An Optimal Dynamic Mechanism
for Multi-Armed Bandit Processes

Sham M. Kakade[*]   Ilan Lobel[†]   Hamid Nazerzadeh[‡§]

October 18, 2010

Abstract

We consider the problem of revenue-optimal dynamic mechanism design in settings where agents' types evolve over time as a function of their (both public and private) experience with items that are auctioned repeatedly over an infinite horizon. A central question here is understanding what natural restrictions on the environment permit the design of optimal mechanisms (note that even in the simpler static setting, optimal mechanisms are characterized only under certain restrictions). We provide a *structural characterization* of a natural "separable" multi-armed bandit environment (where the evolution and incentive structure of the a-priori type is decoupled from the subsequent experience in a precise sense) where dynamic optimal mechanism design is possible. Here, we present the Virtual Index Mechanism, an optimal dynamic mechanism, which maximizes the (long term) *virtual surplus* using the classical Gittins algorithm. The mechanism optimally balances exploration and exploitation, taking incentives into account.

We pay close attention to the applicability of our results to the (repeated) ad auctions used in sponsored search, where a given ad space is repeatedly allocated to advertisers. The value of an ad allocation to a given advertiser depends on multiple factors such as the probability that a user clicks on the ad, the likelihood that the user performs a valuable transaction (such as a purchase) on the advertiser's website and, ultimately, the value of that transaction. Furthermore, some of the private information is learned over time, for example, as the advertiser obtains better estimates of the likelihood of a transaction occurring. We provide a dynamic mechanism that extracts the maximum feasible revenue given the constraints imposed by the need to repeatedly elicit information.

One interesting implication of our results is a certain revenue equivalence between public and private experience, in these separable environments. The optimal revenue is no less than if agents' private experience (which they are free to misreport, if they are not incentivized appropriately) were instead publicly observed by the mechanism.

---

[*]Department of Statistics, Wharton School, University of Pennsylvania – skakade@wharton.upenn.edu. Work performed in part at Microsoft Research.

[†]Stern School of Business, New York University – ilobel@stern.nyu.edu. Work performed in part at Microsoft Research.

[‡]Microsoft Research, New England Lab – hamidnz@microsoft.com

[§]We would like to thank Maher Said for many fruitful discussions.

1 Introduction

Designing mechanisms in *dynamic* environments — in which agents valuations evolve as a function of their "experience" with the allocated items — is a problem which has received much recent interest. One of the most compelling applications here is that of ad auction sponsored search, in which search engines sell the advertisement spaces that appear alongside the search results.

Let us discuss the sponsored search example in more detail: typically, an advertiser places an ad in order to: first, draw a client to visit the advertiser's website (via a click on the displayed ad), and then, subsequently, have the client purchase some product. The expected value that an advertiser obtains from a displayed ad depends on *both* the "click-through rate" (the probability that a user clicks on the ad, sending the user to the advertiser's website) and the "conversion rate" (the probability that the user who visits the website performs a desired transaction, e.g. a purchase). This is a dynamic environment in which both advertisers and the search engine (the mechanism) learn and update their estimates of these rates over time. Observe that a click is a *public experience*, i.e., observed by both the advertiser and the search engine. In contrast, a transaction is only observed by the advertiser — it's *private experience* of the advertiser with the displayed ad. The dynamic challenge here is to design appropriate mechanisms which align incentives such that the search engine and the advertiser share this information for some desired outcome.

In the oft used practical mechanisms, the learning of click-through-rates and conversions-rates have been separated due to this asymmetry of information — (see Mahdian and Tomak [2007], Agarwal et al. [2009] for further discussions on "pay-per-action" pricing schemes). Two fundamental question that arises in this setting is: how much revenue does this asymmetry of information cost the mechanism? How much more revenue would the mechanism be able to obtain, if it were able to monitor the transactions on the advertisers' websites?

In the static setting, the two foremost objectives for a mechanism are either maximizing the social welfare of the buyers (*efficiency*) or the maximizing the revenue of the seller (*optimality*) — though the spectrum of other objectives is large and notable. By extension, these are the natural two objectives to consider for the dynamic setting. With regards to maximizing the future social welfare in a dynamic setting, there is an elegant extension of the efficient (VCG) mechanism applicable to quite general dynamic settings by Parkes and Singh [2003], Bergemann and Välimäki [2007]. This dynamic mechanism seamlessly inherits the core concepts of the static VCG mechanism — namely, charging an agent the externality they impose, which is implemented via dynamic programming ideas. Related dynamic mechanisms include the dynamic budget-balanced efficient mechanism by Athey and Segal [2007], efficient mechanisms for dynamic populations by Cavallo et al. [2007], and non-Bayesian (asymptotically) efficient dynamic mechanisms (see Nazerzadeh et al. [2008], Babaioff et al. [2009]).

With regards to *optimal* dynamic mechanisms in a dynamic setting, the state of affairs is more murky. As we discuss in the next section, while there are detailed characterizations of necessary conditions for which (incentive compatible) dynamic mechanisms must satisfy, there are only a few rather restricted special cases for which optimal mechanisms are characterized (e.g. see Pavan et al. [2008]). To some extent, results for special cases are to be expected, as even in the simpler static mechanism design problem, the efficient (VCG) mechanism is applicable to general settings (e.g. combinatorial auctions with no distributional assumptions) while even the optimal mechanism for selling a single item (provided in the seminal work of Myerson [1981]) is only applicable under certain distributional restrictions

In the more challenging dynamic setting, perhaps the most central question is understanding

what natural restrictions permit the design of optimal mechanisms. This is the focus of this paper, and we provide a certain structural characterization of a "separable" environment (allowing for public and private experience), in which optimal dynamic mechanism design is possible. Our characterization is rather rich in that it permits both a natural stochastic processes (where private and public signals can be discrete or from abstract spaces) and is applicable to certain natural formalizations of the aforementioned sponsored search setting (where both public click-through-rates and private purchase-rates evolve over time). Our construction draws a rather close connection to efficient mechanism design (where our optimal mechanism utilizes the efficient mechanism for a certain affinely transformed social welfare function). Furthermore, we also address the issue of how much revenue is lost due to private signals (rather than public signals, observed by the mechanism) in these separable environments – somewhat surprisingly, there is no loss.

1.1 Contributions

Our main contribution is designing an individually rational and incentive compatible, revenue-optimal mechanism, called the Virtual Index Mechanism, for settings with 1 seller and k (agents) buyers where the environment satisfies certain *separable* properties and evolves according to a *multi-armed bandit process*.

The Virtual Index Mechanism is quite simple. In short, the allocation rule of the mechanism is based on the notion of Gittins indices (see Gittins [1989], Whittle [1982]) and the payment rule is derived by considering a dynamic VCG mechanism (where the social welfare function is transformed under a particular, time-varying affine function).

The allocation assigns to each agent an "index" which is computed based on solely the agent's current state; and at each step the mechanism allocates the item to an agent with the highest index. As in Gittins [1989], Whittle [1982], the key observation is that this computation does not require specifying a policy in terms of the (potentially exponentially many) histories. If all the agents are truthful, then the mechanism maximizes the "virtual surplus"; the idea pioneered in Myerson [1981].

It turns out that the allocation we use also coincides with the efficient dynamic (VCG) mechanism (with respect to a transformed social welfare function). Due to this, the payment rule of our mechanism is rather simple to specify. In fact, one of our technical contributions is using this reduction to dynamic (affine) VCG in the construction of a revenue-optimal mechanism. This connection is useful for two reasons: first, it allows us to reduce the problem of checking incentive compatibility to essentially a one period problem. Second, this VCG pricing allows us to utilize rather general multi-armed bandit processes.

A surprising implication of our result is that the seller does not lose any revenue (under the optimal mechanism) if the experience were private rather than publicly observed. In the context of sponsored search, this implies that (in environments which satisfy our "separability" assumption) the ability to monitor the transactions that occur in the advertisers' own websites does not increase the revenue. An important business insight provided by this result is that pay-per-action mechanisms can be implemented without a loss of revenue if the search engine is able to commit to a long-term contract.

3

1.2 Related Work

The most closely related work to ours is that in Pavan et al. [2008, 2009]. [1] The primary contributions of this work is that it establishes rather detailed *necessary* conditions for dynamic incentive compatibility in both finite horizons (Pavan et al. [2008]) and infinite horizons (Pavan et al. [2009]). In addition, Pavan et al. [2008] also establish a dynamic version of the Revenue Equivalence Theorem in a particular finite horizon setting. With regards to providing dynamic optimal mechanisms, their work only provides mechanisms in somewhat limited special cases — such as when valuations evolve according to a certain auto-regressive $AR(k)$ stochastic processes, and, in Pavan et al. [2009], the value evolution evolves according to a particular additive manner, where each private experience of the agents is assumed to be independent of all previous private experiences (it is allowed to depend only on the number of times the item was previously allocated, a much more restrictive assumption than those provided by our results). We should also note the work in Deb [2008], which provides an optimal mechanism in a restricted setting where the value is Markov in the previous value, among other technical conditions — again, their model does not permit rich dependencies on historical signals. In our sponsored search example, these prior results are not applicable, due to both the multiplicative nature of the value function (as discussed later), and due to that the sequence of experiences are not independent (e.g. with a Bayesian, "Bernoulli" prior on the probabilities of binary "click" or "purchase" events, the experiences are not necessarily independent).

Also, in contrast to Pavan et al. [2008, 2009], we should emphasize that the aim of our work is *not* to characterize necessary conditions which any (incentive compatible) dynamic mechanism must satisfy – our focus is on the optimal mechanism itself. In fact, we do not even utilize the dynamic "envelope" conditions provided by Pavan et al. [2008, 2009], as they require: many detailed technical assumptions; the signals (the experiences) to be real valued; and, often, certain probability kernels to have densities and be differentiable — as such, these conditions either do not hold or are difficult to verify in our setting. Instead, our derivation proceeds from merely static considerations, where we use only incentive compatibility constraints from static mechanism design theory (see Milgrom and Segal [2002]) to establish the expected revenue of any (incentive compatible) mechanism (e.g. the so called "envelope theorem"). Certainly, the sufficient conditions provided by Pavan et al. [2008, 2009] (derived with rather sophisticated proofs) are stronger than than those conditions used here, as they explicitly account for dynamic considerations and are interesting in their own right — one further direction is if these conditions can be used to derive optimal dynamic mechanisms in settings more general than those provided here.

Conceptually, our proof is rather simple: both the connection to dynamic (affine) VCG and our use of only a static "envelope" condition allow us to reduce the proof of dynamic incentive compatibility to essentially a one-period, static problem. However, this one-period verification requires a delicate stochastic coupling argument, where we utilize both the bandit nature and the separability of our stochastic process.

We also briefly mention other notable work here. Vulcano et al. [2002] analyze the problem of optimal dynamic mechanism design in the context of perishable goods (work later expanded on by Pai and Vohra [2008]). These works built upon dynamic programming ideas to extend the classical result of Myerson [1981] to a dynamic setting. A key assumption in both of these papers is that agents' valuations do not evolve over time. Battaglini [2005] studies the question of optimal

---

[1]We note that the work in Pavan et al. [2009] is a recent preliminary draft, and is concurrent to this work.

4

mechanism design in a setting with a single consumer whose private information is given by a 2-state Markov Chain. Ëso and Szentes [2007] obtain a result in a two-period model that is similar in flavor to ours: the buyers do not benefit from obtaining new private information at period 2 if the seller uses a 'handicap' auction, where the handicap is given by the buyer's virtual value at period 1.

1.3 Organization

We organize our paper as follows. In Section 2, we formalize our model, define separable environments, show examples and define the basic notions we use throughout the paper, such as incentive compatibility and optimality of mechanisms. In Section 3, we consider a variant of our model where the mechanism can monitor the private experiences of the agents and describe how to it establishes a bound for the revenue of a mechanism in our setting. In Section 4, we state and explain our main theorem — the optimality of the Virtual Index Mechanism. All proofs are in the Appendix.

2 Preliminaries

2.1 Environment

We consider a setting with 1 seller and k agents (buyers) who are competing for items that are being allocated at every timestep (starting at $t = 1$) over a discrete time infinite horizon. At the start of $t = 1$, agents (privately) learn their initial types. The *initial type* of agent i is a (non-negative) real number $\theta_i \in [0, \overline{\Theta}_i]$, independently distributed according to some given distribution $F_i(\cdot)$. At every subsequent timestep, the state of each agent i is summarized by the tuple of their initial type θ_i and their (subsequent) "experience" with the item — this experience summarizes the type of the agent due to interactions with the item and the experience need not be real valued. More precisely, agent i's state at time t is of the form $(\theta_i, e_{i,t}, \rho_{i,t})$, where the current state of the private experience is denoted by $e_{i,t} \in \mathcal{E}_i$ and the public experience is denoted by $\rho_{i,t} \in \mathcal{P}_i$, where $\mathcal{E}_i$ and $\mathcal{P}_i$ are some (potentially *arbitrary*) set. Here, only the agent observes their private experience, while the public experience is also observed by the mechanism.

We should emphasize that the first type θ is real for reasons similar to that in the static setting — derivations of optimal mechanisms typically involve calculus on real valued types. However, it is *only* this first type that we assume to be real valued (subsequent experience is allowed to live in arbitrary signal spaces). As we specify later, this first type θ has a persistent effect on the incentives (e.g., the values after $t = 1$ could also depend on the initial type).

If agent i is allocated, the state of agent i's public and private experience with the item evolve in a Markovian manner. If i is not allocated, then the experience *does not change* — in this sense, we are dealing with a Markovian "bandit" process. By nature of public information, we assume the public process is completely decoupled from private information, e.g., the probability that the next public experience is ρ_i' conditioned on the current experience being ρ_i is $G(\rho_i'|\rho_i)$. In the most general sense, the evolution of the private experience could depend on the entire current state, e.g., the probability that the next private experience is e_i' conditioned on the current state (θ_i, e_i, ρ_i) is $H(e_i'|\theta_i, e_i, \rho_i)$. However, it turns out that are not able to handle this level of generality, and our structural characterization specifies certain natural restrictions (in the next subsection).

Note that the public experience evolution process only depends on the public times series, while the private experience process is allowed to depend on both private and public experience. We

assume that private experience is only observed by the agent, but the public information is also observed by the mechanism. At time $t = 1$ (prior to the allocation at $t = 1$), for every agent i, $e_{i1} = \emptyset$ is the empty experience, known by both the agent and the mechanism.

The (instantaneous) value of each agent i at time t is a (stationary) function of their current state. In particular, agent i's value is $v_i(\theta_i, e_{i,t}, \rho_{i,t})$ when in state $(\theta_i, e_{i,t}, \rho_{i,t})$ at time t. The expected (future) value of agent i for the item at time t is equal to $\delta^{t-1} v_i(\theta_i, e_{i,t}, \rho_{i,t})$ where δ, $0 < \delta < 1$, is the common discount factor.

2.2 Separable Environments

Characterizing the assumptions which permit optimal mechanisms design is perhaps the most central question — even in the static setting, only in (natural) special cases optimal mechanisms are known. In our dynamic setting, it turns that we are not able to characterize an optimal mechanism in the full generality of the above "bandit" environment. However, our main contribution is specifying a natural "separable" environment, under which we can derive an optimal mechanism. We say that the environment is *separable* if *both* the stochastic process over the types and the value functions themselves are separable, in a precise sense which we now define. Intuitively, the notion of separability decouples the initial (real valued) type θ from the experience, both in terms of the stochastic evolution and the incentive structure.

Definition 2.1. *The stochastic process is said to be separable if the evolution of the private experience is Markovian in the current experience, e.g., the probability that the next private experiences is e_i' conditioned on the current state (θ_i, e_i, ρ_i) is $H(e_i'|e_i, \rho_i)$ (in particular, H does not depend on θ_i).*

We consider two natural classes of separable value functions.

Definition 2.2. *Additively or multiplicatively* separable *value functions are defined as follows:*

- *An* additively separable value *function has the following functional form, for all i, θ_i, e_i, and ρ_i:*

$$v_i(\theta_i, e_i, \rho_i) = A_i(\theta_i, \rho_i) + B_i(e_i, \rho_i)$$

- *A* multiplicatively separable value *is of the form:*

$$v_i(\theta_i, e_i, \rho_i) = A_i(\theta_i) B_i(e_i, \rho_i) - C_i(\rho_i)$$

Taken together, we say that the environment is (additively or multiplicatively) separable.

2.3 Examples of Separable Environments

We now provide two examples of settings for separable value functions, which fall within our framework (and satisfy our assumptions).

Sponsored Search: Consider an auction for a keyword that corresponds to a certain product. Suppose i is an online retailer of such a product who participates in the corresponding sponsored search auction. Every time a user types in the keyword, the ad space is allocated to (at most) one retailer. Every time a user purchases the product from them, the retailer i obtains a value of

6

θ_i (and 0 otherwise). The private experience $e_{i,t}$ describes the retailer's Bayesian belief about the probability of a purchase given a click has occurred. Similarly, the public experience $\rho_{i,t}$ represents the Bayesian belief about the probability of a click occurring given the retailer's ad is shown. Therefore, $v_i(\theta_i, e_{i,t}, \rho_{i,t}) = \theta_i \Pr[\text{purchase}|e_{i,t}, click] \Pr[\text{click}|\rho_{i,t}]$. After each time the ad of retailer i is shown to a user, both the retailer and the search engine update the belief $\rho_{i,t}$ about probability of a click. After each click, the retailer updates their belief $e_{i,t}$ about the probability of a purchase.

Auto-Regressive (AR): The evolution of the valuation of each agent i in an $AR(1)$ model is as follows. The initial value of agent i is given by $v_{i,0} = \theta_i$, and every time the item is allocated to agent i his valuation is updated according to $v_{i,t+1} = A_i v_{i,t} + B_i(e_{i,t}, \rho_{i,t})$. In the AR model considered in Pavan et al. [2008, 2009], the agent's value is updated by adding an independent shock, or a shock that depends only on the previous allocations (e.g., number of times the item was allocated to the agent), but is independent of all private information. Our model allows for the update of the value to depend both on the private and the public experiences. They also consider $AR(k)$ processes, where the valuation is updated according to an affine function of the values at the k previous times the item was allocated to the agent. Our restriction to $AR(1)$ models is without loss of generality since, by augmenting the state space, an $AR(k)$ process can be represented as an $AR(1)$ process (and through an appropriate choice of the function $A_i(\cdot)$).

2.4 Mechanisms, Incentive Constraints, and Optimality

By the Revelation Principle (cf. Myerson [1986]), without loss of generality we can focus on direct mechanisms.[2] A direct mechanism $\mathcal{M}(\mathcal{Q}, \mathcal{P})$ is defined by a pair of an allocation rule $\mathcal{Q}$ and a payment rule $\mathcal{P}$. In a dynamic direct mechanism, at each timestep t, an agent is asked to *report* their current private state pair $(\theta_i, e_{i,t})$ — we denote this report by $(\hat{\theta}_{i,t}, \hat{e}_{i,t})$. Note that a direct mechanism elicits redundant information as the initial type θ_i of an agent remains constant over time, while the mechanism asks the agent to re-report this type every round (similarly, the private experience of an agent does not evolve in a period in which it did not receive an allocation, yet the mechanism asks for re-reports).

We denote the (joint) vector of reports, public state, allocations, and payments at time t by $(\hat{\theta}_t, \hat{e}_t, \rho_t, q_t, p_t)$, where $q_{i,t}$ and $p_{i,t}$ correspond to the allocation and payment of agent i at time t ($q_{i,t} = 1$ if i received the item at time t and 0 otherwise). The history h_t observed by the seller at any given time t includes the past reports, the past public experience, the past allocations, and the past payments (and does not include either the past private experiences or the initial types). The history $h_{i,t}$ observed by an agent i at time time t includes her past initial type, her private experiences, her prior reports, the prior payments, and the prior allocations (and does not include other agents true or reported private experiences or initial types).

In short, at each timestep $t \geq 1$ the following sequence of events occur:

1. Each agent i reports $(\hat{\theta}_{i,t}, \hat{e}_{i,t})$ only to the mechanism.

2. The mechanism allocates the item to an agent $i^\star$, if $q_{i^\star, t} = 1$ (or potentially to no one).

3. Each agent i is charged $p_{i,t}$.

4. Agent $i^\star$'s (private and public) experience evolve.

---

[2] The Revelation Principle implies that an equilibrium outcome in any indirect mechanism can also be induced as an equilibrium outcome of an (incentive compatible) direct mechanism.

We now define the incentive constraints of the mechanism. First, some definitions are in order. A reporting strategy for agent i is a mapping from her type, her private experience state, and the history to a report (of her initial type and the current state of the private experience). Let R denote a joint reporting strategy and R_i denote this strategy for i. For mechanism $\mathcal{M}$, define the (discounted) expected future value and payment of agent i at time t, under (joint) reporting strategy R, conditioned on some event h, as follows:

$$V_{i,t}^{\mathcal{M},R}(h) = \mathbb{E}\left[\sum_{t'=t}^{\infty} \delta^{t'-1} q_{i,t'} v_i(\theta_i, e_{i,t'}, \rho_{i,t'}) \middle| h\right]$$

$$P_{i,t}^{\mathcal{M},R}(h) = \mathbb{E}\left[\sum_{t'=t}^{\infty} \delta^{t'-1} p_{i,t'} \middle| h\right]$$

where the evolution of the process is under $\mathcal{M}$ under reporting strategy $\mathcal{R}$ — conditioned on the event h_t (the expectation is with respect to all variables not conditioned on). For example, for some current state $\theta_i, e_{i,t}, \rho_{i,t}$ for i and history $h_{i,t}$ for agent i, $V_{i,t}^{\mathcal{M},R}(\theta_i, e_{i,t}, \rho_{i,t}, h_{i,t})$ is the expected future value of i conditioned on her knowledge at time t. Similarly, define the (discounted) *expected future utility* for agent i as:

$$U_{i,t}^{\mathcal{M},R}(h) = V_{i,t}^{\mathcal{M},R}(h) - P_{i,t}^{\mathcal{M},R}(h) \tag{1}$$

We say that R_i is a best response to R_{-i} conditioned on event h (for agent i) if the R_i maximizes her utility, e.g., $U_{i,t}^{\mathcal{M},R_i}(h)$ is greater than $U_{i,t}^{\mathcal{M},R_i{}'}(h)$ (for all other $R_i{}'$, where R_{-i} is held fixed). We say the *truthtelling strategy* $\mathcal{T}$ is the reporting strategy under which all agents always reports their initial types and their private experiences truthfully.

We now define incentive compatibility. Roughly speaking, this concept says that as long as all agents are truthful, then no agent ever wants to deviate.

Definition 2.3. *(Incentive Compatibility) A dynamic direct mechanism is* incentive compatible *if, for each agent i, with probability one, truthtelling is a best response (assuming the other agents to be truthful) at each time t with respect to the history of i at time t. Precisely, with probability 1, for all times t and all R_i,*

$$U_{i,t}^{\mathcal{M},\mathcal{T}}(\theta_i, e_{i,t}, \rho_{i,t}, h_{i,t}) \geq U_{i,t}^{\mathcal{M},(R_i, \mathcal{T}_{-i})}(\theta_i, e_{i,t}, \rho_{i,t}, h_{i,t})$$

where the probability is with respect to $\theta_i, e_{i,t}, \rho_{i,t}, h_{i,t}$ sampled under the truthful reporting strategy.

We consider a stronger notion, namely, periodic ex-post incentive compatibility, where best responses hold even on histories where misreports occur (see Definition 4.1).

We also allow the following participation constraint, in which agents may opt out at any time for 0 future utility.

Definition 2.4. *(Individual Rationality) Under an* individually rational *mechanism, for each agent i, with probability 1, truthful agents obtain a non-negative expected future utility assuming the other agents are truthful. Precisely, with probability 1, for all times t and all R_i,*

$$U_{i,t}^{\mathcal{M},\mathcal{T}}(\theta_i, e_{i,t}, \rho_{i,t}, h_{i,t}) \geq 0$$

where the probability is with respect to $\theta_i, e_{i,t}, \rho_{i,t}, h_{i,t}$ sampled under the truthful reporting strategy.

The *expected revenue* of an incentive compatible mechanism under the truthful strategy is the discounted sum of all payments of the agents, i.e.,

$$\text{Rev}^{\mathcal{M}} = \mathbb{E}\left[\sum_{i=1}^{k} P_{i,1}^{\mathcal{M},\mathcal{T}}(\theta)\right] \tag{2}$$

The objective of the seller is to maximize this expected revenue, subject to both the incentive compatibility constraint and rationality constraint. Precisely,

Definition 2.5. *(Optimality) An individually rational and incentive compatible mechanism is optimal if it maximizes the expected revenue among all individually rational and incentive compatible mechanisms.*

3 A Dynamic Constraint From Static Considerations

In this section, we establish a bound on the maximum revenue a seller can obtain given the incentive constraints of the agents. We consider a modified setting that we call the complete dynamic monitoring problem and show that it determines a bound for the seller's revenue in the model defined in Section 2.

Consider the modified setting where the seller can fully monitor the agents' private experiences $\{e_{i,t}\}$, for each agent i and all $t > 0$. The seller still cannot observe the initial types θ_i's, and, as before, agents report $\hat{\theta}_i$ to the mechanism in the initial period. We denote this setting the **complete dynamic monitoring** problem. In this new setting, the mechanism design problem is a completely static one. Therefore, the incentive compatibility results of Milgrom and Segal [2002] for static mechanism design apply here. The following "revenue equivalence" theorem establishes the revenue obtained by an incentive compatible mechanism in the complete dynamic monitoring setting; the proof is in Appendix A.1.

Theorem 3.1. *Assume complete dynamic monitoring. Assume as well that the the partial derivative $\frac{\partial v_i(\theta_i, e_i, \rho_i)}{\partial \theta_i}$ exists for all θ_i, e_i and ρ_i and there exists some $B < \infty$ such that $|\frac{\partial v_i(\theta_i, e_i, \rho_i)}{\partial \theta_i}| \leq B$ for all θ_i, e_i and ρ_i. Then, the revenue $Rev^{\mathcal{M}}$ of any such incentive compatible mechanism $\mathcal{M}$ satisfies*

$$Rev^{\mathcal{M}} = \sum_{i=1}^{k} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t}\psi_i(\theta_i, e_{i,t}, \rho_{i,t}) - U_i^{\mathcal{M}}(0, \theta_{-i})\right] \tag{3}$$

where ψ_i is defined as

$$\psi_i(\theta_i, e_{i,t}, \rho_{i,t}) = v_i(\theta_i, e_{i,t}, \rho_{i,t}) - \frac{1 - F(\theta_i)}{f(\theta_i)}\frac{\partial v_i(\theta_i, e_{i,t}, \rho_{i,t})}{\partial \theta_i} \tag{4}$$

and $U_i^{\mathcal{M}}(0, \theta_{-i})$ is the utility agent i obtains if his type is equal to 0 and the other agents' types are θ_{-i}.

Similarly to Myerson [1981], we refer to ψ_i as the *virtual value*. The right-hand side of Eq. (3) is the *virtual surplus*. Individual rationality is equivalent to the requirement that $U_i^{\mathcal{M}}(0, \theta_{-i}) \geq 0$ for

9

all i and θ_{-i}. Therefore, Eq. (3) implies that for any incentive compatible and individually rational mechanism $\mathcal{M}$ in the complete dynamic monitoring setting,

$$\text{Rev}^{\mathcal{M}} \leq \sum_{i=1}^{k} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \psi_i(\theta_i, e_{i,t}, \rho_{i,t})\right] \tag{5}$$

if the assumptions of Theorem 3.1 hold.

Now consider a direct mechanism $\mathcal{M}$ for the original setting without monitoring which is both incentive compatible and individually rational. The exact same mechanism, including the allocation and payment rules, can be applied in the setting with complete dynamic monitoring. To do so, we simply replace the agent's reported private experiences $\{\hat{e}_{i,t}\}$ by the agent's (monitored) private experiences $\{e_{i,t}\}$ in the input of the allocation and payment rules. Any strategy available to the agents in the setting with complete dynamic monitoring is a feasible strategy in the setting without monitoring (where the agent reports truthfully after the initial report). Therefore, if all other agents are truthful, any profitable deviation from the truthful strategy in the setting with complete dynamic monitoring implies a profitable deviation in the setting without monitoring. Since no such profitable deviations exist in the setting without monitoring, we obtain that the mechanism $\mathcal{M}$ is both incentive compatible and individually rational in the setting with complete dynamic monitoring. Therefore, Eq. (5) establishes an upper bound on the revenue of mechanisms for the setting without monitoring as well.

Corollary 3.1. *Assume the the partial derivative $\frac{\partial v_i(\theta_i, e_i, \rho_i)}{\partial \theta_i}$ exists for all θ_i, e_i and ρ_i and there exists some $B < \infty$ such that $|\frac{\partial v_i(\theta_i, e_i, \rho_i)}{\partial \theta_i}| \leq B$ for all θ_i, e_i and ρ_i. Then, the revenue $Rev^{\mathcal{M}}$ of any incentive compatible, individually rational mechanism $\mathcal{M}$ satisfies*

$$Rev^{\mathcal{M}} \leq \max_{\mathcal{Q} \in \mathbb{Q}} \mathbb{E}\left[\sum_{t=1}^{\infty} \sum_{i=1}^{k} \delta^{t-1} q_{i,t} \psi_i(\theta_i, e_{i,t}, \rho_{i,t})\right], \tag{6}$$

where $\mathbb{Q}$ represents the set of all allocation rules.

The corollary suggests a candidate allocation rule for an optimal mechanism in the setting without monitoring. The maximization problem on the right-hand side of Eq. (6) is a multi-armed bandit problem, where the payoff of arms given by the virtual values. This optimization problem can be solved using Gittins indices (see Gittins [1989], Whittle [1982]). We use this allocation rule in the mechanism we design in the next session.

4 The Virtual Index Mechanism

In this section, we present our main result, an optimal dynamic mechanism, called the *Virtual Index* Mechanism. In short, the allocation rule is as follows: the mechanism assigns to each agent an "index" (computed based on virtual values) and at each step allocates the item to an agent with the highest index. If all the agents are truthful then the mechanism maximizes the revenue as well as the virtual surplus. Furthermore, the mechanism enjoys more desirable incentive constraints — it satisfies stronger notions of incentive compatibility and individual rationality.

Definition 4.1. *(Periodic Ex-post Incentive Compatibility) A dynamic direct mechanism is* periodic ex-post incentive compatible *if for all agents, truth-telling is a best response conditioned on* any

historical event and conditioned on the current state of the other agents (assuming other agents to be truthful in the future). Note here the historical event need not be a truthful history, but is arbitrary.

Definition 4.2. *(Periodic Ex-post Individually Rationality) A dynamic direct mechanism is* periodic ex-post individually rational *if for all agents and conditioned on* any *historical event and conditioned on the current state of the other agents, the agent's expected future utility is non-negative under the truthful strategy (assuming other agents to be truthful in the future).*

These two stronger incentive constraints are ensured by the dynamic VCG mechanism provided by Bergemann and Välimäki [2007]. In our setting, our optimal mechanism also enjoys these properties.

Our main theorem relies on the following:

Assumption 4.1. *(Separable Environment) Assume the environment is (additively or multiplicatively) separable.*

Assumption 4.2. *(Concave Values) Let A_i, B_i and C_i be as defined in Definition 2.2. Assume A_i is differentiable and non-decreasing with respect to θ in both cases (additive and multiplicative); in the additive case, A_i is concave; in the multiplicative case, A_i is log-concave; B_i and C_i are both bounded and non-negative.*

Assumption 4.3. *(Monotone Hazard Rate) The density of $F_i(\cdot)$ exists for every agent i and is denoted by f_i. Also, the inverse hazard rate $\frac{1-F(\theta)}{f(\theta)}$ is decreasing in θ.*

Theorem 4.1. *(Optimality) Suppose Assumptions 4.1, 4.2 and 4.3 hold. Then, the Virtual Index Mechanism (as defined in Figure 1) is optimal. Furthermore, the Virtual Index Mechanism is periodic ex-post incentive compatible and individually rational.*

This theorem has a very surprising implication. The Virtual Index Mechanism is a mechanism designed to maximize revenue in a context where the mechanism has to create appropriate incentives for the agents to reveal private information over time. However, the revenue it produces is *identical* to the mechanism that can (publicly) observe the dynamically evolving private experiences of the agents (since it is also a feasible mechanism and optimal for the problem with complete dynamic monitoring, and by construction, it has the same revenue). Hence, the mechanism's capability to monitor the agents' dynamically evolving private experiences does not yield any revenue for the mechanism. We now formalize this claim.

Corollary 4.1. *Suppose Assumptions 4.1, 4.2 and 4.3 hold. The Virtual Index Mechanism is optimal for the setting with complete dynamic monitoring. Furthermore, the seller obtains the same expected revenue in both the setting with complete dynamic monitoring and without monitoring.*

We now proceed to describe the Virtual Index Mechanism in detail. The analysis of this theorem is contained in Subsection 4.4.

4.1 Virtual Surplus, Social Welfare, and a Fictitious Phase

If agents are truthful, we seek that the allocation rule maximizes the discounted sum of virtual values $\psi_i(\theta_i, e_{i,t}, \rho_{i,t})$, as defined in Eq. (4). However, in order to satisfy incentive constraints, we

---

The Virtual Index Mechanism:

At (fictitious) time $t = 0$,

  Each agent i reports $\hat{\theta}_{i,0}$.

  Each agent i is charged $p_{i,0}$, see Eq. (13).

At each time $t = 1, \ldots$

  Each agent i reports $\hat{\theta}_{i,t}$ and $\hat{e}_{i,t}$.

  Allocate to $i^\star$, an agent with the maximum $\mathcal{G}_i^{\hat{\theta}_{i,0}}(\hat{\theta}_{i,t}, \hat{e}_{i,t}, \rho_{i,t})$, see Eq. (9).

  Charge $i^\star$, $p_{i^\star,t}$, see Eq. (11).

Figure 1: Description of the Virtual Index Mechanism

consider a broader mechanism, which allows the agents to bid at a fictitious $t = 0$ phase — crucially, while we include this fictitious 0-th phase, we should point out that this phase actually occurs at $t = 1$.

Let us first understand the nature of the virtual surplus, under separable values. A key observation is that separability of the value functions implies that the virtual value ψ_i is an affine function of the value, *conditioned* on θ_i. Precisely,

Lemma 4.1. *(Affine Virtual Values) For a separable value function, the virtual value is given by:*

$$\psi_i(\theta_i, e_{i,t}, \rho_{i,t}) = \alpha_i(\theta_i) v_i(\theta_i, e_{i,t}, \rho_{i,t}) + \beta(\theta_i, \rho_{i,t}) \tag{7}$$

where functions $\alpha_i(\theta_i)$ and β_i are defined as follows:

- *(Additive)* $v_i(\theta_i, e_i, \rho_i) = A_i(\theta_i, \rho_i) + B_i(e_i, \rho_i)$

$$
\begin{aligned}
\alpha_i(\theta_i) &= 1 \\
\beta_i(\theta_i, \rho_i) &= -\frac{1 - F(\theta_i)}{f(\theta_i)} \frac{\partial A_i(\theta_i, \rho_i)}{\partial \theta_i};
\end{aligned}
$$

- *(Multiplicative)* $v_i(\theta_i, e_i, \rho_i) = A_i(\theta_i) B_i(e_i, \rho_i) - C_i(\rho_i)$

$$
\begin{aligned}
\alpha_i(\theta_i) &= 1 - \frac{1 - F(\theta_i)}{f(\theta_i)} \frac{A_i'(\theta_i)}{A_i(\theta_i)} \\
\beta_i(\theta_i, \rho_i) &= -\frac{1 - F(\theta_i)}{f(\theta_i)} \frac{A_i'(\theta_i)}{A_i(\theta_i)} C_i(\rho_i).
\end{aligned}
$$

The observation is that once θ_i is fixed, the virtual values are an affine function of the values — though this affine function could vary with time (in the additive $\beta(\cdot)$). Recall, in a static setting,

12

affine transformations of the social welfare function can be implemented via an affinely transformed VCG mechanism. Here, we provide a (time varying) transformation for the dynamic setting, using an affine transformation of the mechanism of Bergemann and Välimäki [2007]. This reduction satisfies the dynamic incentive constraints and leaves us only with the static incentive constraint required for eliciting initial types θ — this is the heart of our technical argument in Lemma 4.4.

The above structure motivates the use of a "fictitious phase", where we break the $t = 1$ phase into two parts. In the first part of this phase (which we label as $t = 0$), the agents make a report r of θ, which we use to specify the functions α and β. From then on, we proceed as an (affinely) transformed VCG mechanism, where agents are allowed to re-report θ (though α and β are pegged to the initial report r). We now specify this precisely.

4.2 Allocation Rule

To find the optimal allocation, we construct the following $(k + 1)$-armed bandit process (where k is the number agents), augmented with a 0-th arm which corresponds to a non-transitioning arm that always pays 0. For a vector $r \in \mathbb{R}_+^k$, define the *weighted social welfare* as:

$$W^r(\theta, e, \rho) = \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} \sum_{i=1}^{k} q_{i,t}\Big(\alpha_i(r_i)v_i(\theta_i, e_{i,t}, \rho_{i,t}) + \beta(r_i, \rho_{i,t})\Big)\Big|\theta, e_1 = e, \rho_1 = \rho\right], \qquad (8)$$

where for $i = 0$, $v_0(\cdot) = \alpha_0(\cdot) = \beta_0(\cdot, \cdot) = 0$. Note that the vector r substitutes θ in the α and β components of the virtual value (cf. Eq. (7)). With this structure, for $r = \theta$, $W^\theta(\theta, e, \rho)$ represents the virtual surplus and the desired revenue of the mechanism (see Corollary 3.1).

We use the initial phase $t = 0$ to allow the agents to set r. Subsequently, the allocation rule we use is the one that maximizes the weighted social welfare for this given r. In particular, for any r, we can find the algorithm that maximizes the weighted social welfare using the Gittins index (see Gittins [1989], Whittle [1982]).

Definition 4.3 (Virtual index). *For each agent i, the virtual index is defined as:*

$$\mathcal{G}_i^{r_i}(\theta_i, e_i, \rho_i) = \max_{\tau_i} \mathbb{E}\left[\frac{\sum_{t=1}^{\tau_i} \delta^{t-1}\xi^{r_i}(\theta_i, e_{it}, \rho_{it})}{\sum_{t=1}^{\tau_i} \delta^{t-1}}\Big|\theta_i, e_{i1} = e_i, \rho_{i1} = \rho_i\right] \qquad (9)$$

where the maximum is taken over all stopping times τ_i and

$$\xi^{r_i}(\theta_i, e_{it}, \rho_{it}) = \alpha_i(r_i)v_i(\theta_i, e_{it}, \rho_{it}) + \beta(r_i, \rho_{it}). \qquad (10)$$

It is well-known that the *Gittins index* policy (that which chooses the arm with highest index) is the (Bayes) optimal algorithm for multi-armed bandit problems. Hence, the allocation rule which chooses the highest virtual index is the optimal algorithm for the $(k + 1)$-armed bandit process where the goal is to maximize the weighted social welfare (where r is some fixed vector).

We now describe the allocation (specified in Figure 4), including reports and allocations. At time $t = 0$, each agent reports $\hat{\theta}_{i,0}$ their initial type θ_i. This initial report is used to set r above — in other words, their initial report determines the weights of the weighted social welfare function which the subsequent allocation tries to maximize. At ever subsequent time $t \geq 1$, they report $\hat{\theta}_{i,t}$ and $\hat{e}_{i,t}$ (the component $\rho_{i,t}$ is observed directly to the mechanism). This "freedom to correct" earlier misreports of θ_i leads to the stronger (periodic ex-post) notion of incentive compatibility.

4.3 Payment Rule

We now construct a payment scheme that makes the mechanism periodic ex-post incentive compatible. It turns out that it is only the $t = 0$ fictitious phase where all agents could potentially make a payment — thereafter, only the agent who is allocated pays. While, as specified in Figure 1, these payments occur before the allocation, this is inconsequential (as they could occur after the first allocation with no change to any of our guarantees).

As mentioned, the mechanism is the reminiscent of a static (affine) VCG mechanism, but with the added dynamic twist of a time-varying, additive offset. In Cavallo et al. [2006], Bergemann and Välimäki [2007], the payment of an agent after the allocation corresponds should be equal to the *externality* he imposes to other agents. In our context, the payment is an affine transformation of the externality imposed. In particular, if at time t the item is allocated to agent i, i pays the following amount:

$$p_{i,t} = \frac{(1-\delta)W_{-i,t}^{\hat{\theta}_0}(\hat{\theta}_t, \hat{e}_t, \rho_t) - \beta_i(\hat{\theta}_{i,0}, \rho_{i,t})}{\alpha_i(\hat{\theta}_{i,0})}, \tag{11}$$

where $W_{-i,t}^r$ is the *optimal virtual surplus* of the other agents. Namely

$$W_{-i,t}^r(\theta, e, \rho) = \max_{\mathcal{Q} \in \mathbb{Q}} \mathbb{E}\left[\sum_{t'=t}^{\infty}\sum_{j \neq i}\delta^{t'-t}q_{j,t'}\xi^{r_j}(\theta_j, e_{j,t'}, \rho_{j,t'})\middle| \theta, e_t = e, \rho_t = \rho\right]$$

where $\mathbb{Q}$ is the set of allocations rules (and j is summing over the other k arms, including the non-paying arm of 0). Also ξ is defined in Eq. (10).

Finally, we specify the payment at time 0, $p_{i,0}$. First define $P_i(\hat{\theta})$ as:

$$P_i(\hat{\theta}) = V_i(\hat{\theta}) - \int_0^{\hat{\theta}_i}\mathbb{E}\left[\sum_{t=1}^{\infty}\delta^{t-1}q_{i,t}^z\frac{\partial v_i(z, e_{i,t}, \rho_{i,t})}{\partial z}\middle| \hat{\theta}_i = z, \hat{\theta}_{-i}\right]dz \tag{12}$$

Note this is the desired payment of agent i conditioned on the vector $\hat{\theta}_0$ of reports of initial type at times 0 — this revenue maximizes the upper bound in Corollary 3.1. The price charged (to each agent i) is $P_i(\hat{\theta}_0)$ with a negative term to offset all expected future payments:

$$p_{i,0} = P_i(\hat{\theta}_0) - \mathbb{E}\left[\sum_{t=1}^{\infty}\delta^{t-1}q_{i,t}p_{i,t}\middle| \hat{\theta}_0\right] \tag{13}$$

This offset allows, in expectation, the revenue to be $P_i(\hat{\theta}_0)$ as desired. The heart of the proof is verifying incentive compatibility with respect to the initial report r.

4.4 Analysis of Theorem 4.1

We now give the outline of the proof of Theorem 4.1 using a series of lemmas. The proofs of these lemmas are given in the Appendix A.2. In the discussion of this subsection, we focus on the issue of incentive compatibility, and address the issue of individual rationality in the proof of the lemmas.

The first step of the proof is to show that the mechanism is periodic ex-post incentive compatible for periods $t \geq 1$, irrespective of the reports of the initial types $\hat{\theta}_0$ at the (fictitious) period 0. Recall that the mechanism implements an efficient allocation with respect to the "weights" that are assigned as a function of the initial reports; the proof technique is similar to Bergemann and Välimäki [2007].

14

Lemma 4.2. *Let Assumption 4.1 hold. Then, for any initial report $\hat{\theta}_0$, the Virtual Index Mechanism is periodic ex-post incentive compatible and individually rational for periods $t \geq 1$.*

The lemma above guarantees that, under the Virtual Index Mechanism, it is always a best response for agents to report their types truthfully regardless of the history, at any time $t \geq 1$ (assuming that other agents will be truthful in the future).

Therefore, we need only concern ourselves with period 0 deviations from the truthful strategy. To obtain incentive compatibility at time 0, we need some notion of monotonicity of the *future allocations* with respect to period 0's report. The following lemma provides this monotonicity result.

Before we show the monotonicity lemma, we introduce some notation. Denote the Virtual Index (immediate) allocation rule by $q^r(\theta, e, \rho)$, where r is the initial report. That is, for each r, we have a (subsequent) allocation rule $q^r(\cdot)$ which, at any time $t \geq 1$ assigns the item based on the reported state $(\hat{\theta}_t, \hat{e}_t, \rho_t)$.

Lemma 4.3. *(Monotonic Allocation) Let Assumptions 4.1, 4.2 and 4.3 hold. Then, for all (joint) states (θ, e, ρ) and any two initial reports r and r', which only differ in the i-th coordinate and $r_i \geq r'_i$, we have for the Virtual Index Mechanism that*

$$q_i^r(\theta, e, \rho) \geq q_i^{r'}(\theta, e, \rho).$$

Note that the lemma above defines an instantaneous notion of monotonicity: it establishes that at any time $t \geq 1$ and any reported state $(\hat{\theta}_t, \hat{e}_t, \rho_t)$, the allocation of the item to agent i is more likely if his first period report $\hat{\theta}_{i,0}$ is higher. The underlying multi-armed bandit stochastic process guarantees that this notion of monotonicity is sufficient for agent i's expected discounted sum of all his future values to be monotonic on $\hat{\theta}_{i,0}$. We, therefore, obtain in the following lemma that the Virtual Index Mechanism is also incentive compatible at period 0. This lemma is the key component of our technical argument.

Lemma 4.4. *Under Assumptions 4.1, 4.2 and 4.3, the Virtual Index Mechanism is both periodic ex-post incentive compatible and individually rational.*

We can now state the proof of our main theorem.

Proof of Theorem 4.1. From Lemma 4.4, we obtain that the Virtual Index Mechanism is both periodic ex-post incentive compatible and individually rational. Hence, From Eq. (13), we get that the revenue produced by the Virtual Index Mechanism is equal to

$$\text{Rev} = \sum_{i=1}^{k} \sum_{t=0}^{\infty} \mathbb{E}[p_{i,t}] = \sum_{i=1}^{k} P_i(\hat{\theta}_0),$$

where $P_i(\cdot)$ is as defined in Eq. (12). This value is equal to the bound given in Corollary 3.1. Therefore, the mechanism is optimal. □

References

Nikhil Agarwal, Susan Athey, and David Yang. Skewed bidding in pay per action auctions for online advertising. *American Economic Review Papers and Proceedings*, 2009.

Susan Athey and Ilya Segal. An efficient dynamic mechanism. *Working paper*, 2007.

Moshe Babaioff, Yogeshwer Sharma, and Aleksandrs Slivkins. Characterizing truthful multi-armed bandit mechanisms. In *ACM Conference on Electronic Commerce*, pages 79–88, 2009.

Marco Battaglini. Long-term contracting with markovian customers. *American Economic Review*, 95(3):637–658, 2005.

Dirk Bergemann and Juuso Välimäki. The dynamic pivot mechanism. *Working paper*, 2007.

Ruggiero Cavallo, David C. Parkes, and Satinder P. Singh. Optimal coordinated planning amongst self-interested agents with private state. In *Proceedings of the 22nd Conference in Uncertainty in Artificial Intelligence*, 2006.

Ruggiero Cavallo, David C. Parkes, and Satinder Singh. Efficient online mechanisms for persistent, periodically inaccessible self-interested agents. *Working paper*, 2007.

Rahul Deb. Optimal contracting of new experience goods. *Working paper*, 2008.

Peter Ëso and Balazs Szentes. Optimal information disclosure in auctions and the handicap auction. *Review of Economic Studies*, 74(3):705–731, 2007.

J. C. Gittins. *Allocation Indices for Multi-Armed Bandits*. Wiley, London, 1989.

Mohammad Mahdian and Kerem Tomak. Pay-per-action model for online advertising. In *Internet and Network Economics, Third International Workshop*, pages 549–557, 2007.

Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.

Roger Myerson. Optimal auction design. *Mathematics of Operations Research*, 1981.

Roger Myerson. Multistage games with communications. *Econometrica*, 54(2):323–358, 1986.

Hamid Nazerzadeh, Amin Saberi, and Rakesh Vohra. Dynamic cost-per-action mechanisms and applications to online advertising. In *Proceedings of the 17th International Conference on World Wide Web*, pages 179–188, 2008.

Mallesh Pai and Rakesh Vohra. Optimal dynamic auctions. *Working paper*, 2008.

David C. Parkes and Satinder P. Singh. An mdp-based approach for online mechanism design. In *Proceedings of the 17th Conference on Neural Information Processing Systems*, 2003.

Alessandro Pavan, Ilya Segal, and Juuso Toikka. Dynamic mechanism design: Incentive compatibility, profit maximization and information disclosure. *Working paper*, 2008.

Alessandro Pavan, Ilya Segal, and Juuso Toikka. Infinite-horizon mechanism design: The independent-schock approach. *Preliminary Draft*, 2009.

Gustavo Vulcano, Garrett van Ryzin, and Costis Maglaras. Optimal dynamic auctions for revenue management. *Management Science*, 48(11):1388–1407, 2002.

Peter Whittle. *Optimization Over Time*, volume 1. Wiley, Chichester, 1982.

A   Appendix

A.1   Analysis of Theorem 3.1

The proof of this theorem follows the outline of Myerson [1981] for establishing incentive compatibility in static mechanism design. In Lemma A.1, we derive an envelope condition that the utility of players participating in incentive compatible mechanism must satisfy. We then use this envelope condition to derive the desired result.

Lemma A.1. *Assume complete dynamic monitoring. Assume as well that the partial derivative $\frac{\partial v_i(\theta_i, e_{i,t}, \rho_{i,t})}{\partial \theta_i}$ exists for all θ_i, $e_{i,t}$ and $\rho_{i,t}$ and there exists some $B < \infty$ such that $|\frac{\partial v_i(\theta_i, e_{i,t}, \rho_{i,t})}{\partial \theta_i}| \leq B$ for all θ_i, $e_{i,t}$ and $\rho_{i,t}$. Then, any incentive compatible mechanism satisfies for all i, θ_i and θ_{-i}, for all players other i being truthful,*

$$U_i^{\mathcal{M}}(\theta_i) - U_i^{\mathcal{M}}(0, \theta_{-i}) = \int_0^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{\partial v_i(z, e_{i,t}, \rho_{i,t})}{\partial z}\bigg| \theta_i = z, \theta_{-i}\right] dz. \qquad (14)$$

Proof. In a direct mechanism in this setting, the agents' only report is $\hat{\theta}$ at the initial period. The mechanism design problem is therefore static and we use the classical results from Milgrom and Segal [2002] (Theorem 2) to obtain our envelope condition. We now show that these conditions are satisfied here.

For any mechanism $\mathcal{M}$ and initial type profile θ, let the expected utility of player i reporting type $\hat{\theta}_i$ is be given by

$$\hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \theta_i | \theta_{-i}) = \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} \left(q_{i,t} v_i(\theta_i, e_{i,t}, \rho_{i,t}) - p_{i,t}(\hat{\theta}, e_{i,t}, \rho_{i,t})\right)\bigg| \theta_{-i}\right].$$

Consider the term $\hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \cdot | \theta_{-i})$ applied to two different values θ_i and θ_i'. Taking the difference between the two and dividing by $\theta_i - \theta_i'$, we obtain

$$\frac{\hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \theta_i | \theta_{-i}) - \hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \theta_i' | \theta_{-i})}{\theta_i - \theta_i'} = \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \left(\frac{v_i(\theta_i, e_{i,t}, \rho_{i,t}) - v_i(\theta_i', e_{i,t}, \rho_{i,t})}{\theta_i - \theta_i'}\right)\bigg| \theta_{-i}\right].$$

Since $|q_{i,t}| \leq 1$ for all $\hat{\theta}$, $e_{i,t}$ and $\rho_{i,t}$, and the partial derivative $\frac{\partial v_i(\theta_i, e_{i,t}, \rho_{i,t})}{\partial \theta_i}$ exists for all $\hat{\theta}$, $e_{i,t}$ and $\rho_{i,t}$ and is bounded by B, we use Lebesgue's Dominated Convergence Theorem to obtain that the partial derivative $\frac{\partial \hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \theta_i | \theta_{-i})}{\partial \theta_i}$ exists for all θ_i and $\hat{\theta}$ and satisfies

$$\frac{\partial \hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \theta_i | \theta_{-i})}{\partial \theta_i} = \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{\partial v_i(\theta_i, e_{i,t}, \rho_{i,t})}{\partial \theta_i}\bigg| \theta_{-i}\right].$$

Furthermore, $|\frac{\partial \hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \theta_i | \theta_{-i})}{\partial \theta_i}| \leq \frac{B}{1-\delta}$ for all $\hat{\theta}_i$ and θ and, therefore, $\hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \cdot | \theta_{-i})$ is absolutely continuous for any $\hat{\theta}_i$ and θ_{-i}. Therefore, the function $\hat{U}_i^{\mathcal{M}}(\hat{\theta}_i, \theta_i | \theta_{-i})$ satisfies the conditions of Milgrom and Segal [2002] Theorem 2, yielding Eq. (14). □

Proof of Theorem 3.1. Consider first the utility $U_i^{\mathcal{M}}(\theta)$ of an agent i under an initial type profile θ, which is given by

$$U_i^{\mathcal{M}}(\theta) - U_i^{\mathcal{M}}(0, \theta_{-i}) = \int_0^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{\partial v_i(z, e_{i,t}, \rho_{i,t})}{\partial z}\bigg|\theta_i = z, \theta_{-i}\right] dz$$

from Eq. (14). Taking the expectation of this term over all possible type profiles θ, we obtain

$$\mathbb{E}[U_i^{\mathcal{M}}(\theta) - U_i^{\mathcal{M}}(0, \theta_{-i})] = \int_0^{\Theta_i} \int_0^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{\partial v_i(z, e_{i,t}, \rho_{i,t})}{\partial z}\bigg|\theta_i = z\right] dz f_i(\theta_i) d\theta_i.$$

Inverting the order of integration,

$$\begin{aligned}
\mathbb{E}[U_i^{\mathcal{M}}(\theta) - U_i^{\mathcal{M}}(0, \theta_{-i})] &= \int_0^{\Theta_i} \int_z^{\Theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{\partial v_i(z, e_{i,t}, \rho_{i,t})}{\partial z}\bigg|\theta_i = z\right] f_i(\theta_i) d\theta_i dz \\
&= \int_0^{\Theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{\partial v_i(z, e_{i,t}, \rho_{i,t})}{\partial z}\bigg|\theta_i = z\right] (1 - F_i(z)) dz.
\end{aligned}$$

By multiplying and dividing the right-hand side of the equation above by the density $f_i(z)$ we obtain an unconditional expectation,

$$\mathbb{E}[U_i^{\mathcal{M}}(\theta) - U_i^{\mathcal{M}}(0, \theta_{-i})] = \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \frac{\partial v_i(\theta_i, e_{i,t}, \rho_{i,t})}{\partial \theta_i}\right]. \tag{15}$$

Now note that the total revenue of the mechanism is given by the sum of the payments from the agents, which themselves are the difference between the value of the allocations to the agents and the utility they obtain, i.e.,

$$\text{Rev}^{\mathcal{M}} = \sum_{i=1}^{k} \mathbb{E}[P_i^{\mathcal{M}}(\theta)] = \mathbb{E}[V_i^{\mathcal{M}}(\theta) - U_i^{\mathcal{M}}(\theta)].$$

Combining the definition of the value $V_i^{\mathcal{M}}(\theta)$ with Eq. (15),

$$\text{Rev}^{\mathcal{M}} = \sum_{i=1}^{k} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} v_i(\theta_i, e_{i,t}, \rho_{i,t}) - \sum_{t=1}^{\infty} \delta^{t-1} q_{i,t} \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} \frac{\partial v_i(\theta_i, e_{i,t}, \rho_{i,t})}{\partial \theta_i} - U_i^{\mathcal{M}}(0, \theta_{-i})\right],$$

yields the desired result by the plugging in the definition of the virtual value ψ. □

A.2 Omitted Proofs

Proof of Lemma 4.1. The claim for additive separability is trivial. In the multiplicatively separable case, we have:

$$
\begin{aligned}
\psi_i(\theta_i, e_i, \rho_i) &= A_i(\theta_i) B_i(e_i, \rho_i) - C_i(\rho_i) - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} A_i'(\theta_i, \rho_i) B_i(e_i, \rho_i) \\
&= \left(A_i(\theta_i) - \frac{1 - F_i(\theta_i)}{f_i(\theta_i)} A_i'(\theta_i, \rho_i) \right) B_i(e_i, \rho_i) - C_i(\rho_i) \\
&= \alpha_i(\theta_i) A_i(\theta_i) B_i(e_i, \rho_i) - C_i(\rho_i) \\
&= \alpha_i(\theta_i) v_i(\theta_i, e_i, \rho_i) + (\alpha_i(\theta_i) - 1) C_i(\rho_i) \\
&= \alpha_i(\theta_i) v_i(\theta_i, e_i, \rho_i) + \beta_i(\theta_i, \rho_i)
\end{aligned}
$$

and the claim follows. $\qquad\square$

Proof of Lemma 4.2. The proof follows the outline of Bergemann and Välimäki [2007], with "virtual values" replacing values as the mechanism's objective. In this proof we assume an initial type report profile $\hat{\theta}_0$ (not necessarily truthful reports) and concern ourselves only with periods $t \geq 1$. Let $W_t^{\mathcal{R}_i}(\theta, e, \rho)$ denote the discounted future virtual surplus, at time $t \geq 1$ and state (θ, e, ρ), when agent i uses reporting strategy $\mathcal{R}_i$ and other agents are truthful.

$$
W_t^{\mathcal{R}_i}(\theta, e, \rho) = \max_{\mathcal{Q} \in \mathbb{Q}} \mathbb{E} \left[\sum_{t'=t}^{\infty} \sum_i \delta^{t'-t} q_{j,t'}^{\mathcal{R}_i} \left(\alpha_j(\hat{\theta}_{j,0}) v_j(\theta_j, e_{j,t'}, \rho_{j,t'}) + \beta(\hat{\theta}_{j,0}, \rho_{j,t'}) \right) \middle| \theta, e_t = e, \rho_t = \rho \right]
$$

where $\mathbb{Q}$ is the set of all allocation rules and $q_{j,t'}^{\mathcal{R}_i}$ denotes the allocation induced by $\mathcal{R}_i$ (other agents are assumed truthful).

Similarly define virtual surplus without agent i:

$$
W_{-i,t}(\theta, e, \rho) = \max_{\mathcal{Q} \in \mathbb{Q}_{-i}} \mathbb{E} \left[\sum_{t'=t}^{\infty} \sum_{j \neq i} \delta^{t'-t} q_{j,t'} \left(\alpha_j(\hat{\theta}_{j,0}) v_j(\theta_j, e_{j,t'}, \rho_{j,t'}) + \beta(\hat{\theta}_{j,0}, \rho_{j,t'}) \right) \middle| \theta, e_t = e, \rho_t = \rho \right],
$$

where $\mathbb{Q}_{-i}$ is the set of allocation rules that never assign items to agent i. Note W_{-i} is defined without reference to $\mathcal{R}_i$ because agent i does not effect the allocation. In addition, let $m_{i,t}$ denote the marginal contribution of agent i, at time t, to the virtual surplus, i.e.,

$$
m_{i,t} = W^{\mathcal{R}_i}(\theta, e_t, \rho_t) - W_{-i}(\theta, e_t, \rho_t) - \delta \mathbb{E}[W^{\mathcal{R}_i}(\theta, e_{t+1}, \rho_{t+1}) - W_{-i}(\theta, e_{t+1}, \rho_{t+1})]. \tag{16}
$$

If the mechanism does not allocate the item to agent i at time t, then $m_{i,t} = p_{i,t} = 0$. But, if the mechanism does allocate the item to agent i at time t, then

$$
W^{\mathcal{R}_i}(\theta, e_t, \rho_t) = \left(\alpha_i(\hat{\theta}_{i,0}) v_i(\theta_i, e_{i,t}, \rho_{i,t}) + \beta(\hat{\theta}_{i,0}, \rho_{i,t}) \right) + \delta \mathbb{E}[W^{\mathcal{R}_i}(\theta, e_{t+1}, \rho_{t+1})].
$$

Since an agent's state does not change in the absence of an allocation, if the item is allocated to i at time t, we have

$$
W_{-i}(\theta, e_t, \rho_t) = W_{-i}(\theta, e_{t+1}, \rho_{t+1}).
$$

19

Hence, we obtain a marginal contribution (cf. Eq. (16)) for agent i at time t of

$$\begin{aligned}
m_{i,t} &= \left(\alpha_i(\hat{\theta}_{i,0})v_i(\theta_i, e_{i,t}, \rho_{i,t}) + \beta(\hat{\theta}_{i,0}, \rho_{i,t})\right) - (1-\delta)W_{-i}(e_t, \rho_t) \\
&= \alpha_i(\hat{\theta}_{i,0})\left(v_i(\theta_i, e_{i,t}, \rho_{i,t}) - p_{i,t}\right),
\end{aligned}$$

where the price $p_{i,t}$ is given in Eq. (11). Noting that $m_{i,t}$ can only be different than zero if $q_{i,t}^{\mathcal{R}_i} = 1$, we obtain that the expected future utility of agent i at time t given reporting strategy $\mathcal{R}_{i,t}$ is

$$\begin{aligned}
U_{i,t}^{\mathcal{R}_i}(\theta, e_{i,t}, \rho_{i,t}) &= \mathbb{E}\left[\sum_{t'=t}^{\infty}\delta^{t'-1}(q_{i,t'}^{\mathcal{R}_i}v_i(\theta_i, e_{i,t}, \rho_{i,t}) - p_{i,t})\right] \\
&= \frac{1}{\alpha_i(\hat{\theta}_{i,0})}\mathbb{E}\left[\sum_{t'=t}^{\infty}\delta^{t'-t}m_{i,t}\right] \\
&= \frac{1}{\alpha_i(\hat{\theta}_{i,0})}\left(W^{\mathcal{R}_i}(\theta, e_t, \rho_t) - W_{-i}(\theta, e_t, \rho_t)\right).
\end{aligned}$$

Note that W_{-i} is independent of all of agent i's reports. Also, $W_{-i}^{\mathcal{R}_i}(\hat{\theta}_t, \hat{e}_t, \rho_t)$ is maximized if i reports truthfully since W is define as the maximum virtual surplus obtained by an allocation with respect to the true (joint) state. Therefore, we obtain that the mechanism is periodic ex-post incentive compatible. Observe as well that $W^{\mathcal{T}} \geq W_{-i}$, where $\mathcal{T}$ denotes the truthful strategy; yielding that the mechanism is also periodic ex-post individually rational. $\square$

Proof of Lemma 4.3. The Virtual Index Mechanism allocates the item to the agent with the highest "virtual index". Therefore, it is sufficient to show that the virtual value ψ_i is non-decreasing on the initial report of the agent. Therefore, it suffices to show that both $\alpha_i(\cdot)$ and $\beta_i(\cdot, \rho)$ are non-decreasing functions of θ_i under the assumptions of the lemma. With this it directly follows that the Virtual Index is monotonic in the initial reports. Let $\eta_i(\theta_i)$ denote the hazard rate, i.e.,

$$\eta_i(\theta_i) = \frac{f_i(\theta_i)}{1 - F_i(\theta_i)}.$$

In the additive case,

$$\beta_i'(\theta_i, \rho_i) = \frac{\eta_i'(\theta_i)}{\eta_i^2(\theta_i)}A_i'(\theta_i) - \frac{1}{\eta_i(\theta_i)}A_i''(\theta_i)$$

where $(\cdot)'$ denotes a partial derivative with respect to θ_i. By the assumptions that A_i is concave and non-decreasing, and the hazard rate is non-negative and increasing, we have that the above is non-negative.

In the multiplicative case, first note that $\alpha_i(\theta_i) = 1 - \frac{1}{\eta_i(\theta_i)}(\log A_i(\theta_i))'$, so

$$\alpha_i'(\theta_i) = \frac{\eta_i'(\theta_i)}{\eta_i^2}\frac{A_i'(\theta_i)}{A_i(\theta_i)} - \frac{1}{\eta_i(\theta_i)}(\log A_i(\theta_i))''$$

which is non-negative by the assumptions. Since $\beta_i = (\alpha_i - 1)C_i$ and since C_i is positive, β_i is also non-decreasing. $\square$

Proof of Lemma 4.4. Let $U_i(\theta)$ be the expected utility of i in under the Virtual Index Mechanism conditioned on the initial types being θ (and everyone behaving truthfully). By construction, see Eq. (12), we have:

$$U_i(\theta) = V_i(\theta) - P_i(\theta) = \int_0^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty}\delta^{t-1}q_{i,t}^z\frac{\partial v_i(z,e_{i,t},\rho_{i,t})}{\partial z}\bigg|\theta_i = z, \theta_{-i}\right]dz \qquad (17)$$

Here, we are slightly abusing notation in that we are only specifying the superscript with respect to z in $q_{i,t}^z$ (it implicitly also depends on θ_{-i}). Observer that $U_i(0) = 0$. Moreover, because v_i is increasing in θ_i, U_i is non-negative. Hence, the mechanism is ex-post individually rational with respect to the (first) report of the initial type. We now prove that the mechanism is ex-post incentive compatible with respect to the (first) report of the initial type.

Let $\tilde{U}_i(\theta;\theta_i')$ be the utility of agent i conditioned on: the initial types being θ; the initial report in the 0-th period of i being θ_i'; where i behaves optimally thereafter; and where all other agents behave truthfully (conditioned on their initial type and report being θ_{-i}). To prove (periodic ex-post) incentive compatibility, we must show that for all θ and all θ_i',

$$U_i(\theta) \geq \tilde{U}_i(\theta;\theta_i').$$

Let θ' equal θ in all coordinates except in coordinate i, where it equals to θ_i'. Lemma 4.2 shows that truthfulness is the optimal continuation strategy from time $t \geq 1$ onwards. Hence,

$$U_i(\theta') = \tilde{U}_i(\theta';\theta_i').$$

Thus, it suffices to show:

$$U_i(\theta) - U_i(\theta') \geq \tilde{U}_i(\theta;\theta_i') - \tilde{U}_i(\theta';\theta_i') \qquad (18)$$

for all θ and θ_i'. Now we write this condition more explicitly.

Now let us consider the mechanism $q^{\theta'}(\cdot)$ from $t \geq 1$, where θ' is fixed. Let $U_i^{\theta'}(\theta)$ be the utility of this mechanism from $t \geq 1$ onwards (where truthful reporting occurs). By Lemma 4.2, we have that $q^{\theta'}$ is an incentive compatible allocation. Hence, the envelope lemma implies:[3]

$$U_i^{\theta'}(\theta) - U_i^{\theta'}(\theta') = \int_{\theta_i'}^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty}\delta^{t}q_{i,t}^{\theta'}\frac{\partial v_i(z,e_{i,t},\rho_{i,t})}{\partial z}\bigg|\theta_i = z, \theta_{-i}\right]dz$$

Now note that at $t = 0$ the charge only depends on the initial report (e.g., θ'). Hence,

$$\tilde{U}_i(\theta;\theta_i') - \tilde{U}_i(\theta';\theta_i') = U_i^{\theta'}(\theta) - U_i^{\theta'}(\theta') = \int_{\theta_i'}^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty}\delta^{t-1}q_{i,t}^{\theta'}\frac{\partial v_i(z,e_{i,t},\rho_{i,t})}{\partial z}\bigg|\theta_i = z, \theta_{-i}\right]dz \quad (19)$$

where now the evolution of the process is under the mechanism $q^{\theta'}$.

Combining Eq (17) and (19), we have that Eq. (18) is equivalent to:

$$\int_{\theta_i'}^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty}\delta^{t-1}q_{i,t}^z\frac{\partial v_i(z,e_{i,t},\rho_{i,t})}{\partial z}\bigg|\theta_i = z, \theta_{-i}\right]dz \geq \int_{\theta_i'}^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty}\delta^{t-1}q_{i,t}^{\theta'}\frac{\partial v_i(z,e_{i,t},\rho_{i,t})}{\partial z}\bigg|\theta_i = z, \theta_{-i}\right]dz$$

---

[3]The envelope lemma requires that $|\frac{\partial v_i(\theta_i,e_i,\rho_i)}{\partial \theta_i}| \leq B$ for some $B < \infty$ for all θ_i, e_i and ρ_i. Assumption 4.2, together with the compact support of θ_i, guarantee that this condition holds.

We can couple the two expectations by sampling the initial types and an infinite sequence of experiences — then each process is evolved on this sample using the appropriate allocations. Thus, this condition is equivalent to:

$$\int_{\theta_i'}^{\theta_i} \mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1}\left(q_{i,t}^z \frac{\partial v_i(z, e_{i,t}, \rho_{i,t})}{\partial z} - q_{i,t}^{\theta'} \frac{\partial v_i(z, e_{i,t}', \rho_{i,t}')}{\partial z}\right)\bigg| \theta_i = z, \theta_{-i}\right] dz \geq 0$$

where $e_{i,t}'$ and $\rho_{i,t}'$ denote the experiences under the allocation $q^{\theta'}$.

Now let τ_k be the time at which the k-th allocation to i occurs under the $q_{i,t}^z$ (assumed to be infinite if the k-th allocation does not occur), and let τ_k' be this time under $q_{i,t}^{\theta'}$. By our coupled process, we have that $e_{i,\tau_k'}'$ equals e_{i,τ_k} for all k. This is because after k allocations, both private experiences have advanced precisely $k-1$ times. Similarly, the same is true for the public experiences. Hence, the above condition is equivalent to:

$$\int_{\theta_i'}^{\theta_i} \mathbb{E}\left[\sum_{k=0}^{\infty}\left(\delta^{\tau_k} - \delta^{\tau_k'}\right) \frac{\partial v_i(z, e_{i,\tau_k}, \rho_{i,\tau_k})}{\partial z}\bigg| \theta_i = z, \theta_{-i}\right] dz$$

since the experiences at τ_k and τ_k' are identical.

Without loss of generality, assume that $\theta > \theta'$. To complete the proof of incentive compatibility, we show that $\tau_k' \geq \tau_k$, using the monotonicity of the allocation (Lemma 4.3). This implies that:

$$q_i^{\theta'}(z, e_t, \rho_t) \leq q_i^z(z, e_t, \rho_t).$$

We proceed inductively. For $k = 1$, note that until the first allocation occurs in either process, the state of i is identical at each timestep under both allocations — this is because the Gittins allocation is an index mechanism (so the states of the other agents advance identically when i is not present). Hence, monotonicity implies that the first allocation under $q^{\theta'}$ cannot occur before that time under q^z. Taking the inductive step, assume that $\tau_{k-1}' \geq \tau_{k-1}$. If $\tau_{k-1}' \geq \tau_k$, then we are done. If not, this implies at time τ_{k-1}', agent i has been allocated precisely $k-1$ times in both processes, and so she has an identical state in either process. Hence, an identical argument to that of the first period shows that the k-th allocation (if it occurs) under $q^{\theta'}$ cannot occur before that under q^z, which completes the proof. $\square$