will consider all three in determining the optimal policy. It is important, though, that all deliberation processes are *independent* (deliberation in one process cannot change the state of another process); otherwise, there will be no reduction to the multi-armed bandit problem. Given this independence, a generalization of Theorem 8.2 immediately follows.[5]

8.5.1 Strategic deliberation

Consider now a setting in which an agent may have one or more deliberation processes that pertain to the value of *other* agents for the resource. This models the setting of strategic deliberation introduced by Larson and Sandholm [2001].[6] Note that the optimal policy might specify "cross-agent" deliberation, with the results of i's research being shared with j (in particular, when i has a better deliberation process than j). One can imagine a "consulting" scenario, where one firm has expertise in optimizing the integration of new technologies into an overarching business plan—expertise that may be of more value applied to *another* company's situation than to its own.

The dynamic-VCG scheme *will not* work here. A subtle condition usually required for the good incentive and IR properties of dynamic-VCG is that the optimal policy for agents other than i does not take any actions that involve agent i. Formally, where Π_{-i} is the set of policies that never specify an action for i, the necessary condition is that $\max_{\pi \in \Pi} V_{-i}(s, \pi) = \max_{\pi \in \Pi_{-i}} V_{-i}(s, \pi)$ (this is strongly related to the analysis, in Chapter 7, Section 7.3, of dynamic-VCG in an interdependent values setting). This condition is not met when the optimal policy has one agent deliberate about another's value. The intuition behind the extension of dynamic-VCG that I present in this section is that the payments make the expected equilibrium payoff to agent i forward from any state equal to the payoff i would receive in the dynamic-VCG mechanism *if his deliberation processes about other agents were actually about himself.* The equilibrium properties then follow immediately from the analysis of the metadeliberation auction in the context of agents with multiple independent deliberation processes only about their own values.

Let p_{ij} denote a deliberation process possessed by agent i pertaining to the value that agent j could obtain from the resource; let $c_{p_{ij}}$ denote the cost (to i) of deliberating on process p_{ij}. For any process p_{ij}, any state $s_{p_{ij}}$ consists of two things: some

[5] Without this independence the dynamic-VCG mechanism will still provide the right incentives, but we will no longer be in a multi-armed bandit world, so the payment scheme will look different and we will not have the reduction that allows for a computationally tractable solution.

[6] But note that our independence assumption precludes results of one agent's deliberation impacting the expected results of another's, though they may concern the same agent's value.

information $\eta(s_{p_{ij}})$ (e.g., the observations of the world acquired from research, or the plan resulting from some computation), and a valuation $v(s_{p_{ij}})$ for j receiving the item given the information content. Let $v_j(\eta(s_{p_{ij}}))$ denote the *actual* value received by j for the information associated with the same state. Allowing for misreports, $v(\hat{s}_{p_{ij}})$ denotes the value that should be achieved by j according to i's state report, $\eta(\hat{s}_{p_{ij}})$ denotes the information content associated with that state report, and $\hat{v}_j(\eta(\hat{s}_{p_{ij}}))$ is a claim made by j about the actual value he achieved. In the mechanism I propose the center computes payments by reasoning about the social value that could be achieved under a policy that is optimal with all agents present, but in which an agent i *cannot take any actions.* I denote this quantity, which is independent of i's state, as $V^{-i}(s_{-i})$, for all $s \in S$.

Definition 8.3 (metadeliberation auction with cross-agent deliberation).

- *Each agent i computes the locally optimal, infinite-horizon Markov chain for every deliberation process it possesses, and reports claims about each MC and its initial local state to the center.*
- *At every time-step t, while the resource has not yet been allocated:*
  1. *For process $p_{\bar{i},\bar{j}}$ activated in the previous time-step, agent i reports a claim $\hat{s}^t_{p_{\bar{i},\bar{j}}}$ about $p_{\bar{i},\bar{j}}$'s current state.*
  2. *The center chooses the action specified by activation of a process p_{ij} with highest Gittins index.*
  3. *If deliberation was performed, agent i pays the center $(1-\gamma)\, V^{-i}(\hat{s}^t_{-i})$.*

     If the item was allocated and $i = j$, j pays the center $V^{-j}(\hat{s}^t_{-j})$. If $i \neq j$, the center communicates $\eta(\hat{s}^t_{p_{ij}})$ to agent j, j communicates $\hat{v}_j(\eta(\hat{s}^t_{p_{ij}}))$ to the center, i pays the center $V^{-i}(\hat{s}^t_{-i}) - \hat{v}_j(\eta(\hat{s}^t_{p_{ij}}))$, and j pays the center $v(\hat{s}^t_{p_{ij}})$.

Theorem 8.3. *Given Assumptions 8.1–8.4, the metadeliberation auction with cross-agent deliberation is truthful, efficient, and IR in within-period ex post Nash equilibrium, and does not run a deficit when agents are truthful.*

Proof sketch. The incentive and IR properties of the mechanism follow from those of the original metadeliberation auction, combined with the following observation: for any process p_{ij} with $i \neq j$, the payment scheme yields a scenario which is, payoff-wise, identical to one in which p_{ij} is a deliberation process pertaining to i's value. If p_{ij} is selected for deliberation then i already pays the cost. If p_{ij} is selected for allocation then i will be paid an amount equal to the actual value yielded from the process (assuming agent j is honest), and j will obtain value 0 (assuming i is honest), since

$v(s^t_{p_{ij}}) = v_j(\eta(s^t_{p_{ij}}))$ by the assumption that beliefs are correct.[7] The mechanism never runs a deficit in equilibrium. Prior to the final allocation step there are no payments that flow to the agents. Then in that final allocation step payments made to the center are $v(s^t_{p_{ij}}) + V^{-i}(\hat{s}^t_{-i}) - v_j(\eta(s^t_{p_{ij}}))$. Given truthful reporting (which, as shown above, is achieved in an ex post equilibrium), this quantity equals $V^{-i}(\hat{s}^t_{-i})$, which is ≥ 0. □

Note that if we conceptualize the value associated with a state in an MDP not as a definite, actual value, but as an *expectation* of the value that will be achieved for allocating in that state, we get very similar results. An analogue of Theorem 8.3 holds in which, given truthful reporting, there is no-deficit (and IR) *in expectation* from every state, at all time-steps, rather than ex post. This is because in the last step we would have that $v(s_{p_{ij}}) = \mathbb{E}[v_j(s^t_{p_{ij}}) \mid s^t_{p_{ij}}]$, for any $s_{p_{ij}}$.

Example 2

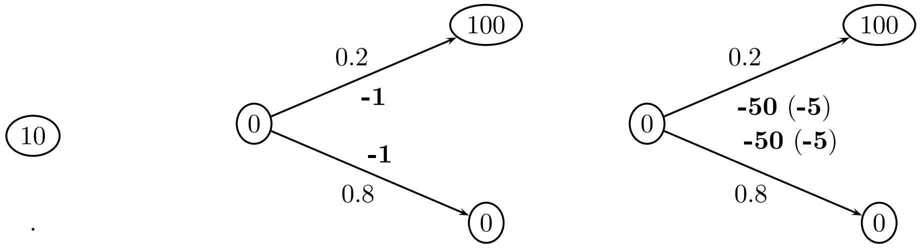
Consider the 2-agent scenario depicted in Figure 8.5, in which one agent has a deliberation process about the other. Agent 1 will obtain value 10 if allocated the resource (there is no deliberation he can do that would change his value), and both agent 1 and agent 2 have a model *about agent 2's value* with one deliberation step, which yields value 100 with probability 0.2 and otherwise yields value 0. Take $\gamma = 0.9$. We will consider two variants.

(a) Consider that agent 2's cost of running his deliberation process is **50** and agent 1's cost is **1**. Agent 1 does not have an incentive to deviate from truthfulness (for instance, by simply claiming agent 2 has the high 100 value without deliberating for him). Agent 1 will be payed the value that agent 2 reports experiencing, given the information obtained from agent 1's deliberation. So agent 1's payment is only based on agent 2's *actual* utility (assuming agent 2 is truthful). If agent 1 reported agent 2 had the high value and didn't communicate corresponding information (e.g., a plan for using the resource), the value agent 2 experiences—and the value agent 1 is payed—would be 0.[8]

(b) Now consider a variant in which agent 2's cost of deliberating is **5** rather than **50**. Agent 2 may know that if he reports truthfully agent 1 will be selected first (since agent 1's deliberation process about agent 2 has a lower cost), and if agent 1's deliberation yields a plan worth value 100 agent 2 will obtain none of the

[7] Note that if an agent i is allocated the item via an agent j's process, both agents are indifferent about their reports during the final allocation stage (this is similar to the fragility of the equilibrium in [Mezzetti, 2004]). Ex post IC and IR are technically maintained as there is only one "possible" true state for j, and it is known to i. There is an alternate payment scheme that avoids this indifference, but in some cases it will lead to a deficit in equilibrium.

[8] Note that this is not an issue of "punishment." Rather, in equilibrium it will *never* be useful to deviate.



(1) Agent 1's process about agent 1

(2) Agent 1's process about agent 2

(3) Agent 2's process about agent 2

Figure 8.5: Agent 1 has one trivial process pertaining to his own valuation. Both agents have processes pertaining to agent 2's valuation: initially the value from allocation is 0; after one deliberation step, with probability 0.2 it increases to 100 and otherwise stays at 0. No further deliberation yields any change. Agent 1's cost is 1, agent 2's cost is -50 (in variant (a)) and -5 (in variant (b)).

surplus. So would he prefer to report cost 0 in order to be asked to perform the deliberation himself first? No. The mechanism specifies that he would be charged *as though both of agent 1's deliberation processes were about agent 1.* So in the first period agent 2 would be charged $(1-\gamma)[\gamma(0.2 \cdot 100 + 0.8 \cdot 10) - 1] = 2.42$ and endure deliberation cost 5. If agent 2's deliberation yields the high value (which it will with probability 0.2) he will obtain the resource (value 100) and make payment $\gamma(0.2 \cdot 100 + 0.8 \cdot 10) - 1 = 24.2$. If it yields low value he gets 0 value and pays 0. Thus agent 2's expected utility from this strategy is $-2.42 - 5 + 0.2 \cdot 0.9 \cdot (100 - 24.2) = 6.224$. But if agent 2 is truthful, he still has a chance for high payoff; recall that the two deliberation processes are *independent*, so the result of one does not bear on what the result of the other will be. In particular, if agent 1 deliberates first agent 2 has expected value $\gamma 0.8(-(1-\gamma)10 + 0.2(\gamma(100 - 10)) + 0.8 \cdot 0 - 5) = 7.344$. (With probability 0.8 agent 1 will find value 0 for agent 2, and then agent 2 is asked to deliberate and with probability 0.2 will achieve value 100, making a payment of 10.) Thus truthfulness is a superior strategy for agent 2.

So this modification of dynamic-VCG specifies cross-agent deliberation exactly when it is socially-optimal. The payments align agents' interests with those of the system as a whole, so each agent's utility maximizing strategy is exactly the strategy that maximizes utility for the system, i.e., truth and obedience.

8.6 Summary

This chapter makes two distinct contributions. First, I provided a proof that the multi-armed bandits problem is suitable for solving multi-agent metadeliberation problems, in this case by careful reduction of the original multi-agent MDP model into a multi-agent Markov chain model. Second, I provided a novel application of the theory of dynamic mechanism design to coordinate deliberative processes of self-interested agents, improving social welfare. This has parallels in work on preference elicitation in settings with costly or bounded elicitation, but is, to my knowledge, the first normative solution in a setting in which information acquisition by participants is incremental rather than instantaneous. I extended the underlying ideas of the dynamic-VCG mechanism to an environment in which it cannot be directly applied because of positive externalities. Remarkably, this does not lead to a budget deficit.

There are many directions for future work in this area, most of which can be considered relaxations of assumptions I made here. Perhaps most exciting would be an extension to the undiscounted setting where agents are completely patient; no index policy is currently known for such settings. Additionally, though we can handle agents with multiple *independent* ways of deliberating, there may be cases in which agents know of many research methods, where the intermediate results of any one method may effect the expected value another method would yield. Again, there is no know computationally tractable solution for such settings, though dynamic-VCG would continue to provide the proper incentives for truthful participation given an optimal policy.

Finally, there is also another class of compelling deliberation scenarios in which deliberation yields *better estimates* of a true valuation; i.e., agents *learn* their valuations through research, rather than increase them by figuring out new uses for a resource. Deriving an efficient way of computing optimal deliberation-allocation policies in such settings would be of great interest; the reduction technique employed in this chapter does not apply.

Chapter 9

Conclusion

This thesis is about trying to make decisions in ways that lead to the realization of as much social welfare as possible. Mechanism design provides a framework for providing incentives for agents to behave in ways that are *socially* optimal, even if they are motivated only by *selfish* concerns. But the theory falls short in important regards; addressing these, the thesis offers two major contributions:

- I provided *redistribution mechanisms* (for both the static and dynamic cases) that achieve drastically improved social welfare properties over previous solutions in important domains.
- I elaborated the theory of *dynamic mechanism design*, an extension of mechanism design from one-shot environments to ones in which a sequence of decisions is to be made over time. I provided fundamental results as well as extensions to settings with a population of agents that changes over time, and an application to resource allocation settings in which agents can improve their values via costly deliberation.

To conclude the thesis I will give a short summary of the discoveries, describe some possible challenges for implementation, and discuss promising avenues for future research.[1]

9.1 An informal summary

Mechanism design, even in typical static settings, has suffered from a significant shortcoming: the solutions proposed often involve large payments from agents to the center. That is, the individuals whose utility the mechanism is employed to serve

[1] The summary here is informal; see either Chapter 1, Section 1.5 or the synopses that begin each chapter for a slightly more structured and complete overview of the results of the thesis.

must transfer away much of the utility they achieve. This fact has often been de-emphasized by positing that the center is "just another agent", so utility he obtains through transfers is still utility enjoyed by "the agents". There are certainly cases where this reasoning is valid—sometimes the goal in decision-making is just to squeeze out all the utility we can from the environment; which person ends up with what portion of it doesn't matter, as long as *someone* is getting the value. Still, in practice this story often just doesn't fit. A group of individuals may agree that maximizing their joint welfare should be given utmost priority, yet they may not be willing to transfer arbitrary amounts of that welfare to a third party for whom they have no concern.

This is where redistribution mechanisms come in, taking the goal of squeezing the most value possible out of the decision-making process *and* maintaining as much of it as possible within the group whose welfare is being maximized via the decision. I found far superior solutions to those that have been proposed prior in terms of social welfare. How? By leveraging structure inherent in the decision-making domain.

In more detail: The VCG mechanism is the classic mechanism design solution. It achieves efficient outcomes in dominant strategies, guarantees (often) that each agent will be no worse off for having participated, and never runs a deficit. I showed that among mechanisms with these properties it *maximizes* the transfers required of the agents and, additionally, in domains without known structure to agent valuations, it also *minimizes* the transfers; so it is unique. But often times valuations *do* have structure that is known, for instance in allocation problems where agents that don't receive items don't obtain any value. I proposed **RM** which, for each agent, computes a "guarantee" on the revenue that will result independent of what valuation that agent reports. The mechanism then runs VCG and gives back (or "redistributes") a portion of the revenue to each agent proportional to the revenue guarantee computed for that agent. The mechanism never sacrifices the incentive or no-deficit properties of VCG, and often yields much better utility for the agents. I showed empirically that for single-item allocation problems with more than a few agents, practically all value is maintained with the group (i.e., payments are close to 0).

❖

Perhaps an even greater weakness of classic mechanism design solutions is the fact that they are designed for *static* settings, where just a single decision is considered at a time. In reality there is practically always a context of future decisions that bears on determining what decision would be optimal now. Should I go on vacation to Italy this month or wait until later? I'll have to decide next summer whether or not to attend my friend's wedding over there... etc. When self-interest and competition are present in such sequential decision-making settings, a static analysis can break down quickly. Should I get to use the car on Friday or should my housemate? What about on Saturday? If we simply hold a Vickrey auction on each day, I might, for instance, gain by reporting a low value on Friday letting my housemate get the car rather than paying the high price for it. Then on Saturday if he no longer needs it I can get it for nothing.

One can imagine a more sophisticated mechanism executed up-front that plans for all time periods, but this will still not work. Why? Because if agents are obtaining *new information* each period (will I finish my task with the car on Friday or need it again Saturday?) then the center must provide incentives for them *each period* to share such information as it becomes known.

I presented *dynamic mechanism design* as a solution framework for such sequential decision making problems with dynamically arriving private information. We saw that many of the results from the static setting carry over, in spirit, to the dynamic case, in that they have rather direct (though more complex) analogues: the dynamic-Groves class of mechanisms extends the Groves class and characterizes the set of mechanisms that are efficient in an ex post equilibrium in dynamic settings; the VCG mechanism has an analogue in Bergemann & Välimäki's dynamic-VCG mechanism, and it is revenue maximizing among efficient mechanisms; the AGV mechanism has an analogue in Athey & Segal's dynamic-balanced mechanism. We even saw that **RM** has an analogue in the dynamic redistribution mechanism (dynamic-RM) for multi-armed bandit settings.

In dynamic settings an additional new challenge arises that isn't present in the static case: what if agents go in and out of contact with the center over time, or what if the make-up of the group itself changes? I provided variants of dynamic-VCG that handle these possibilities given certain assumptions. To handle inaccessible agents, if we assume that all agents must eventually regain contact with the center, then there is a solution that "logs" the payments of dynamic-VCG when an agent is gone and executes them (appropriately scaled for time-discounting) in a lump sum when the agent returns. For the case of changing agent populations, the solution needs to incorporate beliefs about future arrivals and departures of agents in computing payments; when arrivals are independent, the proposed dynamic-VCG variant works.

Finally, in the last chapter I demonstrated how the theory of dynamic mechanism design could be applied to a resource allocation setting. We looked at a one-time, single-item allocation problem with a twist: agents could perform research, at a cost, to potentially increase the value they would obtain from receiving the item. This is a "value discovery" setting where values are constrained to only go up, the idea being that an agent might learn "new uses" for the item without forgetting old ones. How can we implement the sequence of research actions that is optimal before ultimately allocating the resource? It is interesting that this is a *one-time* allocation problem, but it requires a dynamic solution. This is also a prime example of how dynamic mechanism design can be used to provide incentives for agents to do important things (research, here) other than simply reporting their types.

9.2 Something to keep in mind: tractability

If the theory of mechanism design is to be useful in practice, one must be able to determine the nature of the prescriptions it makes in a feasible amount of time. The *computational tractability* of mechanisms—decision policies and transfer payments—is a huge concern. The issue becomes much more acute when we move to a dynamic setting, where computing optimal policies for multi-agent MDPs gets intractable *very* quickly in the worst case (say, when there are more than a few agents with more than a few states in a setting with more than a few decisions to be made sequentially). The equilibrium properties of the mechanisms I presented depend crucially on the center following an efficient policy: the payment structure aligns agent interests with social welfare, but if social welfare isn't being maximized then it's not guaranteed that being truthful will maximize each agent's utility.

One of the things we can do in the face of this conundrum is identify special cases in which optimal policies *are* tractable. This was one of the main motivations for the emphasis I placed on multi-armed bandit settings in the thesis; Gittins showed that in those environments the complexity of computing optimal policies is linear in the number of agents. Fortunately, there are indeed interesting dynamic problems that more or less fit the multi-armed bandits model, and the results for that setting could potentially be put into practice.

But what about domains that do not have structure that allows for tractable computation of optimal policies? Is there still a place for mechanism design? First, we can make statements like "if the mechanism's policy is within ε of optimal, then each agent can gain at most ε from a non-truthful strategy", where one could say the incentive properties are weakened by an amount proportional to the distance of the mechanism's policy from optimality. But I like to think we can do better, and the intuition is as follows: Implementing a Groves payment scheme is *always* computationally tractable when agents can quantify the value they experience (simply ask the agents what value they just experienced, and make payments), which means that interests can *always* be aligned towards maximization of social welfare. Then given this, agents will want to deviate from truth only if they believe doing so will improve social welfare. If all agents agree that the center's chosen decision "heuristic" policy is "best" among any that are known among the group, they will choose to be truthful. A mechanism design approach that explicitly reasons about agent beliefs in this way, and perhaps allows agents to share information about newly discovered decision heuristics with the center, intuitively seems promising.

9.3 Other directions for future work

Redistribution mechanisms

RM is a mechanism applicable to arbitrary domains, but in order to be implemented a revenue-guarantee for each agent must be computed. We saw that in the case of single-item allocation problems (or AON domains, more broadly), a simple algorithm exists—just imagine the agent's value were 0 and see what revenue would result. But this simple solution doesn't generalize. In combinatorial allocation domains a revenue-minimizing report may not specify value 0 for every outcome. I presented a mixed-integer programming formulation for computing revenue-guarantees, but simple solutions, when they exist, are preferable. Since combinatorial allocation problems are so important, it would be worthwhile to try to determine a simple way of computing revenue-guarantees in that space, and also to do an empirical analysis of the performance of **RM** there.

It is also true that designing mechanisms customized to specific settings may allow us to achieve things we can't in a general-purpose mechanism. Guo & Conitzer [2007; 2008c] and Moulin [2007] have achieved positive results in allocation settings with multiple identical items. A custom mechanism for general combinatorial auctions may also yield gains over **RM**.

Another interesting observation is that redistribution mechanisms are somewhat relevant to consideration of *fairness*. In some settings *equitability* and *no-envy* concerns are important—it is often desirable that agents receive a share of the social welfare that is somehow deemed "just", or that no agent would prefer to be in any other agent's shoes. **RM**'s satisfaction of redistribution-anonymity is relevant, but otherwise the mechanism does not explicitly pursue such goals. Still, it is worth pointing out that **RM** improves on VCG in terms of decreasing the discrepancy in the utilities that agents realize. For instance in single-item allocation, agents not allocated the resource obtain positive utility under **RM**. Except for the case of the second highest bidder, all "losing" agents receive a redistribution payment that is weakly bigger than the winner's redistribution payment; this evens things out, to some extent. In the 3-agent example of Figure 1.3, for instance, the utilities under **RM** are 4, 2, and 2.67; under VCG they are 2, 0, and 0. I don't know if achieving less discrepancy in utilities is possible in an efficient mechanism, but it's an interesting question. Even if it means achieving less overall utility for the agents (lower redistribution), in some cases fairness considerations might make it a worthwhile tradeoff.

Dynamic mechanisms

Dynamic mechanism design is a new area, but it's really an extension of the theory built up over the last decades for the static case and thus will probably "mature" relatively quickly. I presented analogues of several fundamental results from the static setting, but important questions remain. The extension of the Green & Laffont [1977] characterization was done within a context of history-independent transfer functions. In a dynamic setting transfers could, in principle, be computed based on entire report histories. The important within-period ex post efficient mechanisms we saw (dynamic-VCG and dynamic-RM) both compute history-independent transfers, but might there be other interesting mechanisms that do not? It seems likely that in moving to a characterization for a history-*dependent* transfer context the dynamic-Groves intuition will persist, but this should be verified and it will be worthwhile to work out the details.

There are important open questions about dynamic redistribution mechanisms. I specified dynamic-RM for multi-armed bandit settings, which are the dynamic analogue of AON domains, but can **RM** be generalized to apply to arbitrary dynamic settings? Interestingly, it may *not* be the case that dynamic-VCG is unique among all within-period ex post efficient, ex post IR, and no-deficit dynamic mechanisms, contrary to the analogous result I provided for VCG and the static setting. Can we specify a mechanism that redistributes dynamic-VCG revenue in any case where doing so is possible without distorting incentives? Can we specify a dynamic redistribution mechanism that is *optimal* under some plausible conditions? Is dynamic-RM optimal for multi-armed bandit worlds when a dynamic analogue of redistribution-anonymity is imposed?

Finally, in Chapter 7 we discussed the implementation of dynamic mechanisms in settings with interdependent types (e.g., where my expected type in the next period given an action may depend on your current type as well as my own). We saw that the dynamic-basic-Groves mechanism provides the right incentives even in these environments, but dynamic-VCG generally does not. An ex ante charge dynamic mechanism could yield ex ante budget balance and IR, but is there a dynamic mechanism that is efficient and IR in within-period ex post Nash equilibrium and is *guaranteed* to not run a deficit?

9.4 A closing thought

It is typical in the development of new theories that certain assumptions are made in order to make analysis tractable. Of course work in mechanism design is no exception. We assumed rational agents that act infallibly to maximize there own utilities; we assumed those utilities have a quasilinear form (which implies no agent cares what transfer payments *other* agents receive); we assumed agents will participate in a mechanism if their expected utility for doing so is not negative (what if it's vanishingly close to 0?), etc. The spirit of these assumptions is grounded in actuality, but they certainly will not always hold true. This is the way we make progress, but I would offer that the problems we wish to solve in this field demand extra diligence in making sure our assumptions actually match the real world. Mechanism design, after all, is a theory about how to organize human behavior, so we should make sure we're considering *actual* human behavior and not just the behavior that is most convenient for our theories.

Bibliography

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

[Athey and Segal, 2007] Susan Athey and Ilya Segal. An efficient dynamic mechanism. Working paper, http://www.stanford.edu/ isegal/agv.pdf, 2007.

[Bailey, 1997] Martin J. Bailey. The demand revealing process: To distribute the surplus. *Public Choice*, 91:107–126, 1997.

[Bellman and Kalaba, 1959] Richard Bellman and Robert Kalaba. A mathematical theory of adaptive control processes. *Proc. of the National Academy of Sciences of the United States of America*, 45(8):1288–1290, 1959.

[Bellman, 1956] Richard Bellman. A problem in the sequential design of experiments. *Sankhya*, 16(3,4):215–220, 1956.

[Bellman, 1957] Richard Bellman. A markovian decision process. *Journal of Mathematics and Mechanics*, 6, 1957.

[Bergemann and Valimaki, 2006] Dirk Bergemann and Juuso Valimaki. Efficient dynamic auctions. Cowles Foundation Discussion Paper 1584, http://cowles.econ.yale.edu/P/cd/d15b/d1584.pdf, 2006.

[Bergemann and Valimaki, 2007] Dirk Bergemann and Juuso Valimaki. Dynamic marginal contribution mechanism. Cowles Foundation Discussion Paper 1616, http://cowles.econ.yale.edu/P/cd/d16a/d1616.pdf, 2007.

[Berry and Fristedt, 1985] D. A. Berry and B. Fristedt. *Bandit problems: sequential allocation of experiments.* Chapman & Hall, London, 1985.

[Bertsimas and Nino-Mora, 1996] D. Bertsimas and J. Nino-Mora. Conservation laws, extended polymatroids and multiarmed bandit problems; a unified approach to indexable systems. *Mathematics Operation Research*, 21:257–306, 1996.

[Boutilier, 1996] Craig Boutilier. Planning, learning and coordination in multiagent decision processes. In *Proceedings of the Conference on Theoretical Aspects of Rationality and Knowledge*, pages 195–210, 1996.

[Castanon *et al.*, 1999] David Castanon, Simon Streltsov, and Pirooz Vakili. Optimality of index policies for a sequential sampling problem. *IEEE Transactions on Automatic Control*, 44(1):145–148, 1999.

[Cavallo and Parkes, 2008] Ruggiero Cavallo and David C. Parkes. Efficient metadeliberation auctions. In *Proceedings of the 26th Annual Conference on Artificial Intelligence (AAAI-08) (to appear)*, 2008.

[Cavallo *et al.*, 2006] Ruggiero Cavallo, David C. Parkes, and Satinder Singh. Optimal coordinated planning amongst self-interested agents with private state. In *Proceedings of the Twenty-second Annual Conference on Uncertainty in Artificial Intelligence (UAI'06)*, 2006.

[Cavallo *et al.*, 2007] Ruggiero Cavallo, David C. Parkes, and Satinder Singh. Online mechanisms for persistent, periodically inaccessible self-interested agents. In *DIMACS Workshop on the Boundary between Economic Theory and Computer Science*, 2007.

[Cavallo, 2006a] Ruggiero Cavallo. Handling self-interest in groups, with minimal cost. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06), Nectar paper track*, 2006.

[Cavallo, 2006b] Ruggiero Cavallo. Optimal decision-making with minimal waste: Strategyproof redistribution of VCG payments. In *Proceedings of the 5th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'06)*, pages 882–889, 2006.

[Cavallo, 2008] Ruggiero Cavallo. Efficiency and redistribution in dynamic mechanism design. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC-08) (to appear)*, 2008.

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

[Cremer *et al.*, 2007] Jacques Cremer, Yossi Spiegel, and Charles Zhoucheng Zheng. Auctions with costly information acquisition. Iowa State University Department of Economics Working Papers Series, 2007.

[Dasgupta and Maskin, 2000] Partha Dasgupta and Eric Maskin. Efficient auctions. *The Quarterly Journal of Economics*, 115(2):341–388, 2000.

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

[de Farias and Roy, 2003] D. P. de Farias and B. Van Roy. The linear programming approach to approximate dynamic programming. *Operations Research*, 51(6):850–856, 2003.

[Duff, 2002] Michael Duff. *Optimal Learning: Computational procedures for Bayes-adaptive Markov decision processes (PhD Thesis).* University of Massassachusetts Amherst, 2002.

[Ephrati and Rosenschein, 1991] E. Ephrati and J. S. Rosenschein. The clarke tax as a consensus mechanism among automated agents. In *Proceedings of the 9th Annual Conference on Artificial Intelligence (AAAI-91)*, pages 173–178, 1991.

[Faltings, 2004] Boi Faltings. A budget-balanced, incentive-compatible scheme for social choice. In *Agent-mediated E-commerce (AMEC'04)*, 2004.

[Feigenbaum *et al.*, 2001] Joan Feigenbaum, Christos H. Papadimitriou, and Scott Shenker. Sharing the cost of multicast transmissions. *Journal of Computer and System Sciences*, 63(1):21–41, 2001.

[Frostig and Weiss, 1999] E. Frostig and G. Weiss. Four proofs of gittins' multiarmed bandit theorem. *Applied Probability Trust*, 1999.

[Gibbard, 1973] Alan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41(4):587–601, 1973.

[Gittins and Jones, 1974] J. C. Gittins and D. M. Jones. A dynamic allocation index for the sequential design of experiments. In *In Progress in Statistics*, pages 241–266. J. Gani et al., 1974.

[Gittins, 1989] J. C. Gittins. *Multi-armed Bandit Allocation Indices.* Wiley, New York, 1989.

[Glazebrook, 1979] K. D. Glazebrook. Stoppable families of alternative bandit processes. *Journal of Applied Probability*, 16:843–854, 1979.

[Green and Laffont, 1977] Jerry Green and Jean-Jacques Laffont. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, 45:427–438, 1977.

[Green and Laffont, 1979] Jerry R. Green and Jean-Jacques Laffont. *Incentives in public decision-making.* North Holland, New York, 1979.

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

[Guo and Conitzer, 2007] Mingyu Guo and Vincent Conitzer. Worst-case optimal redistribution of VCG payments. In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC-07), San Diego, CA, USA*, pages 30–39, 2007.

[Guo and Conitzer, 2008a] Mingyu Guo and Vincent Conitzer. Better redistribution with inefficient allocation. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC-08) (to appear)*, 2008.

[Guo and Conitzer, 2008b] Mingyu Guo and Vincent Conitzer. Optimal-in-expectation redistribution mechanisms. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS-08)*, 2008.

[Guo and Conitzer, 2008c] Mingyu Guo and Vincent Conitzer. Undominated redistribution mechanisms. In *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS-08)*, 2008.

[Hartline and Roughgarden, 2008] Jason D. Hartline and Tim Roughgarden. Optimal mechanism design and money burning. In *Proceedings of the 40th annual ACM symposium on Theory of Computing (STOC'08)*, 2008.

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

[Hurwicz, 1960] Leonid Hurwicz. Optimality and informational efficiency in resource alloca-tion processes. In Arrow, Karlin, and Suppes, editors, *Mathematical Methods inthe Social Sciences*. Stanford University Press, 1960.

[Hurwicz, 1972] Leonid Hurwicz. On informationally decentralized systems. In Radner and McGuire, editors, *Decision and Organization*. North Holland, Amsterdam, 1972.

[Hurwicz, 1975] Leonid Hurwicz. On the existence of allocation systems whose manipulative nash equilibria are pareto optimal. *(presented at the 3rd World Congress of the Econometric Society)*, 1975.

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

[Jackson, 2001] Matthey O. Jackson. A crash course in implementation theory. *Social choice and welfare*, 18:655–708, 2001.

[Jehiel and Moldovanu, 2001] Philippe Jehiel and Benny Moldovanu. Efficient design with interdependent valuations. *Econometrica*, 69:1237–1259, 2001.

[Kaelbling *et al.*, 1996] Leslie Pack Kaelbling, Michael L. Littman, and Andrew W. Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[Katehakis and Veinott, 1987] M. N. Katehakis and A. F. Veinott. The multi-armed bandit problem: decomposition and computation. *Mathematics of Operations Research*, 22(2):262–268, 1987.

[Kearns and Singh, 1998] Michael Kearns and Satinder Singh. Near-optimal reinforcement learning in polynomial time. In *Proceedings of the Fifteenth International Conference on Machine Learning (ICML)*, pages 260–268, 1998.

[Kearns *et al.*, 1999] M. Kearns, Y. Mansour, and A. Ng. A sparse sampling algorithm for near-optimal planning in large markov decision processes. In *Proceedings of the SixteenthInternational Joint Conference on Artificial Intelligence*, pages 1324–1331, 1999.

[Keller and Strazzera, 2002] L. Robin Keller and Elisabetta Strazzera. Examining predictive accuracy among discounting models. *Journal of Risk and Uncertainty*, 24(2):143–160, March 2002.

[Krishna and Perry, 1998] Vijay Krishna and Motty Perry. Efficient mechanism design. Game theory and information, Economics Working Paper Archive at WUSTL, Dec 1998.

[Krishna, 2002] Vijay Krishna. *Auction Theory.* Academic Press, 2002.

[Larson and Sandholm, 2001] K. Larson and T. Sandholm. Costly valuation computation in auctions. In *Eighth Conference of Theoretical Aspects of Knowledge and Rationality (TARK VIII)*, 2001.

[Larson and Sandholm, 2005] K. Larson and T. Sandholm. Mechanism design and deliberative agents. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005)*, 2005.

[Larson, 2006] K. Larson. Reducing costly information acquisition in auctions. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2006.

[Lavi and Nisan, 2000] Ron Lavi and Noam Nisan. Competitive analysis of incentive compatible on-line auctions. In *Proceedings of the 2nd ACM conference on Electronic Commerce*, pages 233–241, 2000.

[Mas-Colell *et al.*, 1995] Andreu Mas-Colell, Jerry R. Green, and Michael D. Whinston. *Microeconomic Theory.* Oxford University Press, USA, 1995.

[Mezzetti, 2004] Claudio Mezzetti. Mechanism design with interdependent valuations: Efficiency. *Econometrica*, 72(5):1617–1626, 2004.

[Milgrom and Weber, 1982] P. R. Milgrom and R. Weber. A theory of auctions and competitive bidding. *Econometrica*, 50:1089–1122, 1982.

[Moulin, 2007] Hervé Moulin. Efficient, strategy-proof and almost budget-balanced assignment. unpublished, 2007.

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

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

[Myerson, 1986] Roger Myerson. Multistage games with communication. *Econometrica*, 54(2):323–358, 1986.

[Nash, 1950] John Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the United States of America*, 36(1):48–49, 1950.

[Osborne and Rubinstein, 1994] Martin J. Osborne and Ariel Rubinstein. *A Course in Game Theory.* MIT Press, Cambridge, MA, 1994.

[Parkes and Singh, 2003] David C. Parkes and Satinder Singh. An MDP-based approach to Online Mechanism Design. In *Proceedings of the 17th Annual Conf. on Neural Information Processing Systems (NIPS'03)*, 2003.

[Parkes *et al.*, 2001] David C. Parkes, J. R. Kalagnanam, and M. Eso. Achieving budget-balance with Vickrey-based payment schemes in exchanges. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 1161–1168, 2001.

[Parkes, 2001] David C. Parkes. *Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency.* PhD Thesis, Department of Computer and Information Science, University of Pennsylvania, 2001.

[Parkes, 2005] David C. Parkes. Auction design with costly preference elicitation. *Annals of Mathematics and AI*, 44:269–302, 2005.

[Parkes, 2006] David C. Parkes. Iterative combinatorial auctions. In Peter Cramton, Yoav Shoham, and Richard Steinberg, editors, *Combinatorial Auctions.* MIT Press, 2006.

[Petcu *et al.*, 2006] Adrian Petcu, Boi Faltings, and David C. Parkes. MDPOP: Faithful distributed implementation of efficient social choice problems. In *Proceedings 5th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'06)*, pages 1397–1404, 2006.

[Porter *et al.*, 2004] R. Porter, Y. Shoham, and M. Tennenholtz. Fair imposition. *Journal of Economic Theory*, 118(2):209–228, Oct 2004.

[Puterman, 1994] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley-Interscience, April 1994.

[Satterthwaite, 1975] Mark A. 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.

[Schweitzer and Seidmann, 1985] P. J. Schweitzer and A. Seidmann. Generalized polynomial approximations in markovian decision processes. *Journal of Mathematical Analysis and Applications*, 110(6):568–582, 1985.

[Segal and Toikka, 2007] Ilya Segal and Juuso Toikka. Revenue equivalence, profit maximization, and transparency in dynamic mechanisms. Working paper, http://www.stanford.edu/ isegal/req.pdf, 2007.

[Shneidman and Parkes, 2004] Jeffrey Shneidman and David C. Parkes. Specification faithfulness in networks with rational nodes. In *Proceedings of the 23rd ACM Symposium on Principles of Distributed Computing*, pages 88–97, 2004.

[Sutton and Barto, 1998] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction.* The MIT Press, Cambridge, MA, 1998.

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

[von Neumann and Morgenstern, 1944] John von Neumann and Oska Morgenstern. *Theory of games and economic behavior.* Princeton University Press, 1944.

[Watkins, 1989] Christopher Watkins. *Learning from delayed rewards.* PhD Thesis, University of Cambridge, Psychology Department, University of Cambridge, 1989.

[Weber, 1992] Richard Weber. On the gittins index for multiarmed bandits. *The Annals of Applied Probability*, 2(4):1024–1033, Nov 1992.

[Weitzman, 1979] M. L. Weitzman. Optimal search for the best alternative. *Econometrica*, 47:641–654, 1979.

[Whittle, 1980] P. Whittle. Multi-armed bandits and the gittins index. *Journal of the Royal Statistical Society, Series B (Methodological)*, 42(2):143–149, 1980.