Online Mechanisms

Algorithmic Game Theory

Edited by
Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay Vazirani

Contents

1

Online Mechanisms

David C. Parkes

Abstract

Online mechanisms extend the methods of mechanism design to dynamic environments with multiple agents and private information. Decisions must be made as information about types is revealed online and without knowledge of the future in the sense of online algorithms. We first consider single-valued preference domains and characterize the space of decision policies that can be truthfully implemented in a dominant strategy equilibrium. Working in a model-free environment we present truthful auctions for domains with expiring items and limited-supply items. Turning to a more general preference domain, and assuming the existence of a probabilistic model for agent types, we define a dynamic Vickrey-Clarke-Groves mechanism that is efficient and Bayes-Nash incentive compatible. We close with some thoughts about future research directions in this area.

1.1 Introduction

The decision problem in many multi-agent problem domains is inherently dynamic rather than static. Consider, for instance, the following environments:

- Selling seats on an airplane to buyers arriving over time.
- Allocating computational resources (bandwidth, CPU, etc.) to jobs arriving over time.
- Selling adverts on a search engine to a possibly changing group of buyers and with uncertainty about the future supply of search terms.
- Allocating tasks to a dynamically changing team of agents.

In each of these settings at least one of the following is true: either agents are dynamically arriving or departing, or there is uncertainty about the set

of feasible decisions in the future. These dynamics present a new challenges when seeking to sustain good system-wide decisions in multi-agent systems with self-interested agents.

This chapter introduces the problem of *online mechanism design* (online MD), which generalizes the theory of computational mechanism design to apply to dynamic problems. Decisions must be made dynamically and without knowledge of future agent types or future decision possibilities, in the sense of online algorithms.

1.1.1 Example: Dynamic Auction with Expiring Items

Consider a dynamic auction model with discrete time periods $T = \{1, 2, \dots, \}$ and a single indivisible item to allocate in each time period. The type of an agent $i \in \{1, \dots, N\}$ is denoted $\theta_i = (a_i, d_i, w_i) \in T \times T \times \mathbb{R}_{>0}$. Agent i has arrival time a_i, departure time d_i, value w_i for an allocation of a single unit of the item in some period $t \in [a_i, d_i]$, and wants at most one unit. This information is all private to an agent. We refer to this as the *canonical expiring items environment*.

The arrival time has a special meaning: it is the first period in which information about the type of this agent can be made available to the auction. (We say "can be made available" because a self-interested agent may choose to delay its report.) Assume quasi-linear utility, with utility $w_i - p$ when the item is allocated in some $t \in [a_i, d_i]$ and payment p is collected from the agent. Consider the following naive generalization of the Vickrey auction to this dynamic environment:

Auction 1. A bid from an agent is a claim about its type, $\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \hat{w}_i)$, necessarily made in period $t = \hat{a}_i$. Then: in each period t, allocate the item to the highest unassigned bid, breaking ties at random. Collect payment equal to the second-highest unallocated bid in this round.

Example 1.1 Jane sells ice cream and can make one cone each hour. The ice cream melts if it is not sold. There are three buyers, with types $(1, 2, 100), (1, 2, 80)$ and $(2, 2, 60)$, indicating (arrival, departure, value). Buyers 1 and 2 are willing to buy an ice cream in either period 1 or 2 while buyer 3 will only buy an ice cream in period 2. In this example, if every buyer is truthful then buyer 1 wins in period 1 for 80, stops bidding, and buyer 2 wins in period 2 for 60. But buyer 1 can do better. For example, buyer 1 can report type $(1, 2, 61)$, so that buyer 2 wins in period 1 for 61, stops bidding, and then buyer 1 wins for 60 in period 2. Buyer 1 can also

report type $(2, 2, 80)$ and delay its bid until period 2, so that buyer 2 wins for 0 in period 1, stops bidding, and then buyer 1 wins for 60 in period 2.

In a static situation the Vickrey auction is (dominant-strategy) truthful because an agent does not affect the price it faces. But, in a sequential setting an agent can choose the auction in which it participates and thus choose the other agents against which it competes and, in turn, the price faced. In fact, if every agent was *impatient* (with $d_i = a_i$) then prices in future periods are irrelevant and the dominant strategy is to bid truthfully immediately upon arrival. Note also that buyer 1's manipulation relied on a suitable bid from buyer 3 in period 2 and will not always be useful. Nevertheless, this serves to demonstrate the failure of dominant strategy truthfulness.

1.1.2 The Challenge of Online MD

The dynamics of agent arrivals and departures, coupled perhaps with uncertainty about the set of feasible decisions in the future and in general about the state of the environment, makes the problem of online MD fundamentally different from that of standard (offline) MD. Important new considerations in online MD are:

(i) Decisions must be made without information about agent types not yet arrived, coupled perhaps with uncertainty about which decisions will be feasible in future periods.
(ii) Agents can misrepresent their arrival and departure time in addition to their valuation for sequences of decisions. Because of this agent strategies also have a temporal aspect.
(iii) Only limited misreports of type may be available, for instance it may be impossible for an agent to report an earlier arrival than its true arrival.

More generally, online MD can also model settings in which an agent's type is revealed *to itself* over time and with its ability to learn dependent on decisions made by the online mechanism; e.g., a bidder needs to receive a resource to understand its value for the resource.

There are two main frameworks in which to study the performance of online mechanisms. The first is *model-free* and adopts a worst-case analysis and is useful when a designer does not have good probabilistic information about future agent types or about feasible decisions in future periods. The second is *model-based* and adopts an average-case analysis. As a motivating

example, consider a search engine selling search terms to advertisers. This is a data rich environment and it is reasonable to believe that the seller can build an accurate model to predict the distribution on types of buyers, including the process governing arrival and departures.

1.1.3 Outline

In Section 1.2 we present a general model for online MD and introduce the concept of limited misreports. Given this, we define direct-revelation, online mechanisms together with appropriate notions of incentive compatibility. Section 1.3 provides a characterization of truthful online mechanisms in the restricted domain of *single-valued preferences* and gives detailed examples of truthful, dynamic auctions. These auctions are analyzed within the framework of worst-case, competitive analysis. Section 1.4 considers general preference domains, and defines a dynamic Vickrey-Clarke-Groves mechanism, that is efficient and applicable when a model is available and common knowledge to agents. Section 1.5 closes with open problems and future directions.

1.2 Dynamic Environments and Online MD

The basic setting assumes risk neutral agents with quasi-linear utility functions, such that an agent acts to maximize the expected difference between its value from a sequence of decisions and its total payment. Consider discrete time periods $T = \{1, 2, \dots, \}$, indexed by t and possibly infinite. A mechanism makes (and enforces) a sequence of decisions $k = (k^1, k^2, \dots,) \in \mathcal{O}$, with decision k^t made in period t. Let $k^{[t_1,t_2]} = (k^{t_1}, \dots, k^{t_2})$. The decisions made by a mechanism can depend on messages, such as bids, received from agents as well as uncertain events that occur in the environment. For example, in sponsored search the realized supply of search terms determines the feasible allocation of user attention to advertisers.

An agent's type, $\theta_i = (a_i, d_i, w_i) \in \Theta_i$, where Θ_i is the set of possible types for agent i, defines a valuation function $v_i(\theta_i, k) \in \mathbb{R}$ on a sequence of decisions k and is private to an agent. Time periods $a_i, d_i \in T$ denote an agent's arrival and departure period and $v_i(\theta_i, k) = v_i(\theta_i, k^{[a_i,d_i]})$, i.e. an agent's value is invariant to decisions outside of its arrival-departure window. In addition to restricting the scope of decisions that influence an agent's value, the arrival period models the first period at which the agent is able to report its type to the mechanism.

The valuation component $w_i \in \mathbb{W}_i$ of an agent's type, where $\mathbb{W}_i$ denotes

the set of possible valuations, parameterizes the agent's valuation function and can be more expressive than a single real number. For example, in an online combinatorial auction this needs to convey enough information to define substitutes ("I want item A or item B but not both") or complements ("I only want item A if I also get item B") preferences. Nor does the valuation need to be constant across all periods, for instance an agent could discount its future value in future periods $t > a_i$ by discount factor γ^{t-a_i} for $\gamma \in (0, 1)$.

1.2.1 Direct-Revelation Mechanisms

The family of direct-revelation, online mechanisms restrict the message that an agent can send to the mechanism to a single, direct claim about its type. For the most part we consider "closed" mechanisms so that an agent receives no feedback before reporting its type, and cannot condition its strategy on the report of another agent.

The *mechanism state*, $h^t \in H^t$, where H^t is the set of possible states in period t, captures all information relevant to the decision by the mechanism in that period. Let $\omega \in \Omega$ define the set of possible stochastic events that can occur in the environment, such as the realization of uncertain supply. This does not include the types of agents or any randomization within the mechanism itself. Write $\Omega = \Pi_{t \in T} \Omega^t$ and let $\omega^t \in \Omega^t$ denote the information about ω that is revealed in period t. Similarly, let θ^t denote the set of agent types reported in period t. Given this, it is convenient to define $h^t = (\theta^1, \ldots, \theta^t; \omega^1, \ldots, \omega^t; k^1, \ldots, k^{t-1})$. In practice the state of will be represented by a small, sufficient statistic of this information. The state space $H = \bigcup_t H^t$ may be finite, countably infinite, or continuous. This depends, in part, on whether agent types are discrete or continuous. Let $K(h^t)$ denote the set of all feasible decisions in the current time period, assumed finite for all h^t. Let $I(h^t)$ denote the set of active agents in state h^t, i.e. those agents for which $t \in [a_i, d_i]$.

Definition 1.2 (direct-revelation online mechanism) A direct-revelation online mechanism, $M = (\pi, x)$, restricts each agent to making a single claim about its type, and defines *decision policy* $\pi = \{\pi^t\}^{t \in T}$ and *payment policy*, $x = \{x^t\}^{t \in T}$, where decision $\pi^t(h^t) \in K(h^t)$ is made in state h^t and payment $x_i^t(h^t) \in \mathbb{R}$ is collected from each agent $i \in I(h^t)$.

Decision policy π may be stochastic. The payment policy may collect payments from an agent across multiple periods. For notational convenience,

we let $\pi(\theta, \omega) = (k^1, k^2, \ldots)$ denote the sequence of decisions, and $p_i(\theta, \omega) \in \mathbb{R}$ denote the total payment collected from agent i, given type profile θ and a realization of uncertain events $\omega \in \Omega$.

Example 1.3 Consider the canonical expiring items environment. The state h^t can be defined as a list of reported agent types that are present in period t, indicating whether each agent is already allocated or not. Decision $k \in K(h^t)$ decides whether to allocate the item in the current period to some agent that is present and unallocated.

Limited misreports constrain the strategy space available to agents in direct-revelation, online mechanisms:

Definition 1.4 (limited misreports) Let $C(\theta_i) \subseteq \Theta_i$ for $\theta_i \in \Theta_i$ denote the set of available misreports to an agent with true type θ_i.

In the standard model adopted in offline MD, it is typical to assume $C(\theta_i) = \Theta_i$. We shall assume *no early-arrival* misreports, with $C(\theta_i) = \{\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \hat{w}_i) : a_i \leq \hat{a}_i \leq \hat{d}_i, \hat{w}_i \in \mathbb{W}_i\}$; i.e. agent i cannot report an earlier arrival because it does not know its type (or about the mechanism) until a_i. Sometimes, we shall also assume *no late-departure* misreports, which together with no early arrivals provides $C(\theta_i) = \{\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \hat{w}_i) : a_i \leq \hat{a}_i \leq \hat{d}_i \leq d_i, \hat{w}_i \in \mathbb{W}_i\}$. For example, we could argue that it is not credible to claim to have value for a ticket for a last minute Broadway show after 5pm because the auctioneer knows that it takes at least 2 hours to get to the theater and the show starts at 7pm.

We restrict attention to mechanisms that are either dominant-strategy or Bayes-Nash incentive compatible. Let $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots)$, $\Theta_{-i} = \Pi_{j \neq i} \Theta_j$ and $C(\theta_{-i}) = \Pi_{j \neq i} C(\theta_j)$ and consider misreports $\theta_i \in C(\theta_i)$.

Definition 1.5 (DSIC) Online mechanism $M = (\pi, x)$ is *dominant-strategy incentive-compatible* (DSIC) given limited misreports C if

$$v_i(\theta_i, \pi(\theta_i, \theta'_{-i}, \omega)) - p_i(\theta_i, \theta'_{-i}, \omega) \geq v_i(\theta_i, \pi(\hat{\theta}_i, \theta'_{-i}, \omega)) - p_i(\hat{\theta}_i, \theta'_{-i}, \omega),$$

for all $\hat{\theta}_i \in C(\theta_i)$, all θ_i, all $\theta'_{-i} \in C(\theta_{-i})$, all $\theta_{-i} \in \Theta_{-i}$, all $\omega \in \Omega$.

It will be convenient to also adopt the terminology *truthful* in place of DSIC. The concept of DSIC is very strong: it says than an agent maximizes its utility by reporting its true type whatever the reports of other agents and for all stochastic events ω. When the decision policy is stochastic then DSIC requires that the *expected utility* is maximized from a truthful report, whatever the reports of other agents and for all stochastic events ω.

A randomized mechanism (i.e., one with a stochastic policy) is said to satisfy ***strong-truthfulness*** when truthful reporting is a dominant strategy for all random coin flips by the mechanism, and for all external stochastic events ω.

For Bayes-Nash incentive compatibility (BNIC), assume in addition that all agents know the correct probabilistic model of the distribution on types and uncertain events, and that this is common knowledge.

Definition 1.6 (BNIC) Online mechanism $M = (\pi, x)$ is *Bayes-Nash incentive-compatible* (BNIC) given limited misreports C if

$$\mathbb{E}\big\{v_i(\theta_i, \pi(\theta_i, \theta_{-i}, \omega)) - p_i(\theta_i, \theta_{-i}, \omega)\big\} \geq \mathbb{E}\big\{v_i(\theta_i, \pi(\hat{\theta}_i, \theta_{-i}, \omega)) - p_i(\hat{\theta}_i, \theta_{-i}, \omega)\big\},$$

for all $\hat{\theta}_i \in C(\theta_i)$, all θ_i, where the expectation is taken with respect to the distribution on types θ_{-i}, and stochastic events ω, and any randomization within the policy.

BNIC is a weaker solution concept than DSIC because it requires only that truth revelation is a best-response when other agents are also truthful, and in expectation given the distribution on agent types and on stochastic events in the environment.

1.2.2 Remark: The Revelation Principle

Commonly held intuition from offline MD might suggest that focusing on the class of incentive compatible, direct-revelation online mechanisms is without loss of generality. However, if agents are unable to send messages to a mechanism in periods $t \notin [a_i, d_i]$ then this is not true:

Example 1.7 (failure of the revelation principle) Consider the model with no early-arrival misreports but allow for late-departure misreports. Consider two time periods $T = \{1, 2\}$, a single unit of an indivisible item to allocate in either period and an environment with a single agent. Denote the type of the agent (a_i, d_i, w_i) with $w_i > 0$ to denote its value for the item if allocated in period $t \in [a_i, d_i]$. Suppose possible types are $(1, 1, 1)$ or $(1, 2, 1)$. Consider an indirect mechanism that allows an agent to send one of messages $\{1, 2\}$ in period 1 and $\{1\}$ in period 2. Let ϕ denote a null message. Consider decision policy: $\pi^1(1) = 0, \pi^1(2) = 1, \pi^2(1, z) = \pi^2(2, z) = 0$, for $z \in \{1, \phi\}$, writing the state as the sequence of messages received and decision $k^t \in \{0, 1\}$ to indicate whether or not the agent is allocated in period $t \in \{1, 2\}$. Consider payment policy: $x^1(1) = x^2(1, \phi) = x^2(1, 1) = 0$, $x^1(2) = 3, x^2(2, 1) = -2.01, x^2(2, \phi) = 0$. Type $(1, 1, 1)$ will report message

1 in period 1 because reporting message 2 is not useful and it cannot report messages (2,1). Type $(1, 2, 1)$ will report messages (2,1) and has no useful deviation. This policy cannot be implemented as a DSIC direct-revelation mechanism because type $(1, 2, 1)$ is allocated in period 1 for payment 0.99, and so type $(1, 1, 1)$ (which is unallocated if truthful) will want to report type $(1, 2, 1)$.

The revelation principle fails in this example because the indirect mechanism prevents the agent from claiming a later departure than its true departure. In fact, the revelation principle continues to hold when misreports are limited to no-late departures in addition to no-early arrivals. A form of the revelation principle can also be recovered by introducing simple "heartbeat" messages into a direct-revelation mechanism, whereby an agent still makes a single report about its type but must also send a non-informative heartbeat message in every period $t \in [\hat{a}_i, \hat{d}_i]$.† We leave the derivation of this "revelation principle plus heartbeat" result as an exercise.

With this in hand, and in keeping with the current literature on online mechanisms, we will focus on incentive-compatible, direct revelation online mechanisms in this chapter.

1.3 Single-Valued Online Domains

In this section we develop a methodology for the design of DSIC online mechanisms in the restricted domain of *single-valued* preferences. We identify the central role of monotonic decision policies in the design of truthful online mechanisms. The methodology is illustrated in the design of a dynamic auction for two environments: (a) allocating a sequence of expiring items, (b) allocating a single, indivisible item in some period while adapting to information about agent types. Both auctions are model-free and we use competitive analysis to study their efficiency and revenue properties. We close the section with remarks that seek to situate the study of truthful online mechanisms in the context of the wider mechanism design literature.

1.3.1 Truthfulness for Single-Valued Preference Domains

An agent with single-valued preferences has the same value, r_i, whenever any of a set of interesting decisions is made in some period $t \in [a_i, d_i]$, and has value for at most one such decision. For example, in the single-

† Thanks to Bobby Kleinberg for suggesting this interpretation.

item allocation problems considered earlier an agent's interesting set was all decisions that allocate an item to the agent.

Let $\mathcal{L}_i = \{L_1, \ldots, L_m\}$ describe a language for defining interesting sets for agent i, where $L \subseteq K = \bigcup_h K(h)$, for any $L \in \mathcal{L}_i$, and defines a subset of single-period decisions. Define partial order $\succeq_L$ on $\mathcal{L}_i$. The valuation component $w_i \in \mathbb{W}_i$ of an agent's type, $\theta_i = (a_i, d_i, w_i)$, defines $w_i = (r_i, L_i)$ with $\mathbb{W}_i = \mathbb{R} \times \mathcal{L}_i$. This picks out the interesting set and defines the value on decisions in that set:

Definition 1.8 (single-valued) A single-valued online domain is one where each agent i has a type $\theta_i = (a_i, d_i, (r_i, L_i))$, with reward $r_i \in \mathbb{R}$ and interesting set $L_i \in \mathcal{L}_i$, where type θ_i defines valuation:

$$v_i(\theta_i, k) = \begin{cases} r_i & \text{, if } k^t \in \{L : L \succeq_L L_i, L \in \mathcal{L}_i\} \text{ for some } t \in [a_i, d_i] \\ 0 & \text{, otherwise,} \end{cases} \quad (1.1)$$

To keep things simple we assume that *the set of interesting decisions is known by the mechanism* and thus the private information is restricted to arrival, departure and its value for a decision. We comment on how to relax this assumption at the end of the section. Given the *known interesting-set* assumption, define a partial-order $\preceq_\theta$ on types:

$$\theta_1 \preceq_\theta \theta_2 \equiv (a_1 \geq a_2) \wedge (d_1 \leq d_2) \wedge (r_1 \leq r_2) \wedge (L_1 = L_2) \quad (1.2)$$

This will be sufficient because we will not need to reason about misreports of interesting set L_i. Consider the following example:

Example 1.9 (single-valued combinatorial auction) Multiple units of indivisible, heterogeneous items G, are in uncertain supply and cannot be stored from one period to the next. Consider single-valued preferences, where interesting set $L_i \in \mathcal{L}_i$ has an associated bundle $S(L_i) \subseteq G$, and characterizes all single-period decisions that allocate agent i bundle $S(L_i)$, irrespective of the allocation to other agents. Define partial order $L_1 \succeq_L L_2 \equiv S(L_1) \supseteq S(L_2)$ for all $L_1, L_2 \in \mathcal{L}_i$. Agent i with type $\theta_i = (a_i, d_i, (r_i, L_i))$ has value r_i when decision k^t allocates a bundle containing at least $S(L_i)$ items to the agent in some period $t \in [a_i, d_i]$.

The subsequent analysis is developed for *deterministic* policies. We adopt shorthand $\pi_i(\theta_i, \theta_{-i}, \omega) \in \{0, 1\}$ to indicate whether policy π makes an interesting decision for agent i with type θ_i in some period $t \in [a_i, d_i]$, fixing type profile θ_{-i} and stochastic (external) events $\omega \in \Omega$. Since we are often considering auction domains, we may also refer to an interesting decision for an agent as an *allocation* to the agent. The analysis immediately applies to

the case of stochastic policies when coupled with strong-truthfulness.‡ We elaborate more on stochastic policies at the end of the section.

Definition 1.10 (critical value) The critical-value for agent i given type $\theta_i = (a_i, d_i, (r_i, L_i))$ and deterministic policy π in a single-valued domain, is defined as:

$$v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega) = \begin{cases} \min r'_i \text{ s.t. } \pi_i(\theta'_i, \theta_{-i}, \omega) = 1 \text{ for } \theta'_i = (a_i, d_i, (r'_i, L_i)) \\ \infty, \text{ if no such } r'_i \text{ exists,} \end{cases} \tag{1.3}$$

where types θ_{-i} and stochastic events $\omega \in \Omega$ are fixed.

Definition 1.11 (monotonic) Deterministic policy π is monotonic if $(\pi_i(\theta_i, \theta_{-i}, \omega) = 1) \wedge (r_i > v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)) \Rightarrow \pi_i(\theta'_i, \theta_{-i}, \omega) = 1$ for all $\theta'_i \succ_\theta \theta_i$, for all θ_{-i}, all $\omega \in \Omega$.

The "strict profit" condition, $r_i > v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$, is added to prevent weak indifference when $\theta'_i \succ_\theta \theta_i$ and $r'_i = r_i$, and is redundant when $r'_i > r_i$. Say that an arrival-departure interval $[a'_i, d'_i]$ is *tighter* than $[a_i, d_i]$ if $a'_i \geq a_i$ and $d'_i \leq d_i$, and weaker otherwise.

Lemma 1.12 *The critical value to agent i is independent of reward r_i and (weakly) monotonically increasing in tighter arrival-departure intervals, given a deterministic, monotonic policy.*

Proof Fix some θ_{-i}, $\omega \in \Omega$. Assume for contradiction that $\theta'_i \preceq_\theta \theta_i$, so that $a'_i \geq a_i$ and $d'_i \leq d_i$, but $v^c_{(a'_i,d'_i,L_i)}(\theta_{-i}, \omega) < v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$. Modify the reward of type $\theta'_i = (a'_i, d'_i, (r'_i, L_i))$ such that $r'_i := v^c_{(a'_i,d'_i,L_i)}(\theta_{-i}, \omega)$ and modify the reward of type $\theta_i = (a_i, d_i, (r_i, L_i))$ such that $r_i := v^c_{(a'_i,d'_i,L_i)}(\theta_{-i}, \omega)$. Now, we still have $\theta'_i \preceq_\theta \theta_i$, but $\pi_i(\theta'_i, \theta_{-i}, \omega) = 1$ while $\pi_i(\theta_i, \theta_{-i}, \omega) = 0$ and a contradiction with monotonicity. □

Theorem 1.13 *A monotonic, deterministic decision policy π can be truthfully implemented in a domain with (known interesting set) single valued preferences, and no early-arrival or late-departure misreports.*

Proof Define payment policy $x^t_i(h^t) = 0$ for all $t \neq \hat{d}_i$, and with

$$x^t_i(h^t) = \begin{cases} v^c_{(\hat{a}_i,\hat{d}_i,L_i)}(\hat{\theta}_{-i}, \omega) & \text{, if } \pi_i(\hat{\theta}_i, \hat{\theta}_{-i}, \omega) = 1 \\ 0 & \text{, otherwise} \end{cases} \tag{1.4}$$

‡ It is convenient for this purpose to consider the random coin flips of a policy as included in stochastic events ω so that no notational changes are required.

when $t = \hat{d}_i$. This critical-value payment is collected upon departure. Fix θ_{-i}, $\theta_i = (a_i, d_i, (r_i, L_i))$, and $\omega \in \Omega$, assume agent i is truthful and proceed by case analysis. (a) If agent i is not allocated, $v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega) > r_i$ and to be allocated the agent must report some $\theta'_i \succ_\theta \theta_i$ which it can only do with a report $\theta'_i = (a_i, d_i, (r'_i, L_i))$, and $r'_i > r_i$, by limited misreports. But since the critical value is greater than its true value r_i it will have negative utility if it wins for r'_i. (b) If agent i is allocated, its utility is non-negative since $v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega) \leq r_i$ and it does not want to report a type for which it would not be allocated. Consider any report $\theta'_i \in C(\theta_i)$ for which the agent continues to be allocated. But, the critical value for θ'_i is (weakly) greater than for θ_i since it is independent of the reported reward r'_i and weakly increasing for an alternate arrival-departure interval since it must be tighter by limited misreports, and then by appeal to Lemma 1.12. □

We turn now to identifying *necessary* conditions for truthfulness. An online mechanism satisfies *individual rationality* (IR) when every agent has non-negative utility in equilibrium. This is required when agents cannot be forced to participate in the mechanism.

Lemma 1.14 (critical payment) *In a single-valued preference domain, any truthful online mechanism that is defined for a deterministic decision policy and satisfies IR must collect a payment equal to the critical value from each allocated agent.*

Proof Fix θ_{-i} and $\omega \in \Omega$. Payment $p_i(\theta_i, \theta_{-i}, \omega)$, made by agent i contingent on successful allocation, cannot depend on reward r_i because if $p_i(\theta_i, \theta_{-i}, \omega) < p_i(\theta'_i, \theta_{-i}, \omega)$ for $\theta_i = (a_i, d_i, (r_i, L_i))$ and $\theta'_i = (a_i, d_i, (r'_i, L_i))$ and $r'_i \neq r_i$ and $\min(r'_i, r_i) \geq v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$ then an agent with type θ'_i should report type θ_i. Fix type θ_i such that $\pi_i(\theta_i, \theta_{-i}, \omega) = 1$. Now, if $p_i(\theta_i, \theta_{-i}, \omega) < v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$ then an agent with type $\theta'_i = (a_i, d_i, (r'_i, L_i))$ and $p_i(\theta_i, \theta_{-i}, \omega) < r'_i < v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$ should report θ_i. This is possible even with negative payment $p_i(\theta_i, \theta_{-i}, \omega)$ as long as rewards can also be negative. On the other hand, if $v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega) < p_i(\theta_i, \theta_{-i}, \omega)$ then the mechanism fails IR for an agent with type $\theta'_i = (a_i, d_i, (r'_i, L_i))$ and $v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega) < r'_i < p_i(\theta_i, \theta_{-i}, \omega)$. □

Say that a domain satisfies *reasonable misreporting* when an agent with type θ_i has available *at least* misreports $\theta'_i \in C(\theta_i)$ with $a'_i \geq a_i$, $d'_i \leq d_i$ and any reward r'_i.

Theorem 1.15 *In a (known interesting set) single-valued preference domain with reasonable misreporting, then any deterministic policy π that can be truthfully implemented in an IR mechanism that does not pay unallocated agents must be monotonic.*

Proof Fix θ_{-i}, $\omega \in \Omega$. Assume, for contradiction, that $\theta_i \prec_\theta \theta_i'$ with $\theta_i = (a_i, d_i, (r_i, L_i))$ and $\theta_i' = (a_i', d_i', (r_i', L_i))$, but $\pi_i(\theta_i, \theta_{-i}, \omega) = 1$, value $r_i > v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$ and $\pi_i(\theta_i', \theta_{-i}, \omega) = 0$. Consider type $\theta_i'' = (a_i, d_i, (v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega), L_i))$. We must have $p_i(\theta_i, \theta_{-i}, \omega) = p_i(\theta_i'', \theta_{-i}, \omega) \leq v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$ where the equality is by truthfulness and the inequality is by IR. Thus, agent i with type θ_i must have strictly positive utility in the mechanism. On the other hand, the agent with type $\theta_i' \succ_\theta \theta_i$ is not allocated, makes non-negative payment and has (weakly) negative utility. But, an agent with type θ_i' can report θ_i, which presents a contradiction with truthfulness. □

The restriction that losing agents do not receive a payment plays an important role. To see this, consider a domain with no late-departure misreports, fix θ_{-i}, and consider a single-item valuation with possible types $\Theta_i = \{(1, 1, \$10), (1, 2, \$10)\}$. Policy $\pi_i((1, 1, \$10), \theta_{-i}) = 1$ and $\pi_i((1, 2, \$10), \theta_{-i}) = 0$ is non-monotonic, but can be truthfully implemented with payments $p_i((1, 1, \$10), \theta_{-i}) = 8$ and $p_i((1, 2, \$10), \theta_{-i}) = -100$.

Monotonic-Late. The sufficiency result can be generalized to a domain with arbitrary misreports of departure. For a particular θ_{-i}, $\omega \in \Omega$ and type $\theta_i = (a_i, d_i, (r_i, L_i))$, define the *critical departure*, $d^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$, as the earliest departure $d_i' \leq d_i$ for which $v^c_{(a_i,d_i',L_i)}(\theta_{-i}, \omega) = v^c_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$. This is the earliest departure time that agent i could have reported without increasing the critical value. Given this we say that policy π is *monotonic-late* if it is monotonic and if no interesting decision is made for agent i before its critical departure period. A monotonic-late, deterministic decision policy π can be truthfully implemented in a domain with no early-arrival misreports but arbitrary misreports of departure. Moreover, this requirement of monotonic-late is necessary for truthfulness in this environment.

1.3.2 Example: A Dynamic Auction with Expiring Items

For our first detailed example we revisit the problem of selling an expiring item, such as ice cream, time on a shared computer, or network resources, to dynamically arriving buyers. This is the canonical expiring items environ-

ment. Assume for notational convenience that the time horizon is finite. We design a strongly-truthful online auction that includes random tie-breaking and satisfies monotonicity however ties are broken.

We assume no early-arrival and no late-departure misreports. The no late-departure assumption can be readily motivated in physical environments. For ice cream, think about a tour group that will be leaving at a designated time so that it is not credible to claim a willingness to wait for an ice cream beyond that period. For network resources, such as an auction for access to WiFi bandwidth in a coffee house, think about requiring a user to be present for the entire period of time reported to the mechanism. A technical argument for why we need this assumption is also provided below.

The assumption of no late-departures can be dispensed with, while still retaining truthfulness, in environments in which it is possible to schedule a resource in some period before an agent's reported departure, but withhold access to the benefit from the use of the resource until the reported departure; e.g., in grid computing, jobs can run on the machine but the result then held until reported departure.

Competitive Analysis. We perform a worst-case analysis and consider the performance of the mechanism given a sequence of types that are generated by an "adversary" whose task it is to make the performance as bad as possible. Of particular relevance is the method of *competitive analysis*, typically adopted in the study of online algorithms. The following question is asked: *how effectively does the performance of the online mechanism "compete" with that of an offline mechanism that is given complete information about the future arrival of agent types?* Again, this is asked in the worst-case, for a suitably adversarially-defined input.

Competitive analysis is most easily justified when the designer does not have a good model of the environment. As a motivating example, consider selling a completely new product or service, for which it is not possible to conduct market research to get a good model of demand. Competitive analysis can also lead to mechanisms that enjoy good average-case performance in practice, provide insight into how to design robust mechanisms, and produce useful "lower-bound" analysis. A lower-bound for a problem makes a statement about the best possible performance that can be achieved by *any* mechanism. Online mechanisms are of special interest when their realized performance matches the lower bound.

In performing competitive analysis, one needs to define: an optimality criterion; a model of the power of the adversary is selecting worst-case inputs; and an offline benchmark, defined with perfect information about the future.

We are interested in the efficiency of a dynamic auction for expiring items and adopt as our optimality criterion the value of the best possible offline allocation. This can be computed as:

$$V^*(\theta) = \max_{x,y} \sum_{i=1}^{N} y_i w_i \tag{1.5}$$

$$\text{s.t.} \quad \sum_{t=a_i}^{d_i} x_{it} \geq y_i, \quad \forall i \in \{1, \ldots, N\} \tag{1.6}$$

$$\sum_{i:t\in[a_i,d_i]} x_{it} \leq 1, \quad \forall t \in T, \tag{1.7}$$

where $y_i \in \{0, 1\}$ indicates whether bid i is allocated and $x_{it} \in \{0, 1\}$ indicates the period in which it is allocated.§ For our adversarial model, we consider a powerful adversary that is able to pick arbitrary agent types, including both the value, arrival and departure of agents. Let $z \in \mathcal{Z}$ denote the set of inputs available to the adversary and θ_z the corresponding type profile. An online mechanism is *c-competitive for efficiency* if:

$$\min_{z\in\mathcal{Z}} \mathbb{E}\left\{ \frac{\text{Val}(\pi(\theta_z))}{V^*(\theta_z)} \right\} \geq \frac{1}{c}, \tag{1.8}$$

for some constant $c \geq 1$. Such a mechanism is guaranteed to achieve within fraction $\frac{1}{c}$ of the value of the optimal offline algorithm, whatever the input sequence. The expectation allows for stochastic policies and can also allow for the use of randomization in defining the power of the adversary (we will see this in the next section). Competitive ratio c is referred to as an *upper-bound* on the online performance of the mechanism.

Now consider the following modification to Auction 1:

Auction 2. A bid from an agent is a claim about its type, $\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \hat{w}_i)$, necessarily made in period $t = \hat{a}_i$.

(i) In each period, t, allocate the item to the highest unassigned bid, breaking ties at random.
(ii) Every allocated agent pays its critical-value payment, collected upon its reported departure.

The auction is the same as Auction 1 except for the payment rule, which now charges the critical value rather than the second price in the period in which an agent wins. We refer to this as a "greedy auction" because the

§ Note that the integer program allows the possibility of allocating more than one item to a winning bid but that this does not change the value of the objective and is not useful.

decision policy myopically maximizes value in each period. When every bidder is impatient then the auction reduces to a sequence of Vickrey auctions (i.e. Auction 1.)

Example 1.16 Consider the earlier example, with three agents and types $\theta_1 = (1, 2, 100), \theta_2 = (1, 2, 80)$ and $\theta_3 = (2, 2, 60)$ and one item to sell in each period. Suppose all three agents bid truthfully. The greedy allocation rule sells to agent 1 in period 1 and then agent 2 in period 2. Agent 1's payment is 60 because this is the critical value for arrival-departure $(1, 2)$ given the bids of other agents. (A bid of just above 60 would allow the agent to win, albeit in period 2 instead of period 1.) Agent 2's payment is also 60.

Theorem 1.17 *Auction 2 is strongly-truthful and 2-competitive for efficiency in the expiring-items environment with no early arrival and no late departure misreports.*

Proof Suppose that random tie-breaking is invariant to reported arrival and departure. The auction is strongly truthful because the allocation function is monotone: if agent i wins in some period $t \in [a_i, d_i]$ then it continues to win either earlier or in the same period for $w_i' > w_i$, and for $a_i' < a_i$ or $d_i' > d_i$. For competitiveness, consider a set of types θ and establish that the greedy online allocation rule is 2-competitive by a "charging argument". For any agent i that is allocated offline but not online, charge its value to the online agent that was allocated in period t in which agent i is allocated offline. Since agent i is not allocated online it is present in period t, and the greedy rule allocates to another agent in that period with at least as much value as agent i. For any agent i that is allocated offline and also online, charge its value to itself in the online solution. Each agent that is allocated in the online solution is charged at most twice, and in all cases for a value less than or equal to its own value. Therefore the optimal offline value $V^*(\theta)$ is at most twice the value of the greedy solution. □

There is actually a 1.618-competitive online algorithm for this problem but it is not monotonic and cannot be implemented truthfully. In fact, there is a matching lower bound for the problem of achieving efficiency and truthfulness:

Theorem 1.18 *No truthful, IR and deterministic online auction can obtain a $(2-\epsilon)$-approximation for efficiency in the expiring items environment with no early-arrival and no late-departure misreports, for any constant $\epsilon > 0$.*

Proof Fix $\epsilon > 0$, consider $T = \{1, 2\}$ and construct the following three scenarios: (i) Consider agents $\theta_1 = (1, 1, q(1+\delta)), \theta_2 = (1, 2, q)$ and choose $0 < \delta < \frac{\epsilon}{1-\epsilon}$ so that $\frac{q(1+\delta)}{q(2+\delta)} < \frac{1}{2-\epsilon}$ and the auction must allocate to both agents to be $(2-\epsilon)$-competitive. Let $q \geq v^c_{(1,1)}(\theta_2)$ (dropping dependence on ω because there are no stochastic events to consider), so that agent 1 must have strictly positive utility since the price is independent of reported value (for truthfulness) and less than or equal to $v^c_{(1,1)}(\theta_{-1})$ for IR. (ii) As in (i) except $\theta_1 \to \theta'_1 = (1, 2, q(1+\delta))$ and a new type $\theta_3 = (2, 2, \infty)$ is introduced. Agent 1 must be allocated else it can report type θ_1. Moreover, agent 1 must be allocated in period 1 because otherwise the mechanism cannot compete when θ_3 arrives. Agent 2 is not allocated. (iii) As in (i) except $\theta_1 \to \theta'_1 = (1, 2, q(1+\delta))$ and $\theta_2 \to \theta'_2 = (1, 1, q)$. The auction must allocate to both agents to be $(2-\epsilon)$-competitive. Further assume that $q > v^c_{(1,1)}(\theta'_1)$, which is without loss of generality because if $q = v^c_{(1,1)}(\theta'_1)$ then we can repeat the analysis with $q' = \alpha q$ for $\alpha > 1$ replacing q throughout. But now agent 2 with type θ'_2 has strictly positive utility since its payment is no greater than its critical value and the auction is not truthful in scenario (ii) because agent 2 can benefit by deviating and reporting θ'_2. □

The following provides a technical justification for why the no late-departure misreports assumption is required in this environment:

Theorem 1.19 *No truthful, IR and deterministic online auction can obtain a constant approximation ratio for efficiency in the expiring items environment with no early-arrival misreports but arbitrary misreports of departure.*

Proof Consider M periods. Fix θ_{-i}. Fix $v^c_{(1,1)}(\theta_{-i}) < \infty$ (dropping dependence on ω because there are no stochastic events to consider). First show that any agent with type $\theta_i = (1, M, w_i)$ for $w_i > v^c_{(1,M)}(\theta_{-i})$ must be allocated in period 1. For this, first show that $v^c_{(1,M)}(\theta_{-i}) = v^c_{(1,1)}(\theta_{-i})$. Construct $\theta'_i = (1, M, w'_i)$ with $w'_i = v^c_{(1,1)} + \epsilon$, some $\epsilon > 0$. By truthfulness and thus monotonicity we have $v^c_{(1,M)}(\theta_{-i}) \leq v^c_{(1,1)}(\theta_{-i})$ and agent i must be allocated. Moreover, it must be allocated in period 1 else an adversary can generate $M-1$ bids $\{(t, t, \beta^{t-1})\}$ for large $\beta > 0$ and $t \in \{2, \ldots, M\}$, all of which must be accepted for the auction to be constant competitive. But in this case the agent should deviate and report $(1, 1, w'_i)$, and be allocated in period 1 with payment $v^c_{(1,1)}(\theta_{-i}) < w'_i$ and have positive utility. Since type $(1, M, w'_i)$ is allocated in period 1 we must have $v^c_{(1,M)}(\theta_{-i}) = v^c_{(1,1)}(\theta_{-i})$ by truthfulness and the critical-payment lemma else type $(1, 1, w'_i)$ can deviate and report $(1, M, w'_i)$ and do better. Consider again type $(1, M, w_i)$,

we now have $w_i > v^c_{(1,M)}(\theta_{-i}) \Rightarrow w_i > v^c_{(1,1)}(\theta_{-i})$ and the agent must be allocated in period 1. To finish the proof, now construct type profile $\theta = \{(1, M, q_1), \ldots, (1, M, q_M)\}$ with $q_1, \ldots, q_m$ unique values drawn from $[q, q+\delta]$ for some $q > 0$ and $\delta > 0$. For any i, we must have $v^c_{(1,1)}(\theta_{-i}) < \infty$ else the mechanism is not competitive because the adversary could replace type i with $\theta'_i = (1, 1, w''_i)$ and some arbitrarily large w''_i. We can also assume $q_i \geq v^c_{(1,M)}(\theta_{-i}) \Rightarrow q_i > v^c_{(1,M)}(\theta_{-i})$, which can always be achieved by a slight upwards perturbation of any value $q_i = v^c_{(1,M)}(\theta_{-i})$. Finally, the online mechanism can allocate at most one of these bids since any bid allocated must be allocated in period 1 and can achieve value at most $q + \delta$ while the efficient offline allocation has value $V^*(\theta) \geq Mq$. Thus, no constant approximation is possible because M can be selected to be arbitrarily large. □

1.3.3 Example: An Adaptive, Limited-Supply Auction

For our second detailed example, we consider an environment with a single, indivisible item to be allocated to one of N agents. Each agent's type is still denoted $\theta_i = (a_i, d_i, w_i) \in T \times T \times \mathbb{R}_{>0}$, with w_i denoting the agent's value for the item. This first into the known interesting-set model. We assume no early-arrival misreports but will allow arbitrary misreports of departure. Our goal is to define an adaptive auction with good revenue and efficiency properties, again in an adversarial setting.

We relate this dynamic auction problem to the classical *secretary problem*, a well studied problem in optimal stopping theory:

The Secretary Problem. An interviewer meets with each from a pool of N job applicants in turn. The total number of applicants is known. Each applicant has a quality and the interviewer learns, upon meeting, the relative rank of each applicant amongst those already interviewed and must make an irrevocable decision about whether or not to hire the applicant. The goal is to hire the best applicant. By the "random-ordering hypothesis", an adversary can choose an arbitrary set of N qualities but cannot control the assignment of quality to applicant, rather this is sampled uniformly at random and without replacement from the set. The online problem is to design a stopping rule that maximizes the probability of hiring the highest rank applicant, in the worst-case for all possible adversarially-selected inputs. Say that a *candidate* is the most qualified of all applicants seen so far. The *optimal* policy (i.e. the policy that maximizes the probability of selecting the best applicant, in the worst case) is to interview the first $t - 1$

applicants and then hire the next candidate (if any), where t is defined by:

$$\sum_{j=t+1}^{N} \frac{1}{j-1} \leq 1 < \sum_{j=t}^{N} \frac{1}{j-1}. \tag{1.9}$$

For instance, with $N = 10,000$ the optimal t is 3680, i.e. sample 3679 applicants and then accept the next candidate. As $N \to \infty$, the probability of hiring the best applicant approaches $1/e$, as does the ratio t/N, and the optimal policy in this big N limit is to sample the first $\lfloor N/e \rfloor$ applicants and then immediately accept any subsequent candidate.

We can reinterpret the secretary problem in the auction context. Bidders, unlike the applicants in the classic model, are strategic and can misrepresent their value and time their entry into the market. Bidders also have both an entry and an exit time. We modify the adversarial model in the secretary problem while retaining the random-ordering hypothesis: an adversary picks a set of values and a set of arrival-departure intervals and agent types are then defined by sampling uniformly at random and without replacement from each set. By an averaging argument, our results for randomly-ordered inputs imply the same (upper-bound) competitive-ratio analysis when the bids consist of i.i.d. samples from an unknown distribution.

In addition to efficiency, we can also consider revenue as an optimality criterion. The auction's revenue for type profile θ is defined as $\text{Rev}(p(\theta)) = \sum_i p_i(\theta)$, where notation $p_i(\theta)$ denotes the (expected) payment by agent i given type profile θ. Notation $\omega \in \Omega$ is suppressed because there are no external stochastic events in the problem. For an offline benchmark we consider the revenue from an offline Vickrey auction and define $R^*(\theta)$ as the second-highest value in type profile θ. An online mechanism is *c-competitive for revenue* if:

$$\min_{z \in \mathcal{Z}} \mathbb{E} \left\{ \frac{\text{Rev}(p(\theta_z))}{R^*(\theta_z)} \right\} \geq \frac{1}{c}, \tag{1.10}$$

where $z \in \mathcal{Z}$ is the set of inputs available to an adversary, in this case choosing the two sets described above, and the expectation here is taken with respect to the random choice of the sampling process that matches values with arrival-departure intervals.

The optimal policy for the secretary problem has a *learning phase* followed by an *accepting phase*. For a straw-man online auction interpretation, consider: *observe the first $\lfloor N/e \rfloor$ reports and then price at the maximal value received so far, and sell to the first agent to subsequently report a value*

greater than this price. Break ties at random. The following example shows that this fails to be truthful.

Example 1.20 Consider six agents, with types $\theta_i = (a_i, d_i, w_i)$ and $\theta_1 = (1, 7, 6)$, $\theta_2 = (3, 7, 2)$, $\theta_3 = (4, 8, 4)$, $\theta_4 = (6, 7, 8)$, and agents 5 and 6 arriving in later periods. The transition to the accepting phase occurs after $\lfloor 6/e \rfloor = 2$ bids. Agent 4 wins in period 6 and makes payment 6. If agent 1 reports $\theta_1' = (5, 7, 6)$ then it wins in period 5, for payment 4.

The auction is truthful when all agents are impatient ($a_i = d_i$) but is not monotonic with respect to arrival time (as the above example illustrates) and fails to be truthful in general. Consider instead the following simple variation:

Auction 3. A bid from an agent is a claim about its type, $\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \hat{w}_i)$, necessarily made in period $t = \hat{a}_i$.

(i) (Learning): In period τ in which the $\lfloor N/e \rfloor th$ bid is received let $p \geq q$ be the top two bid values received so far.
(ii) (Transition): If an agent bidding p is still present in period τ then sell to that agent (breaking ties at random) at price q.
(iii) (Accepting): Else, sell to the next agent to bid a price at least p (breaking ties at random), collecting payment p.

Theorem 1.21 *Auction 3 is strongly-truthful in the single-unit, limited supply environment with no early-arrival misreports.*

Proof Assume that the method used to break ties is independent of the reported departure time of an agent. Fix θ_{-i}. Monotonicity is established by case analysis on type θ_i: (a) If d_i is to the left of the transition the agent is not allocated and monotonicity trivially holds. (b) If $[a_i, d_i]$ spans the transition, agent i does not trigger the transition, and it wins with $w_i > q$ then there is no tie-breaking and the agent continues to win for an earlier arrival or later departure (because this changes nothing about the price it faces when the transition occurs), and continues to win with a higher value. (c) If arrival, a_i, is after the transition and agent i wins with $w_i > p$ (and perhaps winning a random selection over another agent j arriving in the same period also with $w_j > p$) then it continues to win with an earlier arrival (even one that occurs before the transition because its value will define p), with a later departure (because tie-breaking is invariant to reported departure) and with a higher value. (d) If the agent triggers the

transition and wins with $w_i > q$ then its value $w_i = p$, there was no tie to break, and the agent continues to win for an earlier arrival (although at some point the transition will be triggered by the next earliest agent to arrive), for a higher value, and is unaffected by a later departure. The payment is the critical value, namely q in case (b) and (d) and p in case (c). Moreover, the policy is monotonic-late: in case (b) the critical value is infinite for all departures before the transition but constant with respect to departure otherwise and the critical departure period is that of the transition; in cases (c) and (d) the critical value payment is independent of departure time and the critical departure period is equal to the arrival period. □

Example 1.22 Return to the earlier example with six agents and types $\theta_1 = (1, 7, 6)$, $\theta_2 = (3, 7, 2)$, $\theta_3 = (4, 8, 4)$, $\theta_4 = (6, 7, 8)$, with agents 5 and 6 arriving in later periods. The transition to the accepting phase occurs upon the arrival of agent 2. Then $p = 6, q = 2$ and agent 1 wins for 2. Consider instead that $\theta_1' = (1, 2, 6)$. The transition still occurs upon the arrival of agent 2 but now the item is sold in period 6 to agent 4 for a payment of 6. An agent with true type θ_1' does not want to report θ_1 because of the monotonic-late property: although it would win it would not be allocated until period 3, and this is after its true departure.

Theorem 1.23 *Auction 3 is $e+o(1)$-competitive for efficiency and $e^2+o(1)$-competitive for revenue in the single-unit, limited supply environment in the limit as $N \to \infty$.*

Proof Let $\tau = \lfloor N/e \rfloor$. For efficiency, our competitive ratio is at least as great as the probability of selling to the highest value agent. Conditioned on selling at the transition, the probability that we sell to the highest value agent is at least $\frac{\lfloor N/e \rfloor}{N} = 1/e - o(1)$. Conditioned on selling after the transition, the probability of this event is $1/e - o(1)$ according to the analysis of the classical secretary problem. For revenue, our competitive ratio is at least as great as the probability of selling to the highest value agent at a price equal to the second-highest bid. Conditioned on selling at the transition, the probability of this event is $(1/e)^2 - o(1)$ (i.e., the probability that both the highest and second-highest value agents arrive before period τ). Conditioned on selling after the transition, the probability of this event is $(1/e)(1 - 1/e) - o(1)$, i.e. the probability that the second-highest value agent arrives before τ and the highest value agent arrives after τ. The unconditional probability of selling to the highest value agent at the second-highest price is a weighted

average of the two conditional probabilities computed above, hence it is at least $(1/e)^2 - o(1)$. □

The random-ordering hypothesis has a critical role in this analysis: there is no constant competitive mechanism in this environment for the adversarial model adopted in our analysis of the expiring items environment.

For the secretary problem it is well known that no stopping rule can achieve asymptotic success probability better than $1/e$. The same lower bound can be established in our setting, even though the mechanism has richer feedback (i.e., it sees numbers not ranks) and even though an allocation to any bidder, and not just to the highest-rank bidder, contributes to the expected efficiency. The proof of this result is beyond the scope of this chapter.¶

1.3.4 Remarks

We end this section with some general remarks that mostly seek to place the study of online mechanisms in single-valued preference domains in the broader context of computational mechanism design.

Ex post IC. A mechanism is ex post IC if truth revelation is a best-response contingent on other agents being truthful, and whatever the types of other agents (and thus for all possible futures in the context of online MD). In offline mechanisms the solution concepts of ex post incentive compatible (EPIC) and DSIC are equivalent with private value types. This equivalence continues to hold for *closed* online mechanisms, that provide no feedback to an agent before it submits a bid. However, an online mechanism that provides feedback, for instance prices, or in an extreme case current standing bids, loses this property. The report of an agent can now be conditioned on the reports of earlier agents, and monotonicity provides EPIC but not necessarily DSIC. Consider again Auction 2 in the expiring items environment, with true types $\theta_1 = (1, 2, 100), \theta_2 = (1, 2, 80)$ and $\theta_3 = (2, 2, 60)$. If the bids are public then a possible (crazy) strategy of agent 3 is to condition its bid as possible: "bid (2,2,1000) if a bid of (1,2,100) is received or bid (2,2,60) otherwise." Agent 1 will now pay 60 if it bids truthfully, but would pay 60 with a bid of (1,2,90). Nevertheless, truthful bidding is a best-response when other agents bid truthfully.

¶ One shows that for any stopping rule there is some distribution that is hard in the sense that the second-highest value in the sequence is much less than the highest value with high probability. Given this, the expected efficiency ratio of the allocation is determined, to first order, by the probability of awarding the item to the highest bidder.

Simple price-based online auctions. One straightforward method to construct truthful online auctions for known-set, single-valued environments is to define an agent-independent *price schedule* $q_i^t(L, \theta_{-i}, \omega) \in \mathbb{R}$ to agent i for interesting decision set $L \in \mathcal{L}_i$, given stochastic events $\omega \in \Omega$, where $q_i^t(L, \theta_{-i}, \omega)$ defines the price for a decision in set L in period t. Given this, define payment $p_{(a_i,d_i,L_i)}(\theta_{-i}, \omega) = \min_{t \in [a_i,d_i]} q_i^t(L_i, \theta_{-i}, \omega)$ and let $t^*_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$ denote the first period $t \in [a_i, d_i]$ in which $q_i^t(L_i, \theta_{-i}, \omega) = p_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$. Then, decision policy π that allocates to agent i with type $\theta_i = (a_i, d_i, (r_i, L_i))$ if and only if $r_i \geq q_i^t(L_i, \theta_{-i}, \omega)$ in some $t \in [a_i, d_i]$, with the allocation period $t \geq t^*_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$, is monotonic-late and the associated critical-value payment is just $p_{(a_i,d_i,L_i)}(\theta_{-i}, \omega)$. Working with price schedules is quite natural in many domains but not completely general:

Example 1.24 Consider the canonical expiring items environment. Fix θ_{-i}, and consider a monotonic-late policy π with critical-value $v^c_{(1,2)}(\theta_{-i}) = 20, v^c_{(1,1)}(\theta_{-i}) = v^c_{(2,2)}(\theta_{-i}) = 30$ (dropping dependence on ω because there are no stochastic events to consider). This policy allocates to type $\theta_i = (1, 2, 25)$ in period 2 but not type $\theta'_i = (1, 1, 28)$ or $\theta'_i(2, 2, 28)$. No simple price schedule corresponds to this policy, because it would require $q_i^1(\theta_{-i}) > 28, q_i^2(\theta_{-i}) > 28$ but $\min(q_i^1(\theta_{-i}), q_i^2(\theta_{-i})) \leq 25$.

The role of limited misreports. Consider again the above example. The price on an allocation to agent i in period 2 depends on its report: if the agent's type is $\theta_i = (2, 2, w_i)$ then the price is 30 but if the agent's type is $\theta_i = (1, 2, w_i)$ then the price is 20. This is at odds with the principle of "agent-independent prices" that drives the standard analysis of truthful mechanisms. The example also fails *weak-monotonicity*, which is generally necessary for truthfulness.‖

What is going on? In both cases, the reason for this departure from the standard theory for truthful mechanism design is the existence of limited misreports. The auction would not be truthful with early-arrival misreports because an agent with type $(2, 2, 28)$ could usefully deviate and report $(1, 2, 28)$. For limited misreports $C(\theta_i) \subseteq \Theta_i$ that satisfy *transitivity* (which holds for the no-early arrival and no-late departure assumptions that are motivated in online MD), so that $\theta'_i \in C(\theta_i)$ and $\theta''_i \in C(\theta'_i)$ implies $\theta''_i \in C(\theta_i)$, the payment $\tilde{p}_i(k, \theta_i, \theta_{-i}, \omega)$ collected from

‖ A social choice function $f : \Theta \to \mathcal{O}$ satisfies weak-monotonicity if and only if for any $\theta_i \in \Theta_i$, agent i, and $\theta_{-i} \in \Theta_{-i}$, then $f(\theta_i, \theta_{-i}) = a$ and $f(\theta'_i, \theta_{-i}) = b$ implies that $v_i(b, \theta'_i) - v_i(b, \theta_i) \geq v_i(a, \theta'_i) - v_i(a, \theta_i)$. In the example, when agent i changes its type from $(1, 2, 25)$ to $(2, 2, 28)$ it increases its relative value for an allocation in period 2 over no allocation, but the decision policy switches away from allocating to the agent in period 2.

agent i conditioned on outcome $k \in \mathcal{O}$, must satisfy $\tilde{p}_i(k, \theta_i, \theta_{-i}, \omega) = \min\left\{\tilde{p}_i(k, \hat{\theta}_i, \theta_{-i}, \omega) : \hat{\theta}_i \in C(\theta_i), \pi(\hat{\theta}_i, \theta_{-i}, \omega) = k\right\}$, or ∞ if no such $\hat{\theta}_i$ exists, for all i, all $k \in \mathcal{O}$ and all $\omega \in \Omega$. Limited dependence on the reported type is possible as long as the price is independent across available misreports. For unlimited misreports we recover the standard requirement that prices are agent-independent. So, the temporal aspect of online MD is both a blessing and a curse: on one hand we can justify limited misreports and gain more flexibility in pricing and in the timing of allocations, on the other hand decisions must be made in ignorance about future types.

Relaxing the known interesting-set assumption. We assumed that the interesting set $L_i \in \mathcal{L}_i$ was known by the mechanism. Domains in which the interesting set is *private information* to an agent can be handled by making the following modifications to the framework:

(i) Require that agent i's domain of interesting sets $\mathcal{L}_i = \{L_1, \ldots, L_m\}$, defines *disjoint* sets, so that $L_1 \cap L_2 = \emptyset$ for all $L_1, L_2 \in \mathcal{L}_i$.
(ii) Require that a decision policy π is *minimal*, so that it never makes decision $k^t \in L$ for some $L \succ_L L_i$ in some period $t \in [a_i, d_i]$, given reported type $\theta_i = (a_i, d_i, (r_i, L_i))$.
(iii) Extend the partial-order, so that

$$\theta_1 \preceq_\theta \theta_2 \equiv (a_1 \geq a_2) \wedge (d_1 \leq d_2) \wedge (r_1 \leq r_2) \wedge (L_1 \succeq_L L_2), \quad (1.11)$$

and adopt this partial order in defining monotonicity.

Given these modifications the general methods developed above for the analysis of online mechanisms continue to hold. For instance, a monotonic, minimal and deterministic policy continues to be truthful when combined with critical-value payments, and monotonicity remains necessary for truthfulness amongst minimal, deterministic policies. This is left as an exercise.

The requirement that interesting sets are disjoint can significantly curtail the generality of preference domains that can be modeled. It is especially hard to model substitutes preferences, for instance indifference across a set of items. Suppose the items are fruit, with $G = \{apple, banana, pear, lime, lemon\}$. With known interesting sets, we can model an agent with a type that defines a value for receiving an item from any subset of the domain G. We must now assume there is some partition, for instance into $\{\{apple, pear\}, \{banana\}, \{lime, lemon\}\}$ so that the agent either has the same value for an apple or a pear and no value for anything else, or a value for a banana and no value for anything else, or a value for a lime and a lemon but no value for anything else.

Stochastic policies. Stochastic decision policies can be important, both algorithmically (many computational methods for online decision use a probabilistic model to sample possible state trajectories) and also to allow for tie breaking while retaining the anonymity of policies.

So far we have handled this by requiring *strong*-truthfulness. More generally, a stochastic mechanism is DSIC when truthful reporting maximizes expected utility for an agent (with the expectation defined with respect to randomization in the policy), and for all reports of other agents and all *external* stochastic events, $\omega \in \Omega$. To handle this, let $\pi_i(\theta_i, \theta_{-i}, \omega) \in [0,1]$ denote the probability that agent i receives an interesting decision ("is allocated"), given type θ_i, types θ_{-i} and (external) stochastic events ω. The appropriate generalization of monotonicity to this environment requires, for every $\theta_i = (a_i, d_i, (r_i, L_i))$, all θ_{-i}, all $\omega \in \Omega$, that

$$\pi_i((a_i, d_i, (r_i, L_i)), \theta_{-i}, \omega) \geq \pi_i((a_i, d_i, (r_i', L_i)), \theta_{-i}, \omega), \quad \forall r_i \geq r_i', \quad (1.12)$$

and

$$\int_{x=0}^{r_i} \pi_i((a_i, d_i, (x, L_i)), \theta_{-i}, \omega) \mathrm{d}x \geq \int_{x=0}^{r_i} \pi_i((a_i', d_i', (x, L_i)), \theta_{-i}, \omega) \mathrm{d}x, \quad (1.13)$$

for all $a_i' \geq a_i$, $d_i' \leq d_i$. The critical value payment becomes:

$$v^{\mathcal{C}}_{(a_i, d_i, (r_i, L_i))}(\theta_{-i}, \omega) = \pi_i(\theta, \omega) r_i - \int_{x=0}^{r_i} \pi_i((a_i, d_i, (x, L_i)), \theta_{-i}, \omega) \mathrm{d}x \quad (1.14)$$

These definitions of monotonicity and critical-value payment reduce to the earlier cases when the policy is deterministic.

Theorem 1.25 *A stochastic decision policy π can be implemented in a truthful, IR mechanism that does not pay unallocated agents in a domain with (known interesting set) single-valued preferences and no early-arrival or late-departure misreports if and only if the policy is monotonic according to (1.12) and (1.13).*

The payment collected from allocated agents is the critical-value payment. The following example illustrates a stochastic policy that satisfies this monotonicity requirement.

Example 1.26 Consider a domain with no early arrival and no late departure misreports, two time periods $T = \{1, 2\}$, fix θ_{-i}, and consider agent i with a single-item valuation and possible types $\Theta_i =$

$\{(1,1,w_i),(1,2,w_i),(2,2,w_i)\}$. For impatient type $(1,1,w_i)$, consider policy

$$\pi_i((1,1,w_i),\theta_{-i}) = \begin{cases} 0 & \text{, if } w_i \leq 8 \\ \frac{w_i-8}{2} & \text{, if } 8 < w_i \leq 10 \\ 1 & \text{, otherwise} \end{cases} \tag{1.15}$$

Solving for the critical value payment (1.14), we find:

$$v^c_{(1,1,w_i)}(\theta_{-i}) = \begin{cases} 0 & \text{, if } w_i \leq 8 \\ \frac{w_i^2}{4} - 16 & \text{, if } 8 < w_i \leq 10 \\ 9 & \text{, otherwise} \end{cases} \tag{1.16}$$

The policy and critical value payment is defined identically for type $(2,2,w_i)$. For patient type $(1,2,w_i)$, consider policy

$$\pi_i((1,2,w_i),\theta_{-i}) = \begin{cases} \frac{w_i}{20} & \text{, if } 0 \leq w_i \leq 10 \\ \frac{w_i-5}{10} & \text{, if } 10 < w_i \leq 15 \\ 1 & \text{, otherwise.} \end{cases} \tag{1.17}$$

and the critical value payment, from (1.14), is:

$$v^c_{(2,2,w_i)}(\theta_{-i}) = \begin{cases} \frac{w_i^2}{40} & \text{, if } 0 \leq w_i \leq 10 \\ \frac{w_i^2}{20} - \frac{5}{2} & \text{, if } 10 < w_i \leq 15 \\ 8.75 & \text{, otherwise.} \end{cases} \tag{1.18}$$

Notice that $\pi_i((1,1,10),\theta_{-i}) = 1$ and $\pi_i((1,2,10)) = 0.5$, contradicting more simplistic notions of monotonicity, but that truthfulness is retained because $v^c_{(1,1,10)}(\theta_{-i}) = 9$ while $v^c_{(1,2,10)}(\theta_{-i}) = 2.5$. Although type $(1,2,10)$ can misreport to $(1,1,10)$ and be allocated with certainty, it prefers to report $(1,2,10)$ because its expected utility is $(0.5)(10-2.5)+(0.5)(0) > (1.0)(10-9)$. We leave as an exercise to check that these policies satisfy monotonicity, with $\int_{x=0}^{w_i} \pi_i((1,2,x),\theta_{-i})\mathrm{d}x \geq \int_{x=0}^{w_i} \pi_i((1,1,x),\theta_{-i})$ for all w_i.

We make a final remark about stochastic policies. In an environment with a probabilistic model that is common knowledge, and that defines both a probability distribution for agent types and for stochastic events $\omega \in \Omega$, we can settle for a weaker monotonicity requirement in which (1.12) and (1.13) are satisfied in expectation, given the model. However, this provides BNIC but not DSIC since monotonicity may not hold out of equilibrium when other agents are not truthful, since the probabilistic model of agent types upon which monotonicity is predicated will then be incorrect.

1.4 Bayesian Implementation In General Online Domains

In this section we focus on Bayesian implementation of expected value-maximizing policies in environments in which the designer and every agent has a correct, probabilistic model for types and uncertain events, and this is common knowledge. We consider the goal of value-maximization and present a dynamic variation of the offline Vickrey-Clarke-Groves (VCG) mechanism. This will involve computing expected value maximizing sequential decision policies and raise a number of computational challenges. We will see that the dynamic VCG mechanism is BNIC rather than DSIC, with incentive-compatibility contingent on future on-equilibrium play by all participants.

1.4.1 A General Model

A Markov decision process (MDP) provides a useful formalism for defining online mechanisms in model-based environments with general agent preferences. An MDP model $(H, K, \mathcal{P}, R)$ is defined for a set of states H, feasible decisions $K(h)$ in each state, a *probabilistic transition function* $\mathcal{P}(h^{t+1}|h^t, k^t)$ on the next state given current state and decision (with $\sum_{h' \in H^{t+1}} \mathcal{P}(h'|h^t, k^t) = 1$) and a *reward function* $R(h^t, k^t) \in \mathbb{R}$ for decision k^t in state h^t. The Markov property requires that feasible decisions, transitions and rewards depend on previous states and actions only through the current state. It is achieved here, for example, by defining $h^t \in H^t = (\theta^1, \ldots, \theta^t; \omega^1, \ldots, \omega^t; k^1, \ldots, k^{t-1})$, so that the state captures the complete history of types, stochastic events, and decisions. In practice a short summarization of state h^t is often sufficient to retain the Markov property.

Given a social planner interested in maximizing total value, then define reward $R(h^t, k^t) = \sum_{i \in I(h^t)} R_i(h^t, k^t)$, with $I(h^t)$ used to denote the set of agents present in state h^t and agent i's reward $R_i(h^t, k^t)$ defined so that $v_i(\theta_i, k) = \sum_{t=a_i}^{d_i} R_i(h^t, k^t)$ for all sequences of decisions k. For finite time horizons, the expected value of policy π in state h^t is $V^\pi(h^t) = \mathbb{E}_\pi\big[\sum_{\tau=t}^{|T|} R(h^\tau, \pi^\tau(h^\tau))\big]$, where the expectation is taken with respect to the transition model and given the state-dependent decisions implied by policy π. For infinite time horizons, a standard approach is to define a *discount factor* $\gamma \in (0, 1)$, so that the expected discounted value of policy π in state h^t is $V^\pi(h^t) = \mathbb{E}_\pi\big[\sum_{\tau=t}^{\infty} \gamma^{\tau-t} R(h^\tau, \pi^\tau(h^\tau))\big]$. This makes sense in a multi-agent environment when every agent has the same discount factor γ.

Given MDP value, $V^\pi(h^t)$, then the optimal policy π^* maximizes this value, $V^\pi(h^t)$, in every state h^t. For instance, in the finite time-horizon

(no discounting) setting, the *optimal MDP-value function*, V^*, is defined to satisfy recurrence:

$$V^*(h) = \max_{k \in K^t(h)} [R(h,k) + \sum_{h' \in H^{t+1}} \mathcal{P}(h'|h,k)V^*(h')], \tag{1.19}$$

for all time t and all $h \in H^t$. Given this, the optimal decision policy can then be defined as:

$$\pi^*(h \in H^t) \in \arg\max_{k \in K^t(h)} [R(h,k) + \sum_{h' \in H^{t+1}} \mathcal{P}(h'|h,k)V^*(h')]. \tag{1.20}$$

Of course, the type information within the state is private to agents and we will need to provide incentive compatibility so that the policy has the correct view of the current state.

Example 1.27 The definition of state, feasible decision and agent type is as in Example 1.3. The transition function $\mathcal{P}(h^{t+1}|h^t, k^t)$ is constructed to reflect a probabilistic model of new agent arrivals, and also the allocation decision. The MDP reward function, $R(h^t, k^t)$, can be defined with $R(h^t, k^t) = w_i$ if decision k^t allocates the item to agent i, for some agent i present in the state, and zero otherwise.

1.4.2 A Dynamic Vickrey-Clarke-Groves Mechanism

For concreteness, consider an environment with a finite time horizon and no discounting, and with the optimal MDP value $V^*(h)$ defined as the total expected reward from state h until the time horizon. We make some remarks about how to handle an infinite time horizon in Section 1.4.3. Consider the following dynamic VCG mechanism.†† We assume that the decisions and reports in previous periods $t' < t$ are all *public* in period t, although similar analysis holds without this.

Auction 4. The *dynamic VCG mechanism* for the finite time horizon and no-discounting online MD environment works as follows:

(i) Each agent, i, reports a type $\hat{\theta}_i$ in some period $\hat{a}_i \geq a_i$.
(ii) Decision policy: Implement optimal policy π^*, which maximizes the total expected value, assuming the current state as defined by agent reports is the true state.

†† The mechanism is presented in the no early-arrival misreports model but remains BNIC without this assumption.

(iii) Payment policy: In an agent's reported departure period, $t = \hat{d}_i$, collect payment

$$x_i^t(h^t) = v_i(\hat{\theta}_i, \pi^*(\theta^{\leq t}, \omega^{\leq t})) - \left[V^*(h^{\hat{a}_i}) - V^*(h_{-i}^{\hat{a}_i})\right], \qquad (1.21)$$

where $\pi^*(\theta^{\leq t}, \omega^{\leq t})$ denotes the sequence of decisions made up to and including period t based on types $\theta^{\leq t}$ and stochastic events $\omega^{\leq t}$, $V^*(h^t)$ is the optimal MDP value in state h^t, and h_{-i}^t defines the (counterfactual) MDP state constructed to be equal to h^t but removing agent i's type from the state. The payment is zero otherwise.

Agent i's payment is its ex post value discounted by term $(V^*(h^{\hat{a}_i}) - V^*(h_{-i}^{\hat{a}_i}))$, which is the expected marginal value it contributes to the system as estimated upon its arrival and based on its report. With this, the expected utility to agent i when reporting truthfully is equal to the expected marginal value that it contributes to the multi-agent system through its presence.

For incentive-compatibility, we need the technical property of *stalling*, which requires that the expected value of policy π^* cannot be improved (in expectation) by delaying the report of an agent:‡‡

Theorem 1.28 *The dynamic VCG mechanism, coupled with a policy that satisfies stalling, is Bayes-Nash incentive compatible (BNIC) and implements the expected-value maximizing policy, in a domain with no early-arrival misreports but arbitrary misreports of departure.*

Proof Consider the expected utility (defined with respect to its information in period a_i) to agent i for misreport $\hat{\theta}_i \in C(\theta_i)$. Let $c \geq 0$ denote the number of periods by which agent i misreports its arrival time. The expected utility is:

$$\underbrace{\mathbb{E}_{\pi^*}\left\{v_i(\theta_i, \pi^*(h^{a_i}))|\hat{\theta}_i\right\}}_{\text{(A)}} \quad \underbrace{+\mathbb{E}_{\pi^*}\Big\{\sum_{t=a_i+c}^{|T|} R_{-i}(h^t, \pi^*(h^t))\Big\}}_{\text{(B)}} \quad \underbrace{-\mathbb{E}_{\pi^*}\left\{V^*(h_{-i}^{a_i+c})\right\}}_{\text{(C)}}$$

Term (A) denotes the expected value to agent i given its misreport. Term (B), which denotes the total expected value to other agents forward from reported arrival, a_i+c, given agent i's misreport, corresponds to the expected value of terms $\{-v_i(\hat{\theta}_i, \pi^*(\theta^{\leq \hat{d}_i}, \omega^{\leq \hat{d}_i})) + V^*(h^{\hat{a}_i})\}$ in the payment. Notation R_{-i} denotes the total reward that accrues due to all agents except agent i. Term (C), which denotes the total expected value to other agents forward

‡‡ This is typically reasonable, for example any optimal policy that is able to delay for itself any decisions that pertain to the value of an agent will automatically satisfy stalling.

from period $a_i + c$, but with agent i removed, corresponds to the final term in the payment. Now, add term $\mathbb{E}_{\pi^*}\{\sum_{t=a_i}^{a_i+c-1} R_{-i}(h^t, \pi^*(h^t))\}$ to term (B) and subtract it again from term (C). The adjusted term (C') is now agent-independent and can be ignored for the purpose of establishing BNIC. Term (A) combined with adjusted term (B') is the expected value to all other agents forward from period a_i, plus the expected true value to agent i. Agent i's best response is to report its true type (and immediately upon arrival) because the policy π^* is defined to maximize (A)+(B') when the other agents are truthful, i.e. in a Bayes-Nash equilibrium. □

It bears repeating that truth telling is not a dominant strategy equilibrium. We only have BNIC because the correctness of the policy depends on the center having the correct model for the distribution on agent types. Without the correct model, the policy is not optimal in expectation and an agent with beliefs different from that of the center may be able to improve (its belief about) the expected utility it will receive by misreporting its type and thus misrepresenting the state.§§

1.4.3 Remarks

We end this section with some general remarks that touch on the computational aspects of planning in model-based environments, and also describe a couple of additional environments in which dynamic VCG mechanisms can be usefully applied.

Infinite time horizon and discounting. The dynamic VCG mechanism can be extended to handle an infinite time horizon when every agent has a common discount factor. Rather than collect a payment once, upon departure, a payment can be collected from agent i in each period, so as to align its utility stream with the expected, marginal stream of value that it contributes through its presence in the multi-agent system.

Computational notes. Many algorithms exist to compute optimal decision policies in MDPs. These include dynamic programming, value iteration, policy iteration, and LP-based methods. However, the state space

§§ The additional property of (ex post) IR is ensured when the environment satisfies *agent-monotonicity*, which requires that introducing an agent increases the MDP value of any state. The payments collected by the mechanism are non-negative in expectation (ex ante BB) when the environment satisfies *no positive externalities*, which requires that the arrival of an agent does not have a positive expected effect on the total value of the *other* agents.

and action space for real-world online MD problems is large and approximations will typically be required. One appealing method is to couple the VCG mechanism with an online, sampling-based approximation algorithm. Rather than compute *a priori* an entire policy for every possible state one can determine the next decision to make in state h^t by approximating the decision problem forward from that state. Given an ϵ-approximation, the dynamic VCG mechanism is ϵ-BNIC, in the sense that no agent can gain more than some amount $\epsilon > 0$ (that can be made arbitrarily small) by deviating from truthful reporting, as long as the other agents are truthful and an ϵ-accurate estimate of the optimal MDP value is also available.

One class of online, sparse-sampling algorithms work by building out a sample tree of future states based on decisions that could be made by the policy forward to some look-ahead horizon. These algorithms have run time that is independent of the size of the state space but scales exponentially in the number of decisions and in the look-ahead horizon. More recently, a family of *stochastic online combinatorial optimization* algorithms have been proposed that seem especially applicable to online MD environments. The algorithms solve a sub-class of MDPs in which the realization of uncertainty is independent of any decision. This is a natural assumption for truthful dynamic auctions: the decisions made by an IC mechanism will not affect the reports of agents, and thus the realization of new types is independent of allocation decisions.

Strategic learning. A variant on the dynamic VCG mechanism can be used to support optimal, coordinated learning amongst a fixed population of self-interested agents. Suppose that in addition to influencing the reward received by an agent in each time period, the decisions made by a mechanism also reveal *information* that an agent can use to update its belief about its type, i.e. types are revealed online. A simple model of this is given by a multi-agent variation on the classical multi-armed bandits problem. Each agent owns an "arm" and receives a reward when its arm is activated, sampled from a stationary distribution. The reward signals are privately observed and allow an agent to update its model for the reward on its arm. In a setting with an infinite time horizon and discounting, one can use Gittins' celebrated index policy to characterize an efficient online policy that makes the optimal tradeoff between exploitation and exploration. In the presence of self-interest, a variant on the dynamic VCG mechanism can provide incentives to support truthful reporting of reward signals by each agent, and thus implement the efficient learning policy.

1.5 Conclusions

We briefly consider some of the many possible future research directions in this area of online mechanism design:

- Revenue: Little work exists on the design of revenue-maximizing online mechanisms in model-based environments. For example, the problem of designing an analog to Myerson's optimal auction is currently only partially solved, even in the very simplest of online settings.
- Learning by the center: It is interesting to allow the mechanism to improve its probabilistic model of the distribution on agent types across time, while retaining incentive compatibility along the path of learning, and seek to converge to an efficient or revenue-optimal mechanism.
- Alternative solution concepts: Introduce weaker solution concepts than DSIC that avoid the strong common knowledge assumptions that are required to justify BNIC analysis. These could include, for instance, set Nash equilibria, implementation in undominated strategies, or implementation in min-max-regret equilibria and other robust solution concepts.
- Endogenous information: Extend online MD to domains in which decisions made by the mechanism affect the information available to agents about their types; i.e., cast online MD as a general problem of coordinated learning by self-interested agents in an uncertain environment.
- Richer domains: The current work on *dominant-strategy* implementation is limited to single-valued preference domains with quasi-linear utilities. Simple generalizations, such as to an environment in which some agents want an apple, some a banana, and some are indifferent across an apple and a banana do not satisfy the partition requirement on the structure of interesting sets and remain unsolved. Similar complications occur when one incorporates budget constraints, or generalizes to interdependent valuations. With time, perhaps progress can be made on the problem of online *combinatorial* auctions and exchanges in their full generality.

Exercises

1.1 Prove that the revelation principle holds with no early-arrival and no late-departure misreports and prove the "revelation principle + heartbeats" result in combination with no early-arrival misreports.

1.2 Consider a (known interesting set) single-valued preference domain with no late-departure misreports. Show that any decision policy π that can be truthfully implemented by an IR mechanism, and does

not pay unallocated agents, must be monotonic-early (for a suitable definition of monotonic-early).

1.3 Prove that the approach outlined to constructing truthful online auctions in terms of an agent-independent price schedule $q_i^t(L, \theta_{-i}, \omega)$ induces a monotonic-late decision policy and critical-value payments. How would you modify the construction for an environment with both no early-arrival and no late-departure misreports?

1.4 Construct an example to show that the greedy auction in the expiring items setting has an arbitrarily bad competitive ratio with respect to offline VCG revenue.

1.5 Establish that the self-consistency property on prices in Section 1.3.4, coupled with the condition that a mechanism selects an outcome that maximizes utility for every agent at these prices is sufficient for truthfulness. Prove that the condition reduces to agent-independent prices for unrestricted misreports.

1.6 Prove that modifications (i–iii) in Section 1.3.4 are sufficient to achieve truthfulness with agents with unknown interesting sets, together with no early-arrival and no late-departure misreports and a critical-value payment. What could break if the interesting sets are not disjoint, or if the policy is not minimal?

1.7 Show that the stochastic policy outlined in Example 1.26 satisfies monotonicity conditions (1.12) and (1.13).

1.8 Define a dynamic VCG mechanism that works for infinite time horizon and agents with a common, known discount factor $\gamma \in (0, 1)$.

Notes

Lavi and Nisan [LN00] coined the term *online auction* and initiated the study of truthful mechanisms in dynamic environments within the computer science literature. Friedman and Parkes [FP03] later coined the term *online mechanism design*. The characterization of monotonicity requirements for truthful online mechanisms in single-valued domains is based on Hajiaghayi et al. [HKMP05], with extensions to single-valued preferences building on Babaioff et al. [BLP05].¶¶ Weak-monotonicity and its role in truthful mechanism design is discussed in Bikhchandani et al. [BCL$^+$06].

The discussion of the secretary problem and adaptive truthful auctions in the single-item setting is based on Hajiaghayi et al. [HKP04]; see [BIK07]

¶¶ The original paper by Hajiaghayi, Kleinberg, Mahdian and Parkes [HKMP05] mischaracterized the monotonicity requirement that is necessary for the truthful implementation of stochastic policies. This was originally brought to the attention of the authors by R.Vohra. The corrected analysis (presented here) is due to M.Mahdian.

for a recent extension and [GM66, Dyn63] for classic references. The discussion of online mechanisms for expiring items is based on Hajiaghayi et al. [HKMP05], and the negative result is due to Lavi and Nisan [LN05] (who also adopted an alternate solution concept in their analysis); see also [NPS03, Por04, JP06] and Awerbuch et al. [AAM03]. Additional models of dynamic auctions in the computer science literature include: unlimited supply, digital goods [BYKW02, BKRW03, BH05], two-sided auctions with both buyers and sellers [BP05, BSZ06], and interdependent value environments [CIP06].

Moving to the model-based framework, the discussion of the dynamic VCG mechanism is based on Parkes and Singh [PS03, PSY04]. Related concepts are discussed in Bergemann and Välimäki [BV06b] and Athey and Segal [AS06], whose work along with that of Cavallo et al. [CPS06] and Bapna and Weber [BW06] pertains to a model of strategic learning; see also [BV03, BV06a]. Pai and Vohra [PV06] advance the study of revenue-optimal online mechanisms in model-based environments, and together with Gallien [Gal06] work to extend Myerson's [Mye81] optimal auction to dynamic environments. The observation about the failure of the revelation principle, the example to illustrate the role of non-negative payments, as well as inspiration for the extended example of a truthful, stochastic policy are due to Pai and Vohra. For references on online algorithms and methods for solving sequential decision problems see [BEY98, HB06, Put94, KMN99].

Acknowledgments

Many thanks to Florin Constantin, Bobby Kleinberg, Mallesh Pai, and Rakesh Vohra for providing detailed and constructive comments on an earlier draft, and to my collaborators in this work, including Jonathan Bredin, Ruggiero Cavallo, Florin Constantin, Quang Duong, Eric Friedman, Mohammad Hajiaghayi, Adam Juda, Bobby Kleinberg, Mohammad Mahdian, Chaki Ng and Satinder Singh. Parkes is supported in part by National Science Foundation grants IIS-0238147, IIS-0534620 and an Alfred P. Sloan Foundation award.

Bibliography

[AAM03] Baruch Awerbuch, Yossi Azar, and Adam Meyerson. Reducing truth-telling online mechanisms to online optimization. In *Proc. ACM Symposium on Theory of Computing (STOC'03)*, 2003.

[AS06] Susan Athey and Ilya Segal. An efficient dynamic mechanism. Technical report, Harvard University and Stanford University, 2006.

[BCL+06] S. Bikhchandani, S. Chatterji, R. Lavi, A. Mu'alem, N. Nisan, and A. Sen. Weak monotonicity characterizes deterministic dominant strategy implementation. *Econometrica*, pages 1109–1132, 2006.

[BEY98] A Borodin and R El-Yaniv. *Online Computation and Competitive Analysis*. Cambridge University Press, 1998.

[BH05] Avrim Blum and Jason Hartline. Near-optimal online auctions. In *Proceedings of the 16th Annual ACM-SIAM symposium on Discrete algorithms*, 2005.

[BIK07] Moshe Babaioff, Nicole Immorlica, and Robert Kleinberg. Matroids, secretary problems, and online mechanisms. In *Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*, 2007.

[BKRW03] Avrim Blum, Vijar Kumar, Atri Rudra, and Felix Wu. Online learning in online auctions. In *Proceedings of the 14th Annual ACM-SIAM symposium on Discrete algorithms*, 2003.

[BLP05] Moshe Babaioff, Ron Lavi, and Elan Pavlov. Mechanism design for single-value domains. In *Proc. 20th National Conference on Artificial Intelligence 2005 (AAAI'05)*, pages 241–247, 2005.

[BP05] Jonathan Bredin and David C. Parkes. Models for truthful online double auctions. In *Proc. 21st Conference on Uncertainty in Artificial Intelligence (UAI'2005)*, pages 50–59, 2005.

[BSZ06] Avrim Blum, Tuomas Sandholm, and Martin Zinkevich. Online algorithms for market clearing. *Journal of the ACM*, 2006. To appear.

[BV03] Dirk Bergemann and Juuso Välimäki. Dynamic common agency. *Journal of Economic Theory*, 11:23–48, 2003.

[BV06a] Dirk Bergemann and Juuso Välimäki. Dynamic price competition. *Journal of Economic Theory*, 127:232–263, 2006.

[BV06b] Dirk Bergemann and Juuso Välimäki. Efficient dynamic auctions. Technical Report Cowles Foundation Discussion Paper No. 1584, Yale University, 2006.

[BW06] Abhishek Bapna and Thomas A Weber. Efficient dynamic allocation with uncertain valuations. Technical report, Stanford University, 2006.

[BYKW02] Z. Bar-Yossef, K.Hildrum, and F. Wu. Incentive-compatible online auctions for digital goods. In *Proc. 13th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2002.

[CIP06] Florin Constantin, Takayuki Ito, and David C. Parkes. Online auctions for bidders with interdependent values. Technical report, Harvard University, 2006.

[CPS06] Ruggiero Cavallo, David C. Parkes, and Satinder Singh. Optimal coordinated learning among self-interested agents in the multi-armed bandit problem. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI'2006)*, Cambridge, MA, 2006.

[Dyn63] E B Dynkin. The optimum choice of the instant for stopping a Markov process. *Sov. Math. Dokl.*, 4:627–629, 1963.

[FP03] Eric Friedman and David C. Parkes. Pricing WiFi at Starbucks– Issues

in online mechanism design. In *Fourth ACM Conf. on Electronic Commerce (EC'03)*, pages 240–241, 2003.

[Gal06] Jeremie Gallien. Dynamic mechanism design for online commerce. *Operations Research*, 54:291–310, 2006.

[GM66] J. Gilbert and F. Mosteller. Recognizing the maximum of a sequence. *J. Amer. Statist. Assoc.*, 61(313):35–73, 1966.

[HB06] Pascal Van Hentenryck and Russell Bent. *Online Stochastic Combinatorial Optimization*. MIT Press, 2006.

[HKMP05] Mohammad T. Hajiaghayi, Robert Kleinberg, Mohammad Mahdian, and David C. Parkes. Online auctions with re-usable goods. In *Proc. ACM Conf. on Electronic Commerce*, pages 165–174, 2005.

[HKP04] Mohammad T. Hajiaghayi, Robert Kleinberg, and David C. Parkes. Adaptive limited-supply online auctions. In *Proc. ACM Conf. on Electronic Commerce*, pages 71–80, 2004.

[JP06] Adam Juda and David Parkes. The sequential auction problem on eBay: An empirical analysis and a solution. In *Proc. 7th ACM Conf. on Electronic Commerce (EC'06)*, pages 180–189, 2006.

[KMN99] Michael Kearns, Yishay Mansour, and Andrew Y Ng. A sparse sampling algorithm for near-optimal planning in large Markov Decision Processes. In *Proc. 16th Int. Joint Conf. on Artificial Intelligence*, pages 1324–1331, 1999. To appear in Special Issue of *Machine Learning*.

[LN00] Ron Lavi and Noam Nisan. Competitive analysis of incentive compatible on-line auctions. In *Proc. 2nd ACM Conf. on Electronic Commerce (EC-00)*, 2000.

[LN05] Ron Lavi and Noam Nisan. Online ascending auctions for gradually expiring goods. In *Proc. of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2005.

[Mye81] Robert B Myerson. Optimal auction design. *Mathematics of Operation Research*, 6:58–73, 1981.

[NPS03] Chaki Ng, David C. Parkes, and Margo Seltzer. Virtual Worlds: Fast and Strategyproof Auctions for Dynamic Resource Allocation. In *Fourth ACM Conf. on Electronic Commerce (EC'03)*, pages 238–239, 2003.

[Por04] Ryan Porter. Mechanism design for online real-time scheduling. In *Proc. ACM Conf. on Electronic Commerce (EC'04)*, 2004.

[PS03] David C. Parkes and Satinder Singh. An MDP-based approach to Online Mechanism Design. In *Proc. 17th Annual Conf. on Neural Information Processing Systems (NIPS'03)*, 2003.

[PSY04] David C. Parkes, Satinder Singh, and Dimah Yanovsky. Approximately efficient online mechanism design. In *Proc. 18th Annual Conf. on Neural Information Processing Systems (NIPS'04)*, 2004.

[Put94] M. L. Puterman. *Markov decision processes : discrete stochastic dynamic programming*. John Wiley & Sons, New York, 1994.

[PV06] Mallesh Pai and Rakesh Vohra. Optimal dynamic auctions. Technical report, Kellogg School of Management, Northwestern University, 2006.