Efficiency and Redistribution in Dynamic Mechanism Design

Ruggiero Cavallo
SEAS, Harvard University
33 Oxford St.
Cambridge, MA 02138
cavallo@eecs.harvard.edu

ABSTRACT

The emerging area of dynamic mechanism design seeks to achieve desirable equilibrium outcomes in multi-agent sequential decision-making problems with self-interest. Here we take the goal of maximizing social welfare. We start by extending the characterization result of Green & Laffont [1977] to a dynamic setting, defining the dynamic-Groves class of dynamic mechanisms and showing that it exactly corresponds to the set of mechanisms that are efficient (social welfare maximizing) and incentive compatible in an ex post equilibrium. The dynamic-VCG mechanism of Bergemann & Välimäki [2006] is a dynamic analogue of the static VCG mechanism and is efficient, incentive compatible, and individual rational in an ex post equilibrium; we use our characterization result to show here that it is also revenue maximizing among all dynamic mechanisms with these properties. In other words, dynamic-VCG maximizes the payments required of the agents and thus, while perhaps desirable for an auctioneer seeking high revenue, is in fact *worst* when maximizing agent utility is the goal. We then build on recent work on static redistribution mechanisms (see [Cavallo, 2006]) to design a dynamic redistribution mechanism for multi-armed bandit settings (e.g., the repeated allocation of a single good) that returns much of the revenue under dynamic-VCG back to the agents, while maintaining the same efficiency, incentive compatibility, individual rationality, and no-deficit properties. We conclude with a numerical analysis, demonstrating empirically that this redistribution mechanism typically comes close to perfect budget balance.

Categories and Subject Descriptors

J.4 [**Social and Behavioral Sciences**]: Economics; I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—Multiagent systems

General Terms

Economics, Theory

Keywords

Mechanism design, Social welfare, Redistribution

EC'08, July 8–12, 2008, Chicago, Illinois, USA.

1. INTRODUCTION

In this paper we consider settings in which a group of self-interested parties faces a series of decisions to be made *sequentially over time*, with new private information potentially obtained after each decision. We seek to maximize the social welfare generated by the sequence of decisions.

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. In a scenario like this, the government probably does not want to extract large payments from the communities that use the clinic—it is a public good—but, as we will see, certain payment mechanisms are desirable in that they can elicit honest evaluation and reporting of needs.

The emerging area of dynamic mechanism design (DMD) exists to address such problems. The core ideas of DMD extend the relatively mature theory of static mechanism design (MD) for one-shot settings. In the static case, a mechanism consists of a choice function that maps agent reports of private information to outcomes, and a monetary transfer function, i.e., payments imposed on the agents that serve the purpose of aligning incentives towards the mechanism designer's goal. A *dynamic mechanism* is different in that it chooses an outcome every time-period, incorporating reports of new private information, and can also specify transfer payments each period (see [Parkes, 2007] for a recent survey).

In this paper we address some core issues in dynamic mechanism design, motivated by the goal of maximizing social welfare; we find that important static-setting results have natural (though more complex) extensions to dynamic settings. We start by defining a class of "dynamic-Groves" mechanisms that generalizes the Groves class for static settings; every dynamic-Groves mechanism specifies a transfer function such that future payments to each agent—in expectation given that other agents participate truthfully—equal the value the other agents will obtain going forward, minus some constant. We prove that this class fully characterizes the set of mechanisms that implement the efficient decision policy in a truthtelling ex post Nash equilibrium.

We then analyze the dynamic-VCG mechanism, which was recently derived by Bergemann & Välimäki [2006] and is the natural analogue of the VCG mechanism for static settings. Under dynamic-VCG each agent's equilibrium expected payoff equals the amount it contributes to social welfare. Dynamic-VCG has several very nice properties: it is efficient, incentive compatible (IC), and individual rational (IR) in an ex post Nash equilibrium and never runs a deficit. In fact, we show in this paper that among all efficient, IC and IR mechanisms it *maximizes revenue*, i.e., requires the largest possible payments from the agents for every possible instance. In some settings this may be considered a good thing, for instance when an auctioneer seeks to implement efficient decisions and at the same time extract as much of the value as possible.

That said, here we take seriously the fact that in many cases the payments required of agents under dynamic-VCG are not desirable, but are rather a "cost of implementation" for a mechanism that is both efficient and no-deficit. When a group of agents simply wants to make welfare maximizing decisions, the mechanism payments are just a tool for eliciting truthful reporting of private information, and in fact detract from social welfare. In other words, they're waste. Bailey [1997], Cavallo [2006], and others have specified so-called "redistribution mechanisms" for the static setting, in which large portions of the payments required under VCG are redistributed back to agents without diminishing the incentive properties. Here we apply the same idea to dynamic settings that can be modeled as *multi-armed bandits*, e.g., the repeated allocation of a single good, and find that the vast majority of revenue under dynamic-VCG can be returned to the agents. An array of important sequential decision problems fit the multi-armed bandits model, including the mobile health clinic scenario we described above.

2. SETUP AND BACKGROUND

We consider settings in which there is a group I of agents $\{1, 2, \ldots, n\}$, and a sequence of K decisions or "actions" to be taken, one per time-step (where K is potentially infinite). At each time-step, each agent has some private information that determines the value it 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.

To formalize these notions we use the Markov decision process (MDP) framework. There is an exogenously defined space of actions A and space of *types* Θ_i for each agent i. Each 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, $s_{\theta_i^t} \in S_i$ is the "current" local state. $r_{\theta_i^t} : S_i \times A \rightarrow \Re$ is the value (or "reward") function, with $r_{\theta_i^t}(s_i, a)$ denoting the immediate value that i obtains if action a is taken when i is in local state s_i. $\tau_{\theta_i^t} : S_i \times A \times S_i \rightarrow \Re$ 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 an MDP for agent i.

Note that this set-up places us in a private values setting, excluding scenarios where an agent's value for an action depends on the private information of some other agent. But we 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 at t were θ_j his current state would be some θ_j''.[1]

We will apply dynamic mechanism design to such settings, in which a coordinator or *center* elicits reports from agents regarding private types in every period, and then takes an action. We let $\theta_c^t \in \Theta_c$ denote the center's "type" at time t, i.e., a representation of any information known to the center at t. We then denote the joint type-space $\Theta = \Theta_c \times \Theta_1 \times \ldots \times \Theta_n$. As a notational simplification, we will write $\tau(\theta, a)$ for the random variable representing the joint type in Θ that results when action a is taken in joint type θ. We use $-i$ to denote $I \setminus \{i\}$ (e.g., θ_{-i}^t for the profile of types for agents other than i at time t), and $r(\theta, a)$ to denote the immediate value of taking action a in joint state s_θ. We write $r_i(\theta_i, a) = r_{\theta_i}(s_{\theta_i}, a)$, $r(\theta, a) = \sum_{i \in I} r_i(\theta_i, a)$, and $r_{-i}(\theta_{-i}, a) = \sum_{j \in I \setminus \{i\}} r_j(\theta_j, a)$. We 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 $\gamma^t x$.

Formally, a dynamic mechanism is a tuple (π, T), where decision policy $\pi : \Theta \rightarrow A$ maps a joint reported type to an action,[2] and $T = (T_1, \ldots, T_n)$, where each $T_i : \Theta \rightarrow \Re$ maps a joint reported type to a monetary payment delivered *from* the center *to* agent i. As we will see, certain payment schemes will succeed in aligning interests towards execution of certain decision policies, such that each agent will be best off participating truthfully given the center's policy. We will focus on the socially optimal or *"efficient"* decision policy π^* (defined formally below). Agents report types according to *strategies*. We let $\sigma_i : \Theta_i \rightarrow \Theta_i$ denote a reporting strategy for agent i.[3] We let $\sigma = (\sigma_1, \ldots, \sigma_n)$, and for any $\theta \in \Theta$, $\sigma(\theta) = (\theta_c, \sigma_1(\theta_1), \ldots, \sigma_n(\theta_n))$.

Note that in a dynamic mechanism the report history is not relevant to determining the optimal decision policy (when agents are truthful). For simplicity, here we also assume a context of *history-independent transfers*—which can be modeled by simply assuming θ_c^t does not track types reported in periods previous to t—and leave a more thorough analysis considering history-dependent transfers to a future extended version of the paper.

[1] This can be modeled explicitly by 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 i's next type—this allows a coupling (only through the realization of random events) of agent type transitions. But for simplicity of exposition we omit φ from the notation going forward.

[2] We 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 and τ are constant), but formally a dynamic mechanism will allow each agent to report its entire type in every period (allowing for the possibility of an agent i being truthful *in the future* from a time in which he has misreported r_i or τ_i).

[3] Though an agent may be aware of its entire *history* of types (not just the current type), this formulation is without loss of generality, as whenever histories (or reported histories) may play a role in an agent's strategy we can consider that typespaces are defined with each local state θ_i^t containing a representation of i's entire state history through time t.

We 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 values to be obtained by agent i in the future given true joint type θ^t, decision policy π, and reporting strategy profile σ. Algebraically,

$$V_i(\theta^t, \pi, \sigma) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} r_i(\theta_i^k, \pi(\sigma(\theta^k))) \,\Big|\, \theta^t, \pi, \sigma\Big] \quad (1)$$

  Here and in all other places in this paper, the expectation is taken over future true types of the agents (θ^k for $k > t$) given the decision policy and strategies, and is based on current *true type* θ^t (not the reported type). When we omit σ_i or σ_{-i}, we intend that the truthful strategy is followed. When we omit π, we intend that the expectation is based on execution of π^*. So, for example, $V_i(\theta^t)$ is the expected utility to i given joint type θ^t, truthful reporting by all agents, and execution of π^*. Letting Π be the set of all possible decision policies,

$$\forall \theta^t \in \Theta,\ \pi^* = \arg\max_{\pi \in \Pi} \sum_{i \in I} V_i(\theta^t, \pi) \quad (2)$$

- 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., $\arg\max_{\pi \in \Pi} \sum_{j \in I\setminus\{i\}} V_j(\theta^t, \pi)$). We 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,

$$\begin{aligned} V_{-i}(\theta^t_{-i}) &= V_{-i}(\theta^t_i, \theta^t_{-i}, \pi^*_{-i}, \sigma_i) & (3)\\ &= \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} r_{-i}(\theta^k_{-i}, \pi^*_{-i}(\theta^k_{-i})) \,\Big|\, \theta^t, \pi^*_{-i}\Big] \end{aligned}$$

- 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):

$$\mathcal{T}_i(\theta^t, \pi, \sigma) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} T_i(\sigma(\theta^k)) \,\Big|\, \theta^t, \pi, \sigma\Big] \quad (4)$$

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

We assume quasilinear utility throughout the paper, which, given this notation, 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:

$$V_i(\theta^t, \pi, \sigma) + \mathcal{T}_i(\theta^t, \pi, \sigma) \quad (5)$$

The goal in dynamic mechanism design is to achieve implementation of desirable decision policies—typically, the efficient policy π^*—in a game theoretic equilibrium. We take as our solution concept the strong *within-period ex post Nash equilibrium*, where there is a strategy profile in which each agent maximizes its *payoff* (expected discounted utility) by playing the equilibrium strategy, given that the other agents do, for *every possible joint type*.

Definition 1. (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,*

$$\begin{aligned} &V_i(\theta^t, \pi, \sigma_i, \sigma_{-i}) + \mathcal{T}_i(\theta^t, \pi, \sigma_i, \sigma_{-i}) & (6)\\ &\geq V_i(\theta^t, \pi, \sigma'_i, \sigma_{-i}) + \mathcal{T}_i(\theta^t, \pi, \sigma'_i, \sigma_{-i}) & (7) \end{aligned}$$

A mechanism is *incentive compatible (IC)* in this equilibrium if each agent maximizes its 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. For clarity, we provide the full formal descriptions of these two concepts:

Definition 2. (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 θ^t, and for all σ_i,*

$$V_i(\theta^t, \pi) + \mathcal{T}_i(\theta^t, \pi) \geq V_i(\theta^t, \pi, \sigma_i) + \mathcal{T}_i(\theta^t, \pi, \sigma_i) \quad (8)$$

Definition 3. (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 $\theta^t \in \Theta$,*

$$V_i(\theta^t, \pi, \sigma) + \mathcal{T}_i(\theta^t, \pi, \sigma) \geq 0 \quad (9)$$

It may initially surprise some that we are in a private values setting, yet ex post Nash equilibrium is distinct from dominant strategy. To see why, consider a dynamic setting in which an agent i will misreport type information in the current period, 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 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. Intuitively, it says if an agent i knew "everything that is knowable"[4] (i.e., other agents' true current types, whatever they are, but *not* future state transitions), i would want to report honestly as long as other agents do.

2.1 Related work

There is a wealth of related work in static mechanism design that we build on in this paper, and also several relevant recent developments in dynamic mechanism design. Cavallo, Parkes, and Singh [2006] give one of the first treatments of the DMD problem, and provide a mechanism that is efficient and IC in within-period ex post Nash equilibrium. The underlying idea is an extension of the core intuition of the basic Groves mechanism for static settings [Groves, 1973]. Cavallo et al. also provide a mechanism that is no-deficit *ex ante*, but with the individual rationality property also weakened to ex ante (i.e., in expectation from the beginning of execution).

In a key development, Bergemann & Välimäki [2006] provide a dynamic analogue of the celebrated VCG mechanism

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

for static settings. Their *dynamic-VCG* mechanism (which we present and analyze in detail in Section 4) is efficient, IC, and IR in within-period ex post Nash equilibrium, and never runs a deficit, thus improving on the result of Cavallo et al. by significantly strengthening the IR property. Cavallo, Parkes, and Singh [2007] extend dynamic-VCG to settings where the population of agents changes over time, or where agents periodically go out of communication and cannot make or receive transfers. Ieong et al. [2007] also study welfare maximization in a multi-stage model.

Athey & Segal [2007], recognizing that revenue is undesirable in many circumstances, provide a mechanism that is strongly budget-balanced. Their mechanism is a successful extension of the AGV mechanism for static settings [D'Aspermont and Gerard-Varet, 1979; Arrow, 1979], but the weaknesses of AGV also carry over or become more acute in the dynamic setting. Specifically, the mechanism achieves a weaker equilibrium (Bayes-Nash) and the IR property is significantly diminished, effectively back to that of [Cavallo *et al.*, 2006] in which agents will "sign up" at the beginning of the mechanism, but may wish to drop out depending on the types that are realized during execution.

In this paper we provide a dynamic setting analogue to the characterization result of Green & Laffont [1977], who showed that for unrestricted type spaces the Groves class of mechanisms for static settings completely characterizes the set of strategyproof and efficient mechanisms. This helps guide the way in searching for mechanisms that have desirable budget properties while achieving efficiency in the strong within period ex post Nash equilibrium. In Section 5 we present a mechanism for multi-armed bandit settings that does not sacrifice the efficiency, IC, IR, or no-deficit properties of dynamic-VCG, yet yields significantly greater payoff to the agents. Our approach borrows directly from the analysis of [Cavallo, 2006], where much of the VCG revenue is "redistributed" back to the agents in important settings. While that redistribution mechanism can be applied to arbitrary static problems, it has a particularly simple and elegant form in the case of single-item allocation problems, and in that specific setting it coincides with mechanisms specified earlier by Bailey [1997] and Porter et al. [2004].

For arbitrary decision problems, the mechanism of [Cavallo, 2006] is *optimal* (i.e., redistributes the most revenue to the agents, for *any* set of agent types) when a rather strong fairness constraint is imposed. Guo & Conitzer [2007] relax that constraint and find mechanisms that yield even more payoff to the agents in some cases, though their mechanisms are only applicable to multi-unit auctions settings; they do a worst-case analysis and find a "worst-case optimal" mechanism for such environments. In a related vein, Moulin [2007] also optimizes for worst-case performance in multi-unit auctions, but uses a different performance metric. In more recent work for the same setting, Guo & Conitzer [2008] consider redistribution that leverages a prior distribution over agent valuations. Hartline & Roughgarden [2008] consider a setting in which transfers are not possible.

3. EFFICIENT INCENTIVE COMPATIBLE MECHANISMS

We start by pursuing a characterization of dynamic mechanisms that are efficient and incentive-compatible in within-period ex post Nash equilibrium. This will define the terrain, allowing us to focus our analysis in pinpointing particular mechanisms with other desirable qualities. Our methods in this section build on and follow closely the analysis of Groves [1973] in defining the class, and Green & Laffont [1977] in proving the characterization. Consider the following class of "dynamic-Groves" mechanisms, which we 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 outcome, minus some quantity beyond its influence.

DEFINITION 4. (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 $C_i : \Theta \rightarrow \mathcal{R}$ such that, letting $\mathcal{C}_i(\theta^t, \sigma_i) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} C_i(\sigma_i(\theta_i^k), \theta_{-i}^k)) \,|\, \theta^t, \pi^*, \sigma_i]$:*

$$\mathcal{T}_i(\theta^t, \sigma_i) = V_{-i}(\theta^t, \sigma_i) - \mathcal{C}_i(\theta^t, \sigma_i), \qquad (10)$$

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

$$\mathcal{C}_i(\theta^t, \sigma_i') = \mathcal{C}_i(\theta^t, \sigma_i'') \qquad (11)$$

The defining attribute of a dynamic-Groves mechanism is 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 (when truthful) from decisions based on those reports:

LEMMA 1. *A dynamic mechanism (π^*, T) is a dynamic-Groves mechanism if and only if $\forall i \in I, \theta^t \in \Theta, \sigma_i', \sigma_i''$:*

$$\mathcal{T}_i(\theta^t, \sigma_i') - \mathcal{T}_i(\theta^t, \sigma_i'') = V_{-i}(\theta^t, \sigma_i') - V_{-i}(\theta^t, \sigma_i'') \qquad (12)$$

PROOF. First, it is obvious that any dynamic-Groves mechanism satisfies (12). Now, for any mechanism (π^*, T) there is some $C_i : \Theta \rightarrow \Re$ 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 (12). Then, substituting for $\mathcal{T}$ with V_{-i} and such a $\mathcal{C}$ in (12),

$$\begin{aligned}&(V_{-i}(\theta^t, \sigma_i') - \mathcal{C}(\theta^t, \sigma_i')) - (V_{-i}(\theta^t, \sigma_i'') - \mathcal{C}(\theta^t, \sigma_i''))\\ &= V_{-i}(\theta^t, \sigma_i') - V_{-i}(\theta^t, \sigma_i'') \qquad (13)\end{aligned}$$

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

We now show that this fact implies that all dynamic-Groves mechanisms are efficient and incentive compatible in within-period ex post Nash equilibrium.

THEOREM 1. *All dynamic-Groves mechanisms are efficient and incentive compatible in within-period ex post Nash equilibrium.*[5]

PROOF. By Lemma 1, for any dynamic-Groves mechanism (π^*, T), for any θ^t and σ_i:

$$\begin{aligned}&(V_i(\theta^t) + \mathcal{T}_i(\theta^t)) - (V_i(\theta^t, \sigma_i) + \mathcal{T}_i(\theta^t, \sigma_i)) && (14)\\ =\ &(V_i(\theta^t) + V_{-i}(\theta^t)) - (V_i(\theta^t, \sigma_i) + V_{-i}(\theta^t, \sigma_i)) && (15)\\ =\ &V(\theta^t) - V(\theta^t, \sigma_i) && (16)\\ \geq\ &0 && (17)\end{aligned}$$

[5] This theorem is essentially a recasting of Lemma 1 of [Cavallo *et al.*, 2007].

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 $\forall \theta \in \Theta$, $\pi(\theta) = \pi^*(\sigma_i(\theta_i), \theta_{-i})$. □

It is more involved to establish that every mechanism that is efficient and incentive compatible in within-period ex post Nash equilibrium is a dynamic-Groves mechanism; 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.

We 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 partially determined by the realization of *random* events through time $k-1$, and so we let $\mathcal{A}(\theta^t, \sigma)$ be an "action sequence mapping" from the space of possible (given θ^t and σ) random event realizations to a sequence of actions. Given $\dot{\theta}^t, \hat{\theta}^t \in \Theta$, σ', and σ'', then, $\mathcal{A}(\dot{\theta}^t, \sigma') = \mathcal{A}(\hat{\theta}^t, \sigma'')$ means that $\pi^*(\sigma'(\dot{\theta}^t)) = \pi^*(\sigma''(\hat{\theta}^t))$, 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 $(\hat{\theta}^t, \sigma')$, and so on for times $t+2, \ldots, K$.

The proof of our characterization result for dynamic-Groves is simplified by the following lemma, which says that in any within period ex post efficient and IC 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 2. *If a dynamic mechanism (π^*, T) is efficient and incentive compatible in within-period ex post Nash equilibrium, then $\forall i \in I$, $\theta^t_{-i} \in \Theta_{-i}$, and $\dot{\theta}^t_i, \hat{\theta}^t_i \in \Theta_i$,*

$$\pi^*(\dot{\theta}^t_i, \theta^t_{-i}) = \pi^*(\hat{\theta}^t_i, \theta^t_{-i}) \Rightarrow T_i(\dot{\theta}^t_i, \theta^t_{-i}) = T_i(\hat{\theta}^t_i, \theta^t_{-i}) \quad (18)$$

PROOF. Consider an arbitrary mechanism (π^*, T) for which there exists an agent i and types $\dot{\theta}^t_i$, $\hat{\theta}^t_i$, and θ^t_{-i} such that $\pi^*(\dot{\theta}^t_i, \theta^t_{-i}) = \pi^*(\hat{\theta}^t_i, \theta^t_{-i})$ and $T_i(\hat{\theta}^t_i, \theta^t_{-i}) > T_i(\dot{\theta}^t_i, \theta^t_{-i})$. 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 $\hat{\theta}^t_i$ at t since the same action will be taken at t (we use history independence of transfers here). We have:

$$\mathbb{E}\Big[V_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\dot{\theta}^t_i, \theta^t_{-i}))) + \mathcal{T}_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\dot{\theta}^t_i, \theta^t_{-i})))\Big] \quad (19)$$

$$= \mathbb{E}\Big[V_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\hat{\theta}^t_i, \theta^t_{-i}))) + \mathcal{T}_i(\tau(\dot{\theta}^t_i, \theta^t_{-i}, \pi^*(\hat{\theta}^t_i, \theta^t_{-i})))\Big] \quad (20)$$

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

THEOREM 2. *For unrestricted types, if a dynamic mechanism is efficient and incentive compatible in within-period ex post Nash equilibrium then it is a dynamic-Groves mechanism.*

PROOF. Assume for contradiction existence of a mechanism (π^*, T) that is *not* a member of the dynamic-Groves class yet is efficient and IC in within-period ex post Nash equilibrium. By Lemma 1, there is an $i \in I$, joint type $(\hat{\theta}^t_i, \theta^t_{-i})$, strategies σ'_i and σ''_i for agent i, and $\epsilon > 0$ such that:

$$\begin{aligned} &\mathcal{T}_i(\hat{\theta}^t_i, \theta^t_{-i}, \sigma'_i) - \mathcal{T}_i(\hat{\theta}^t_i, \theta^t_{-i}, \sigma''_i) = \\ &V_{-i}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma'_i) - V_{-i}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma''_i) + \epsilon \quad (21) \end{aligned}$$

Consider a type $\dot{\theta}^t_i$ *correlated* with $\hat{\theta}^t_i$ 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 $\hat{\theta}^t_i$. Then, there are strategies $\sigma'_{\hat{\theta}^t_i}$ and $\sigma''_{\hat{\theta}^t_i}$ such that $\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) = \mathcal{A}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma'_i)$ and $\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\hat{\theta}^t_i}) = \mathcal{A}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma''_i)$. Let c be some constant greater than $V_{-i}(\theta^t_{-i})$, and consider that $\dot{\theta}^t_i$ is also such that $\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}) = \mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\hat{\theta}^t_i})$ and, for some $0 < \delta < \epsilon$,

$$\begin{aligned} V_i(\dot{\theta}^t_i, \theta^t_{-i}) &= -V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}) + c + \delta && (22) \\ &= V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\hat{\theta}^t_i}) = -V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\hat{\theta}^t_i}) + c + \delta, && (23) \\ V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) &= -V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) + c, && (24) \end{aligned}$$

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 $\mathcal{A}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma''_i)$ or $\mathcal{A}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma'_i)$ is -1 times the other agents' combined expected value (V_{-i}) for that mapping, plus c. 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.

$\mathcal{A}(\dot{\theta}^t_i, \theta^t_{-i}) = \mathcal{A}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma''_i)$ combined with Lemma 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 $\hat{\theta}^t_i$ and i follows reporting strategy σ''_i. We have:

$$\begin{aligned} &\mathcal{T}_i(\hat{\theta}^t_i, \theta^t_{-i}, \sigma'_i) - \mathcal{T}_i(\hat{\theta}^t_i, \theta^t_{-i}, \sigma''_i) && (25) \\ &= \mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) - \mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}) && (26) \\ &= V_{-i}(\hat{\theta}^t_{-i}, \theta^t_{-i}, \sigma'_i) - V_{-i}(\hat{\theta}^t_i, \theta^t_{-i}, \sigma''_i) + \epsilon && (27) \\ &= V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) - V_{-i}(\dot{\theta}^t_i, \theta^t_{-i}, \sigma''_{\hat{\theta}^t_i}) + \epsilon && (28) \\ &= V_i(\dot{\theta}^t_i, \theta^t_{-i}) - \delta - V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) + \epsilon, && (29) \end{aligned}$$

from which we can see that:

$$\begin{aligned} &\mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) + V_i(\dot{\theta}^t_i, \theta^t_{-i}, \sigma'_{\hat{\theta}^t_i}) \\ &> \mathcal{T}_i(\dot{\theta}^t_i, \theta^t_{-i}) + V_i(\dot{\theta}^t_i, \theta^t_{-i}) \quad (30) \end{aligned}$$

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

THEOREM 3. *For unrestricted types, a dynamic mechanism is efficient and incentive compatible in within-period ex post Nash equilibrium if and only if it is a dynamic-Groves mechanism.*

PROOF. Follows immediately from Lemma 1 and Theorems 1 and 2. □

4. DYNAMIC-VCG AND REVENUE MAXIMIZATION

The results of the previous section provide a complete mapping of the space of possible mechanisms we can consider if we require efficiency and incentive compatibility in a within period ex post Nash equilibrium. But there are additional criteria that will typically be applied to 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. It is typically also important that a mechanism have the *no-deficit* property, i.e., net payments made by the center should be less than or equal to 0. This is important for the feasibility of the mechanism; when the no-deficit property does not hold, the mechanism designer may require an external budget.

Bergemann & Välimäki's dynamic-VCG mechanism, we will see, is efficient, IC, and IR in within-period ex post Nash equilibrium, and is also no-deficit. We will demonstrate efficiency and IC by showing that dynamic-VCG is a dynamic-Groves mechanism and then referring to Theorem 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 some settings a mechanism designer may seek to implement a mechanism in which revenue is high, extracting as much value as possible; we will show that dynamic-VCG is optimal here (if efficiency is required). However, we also note that in many decision problems "revenue" is really just *waste*—in Section 5 we improve on dynamic-VCG in this regard, seeking to *minimize* rather than maximize revenue.

DEFINITION 5 (DYNAMIC-VCG). *[Bergemann and Valimaki, 2006] Decision policy π^* is executed and, $\forall i \in I, \theta^t \in \Theta$:*

$$T_i(\theta^t) = r_{-i}(\theta^t_{-i}, \pi^*(\theta^t)) + \gamma \mathbb{E}[V_{-i}(\tau(\theta^t_{-i}, \pi^*(\theta^t)))] - V_{-i}(\theta^t_{-i}) \tag{31}$$

Recall that $V_{-i}(\theta^k_{-i})$ denotes $V_{-i}(\theta^k_i, \theta^k_{-i}, \pi^*_{-i}, \sigma_i)$, and thus $\mathbb{E}[V_{-i}(\tau(\theta^t_{-i}, \pi^*(\theta^k)))]$ denotes the expected value that agents other than i would obtain from a policy that is optimized for them from the joint type that results when the *socially* optimal policy 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 its current type report inhibits other agents from obtaining value in the present and in the future.[6]

Besides Theorem 6 regarding revenue maximization, which is wholly original to this paper, the properties we observe regarding dynamic-VCG were previously known or follow easily from the analysis of [Bergemann and Valimaki, 2006] or [Cavallo *et al.*, 2007]. Bergemann & Välimäki [2006] provide a direct proof that dynamic-VCG is efficient and IC in equilibrium, and Cavallo et al. [2007] follow with a different, simple proof. We essentially present the core of the [Cavallo *et al.*, 2007] proof here with some minor modifications, but note that our analysis of dynamic-Groves mechanisms allows us 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. We will show that it is, by observing that when other agents are truthful the expected sum, over time, of the first term in (31) 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 4. *The dynamic-VCG mechanism is a dynamic-Groves mechanism.*

PROOF. *(derived from [Cavallo* et al.*, 2007])* 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 denote the true joint type at time $k \geq t$ given other agents are truthful, i follows σ_i, and π^* is executed from θ^t. We have:

$$\mathcal{T}_i(\theta^t,\sigma_i) = \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}(r_{-i}(\theta_{-i}, \pi^*(\sigma_i(\theta^k_i), \theta^k_{-i})) + \gamma\, V_{-i}(\theta^{k+1}_{-i}) - V_{-i}(\theta^k_{-i}))\Big] \tag{32}$$

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

$$= V_{-i}(\theta^t, \sigma_i) - \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t}(V_{-i}(\theta^k_{-i}) - \gamma V_{-i}(\theta^{k+1}_{-i}))\Big] \tag{33}$$

Expanding out the summation, then extracting $V_{-i}(\theta^t_{-i})$ out from the first summation and canceling out (noting that $V_{-i}(\theta^{K+1}_{-i})$ necessarily equals 0), we see this:

$$= V_{-i}(\theta^t, \sigma_i) - \mathbb{E}\Big[\sum_{k=t}^{K} \gamma^{k-t} V_{-i}(\theta^k_{-i}) - \gamma \sum_{k=t}^{K} \gamma^{k-t} V_{-i}(\theta^{k+1}_{-i})\Big] \tag{34}$$

$$= V_{-i}(\theta^t, \sigma_i) - V_{-i}(\theta^t_{-i}) - \mathbb{E}\Big[\gamma \sum_{k=t}^{K-1} \gamma^{k-t} V_{-i}(\theta^{k+1}_{-i}) - \gamma \sum_{k=t}^{K} \gamma^{k-t} V_{-i}(\theta^{k+1}_{-i})\Big] \tag{35}$$

$$= V_{-i}(\theta^t, \sigma_i) - V_{-i}(\theta^t_{-i}) \tag{36}$$

Thus dynamic-VCG is a dynamic-Groves mechanism, as we have shown that, letting $C_i(\theta^t) = V_{-i}(\theta^t_{-i}) - \gamma \mathbb{E}[V_{-i}(\tau(\theta^t_{-i}, \pi^*(\theta^t)))]$, $\mathcal{C}_i(\theta^t, \sigma_i) = V_{-i}(\theta^t_{-i})$ for any σ_i. □

Theorems 3 and 4 together yield:

COROLLARY 1. *The dynamic-VCG mechanism is within-period ex post incentive compatible.*

The following statements about expected equilibrium utilities follow immediately from the proof of Theorem 4:

COROLLARY 2. *Utility to any agent i in the truthful equilibrium under dynamic-VCG, in expectation forward from any θ^t, is $V(\theta^t) - V_{-i}(\theta^t_{-i})$.*

By optimality of π^*, $V(\theta^t) \geq V_{-i}(\theta^t_{-i})$ for all θ^t and i, so we have:

COROLLARY 3. *The dynamic-VCG mechanism is within-period ex post individual rational.*

[6] Bergemann & Välimäki have alternately referred to the mechanism as the "dynamic marginal contribution mechanism".

Corollary 4. *Social utility in the truthful equilibrium under the dynamic-VCG mechanism, in expectation forward from any θ^t, is $n \cdot V(\theta^t) - \sum_{i \in I} V_{-i}(\theta^t_{-i})$.*

Corollary 5. *Revenue in the truthful equilibrium under the dynamic-VCG mechanism, in expectation forward from any θ^t, is $\sum_{i \in I} V_{-i}(\theta^t_{-i}) - (n-1)V(\theta^t)$.*

Theorem 5. *The dynamic-VCG mechanism never runs a deficit, even when agents play off-equilibrium strategies.*

Proof. By optimality (for agents other than i) of π^*_{-i}, for any type θ^t and any reporting strategies σ_i and σ_{-i}, $r_{-i}(\sigma_{-i}(\theta^t_{-i}), \pi^*(\sigma(\theta^t))) + \gamma\mathbb{E}[V_{-i}(\tau(\sigma_{-i}(\theta^t_{-i}), \pi^*(\sigma(\theta^t))))] \leq V_{-i}(\sigma_{-i}(\theta^t_{-i}))$. Thus the net payment to each agent *in every time period* is at most 0 and a deficit can never result. □

We now show that if individual rationality is required in addition to efficiency and incentive compatibility, no mechanism yields more revenue in expectation than dynamic-VCG in a truthful ex post equilibrium, for *any* joint type.

Theorem 6. *For unrestricted types, among all 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: $\sum_{i \in I}[V_{-i}(\theta^t_{-i}) - V_{-i}(\theta^t)]$. Consider any dynamic-Groves mechanism (π^*, T) that yields more revenue (this is without loss of generality by Theorem 3). This mechanism must define $C_1, \ldots, C_n$ such that $\sum_{i \in I} \mathcal{C}_i(\theta^t) > \sum_{i \in I} V_{-i}(\theta^t_{-i})$, 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:

$$\mathcal{C}_i(\theta^t) > V_{-i}(\theta^t_{-i}) \tag{37}$$

Recall that $\mathcal{C}_i(\theta^t)$ must be independent of i's type reports, and thus independent of i's actual sequence of 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:

$$V(\theta^t) - \mathcal{C}_i(\theta^t) = V_{-i}(\theta^t_{-i}) - \mathcal{C}_i(\theta^t) \tag{38}$$

$$< V_{-i}(\theta^t_{-i}) - V_{-i}(\theta^t_{-i}) = 0, \tag{39}$$

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

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. In the static setting for unrestricted valuations the answer is no—VCG is simultaneously revenue maximizing and revenue minimizing, i.e., it is the *only* mechanism with these desirable properties (see [Cavallo, 2006], Proposition 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. We will follow the same approach here, looking, for instance, at dynamic settings in which a single item is to be allocated *repeatedly.* This domain and others fall in the category of multi-armed bandit settings.

5. REDISTRIBUTING REVENUE IN MULTI-ARMED BANDIT SETTINGS

Multi-armed bandit (MAB) problems, so-called due to an analogy that can be drawn to the problem of playing a set of slot-machines with distinct payout rates, are sequential decision-making problems with a strong factorization of the state space. Specifically, there are n Markov processes, one of which may be activated at any given time-step. When a process is activated, a reward is obtained that depends only on the local state of that process, and the process's state changes (all other processes' states remain unchanged).

Among the many good reasons to consider multi-armed bandit problems are: the interesting real-world problems that more or less fit the restrictions of the model; the elegance of the solutions we can achieve; and perhaps most importantly, the computational tractability of actually computing 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:

Theorem 7. [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 $\pi^*(s) = \arg\max_i \nu(s_i)$, $\forall s \in S$.*

An array of real-world decision problems can be modeled as multi-armed bandit scenarios, and there is a natural multi-agent interpretation: 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 (unrestricted types) MDP model we've used for the entire paper, 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. We described one such setting (allocation of a mobile health clinic) in the introduction. 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.

In multi-agent MAB domains, the dynamic-VCG payment structure reduces to a very simple form. Because agents that are not "activated" (e.g., allocated the resource) at any given time do not undergo a state change, their marginal contributions (and thus their payments) are 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.

Definition 6 (Dynamic-VCG in MAB worlds). *[Bergemann and Valimaki, 2006] Decision policy π^* is executed and, $\forall i \in I, \theta^t \in \Theta$:*

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

Observe that the expected revenue generated by dynamic-VCG in a MAB setting is quite large. At the end of this section we 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). We now address that issue by proposing a *dynamic redistribution mechanism.*[7]

For all time-periods t and 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 $j = \pi^*_{-i}(\theta^t)$).

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., $W(\theta^t, \pi^*) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} w(\theta^k, \pi^*) \,|\, \theta^t, \pi^*]$, where θ^k for $k > t$ is a random variable representing the joint type at time k. 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 given θ^t if agent i were not present in the system. We now use these concepts to define dynamic redistribution mechanism *dynamic-RM*:

Definition 7 (Dynamic-RM). *Decision policy π^* is executed and, $\forall i \in I, \theta^t \in \Theta$:*

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

where:

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

The mechanism is dynamic-VCG plus a revenue "redistribution payment". As we will see in Theorem 8, this payment is defined such that the expected sum of redistribution over time to each agent i—no matter what i's strategy—is a constant fraction of the expected revenue that would have resulted if i were not present in the system.

Theorem 8. *Dynamic-RM is efficient and incentive compatible in within-period ex post Nash equilibrium.*

Proof. Since dynamic-VCG is a dynamic-Groves mechanism, by Theorem 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 $\mathcal{Z}(\theta^t, \sigma_i) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} Z_i(\sigma_i(\theta^k_i), \theta^k_{-i}) \,|\, \theta^t, \pi^*, \sigma_i]$:

$$\mathcal{Z}(\theta^t, \sigma'_i) = \mathcal{Z}(\theta^t, \sigma''_i) \tag{40}$$

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:

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

where $\tau(\theta_{-i}, \pi^*_{-i}, \theta'_{-i})$ is 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 $\frac{1}{n} Y_h(\theta^t_{-i}, t)$ 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, $Y_{h^1}(\theta^t_{-i}, t) = \mathbb{E}[\sum_{k=t}^{K} \gamma^{k-t} w(\theta^k_{-i}, \pi^*_{-i}) \,|\, \theta^t_{-i}, \pi^*_{-i}] = W(\theta^t_{-i}, \pi^*_{-i})$. We will now show that for all t, θ^t, and i, for any indicator function h,

$$Y_h(\theta^t_{-i}, t) = Y_{h^1}(\theta^t_{-i}, t) = W(\theta^t_{-i}, \pi^*_{-i}) \tag{41}$$

Take arbitrary t, θ^t, i, and h, and assume for contradiction that $\exists \epsilon > 0$ s.t. $|Y_h(\theta^t_{-i}, t) - Y_{h^1}(\theta^t_{-i}, t)| \geq \epsilon$. Now consider the greatest 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 $\dot{\theta}^{k_h}_{-i}$ associated with a $(k_h - t)^{th}$ expansion of Y_h, given θ^t_{-i} and h. We have that $Y_h(\dot{\theta}^{k_h}_{-i}, k_h) - Y_{h'}(\dot{\theta}^{k_h}_{-i}, k_h)$

$$= (1-\gamma)W(\dot{\theta}^{k_h}_{-i}, \pi^*_{-i}) + \gamma \mathbb{E}\Big[\sum_{k=0}^{K} \gamma^k w(\dot{\theta}^{k_h+k}_{-i}, \pi^*_{-i})\Big] \tag{42}$$

$$- \mathbb{E}\Big[\sum_{k=0}^{K} \gamma^k w(\dot{\theta}^{k_h+k}_{-i}, \pi^*_{-i})\Big]$$

$$= (1-\gamma)W(\dot{\theta}^{k_h}_{-i}, \pi^*_{-i}) + \gamma W(\dot{\theta}^{k_h}_{-i}, \pi^*_{-i}) - W(\dot{\theta}^{k_h}_{-i}, \pi^*_{-i}) \tag{43}$$

$$= 0 \tag{44}$$

Note that for an indicator $h^{1'}$ that has $h^{1'}(k) = 1$ for all $k \geq k_h$, $Y_{h^{1'}}(\dot{\theta}^{k_h}_{-i}, k_h) = \mathbb{E}[\gamma^k \sum_{k=0}^{K} w(\theta^{k_h+k}_{-i}, \pi^*_{-i}) \,|\, \dot{\theta}^{k_h}_{-i}, \pi_{-i}]$. This allows the move to (42). The move from (42) to (43) 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 we 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" 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 $\gamma^{k_h} |Y_{\hat{h}}(\theta^{k_h}_{-i}, k_h) - Y_h(\theta^{k_h}_{-i}, t)| < \epsilon$ 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 $|Y_h(\theta^t_{-i}, t) - Y_{h^1}(\theta^t_{-i}, t)| < \epsilon$. This contradicts our assumption that $|Y_h(\theta^t_{-i}, t) - Y_{h^1}(\theta^t_{-i}, t)| \geq \epsilon$. Since ϵ was chosen arbitrarily, this proves the validity of (41).

Note again that any agent i's only influence on its 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 (41) implies that for any reporting strategies σ'_i and σ''_i,

$$\mathcal{Z}(\theta^t, \sigma'_i) = \mathcal{Z}(\theta^t, \sigma''_i) = \frac{1}{n} W(\theta^t_{-i}, \pi^*_{-i}) \tag{45}$$

□

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

Theorem 9. *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 for every agent i, for all θ^t, for any σ_i, $\mathcal{Z}_i(\theta^k, \sigma_i) \geq 0$.

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 10. *Dynamic-RM is no-deficit.*

Proof. Since dynamic-VCG is no-deficit, it is sufficient to show that for every θ^t:

$$\sum_{i \in I} Z_i(\theta^t) \leq w(\theta^t, \pi^*) = (1-\gamma)\, V_{-i}(\theta^t_{-i}) \qquad (46)$$

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 $\forall i \neq \pi^*(\theta^t)$:

$$n \cdot Z_i(\theta^t) = w(\theta^t_{-i}, \pi^*_{-i}) \leq w(\theta^t, \pi^*), \qquad (47)$$

where the inequality holds simply by observation that $V_{-i}(\theta^t_{-i}) \leq V(\theta^t)$, $\forall \theta^t \in \Theta$. To finish the proof we must show that for $i = \pi^*(\theta^t)$, $n \cdot Z_i(\theta^t) \leq w(\theta^t, \pi^*)$, i.e., that $(1-\gamma) W(\theta^t_{-i}, \pi^*_{-i}) \leq (1-\gamma) V_{-i}(\theta^t_{-i})$, or,

$$W(\theta^t_{-i}, \pi^*_{-i}) \leq V_{-i}(\theta^t_{-i}) \qquad (48)$$

But this holds immediately by within period ex post individual rationality of dynamic-VCG (Theorem 5)—if, in a world without some agent i, the expected discounted payments made to the center were more than the expected value obtained by the agents, some agent would necessarily expect to pay more than the value he obtained from the decision policy. The theorem follows. □

Theorem 11. *Utility to any agent i in the truthful equilibrium under dynamic-RM, in expectation from any θ^t, is:*

$$V(\theta^t) - V_{-i}(\theta^t_{-i}) + \frac{1}{n} \sum_{j \in I \setminus \{i\}} \left[V_{-i,j}(\theta^t_{-i,j}) - V_{-i,j}(\theta^t_{-i}) \right] \quad (49)$$

Proof. From equation (45), in dynamic-RM the expected utility to agent 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 i were not in the system.

From Corollary 5, under dynamic-VCG given any θ^t expected revenue going forward in the truthful equilibrium equals $\sum_{j \in I} V_{-j}(\theta^t_{-j}) - (n-1)V(\theta^t)$, i.e., $\sum_{j \in I} [V_{-j}(\theta^t_{-j}) - V_{-j}(\theta^t)]$. So $W(\theta^t_{-i}, \pi^*_{-i})$ can be written:

$$\sum_{j \in I \setminus \{i\}} \left[V_{-i,j}(\theta^t_{-i,j}) - V_{-i,j}(\theta^t_{-i}) \right] \qquad (50)$$

Adding the payoff under dynamic-VCG (see Corollary 2) and $\frac{1}{n}$ times (50) yields (49). □

Corollary 6. *Social utility in the truthful equilibrium under dynamic-RM, in expectation forward from any θ^t, is:*

$$n \cdot V(\theta^t) - \frac{1}{n} \sum_{i \in I} \left[(2n-2) V_{-i}(\theta^t_{-i}) + \sum_{j \in I \setminus \{i\}} V_{-i,j}(\theta^t_{-i,j}) \right] \quad (51)$$

Corollary 7. *The social utility gain from redistribution in the truthful equilibrium, in expectation from any θ^t, is:*

$$\frac{1}{n} \sum_{i \in I} \sum_{j \in I \setminus \{i\}} \left[V_{-i,j}(\theta^t_{-i,j}) - V_{-i,j}(\theta^t_{-i}) \right] \qquad (52)$$

5.1 Empirical analysis

We ran a numerical analysis to determine what the analytical results for social welfare improvement brought by dynamic-RM map to on plausible problem instances. The punchline is that our simulations demonstrate 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.

We 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 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 it observes a success its α parameter is updated to $\alpha+1$, and if it observes a failure its β is updated to $\beta+1$.

We generated agent types by selecting a number x between 2 and 20 uniformly at random for the number of "prior observations" $(\alpha+\beta)$, and then selecting α uniformly at random from 1 to $x-1$, with $\beta = x - \alpha$. Essentially, this generates a uniform distribution over prior knowledge in the agent population, and a uniform distribution over valuation levels.[8] We 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 we computed[9] 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 4), and the expected percentage retained under dynamic-RM (see Corollary 6). We computed results for several different discount factors (γ), but there were not major differences. Figure 1 plots the results under each mechanism for a range of different population sizes, with $\gamma = 0.8$. For each population size we computed 100 samples and took the average.

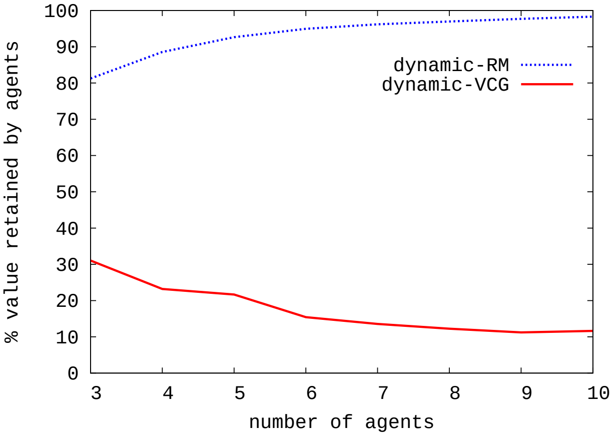

Figure 1: Comparison of the percentage of value from the socially optimal sequence of decisions retained by the agents under dynamic-VCG and dynamic-RM. $\gamma = 0.8$; average over 100 samples for each agent population size.

[8] We would expect dynamic-RM to perform even better on other distributions over agent types, as "similarity" of agent valuations allows greater redistribution in general. We verified this experimentally for a normal distribution, finding $\sim 95\%$ value retained even with just 4 agents.

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

6. CONCLUSION

In this paper we sought to make progress towards understanding how social welfare can be maximized among a group of self-interested agents in sequential decision-making problems. We made three main contributions:

1) we specified the dynamic-Groves class of mechanisms, and proved that it characterizes the set of dynamic mechanisms that are efficient and incentive compatible in within-period ex post Nash equilibrium;
2) we used this characterization to analyze Bergemann & Välimäki's dynamic-VCG mechanism, and proved that it is revenue maximizing (payoff minimizing for the agents) among all IR and no-deficit mechanisms in this class;
3) we proposed the dynamic-RM mechanism for settings that can be modeled as multi-armed bandits (e.g., repeated single-item allocation problems), which redistributes revenue under dynamic-VCG such that the vast majority of value yielded by a sequence of decisions is typically maintained within the set of agents.

Our motivation for dynamic-RM is that in some important settings dynamic-VCG can be considered "wasteful", as the value of decisions is largely not kept within the population of agents. Athey & Segal's [2007] mechanism keeps all value within the group but sacrifices on the equilibrium and, as importantly, on the IR property. A mechanism that is not generally IR in every time-period (theirs is not) raises significant questions about implementability. For repeated single-item allocation settings, dynamic-RM maintains the strong efficiency, IC, IR, and no-deficit properties of dynamic-VCG, while typically obtaining near-perfect budget-balance.

Bergemann & Välimäki 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 this condition, 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.

There are many important directions for further research. For instance, our characterization of the space of efficient and IC dynamic mechanisms was for unrestricted valuations; presumably this result can be strengthened to more restricted classes (as [Holmstrom, 1979] did for the static Groves class). Considering a model in which per-period payments are allowed to depend on entire report histories will also be a worthy extension. There are also many additional interesting questions about redistribution mechanisms. Is there a natural generalization of dynamic-RM to domains beyond those that can be modeled as multi-armed bandits? Is dynamic-RM "optimal" in the strong sense that [Cavallo, 2006] showed of the static version when the analogous fairness constraint is imposed? We suspect the answer to the latter question 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.

7. ACKNOWLEDGMENTS

I am extremely grateful to David Parkes for key input and guidance in the development of this work.

8. REFERENCES

[Arrow, 1979] Kenneth J. Arrow. The property rights doctrine and demand revelation under incomplete infor mation. 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.

[Bergemann and Valimaki, 2006] Dirk Bergemann and Juuso Valimaki. Efficient dynamic auctions. Cowles Foundation Discussion Paper 1584, 2006.

[Cavallo *et al.*, 2006] Ruggiero Cavallo, David C. Parkes, and Satinder Singh. Optimal coordinated planning amongst self-interested agents with private state. In *Proc. 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, 2006] Ruggiero Cavallo. Optimal decision-making with minimal waste: Strategyproof redistribution of VCG payments. In *Proc. of the 5th Int. Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS'06)*, 2006.

[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.

[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.

[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.

[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 *Proc. of the 8th ACM Conference on Electronic Commerce (EC-07), San Diego, CA, USA*, pages 30–39, 2007.

[Guo and Conitzer, 2008] Mingyu Guo and Vincent Conitzer. Optimal-in-expectation redistribution mechanisms. In *Proc. of the 7th Int. Conf. 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.

[Ieong *et al.*, 2007] Samuel Ieong, Anthony Man-Cho So, and Mukund Sundararajan. Mechanism design for stochastic optimization problems. In *3rd International Workshop on Internet and Network Economics*, pages 269–280, 2007.

[Moulin, 2007] Hervé Moulin. Efficient, strategy-proof and almost budget-balanced assignment. unpublished, 2007.

[Parkes, 2007] David C. Parkes. Online mechanisms. In N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors, *Algorithmic Game Theory.* CUP, 2007.

[Porter *et al.*, 2004] R. Porter, Y. Shoham, and M. Tennenholtz. Fair imposition. *Journal of Economic Theory*, 118:209–228, 2004.