Detecting Profitable Deviations

David Rahman*
University of Minnesota

First draft: April 16, 2009
This version: February 9, 2010

Abstract

This paper finds necessary and sufficient conditions for implementability as well as interim implementability in a quasi-linear context with arbitrary type spaces. By viewing an agent's deviation gains as payments in a hypothetical zero-sum game between a principal and an agent, I show that an allocation is (i) implementable if and only if every profitable deviation is detectable, and (ii) interim implementable if and only if every infinitesimally detectable deviation is at most infinitesimally profitable. I also provide several natural extensions of these results, including complete characterizations of revenue equivalence (both ex post and interim), budget balanced implementation, bargaining with interdependent values, moral hazard, optimal mechanisms, surplus extraction, and revealed stochastic preference.

JEL Classification: D21, D23, D82.
Keywords: mechanism design, implementation, duality.

*I owe many thanks to Sushil Bikhchandani and Christoph Müller for numerous conversations and encouragement. Please send any comments to dmr@umn.edu.

1 Introduction

Understanding implementable allocations is an essential topic of mechanism design. This problem has been addressed in many general as well as specific contexts. In this paper I will restrict attention to the quasi-linear setting but otherwise keep the model as general as possible. Specifically, I characterize implementable allocations on arbitrary type spaces without placing any restrictions on utilities. Furthermore, I characterize interim implementable allocations on arbitrary type spaces without placing any restrictions on expected utilities. I also explore interesting extensions.

The paper's leading contribution is the intuition behind the main results because it helps to understand the economics of incentive compatibility—just like the planner's problem helps understand its decentralization through prices. Here it is:

Consider the following hypothetical zero-sum game between a principal and an agent. The principal chooses a vector of report-contingent money payments ξ and the agent chooses a reporting strategy π, i.e., a map from types to probability distributions over reports. The principal pays the agent the deviation gains from reporting according to π rather than just planning to tell the truth. These gains arise from both changes in the allocation (call these "gross" gains) and changes in money payments as a result of misreporting. By definition, an allocation is implementable if for some ξ the agent cannot make positive deviation gains. Since the agent can guarantee non-negative gains by reporting truthfully, implementability is equivalent to both the principal and agent receiving a payoff of zero, in other words, the value of this hypothetical game is zero. If this game is finite then by the Minimax Theorem it doesn't matter who goes first. Hence, an allocation is implementable if and only if for any reporting strategy π for the agent there is a payment scheme ξ that makes unprofitable reporting according to π. However, the crucial insight that follows from the Minimax Theorem is that now different ξ's may be chosen to discourage different π's.

If the strategy π is detectable, i.e., it makes the probability distribution over reported types differ from that of actual types, then it is easy to find a ξ that discourages π: a type whose probability of being reported is larger than that with which it realizes is dubbed a "bad" report. Otherwise, call it a "good" report. Now pay the agent for good reports, charge him for bad ones, and increase the wedge between good and bad reports by increasing the scale of this payment scheme until any utility gains are overwhelmed by monetary losses associated with π.

If π is undetectable then the agent receives the same expected money payment whether he reports truthfully or according to π, regardless of ξ. Hence, the deviation gains from π are non-positive if and only if the gross gains from π are non-positive. Call such a π "unprofitable." This yields the finite version of Theorem 1: an allocation is implementable if and only if every undetectable deviation is unprofitable. The main obstacle in establishing Theorem 1 is showing that this intuition holds for arbitrary type spaces, not just finite ones.

Now suppose that the principal observes a signal, such as output or others' types, possibly correlated with the agent's true type, and may pay the agent contingent on the realized signal as well as the agent's report. When such a payment scheme exists to induce truth-telling, we will say that an allocation is interim implementable. Theorem 5 characterizes interim implementability on arbitrary type and signal spaces. If the set of types is finite then the same argument as that for Theorem 1 follows, except that the notion of detectability takes into account the signal's distribution. When the set of types is infinite, Theorem 5 requires a slightly stronger condition than in the finite case, which may be interpreted mathematically as bounded steepness. Strategically, an allocation is interim implementable if and only if every infinitesimally detectable deviation is at most infinitesimally profitable. Intuitively, the problem is this. Consider a sequence of deviations that converge towards being undetectable. If the gross gains from these deviations are positively bounded below then there is an "infinitesimal" deviation with positive profit, and this cannot be discouraged with a payment scheme that also discourages the non-infinitesimal deviations.

An important "windfall" advantage of these results is that they are easy to generalize. This contrasts, for instance, Rochet's (1987) Theorem. (This result is very closely related—indeed, logically equivalent—to Theorem 1, see Corollary 1.) Even though Rochet's Theorem has a beautifully simple, elegant proof, arguably that proof is "too short" because it is ad hoc and difficult to generalize. By contrast, the proofs of Theorems 1 and 5 are systematic and easy to generalize. In addition, Rochet's proof is mathematical rather than strategic, unlike the proofs of Theorems 1 and 5.

To illustrate this advantage, I extend the main results of this paper in a number of interesting directions. With regard to Theorem 1, I provide the following characterizations: (i) ex post revenue equivalence with a strategic interpretation (contrast with the graph-theoretic approach of Heydenreich et al., 2009), and (ii) budget-balanced ex post implementation with interdependent values.

With regard to extending Theorem 5, I characterize: (i) budget-balanced interim implementation, (ii) existence of bargaining solutions with interdependent values, (iii) both full surplus extraction and virtually full surplus extraction on arbitrary type spaces,[1] (iv) interim implementability with moral hazard, (v) interim revenue equivalence[2] with a strategic interpretation even if types are correlated, (vi) revealed stochastic preference.[3]

As a corollary to the duality approach explored in this paper, I obtain a subdifferential characterization of the set of implementing payment schemes for a given allocation, and offer some strategic intuition behind it.[4]

The remainder of the paper is organized as follows. In the next section, I begin by formally stating Theorem 1, which characterizes implementable allocations. I then provide a detailed comparison of Theorem 1 and Rochet's Theorem and characterize the set of implementing payment schemes as a complete lattice. Then I consider the extensions of Theorem 1 described above.

In Section 3 I begin by formally defining the bounded steepness condition that I label as every infinitesimally detectable deviation being at most infinitesimally profitable. Then I state Theorem 5, which characterizes interim implementability in terms of this condition. Finally, I present several examples to illustrate the result and how the characterizing condition relates to bounded steepness. In Section 4, I extend Theorem 5 in the various directions described above.

Omitted proofs and ancillary results are collected in the appendix.

[1] My results contrast the work of Cremer and McLean (1988) and McAfee and Reny (1992a) in important ways. First, nd necessary and su cient conditions for full surplus extraction on arbitrary type spaces. Secondly, I clarify the role that McAfee and Reny's \ nite participation fees" played in their derivation of virtual surplus extraction, as well as provide conditions necessary and su cient conditions for it. Finally, I also describe how detectability is a di erent, much weaker condition than Cremer and McLean's (1988) convex independence.

[2] I characterize both when there is a unique expected payment schedule (subject to a constant) as well as when there is a unique contingent schedule subject to a constant. By contrast, Heydenreich et al. (2009) only characterize unique expected payments assuming that types are independent.

[3] Although in this paper I restrict attention to revealed stochastic preference under the assumption of quasi-linearity (this assumption can be relaxed, see Afriat, 1967, p. 72), I otherwise generalize the work of McFadden and Richter (1990) and McFadden (2005) in several ways. First, I allow for prices, or \budgets" to vary randomly in a way that may be correlated with stochastic preferences. This allows for a richer interpretation of stochastic preferences as those of a population. Secondly, I do not restrict attention to compact metric type spaces.

[4] For a related result in the case of ex post implementation, see Kos and Messner (2009).

2 Ex Post Implementation

Consider an agent with private information parametrized by a *type* t in some set T, interpreted as the collection of all types that the principal deems possible. Let X be a nonempty set of outcomes, and denote by $v(t,x) \in \mathbb{R}$ his utility from outcome $x \in X$ when his type is $t \in T$. An *allocation* is any map $\mathbf{x} : T \to X$. An *incentive scheme* (or simply *scheme*) is any map $\xi : T \to \mathbb{R}$, interpreted as report-contingent linear transfers intended to induce the agent to report his type truthfully.

An allocation $\mathbf{x}$ is called *implementable* if a scheme ξ exists such that

$$v(t,\mathbf{x}(s)) - v(t,\mathbf{x}(t)) \;\leq\; \xi(s) - \xi(t) \qquad\qquad \forall (t,s) \in T \times T. \tag{1}$$

A *finite reporting strategy* (or simply *reporting strategy*) is any map $\pi : S \to \Delta(S)$ defined on some finite subset S of T, where $\pi(s|t)$ is interpreted as the conditional probability that the agent's report equals s when his true type equals t. For instance, the *truthful* reporting strategy θ is defined for every pair (s,t) by $\theta(s|t) = 1$ if $s = t$ and 0 otherwise. A *finite deviation* (or simply *deviation*) is any reporting strategy that isn't truthful, i.e., it lies with positive probability conditional on some type.

A reporting strategy π is called *undetectable* if

$$\sum_{t \in S} \pi(s|t) \;=\; 1 \qquad\qquad \forall s \in S. \tag{2}$$

In other words, π is doubly stochastic. Otherwise, π is called *detectable*. Suppose that types are drawn from S according to the uniform distribution (so the probability of each type equals $1/|S|$).[5] Intuitively, π is undetectable if the probability distribution over reports coincides with that of actual types, $1/|S|$.

A reporting strategy π is **x**-*profitable* if it yields a higher ex ante payoff than truthful reporting assuming that the incentive scheme is identically zero, i.e.,

$$\sum_{(t,s)} \pi(s|t)[v(t,\mathbf{x}(s)) - v(t,\mathbf{x}(t))] \;>\; 0. \tag{3}$$

Otherwise, π is called **x**-*unprofitable*. Intuitively, (3) says that the expected utility from reporting according to π is greater than that from reporting truthfully.

Theorem 1. *A given allocation* $\mathbf{x}$ *is implementable if and only if every* **x**-*profitable deviation is detectable.*

[5] Although the uniform assumption is useful, it is by no means necessary for any of the results.

2.1 Rochet's Theorem

Let us compare Theorem 1 to Rochet's (1987) Theorem, which states that a given allocation $\mathbf{x}$ is implementable if and only if it is *cyclically monotone*, i.e., for every finite cycle $(t_1, \ldots, t_{m+1})$ such that $t_1 = t_{m+1}$,

$$\sum_{k=1}^{m} v(t_{k+1}, \mathbf{x}(t_k)) - v(t_k, \mathbf{x}(t_k)) \;\leq\; 0. \tag{4}$$

Rochet's proof of this result (adapted from Rockafellar, 1970) is remarkable not only for its simplicity, but also because it is constructive: if an allocation is implementable, the proof produces an incentive scheme that implements it. We include it below.

Proof of Rochet's Theorem. For sufficiency, suppose that $\mathbf{x}$ is implementable and let $(t_1, \ldots, t_{m+1})$ be a finite cycle, so $t_1 = t_{m+1}$. By hypothesis, there exists a scheme ξ such that $v(t_{k+1}, \mathbf{x}(t_k)) - v(t_{k+1}, \mathbf{x}(t_{k+1})) \leq \xi(t_k) - \xi(t_{k+1})$ for every $k \in \{1, \ldots, m\}$. Adding up all these inequalities yields $\sum_{k=1}^{m} v(t_{k+1}, \mathbf{x}(t_k)) - v(t_{k+1}, \mathbf{x}(t_{k+1})) \leq 0$, or equivalently $\sum_{k=1}^{m} v(t_{k+1}, \mathbf{x}(t_k)) - v(t_k, \mathbf{x}(t_k)) \leq 0$. Conversely, fix $t_0 \in T$ and define $U(t_0, t) = \sup \sum_{k=1}^{m} v(t_{k+1}, \mathbf{x}(t_k)) - v(t_k, \mathbf{x}(t_k))$, where the sup is with respect to all finite sequences $(t_1, \ldots, t_m)$ such that $t_1 = t_0$ and $t_m = t$. By cyclic monotonicity, $U(t_0, t_0) = 0$. Moreover, $U(t_0, t_0) \geq U(t_0, t) + v(t_0, \mathbf{x}(t)) - v(t, \mathbf{x}(t))$ for all t, so $U(t_0, t)$ is finite. Hence, $U(t_0, t) \geq U(t_0, s) + v(t, \mathbf{x}(s)) - v(s, \mathbf{x}(s))$ for all (t, s). Finally, if $\xi(t) = v(t, \mathbf{x}(t)) - U(t_0, t)$ then $v(t, \mathbf{x}(s)) - v(t, \mathbf{x}(t)) \leq \xi(s) - \xi(t)$. □

To relate cyclic monotonicity with Theorem 1, we will show that a cycle can be interpreted as an undetectable reporting strategy with rational probabilities. In so doing, we provide another characterization of implementability in terms of permutations, which may be thought of as undetectable "pure" reporting strategies.

A *finite permutation* (or simply *permutation*) is any map $\sigma : S \to S$ defined on some finite subset S of T such that σ is both one-to-one and onto. A permutation σ can be written as a pure reporting strategy, π_σ defined pointwise by $\pi_\sigma(s|t) = 1$ if $s = \sigma(t)$ and 0 otherwise. By virtue of σ being a permutation, it follows that for every $t \in S$, (i) there exists a unique $s \in S$ such that $\pi_\sigma(s|t) = 1$, and (ii) there exists a unique $s \in S$ such that $\pi_\sigma(t|s) = 1$. Therefore, π_σ is undetectable.

Corollary 1. *The following statements are equivalent for a given allocation* $\mathbf{x}$*:*
(i) *Every undetectable deviation is* $\mathbf{x}$*-unpro table.*
(ii) $\mathbf{x}$ *is cyclically monotone.*
(iii) *Every permutation is* $\mathbf{x}$*-unpro table.*

Proof. By Rochet's Theorem and Theorem 1, (i) is equivalent to (ii), and (i) is equivalent to (iii) by linearity of $g(w)$ with respect to $g \in \mathbb{R}^{(T\times T)}$ together with the Birkhoff-von Neumann Theorem, which states that the set of doubly stochastic matrices is the convex hull of the set of permutation matrices. □

It is instructive to consider a more direct argument for Corollary 1: (iii) implies (i) by the Birkhoff-von Neumann Theorem. (ii) implies (iii) because a permutation is a finite collection of cycles, each without repetitions, and cyclic monotonicity applied to a permutation implies that it is $\mathbf{x}$-unprofitable. Finally, (i) implies (ii) by representing a cycle as an undetectable reporting strategy with rational probabilities as follows.

Indeed, let $(t_1, \ldots, t_{m+1})$ be a cycle, so $t_1 = t_{m+1}$. Let $S = \{s_1, \ldots, s_\ell\}$, with $\ell \leq m$, be the set of distinct elements in the cycle, and write $[s_j]$ for the number of times that s_j appears in $(t_1, \ldots, t_m)$. Also write $[s_i, s_j]$ for the number of times that s_i appears immediately before s_j in $(t_1, \ldots, t_{m+1})$. Let s_0 be any type that solves $[s_0] = \max_j [s_j]$, and define $\pi(s|t) = [t, s]/[s_0]$ if $s \neq t$ and $1 - \sum_{s\neq t}[t, s]/[s_0]$ otherwise. Clearly, π is a reporting strategy. To see that π is undetectable, notice that since $(t_1, \ldots, t_{m+1})$ is a cycle, $\sum_{s\neq t}[t, s] = \sum_{s\neq t}[s, t]$ for every t: the outflow from t equals the inflow to t. Finally, it is clear that by construction every element of π is a rational number, since every element is obtained as the difference between a natural number and the ratio of two natural numbers.

By Corollary 1, (iii) also characterizes implementability. Since the set of permutations are the extreme points of the set of doubly stochastic matrices, (iii) exploits linearity to provide this alternative characterization by just checking for unprofitability at the extreme points of the set of undetectable deviations. In this respect, (iii) is arguably more "efficient" than (ii) or (i), on the grounds that (iii) requires checking for unprofitability of a strict subset of the reporting strategies in (ii) or (i). In fact, (iii) checks for unprofitability of the smallest such subset of reporting strategies.

2.2 Implementing Allocations

It is not clear from the proof of Theorem 1 how to find a scheme that implements a given allocation. We fill this gap below. Using the notation developed in Section B, consider the following linear programming problem given a pair of types (t_0, t_1):

$$V(t_0, t_1) \;=\; \inf_{\xi\in\mathbb{R}^T} \{\xi(t_1) - \xi(t_0) : v(t, \mathbf{x}(s)) - v(t, \mathbf{x}(t)) \leq \xi(s) - \xi(t)\ \forall (t, s)\}. \qquad \text{(P)}$$

By incentive compatibility, $v(t_0, \mathbf{x}(t_1)) - v(t_0, \mathbf{x}(t_0)) \leq \xi(t_1) - \xi(t_0)$ for every ξ that implements $\mathbf{x}.v$

preference, $W(t,s) + W(s,r) \leq W(t,r)$, so $W(t,r) - W(t,s) \geq W(s,r)$. Therefore, $W(r,s) - W(r,t) \geq W(t,s) \geq v(t,\mathbf{x}(s)) - v(t,\mathbf{x}(t))$ for all (t,s,r). □

Theorem 2 derives a scheme to implement a given allocation from the value of (D). We will now show how it is related to the value of (P).

Corollary 2. *$V(t_0,t_1) = W(t_0,t_1)$ for every pair (t_0,t_1).*

Proof. If $\mathbf{x}$ is not implementable then $V(t_0,t_1) = \infty$. By Theorem 1, there is an $\mathbf{x}$-profitable, undetectable deviation h. Given a feasible dual solution g, $g + \alpha h$ is also feasible for any scalar $\alpha \geq 0$, and $(g+\alpha h)(w) \to \infty$ as $\alpha \to \infty$, so $W(t_0,t_1) = V(t_0,t_1)$. If $\mathbf{x}$ is implementable then $W(t_0,\cdot)$ implements $\mathbf{x}$ by Theorem 2. By weak duality, $W(t_0,t_1) \leq V(t_0,t_1)$, and since $\xi(\cdot|t_0)$ is a feasible primal solution with $W(t_0,t_0) = 0$, it follows that $W(t_0,t_1) \geq V(t_0,t_1)$. □

The results above yield an alternative, constructive proof of Theorem 1, stated below.

Corollary 3. *An allocation $\mathbf{x}$ is implementable if and only if $W(t,t) = 0$ for all t. In this case, the scheme $W(t,\cdot)$ implements $\mathbf{x}$ for every t.*

Proof. This result follows from Theorem 2 and Corollary 2. □

2.3 Revenue Equivalence

The linear programs of Section 2.2 suggest an immediate characterization of revenue equivalence. An (implementable) allocation exhibits *revenue equivalence* if any two incentive schemes that implement it differ by a constant, i.e., for any implementing schemes ξ and ζ there exists $c \in \mathbb{R}$ such that $\xi(t) = \zeta(t) + c$ for all $t \in T$.

By Lemma 1, an allocation exhibits revenue equivalence if and only if the set $\mathscr{F}(t)$ is a singleton for every type t. This is clearly equivalent to $\bigwedge \mathscr{F}(t) = \bigvee \mathscr{F}(t)$.

Lemma 2. *For every type t_1, the incentive scheme $-V(\cdot,t_1)$ implements $\mathbf{x}$, solves (P) above for all t_0, and satisfies $-V(\cdot,t_1) = \bigvee \mathscr{F}(t_1)$.*

Proof. Follows from Lemma 1 and Proposition 1. □

Theorem 3. *An allocation exhibits revenue equivalence if and only if for all (t_0,t_1),*

$$W(t_0,t_1) + W(t_1,t_0) \;=\; 0. \tag{5}$$

Proof. Follows immediately from Proposition 1, Lemma 2 and Corollary 2. □

Theorem 3 provides a dual characterization of revenue equivalence with the following strategic interpretation. $W(t_0, t_1)$ may be interpreted as the maximum profit from deviations that shift the same fixed probability mass from t_0 to t_1. Therefore, revenue equivalence holds if and only if for every (t_0, t_1), this profit is equal to the maximum profit from deviations that shift the same probability mass back from t_1 to t_0.

Mathematically, this result is equivalent to Theorem 1 of Heydenreich et al. (2009). However, the interpretation provided above for Theorem 3 is substantially different from theirs. Indeed, the interpretation here is strategic, whereas the interpretation offered by Heydenreich et al. (2009) is graph-theoretic. Specifically, Heydenreich et al. (2009) show that revenue equivalence is characterized by $U(t_0, t_1) = -U(t_1, t_0)$ for all (t_0, t_1), where U is defined in the proof of Rochet's Theorem (Section 2.1). It is not immediately clear how to interpret the optimization problem that defines U (although the discussion at the end of Section 2.1 suggests one). On the other hand, the dual problem that defines W carries the readily-available interpretation offered in the paragraph above. Finally, Theorem 1 of Heydenreich et al. (2009) does not apply to interim implementation. This issue is discussed in Section 4.7 below, which provides yet another characterization of revenue equivalence.

2.4 Budget Balance

We begin by extending the model to include several agents. Let I be a set of agents, and for every agent $i \in I$, let T_i be the set of i's possible types. Let $T = \prod_i T_i$ be the space of *type profiles*. Let $v_i(t, x) \in \mathbb{R}$ be the utility of agent i from choice x if the type profile in society is t. An allocation $\mathbf{x} : T \to X$ is defined as before. An incentive scheme is now a map $\xi : I \times T \to \mathbb{R}$, and a mechanism is any pair $(\mathbf{x}, \xi)$.

It is well known that Rochet's Theorem can be extended in the context of many agents to so-called ex post implementation. Below we similarly extend Theorem 1, which is a trivial exercise given the previous results. We include it for completeness and because it will help to compare with results on budget balance below and Bayesian implementation in the next section.

Call $(\mathbf{x}, \xi)$ *ex post incentive compatible* (EPIC) if

$$v_i(t, \mathbf{x}(s_i, t_{-i})) - v_i(t, \mathbf{x}(t)) \;\leq\; \xi_i(s_i, t_{-i}) - \xi_i(t) \qquad \forall (i, t_i, s_i, t_{-i}).$$

An allocation $\mathbf{x}$ is *ex post implementable* if there is a scheme ξ that makes it EPIC. Intuitively, a mechanism is EPIC if for every agent, it is optimal to reveal one's true type after observing others' true types. Therefore, an EPIC mechanism will be incentive compatible regardless of one's beliefs about others, since the expected payoff from any reporting strategy is implied the state-by-state payoffs, where in this case a state is any profile of other agents' types.

A *reporting strategy* for agent i is now a map $\pi_i : S \to \Delta(S_i)$ defined on some finite subsets $S_i \subset T_i$ and $S \subset T$, where $\pi_i(s_i|t)$ is the probability i reports s_i when the profile of types is t. Now π_i is *undetectable* if $\sum_{s_i} \pi_i(s_i|t) = \sum_{s_i} \pi_i(t_i|s_i, t_{-i})$ for all t. Given an allocation $\mathbf{x}$, a reporting strategy is called *ex post $\mathbf{x}$-profitable* if

$$\sum_{(t_i,s_i)} \pi_i(s_i|t)[v_i(t, \mathbf{x}(s_i, t_{-i})) - v_i(t, \mathbf{x}(t))] \; > \; 0 \qquad\qquad \forall t_{-i} \in T_{-i}.$$

A deviation by agent i is any reporting strategy by i that isn't truthful.

The following result is an immediate consequence of Theorem 1.

Corollary 4. *A given allocation $\mathbf{x}$ is ex post implementable if and only if every ex post $\mathbf{x}$-profitable deviation is detectable.*

A scheme is *budget balanced* if $\sum_i \xi_i(t) = 0$ for every type profile t. An allocation $\mathbf{x}$ is *ex post implementable with budget balance* if there is a budget balanced scheme ξ such that $(\mathbf{x}, \xi)$ is EPIC. Below, we extend Theorem 1 to characterize budget balanced implementation. For this purpose, assume that $|I| = n \in \mathbb{N}$, i.e., there are finitely many agents. Although not essential, this assumption is useful for the sake of clarity.

A *finite strategy profile* (or simply *strategy profile*) is any family of reporting strategies $\pi = \{\pi_i : i \in I\}$. A *deviation profile* is any strategy profile with at least one deviation. A strategy profile π is called *unattributable* if

$$\sum_{s_i \in S_i} \pi_i(t_i|s_i, t_{-i}) \; = \; \sum_{s_j \in S_j} \pi_j(t_j|s_j, t_{-j}) \qquad\qquad \forall (i, j, t).$$

Otherwise, π is called *attributable*. Intuitively, a strategy profile is unattributable if the probability distribution over reported types is the same across players. After an unattributable deviation, even though the deviation may have been detected, it is impossible to identify the identity of the deviator or any non-deviator. Finally, a strategy profile π is *ex post $\mathbf{x}$-profitable* if

$$\sum_{(i,t,s_i)} \pi_i(s_i|t)[v_i(t, \mathbf{x}(s_i, t_{-i})) - v_i(t, \mathbf{x}(t))] \; > \; 0,$$

where the sum above is also taken with respect to the set of agents.

Theorem 4. *A given allocation* $\mathbf{x}$ *is ex post implementable with budget balance if and only if every ex post* $\mathbf{x}$*-pro table deviation pro le is attributable.*

The results of Sections 2.2 and 2.3 extend easily to the budget balanced setting after suitable modifications. Specifically, define $V_i(t_i^0, t_i^1; t_{-i}) = \inf \xi_i(t_i^1, t_{-i}) - \xi_i(t_i^0, t_{-i})$, where the infimum is taken with respect to incentive schemes that ex post implement $\mathbf{x}$ with budget balance. By the same argument as in Lemma 1, for any $t^0 \in T$, the set of all schemes ξ that implement $\mathbf{x}$ with budget balance and satisfy $\xi_i(t_i^0, t_{-i}) = 0$ for all t_{-i} is a complete lattice. Therefore, given $t^0 \in T$, the scheme $\xi_i(t) = V_i(t_i^0, t_i; t_{-i})$ ex post implements $\mathbf{x}$ with budget balance. This leads to the dual problem below.

$$W_i(t_i^0, t_i^1; t_{-i}) \;=\; \sup_{\lambda \geq 0, \eta} \{\lambda(w) : D\lambda = \eta + \mathbf{1}^i_{t_i^1} - \mathbf{1}^i_{t_i^0}\},$$

where $\mathbf{1}^i_{t_i} \in \mathbb{R}^{I \times T}$ is defined by $\mathbf{1}^i_{t_i^0}(j,t) = 1$ if $j = i$ and $t_i = t_i^0$, and zero otherwise.

Now Proposition 1, Theorem 2, Corollaries 2 and 3, as well as Lemma 2 easily extend with budget balance. Hence, Theorem 4 generalizes cyclic monotonicity to characterize budget-balanced implementation as follows: $W_i(t_i, t_i; t_{-i}) = 0$ for all (i, t).

Finally, Theorem 3 also extends to the case of budget balance with the following generalization. An allocation $\mathbf{x}$ has *budget-balanced revenue equivalence* if for any two schemes ξ and ζ that ex post implement $\mathbf{x}$ with budget balance, $\xi_i(t) = \zeta_i(t) + c_i(t_{-i})$ for all (i, t) and some $c_i(t_{-i}) \in \mathbb{R}$. (Hence, $\sum_i c_i(t_{-i}) = 0$ for all t.)

Corollary 5. *An allocation has budget-balanced revenue equivalence if and only if*

$$W_i(t_i^0, t_i^1; t_{-i}) + W_i(t_i^1, t_i^0; t_{-i}) \;=\; 0 \qquad \forall (i, t_i^0, t_i^1, t_{-i}).$$

The interpretation behind Corollary 5 is almost identical to that of Theorem 3. The only difference is that "attributable" now replaces "detectable." The amount $W_i(t_0, t_1; t_{-i})$ corresponds to the maximum profit from a strategy profile that is unattributable except for a given agent i, whose strategy changes the probability distribution over reports by taking probability mass from t_i^0 to t_i^1. Revenue equivalence with budget balance is equivalent to this profit plus the maximum profit from taking the probability mass back from t_i^1 to t_i^0 being always equal to zero.

3 Interim Implementation

Now consider interim implementation. Intuitively, the principal observes a signal that may be correlated with an agent's type. As a result, it may be easier to implement an allocation if the signal can be used to discern different agent types.

For simplicity (and without loss), we focus on incentives for a single agent. T still denotes the set of possible agent types. Let $(Y, \mathscr{Y})$ be a measurable space of possible signals that the principal may observe, such as output or other agents' reported types. For each type t, let $p(t)$ be a countably additive probability measure on $\mathscr{Y}$ describing the probability of signals given t. X is still the set of choices. Let $v(t, x, y) \in \mathbb{R}$ be the agent's utility from choice x when his type is t and the realized signal is y. An *allocation* is now a map $\mathbf{x} : T \times Y \to X$. A *scheme* is now a map $\xi : T \times Y \to \mathbb{R}$.

An allocation $\mathbf{x}$ is called Y-*interim implementable* if there exists a scheme ξ such that

$$\int_Y [v(t, \mathbf{x}(s, y), y) - v(t, \mathbf{x}(t, y), y)]p(dy|t) \leq \int_Y [\xi(s, y) - \xi(t, y)]p(dy|t) \qquad \forall (t, s).$$

For this system of inequalities to be well defined, we must impose some integrability restrictions on v, $\mathbf{x}$ and ξ. Firstly, we assume that $v(t, \mathbf{x}(s, y), y)$ is a $\mathscr{Y}$-measurable function of y for every (t, s) and $\xi(t, y)$ is also a $\mathscr{Y}$-measurable function of y for all t.

However, this restriction is not enough to avoid integrals that involve $\infty - \infty$. On the other hand, we wish to maintain a general model. We attempt to reconcile this trade-off by making the following assumption, which will be discussed momentarily.

Assumption 1. (i) The integral $w(t, s) = \int_Y [v(t, \mathbf{x}(s, y), y) - v(t, \mathbf{x}(t, y), y)]p(dy|t)$ is well defined for every (t, s), with values in $\mathbb{R} \cup \{-\infty\}$. (ii) Every scheme ξ has the property that $\xi(t, y)$ is a bounded measurable function of y for all t, i.e., $\xi \in B(Y)^T$.

A reporting strategy is still a map $\pi : S \to \Delta(S)$ for some finite subset $S \subset T$. A deviation is an untruthful reporting strategy. Say that π is Y-*undetectable* if

$$\sum_{t \in S} \pi(s|t)p(t) \;=\; p(s) \qquad \forall s \in S.$$

A reporting strategy π is called $\mathbf{x}$-profitable if $w \cdot \pi = \sum_{(t,s)} \pi(s|t)w(t, s) > 0$.

It might be conjectured that an allocation $\mathbf{x}$ is Y-interim implementable if and only if every $\mathbf{x}$-profitable deviation is Y-detectable. However, this conjecture is false. It turns

out that this condition is necessary but not sufficient for Y-interim implementation, in contrast with Theorem 1. (Examples illustrating this lack of sufficiency are provided below.) Intuitively, characterizing interim implementation requires in addition that infinitesimally detectable deviations be at most infinitesimally profitable.

Definition 1. *Every infinitesimally Y-detectable deviation is at most infinitesimally* **x***-profitable* if "every $\mathbf{x}$-profitable deviation is uniformly Y-detectable," i.e.,

$$\mathscr{D} \;:=\; \inf_{\xi} \sup_{\pi} \frac{w \cdot \pi}{|D\pi(\xi)|} \;<\; +\infty,^{6} \tag{4}$$

where $\xi : T \times Y \to \mathbb{R}$ is any real-valued function such that $\xi(t, y)$ is a $p(t)$-integrable function of y for every t, π is a reporting strategy and

$$D\pi(\xi) = \sum_{(t,s)} \int_{Y} \xi(s, y)\pi(s|t)[p(dy|t) - p(dy|s)].$$

Intuitively, not only is every Y-undetectable deviation $\mathbf{x}$-unprofitable, but also the profitability of every deviation is uniformly bounded by its detectability.

Theorem 5. *A given allocation* $\mathbf{x}$ *is Y-interim implementable if and only if every infinitesimally Y-detectable deviation is at most infinitesimally* $\mathbf{x}$*-profitable.*

Before proving Theorem 5, let us discuss its differences with Theorem 1 in the context of some illustrative examples. Notice that all the examples below exhibit the fact that every $\mathbf{x}$-profitable deviation is Y-detectable, yet $\mathbf{x}$ is not Y-interim implementable.

Example 1. Let $T = [0, 1]$ and $Y = \{0, 1\}$. Define $p(0) = [0]$, $p(1) = [1]$ and $p(t) = \frac{1}{2}[0] + \frac{1}{2}[1]$ for all $t \in (0, 1)$, where $[z]$ stands for Dirac measure.[7] For every $t \in (0, 1)$, let π_t be the reporting strategy defined pointwise by $\pi_t(0|t) = t = \pi_t(1|t)$, $\pi_t(t|t) = 1 - t$, $\pi_t(t|0) = t = \pi_t(t|1)$ and $\pi_t(0|0) = 1 - t = \pi_t(1|1)$. It is evident that $D\pi_t(\xi) = \frac{t}{2}[\xi(0, 1) - \xi(0, 0) + \xi(1, 0) - \xi(1, 1)]$ for all $\xi : T \times Y \to \mathbb{R}$. Therefore, $|D\pi_t(\xi)| \to 0$ as $t \to 0$ for every ξ. Define w as follows: $w(t, 0) = 1/t$ for every $t \in (0, 1)$ and $w(r, s) = 0$ for all other (r, s). Clearly, every $\mathbf{x}$-profitable deviation is Y-detectable, since making any profit requires type 0 misreporting to some type $t \in (0, 1)$ with positive probability, and this is Y-detectable. Now the profit from π_t

[6] We adopt the convention that any real number divided by in nity equals zero, and any real number divided by zero equals zero if the numerator is zero and otherwise equals $\pm\infty$ depending on the sign of the numerator.

[7] In other words, $[z](Z) = 1$ if $z \in Z$ and 0 otherwise.

is given by $w \cdot \pi_t = t[w(t,0) + w(t,1) + w(0,t) + w(1,t)] = 1$ for every $t \in (0,1)$. As a result, $\mathscr{D} = +\infty$, so Y-interim implementability fails by Theorem 5. Notice that this argument fails if and only if $w(t,0)$ is bounded above by a constant for all t.

Example 1 identifies an important difference between Theorems 1 and 5, namely that detection and infinitesimal detection are potentially different notions. Although the example above relies on w becoming unbounded, this is by no means a prerequisite for the kind of pathology that it portrays, as the next examples show.

Example 2. Let $T = [0,1]$ and $Y = \{0,1\}$. Define $p(0) = [0]$, $p(1) = [1]$ and $p(t) = (1-t)[0] + t[1]$ for all other t. Given y and any finite subset of types there is only one type with largest probability over y, so every deviation is Y-detectable. For every $k \in \mathbb{N}$, let $t_k = 1/k$ and define π_k by $\pi_k(0|t_k) = (1-t_k)$ and $\pi_k(1|t_k) = t_k$. (Let π_k be honest elsewhere.) By routine calculations, $D\pi_k(\xi) = \frac{1}{k}(1-\frac{1}{k})\Delta\xi$ for all $\xi : T \times Y \to \mathbb{R}$, where $\Delta\xi = \xi(0,1) - \xi(0,0) + \xi(1,0) - \xi(1,1)$. Define w by $w(t,0) = 1$ for all t and $w(t,s) = 0$ for all other (t,s). Clearly, $w \cdot \pi_k = (1-\frac{1}{k})w(t_k,0) = (1-\frac{1}{k})$. Finally, $\lim w \cdot \pi_k / |D\pi_k(\xi)| = \lim (1-\frac{1}{k}) / (\frac{1}{k}(1-\frac{1}{k})|\Delta\xi|) = \lim \frac{k}{|\Delta\xi|} = +\infty$. Therefore, $\mathscr{D} = +\infty$ and Y-interim implementability fails by Theorem 5.

Example 2 above shows that a suitable discontinuity in w is sufficient to prevent an allocation from being interim implementable. Indeed, notice that in the example $w(t,0)$ does not tend to 0 as $t \to 0$, even though $w(0,0) = 0$. However, discontinuity is not necessary for interim implementation to fail, as the next example shows.

Example 3. Consider exactly the same setting and sequence $\{\pi_k\}$ as in Example 2. The only difference here is that now w is defined by $w(0,t) = \sqrt{t}$. It is easy to see that now $w \cdot \pi_k = (1-\frac{1}{k})w(t_k,0) = (1-\frac{1}{k})\frac{1}{\sqrt{k}}$. Simple calculations show that $\lim w \cdot \pi_k / |D\pi_k(\xi)| = \lim ((1-\frac{1}{k})\frac{1}{\sqrt{k}}) / (\frac{1}{k}(1-\frac{1}{k})|\Delta\xi|) = \lim \frac{k}{\sqrt{k}|\Delta\xi|} = +\infty$. Therefore, $\mathscr{D} = +\infty$ and again Y-interim implementability fails by Theorem 5.

Example 3 shows that a failure of Lipschitz continuity in w is enough for interim implementation to fail. However, yet again this is not necessary. The next example highlights that what drives all these failures is not failure of Lipschitz continuity, but rather a lack of bounded steepness between the change in probabilities and the change in payoffs from misreporting.

Example 4. Let $T = [0,1]$ and $Y = \{0,1\}$. Let $p(t) = (1-t^2)[0] + t^2[1]$ for all t. As in Example 2, every deviation is Y-detectable. Given $k \in \mathbb{N}$, let $t_k = 1/k$

and π_k be the reporting strategy defined by $\pi_k(0|t_k) = (1 - t_k^2)$ and $\pi_k(1|t_k) = t_k^2$. By routine calculations, $D\pi_k(\xi) = \frac{1}{k^2}(1 - \frac{1}{k^2})\Delta\xi$ for all $\xi : T \times Y \to \mathbb{R}$, where $\Delta\xi = \xi(0,1) - \xi(0,0) + \xi(1,0) - \xi(1,1)$. Define w by $w(t,0) = t$ for all t and 0 elsewhere. Clearly, $w \cdot \pi_k = (1 - \frac{1}{k^2})w(t_k, 0) = (1 - \frac{1}{k^2})\frac{1}{k}$. After simple calculations, $\lim w \cdot \pi_k / |D\pi_k(\xi)| = \lim((1 - \frac{1}{k^2})\frac{1}{k})/(\frac{1}{k^2}(1 - \frac{1}{k^2})|\Delta\xi|) = \lim \frac{k^2}{k|\Delta\xi|} = +\infty$. Therefore, $\mathscr{D} = +\infty$ and once again Y-interim implementability fails by Theorem 5.

Example 4 exhibits a Lipschitz continuous function w yet Y-interim implementation fails even though every deviation is Y-detectable. Intuitively, this happens here because the "steepness" ratio of changes in payoffs (linear) to changes in probabilities (quadratic) explodes as the deviation becomes infinitesimal. By Theorem 5, interim implementation is equivalent to this steepness being uniformly bounded.

We end this section by showing that Theorem 1 is a special case of Theorem 5 when the principal's signal y is independent of the agent's type t.

Proposition 2. *Given an allocation* $\mathbf{x}$*, suppose that* $p(t)$ *does not depend on* t*. Every infinitesimally* Y*-detectable deviation is at most infinitesimally* $\mathbf{x}$*-profitable if and only if every* $\mathbf{x}$*-profitable deviation is detectable.*

4 Extensions

In this section we discuss several extensions: moral hazard, revealed stochastic preference, surplus extraction, budget balanced implementation, bargaining with interdependent values, optimal mechanisms and finally a "subdifferential" characterization of interim implementing incentive schemes.

4.1 Moral Hazard

The moral hazard problem fits easily into the framework developed above. To see this, consider a prototypical such problem. An agent's possible actions are described by an arbitrary set A. The principal wants the agent to choose some fixed action $a \in A$, but the agent may choose any action $b \in A$. Let Y be another measurable space of verifiable output, and $p(a) \in \Delta(Y)$ the conditional probability of such output. Finally, let $v(a) \in \mathbb{R}$ be the agent's utility from each action a.

An action a is Y-*enforceable* if there is a payment scheme $\xi \in B(Y)$ such that

$$v(b) - v(a) \;\leq\; \int_Y \xi(y)[p(dy|b) - p(dy|a)] \qquad \forall b \in A.$$

A *deviation* in this setting is any $\pi \in \mathbb{R}^{(A)}$ such that $\pi \geq 0$ and $\sum_a \pi(a) = 1$. Such a π is called Y-*undetectable* if

$$p(a) \;=\; \sum_{b \in A} \pi(b)p(b).$$

A deviation π is called a-*profitable* if $\sum_b \pi(b)[v(b) - v(a)] > 0$. Finally, say that *every infinitesimally Y-detectable deviation is at most infinitesimally a-profitable* if

$$\inf_{\xi} \sup_{\pi} \frac{w \cdot \pi}{|D\pi(\xi)|} < +\infty,$$

where $w \in \mathbb{R}^A$ is the vector defined pointwise by $w(b) = v(b) - v(a)$ for all $b \in A$ and $D\pi(\xi) = \sum_b \int_Y \xi(y)[p(dy|b) - p(dy|a)]$. The following result follows easily from previous ones, so its proof is omitted.

Theorem 6. *A given action a is Y-enforceable if and only if every infinitesimally Y-detectable deviation is at most infinitesimally a-profitable.*

Theorem 6 shows how implementability of an allocation is characterized in the same manner under adverse selection as under moral hazard. Under each environment, implementability boils down to detecting profitable deviations from either honesty or obedience. An immediate difference between moral hazard and adverse selection, of course, is that if output is independent of effort then clearly it is impossible to implement any (opportunity) costly effort by the agent.

4.2 Surplus Extraction

In this subsection we show (i) how the notions of detectability introduced in this paper differ substantially from the conditions of Cremer and McLean (1988) and McAfee and Reny (1992a,b), and (ii) how their results can be generalized to arbitrary settings using the tools of this paper. Specifically, we characterize below full surplus extraction and discuss "virtually full" surplus extraction.

We begin by discussing Cremer and McLean's contribution. Cremer and McLean (1988) show that in a setting with finitely many types, agents' conditional probability

vectors exhibit convex independence (defined below) if and only if for any profile of utility functions, every allocation is implementable with an incentive scheme that makes every individual rationality constraint bind. Hence, the scheme extracts all the surplus. Formally, p exhibits *convex independence* if $\sum_s \lambda(t,s)[p(t) - p(s)] = 0$ for every type t and $\lambda \geq 0$ imply that $\lambda \cdot w = 0$ for all w with $w(t,t) = 0$ given t. (Equivalently, convex independence is defined by $p(t) \notin \text{conv}\{p(s) : s \neq t\}$ for all t.)

To see how the notions of detectability introduced in this paper differ from convex independence, consider the following simple example.

Example 5. Let $T = \{0, \frac{1}{2}, 1\}$, $Y = \{0, 1\}$, and $p(t) = t[1] + (1-t)[0]$. In this case, it is clear that convex independence fails, since $p(\frac{1}{2}) = \frac{1}{2}p(0) + \frac{1}{2}p(1)$, and hence $p(\frac{1}{2})$ lies in the convex hull of $\{p(0), p(1)\}$. On the other hand, it is easy to see that every deviation is Y-interim detectable.

Cremer and McLean's result is usually summarized by the slogan "if types are correlated then you can extract the surplus." Example 5 shows that "correlated types" is not enough to extract the surplus. Indeed, types *are* correlated in Example 5 because t and y are clearly correlated, yet convex independence fails. (Think of y, e.g., as others' types.) Therefore the surplus cannot always be extracted.

It is easy to see that detectability is logically a much weaker condition than convex independence. Indeed, that every deviation is detectable may be written as

$$\sum_{s \in S} \pi(t|s)p(s) \;=\; p(t) \quad \forall t \quad \Rightarrow \quad \pi(s|t) = 0 \text{ if } s \neq t. \tag{7}$$

Intuitively, the left-hand side above means that the probability distribution induced by truth-telling coincides with that arising from the reporting strategy π, where the types in S are given the uniform prior distribution.

On the other hand, convex independence may be written as

$$\sum_{s \in S} \pi(s|t)p(s) \;=\; p(t) \quad \forall t \quad \Rightarrow \quad \pi(s|t) = 0 \text{ if } s \neq t. \tag{8}$$

Therefore, convex independence implies that every deviation is interim detectable, so by Example 5, detectability is strictly weaker than convex independence. More importantly, the two conditions above differ in terms of interpretation. Convex independence may be interpreted as a "conditional" detectability criterion. Intuitively, convex independence means that after any given type, every deviation changes the conditional probability distribution over outcomes.

Despite the differences described above, the problem of surplus extraction is easily modeled with the tools developed here. Indeed, we will now generalize Cremer and McLean's result to arbitrary type and signal spaces, without assuming continuity or compactness. Furthermore, these results are derived with simple arguments that rely on duality, as the rest of the results in this paper. This contrasts the work of McAfee and Reny (1992b). They emphasized that their work was not just an application of duality, but rather somewhere "[...] between the Stone-Weierstrass Theorem and a corollary to the Hahn-Banach Theorem." (McAfee and Reny, 1992b, p. 61.)

Formally, p exhibits *virtual convex independence* if $\sum_s \lambda_\delta(\cdot, s)[p(\cdot) - p(s)] \to 0$ weakly[8] and $\lambda_\delta \geq 0$ for all δ imply $\lim \lambda_\delta \cdot w = 0$ for all w with $w(t,t) = 0$ given t. Below we equate virtual convex independence to full surplus extraction. Given an allocation $\mathbf{x}$, say that *all the surplus can be extracted* from $(v, \mathbf{x})$ if there is a scheme ξ that Y-interim implements $\mathbf{x}$ and $\int_Y [v(t, \mathbf{x}(t,y), y) - \xi(t,y)]p(dy|t) = 0$ for all t. We now extend Cremer and McLean's theorem to arbitrary type spaces.

Theorem 7. *All the surplus can be extracted from any given $(v, \mathbf{x})$ if and only if the information structure p exhibits virtual convex independence.*

Theorem 7 above shows that virtual convex independence characterizes full surplus extraction. The requirement of virtual convex independence is significantly stronger than convex independence, as Example 6 below shows. Example 6 presents an information structure where virtual convex independence fails, yet convex independence holds. By Theorem 7, full surplus extraction fails. As will be discussed in detail later, this information structure also fails McAfee and Reny's condition for virtually full surplus extraction.

Example 6. Let $T = [0,1]$ and $Y = \{a, b, c\}$. Define p by $p(0) = [a]$, $p(1) = [b]$ and $p(t) = (1-t)^2[a] + t^2[b] + 2t(1-t)[c]$. As Figure 1 below illustrates, it is clear that p exhibits convex independence.

However, interim implementability may fail. To see this, let π_k be defined exactly as in Example 4. By routine calculations, $D\pi_k(\xi) = (1 - \frac{1}{k})^2(\xi(0,a)\frac{1}{k}(\frac{1}{k} - 2) + \xi(0,b)\frac{1}{k^2} + 2\xi(0,c)\frac{1}{k}(1 - \frac{1}{k})) + \frac{1}{k^2}(\xi(1,a)(1 - \frac{1}{k})^2 - \xi(1,b)(1 - \frac{1}{k^2}) + 2\xi(1,c)\frac{1}{k}(1 - \frac{1}{k}))$. Hence, $|D\pi_k(\xi)| = O(\frac{1}{k})$. Letting $w(t,0) = \sqrt{t}$, we obtain $w \cdot \pi_k = (1 - \frac{1}{k})^2/\sqrt{k}$. It follows that $\lim w \cdot \pi_k / |D\pi_k(\xi)| = +\infty$ for all ξ, so by Theorem 5 interim implementability

[8] Henceforth, *convergence is weak unless otherwise stated.* Thus, $\sum_s \lambda_\delta(\cdot, s) \to 0$ means that $\sum_{(t,s)} \lambda_\delta(t,s) f(t) \to 0$ in $\mathbb{R}$ for every $f \in \mathbb{R}^T$.

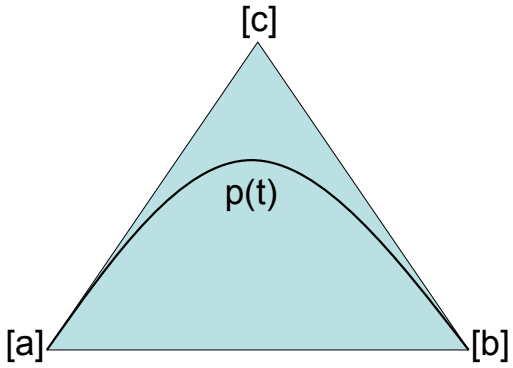

Figure 1: Convex independence holds but virtual convex independence fails.

fails. Therefore, no fraction of the surplus may be extracted incentive compatibly because there is no payment scheme that yields interim implementation.

Example 6 above suggests a weaker characterization of full surplus extraction that *assumes* interim implementability. Formally, say that ***all the surplus can be extracted from*** $(v, \mathbf{x})$ ***assuming interim implementability*** if there is a payment scheme ξ that extracts all the surplus—i.e., such that $\int_Y [v(t, \mathbf{x}(t,y), y) - \xi(t,y)]p(dy|t) = 0$ for all t—and satisfies the following system of inequalities:

$$0 \;\leq\; \int_Y [\xi(s,y) - \xi(t,y)]p(dy|t) \qquad \forall (t,s).$$

Intuitively, this system of inequalities says that ξ does not disrupt any incentive compatibility constraints. Hence, if $\mathbf{x}$ is implementable with some scheme ζ then $\zeta + \xi$ still implements $\mathbf{x}$. In other words, if $\mathbf{x}$ is interim implementable, say with scheme ζ, then it is possible to find another scheme ξ such that $\zeta + \xi$ still interim implements $\mathbf{x}$ and extracts $v - \zeta$ of the surplus. Our next goal is to find a condition on the information structure such that this is the case for every $(v, \mathbf{x})$. When such a condition holds, we can find a ξ such that $\xi + \zeta$ extracts v of the surplus—i.e., all the surplus can be extracted conditional on interim implementability—regardless of v.

Theorem 8. *All the surplus can be extracted from any given* $(v, \mathbf{x})$ *assuming interim implementability if and only if* p *exhibits* asymptotic convex independence*, that is, for any net of reporting strategies* $\{\pi_\delta\}$ *with* $\pi_\delta(t|t) < 1$ *for at most nitely many* t*,* $\sum_s \pi_\delta(s|\cdot)p(s) \to p(\cdot)$ *implies that* $\sum_s \pi_\delta(\cdot|s) \to 1$*.*

We will now discuss Theorem 8, the asymptotic convex independence condition and how they compare with the work of McAfee and Reny (1992a). Firstly, Theorem 8 is

proved in the appendix similarly to Theorem 7, where duality is used to characterize surplus extraction given interim implementability. This condition is interpreted in the proof and shown to be equivalent to asymptotic convex independence.

Asymptotic convex independence is strictly weaker than virtual convex independence and strictly stronger than convex independence. To see this, asymptotic convex independence follows from virtual convex independence by restricting λ_δ to be a strategy, and Example 6 presents an information structure where virtual convex independence fails yet asymptotic convex independence holds (convex independence holds there, too). Moreover, convex independence follows from asymptotic convex independence by restricting attention to constant nets. We now present a simple example where convex independence holds but asymptotic convex independence fails.[9]

Example 7. Let $T = Y = \mathbb{N}$, $p(t+1) = [t]$ for each $t \in \mathbb{N}$, and $p(1) = \sum_k 2^{-k}[k]$.

Theorem 8 extends the work of McAfee and Reny (1992a) in several ways. First of all, Theorem 8 does not require continuity and compactness. Secondly, it can be shown that asymptotic convex independence is equivalent to condition $(*)$ in McAfee and Reny (1992a, p. 404). Indeed, let T be a compact metric space and $\Delta(T)$ the set of Borel probability measures on T. McAfee and Reny's condition $(*)$ is this: Given $t \in T$ and $\mu \in \Delta(T)$, if $p(t) = \int_T p(s)\mu(ds)$ then $\mu = [t]$, where $[t]$ stands for Dirac measure.[10] It is well known that the set of Borel probability measures with finite support is (weak$*$) dense in $\Delta(T)$ (e.g., Aliprantis and Border, 2006, p. 510, Theorem 15.10). Therefore, any $\mu(t) \in \Delta(T)$ is the limit of a sequence[11] of probability measures $\{\pi_m(t)\}$, where each $\pi_m(t)$ has finite support. Hence, the key difference between asymptotic convex independence and condition $(*)$ is between weak and pointwise convergence. In McAfee and Reny's restricted setting, this difference disappears.

Proposition 3. *Suppose that T is a compact metric space and both v and p are continuous. All the surplus can be extracted from any given $(v, \mathbf{x})$ with a continuous scheme assuming interim implementability if and only if condition $(*)$ holds.*

Thirdly, Theorem 8 allows for general, direct revelation mechanisms, whereas McAfee and Reny (1992a) restrict attention to finite "participation fee schedules." Specifically,

[9] I apologize for such a tongue-twisting paragraph.

[10] McAfee and Reny also assume that p is continuous and $p(t)$ has a continuous density for all t.

[11] Since T is compact metric, it is separable, hence (T) is, too, so its topology is rst countable. Hence, without loss we may focus on sequences rather than nets.

they break down the surplus extraction problem into two stages. In the first stage, the agent is offered a finite menu of participation fee schedules, with the promise that his choice of fee schedule will not be used in the subsequent stage. In the second stage, an allocation is implemented incentive compatibly. McAfee and Reny (1992a) leave the second stage implicit, so they study surplus extraction *assuming* interim implementability in the second stage. However, notice that their fee schedules are not direct revelation mechanisms, and in principle incur some loss of generality.

As will be explained below, their fee schedules do in fact incur a loss of generality. By focusing on finite menus, they were not able to attain full surplus extraction, and were forced to settle with "virtually full" surplus extraction. Paraphrasing McAfee and Reny (1992a), say that *virtually all the surplus can be extracted* from $(v, \mathbf{x})$ if for every $\varepsilon > 0$ there exists a scheme ξ that Y-interim implements $\mathbf{x}$ and

$$0 \;\leq\; \int_Y [v(t, \mathbf{x}(t,y), y) - \xi(t,y)] p(dy|t) \;\leq\; \varepsilon \qquad \forall t.$$

Since virtually full surplus extraction bounds the surplus uniformly in t, it was natural for McAfee and Reny (1992a) to restrict v to be bounded. Furthermore, given their finite menus, it was also natural to restrict v to be continuous and T to be compact.

As a result, by using direct revelation mechanisms, Theorem 8 attains full surplus extraction (given interim implementability) in more general settings that McAfee and Reny (1992a) with a condition that in their restricted setting is equivalent. On the other hand, McAfee and Reny (1992a) characterized virtually full surplus extraction. This begs the question, what might characterize virtually full surplus extraction with direct revelation mechanisms? We answer this question in the next result.

Theorem 9. *With direct revelation mechanisms, given any $(v, \mathbf{x})$, virtually all the surplus can be extracted from $(v, \mathbf{x})$ if and only if all the surplus can be extracted from $(v, \mathbf{x})$. This result still holds if interim implementability is assumed.*

Theorem 9 says that with direct mechanisms, any condition that captures full surplus extraction also captures virtually full surplus extraction, and vice versa. Therefore, McAfee and Reny's characterization of virtually full surplus extraction—rather than full extraction—relies squarely on their restriction to finite participation schedules. If they had allowed for general mechanisms, their condition would have lead not just to virtually full surplus extraction, but to full extraction of the surplus.

4.3 Revealed Stochastic Preference

It is well-known that cyclic monotonicity characterizes rationalizable economic behavior in the spirit of Afriat (1967) and others, so that Rochet's Theorem is comparable to Afriat's Theorem of revealed preference. This comparison is formalized by Rochet himself (See Rochet, 1987, pp. 195–196) in a quasi-linear context.[12]

Briefly, recall that an allocation $\mathbf{x} : T \to X$ is implementable if

$$v(t, \mathbf{x}(t)) - \xi(t) \;\geq\; v(t, \mathbf{x}(s)) - \xi(s) \qquad \forall (t, s).$$

By the taxation principle, any two reports that lead to the same choice must cost the same amount of money, i.e., $\mathbf{x}(t) = \mathbf{x}(s)$ implies that $\xi(t) = \xi(s)$ whenever $\mathbf{x}$ is implementable. Hence we may rewrite the previous inequalities as

$$v(t, \mathbf{x}(t)) - \xi(\mathbf{x}(t)) \;\geq\; v(t, \mathbf{x}(s)) - \xi(\mathbf{x}(s)) \qquad \forall (t, s).$$

Reinterpret $v'(t) = -v(t) \in \mathbb{R}^X$ as a vector of "nonlinear prices," $\mathbf{x}(t)$ as a "choice" and t as a parameter indexing price/choice outcomes. Finally, interpret $\xi'(x) = -\xi(x)$ as a utility function over the range of $\mathbf{x}$. By definition, there exists a quasi-linear utility function ξ' that rationalizes every choice $\mathbf{x}(t)$ given prices $v'(t)$ if

$$\xi'(\mathbf{x}(t)) - v'(t, \mathbf{x}(t)) \;\geq\; \xi'(\mathbf{x}(s)) - v'(t, \mathbf{x}(s)) \qquad \forall (t, s).$$

Now it is clear how Rochet's Theorem and Afriat's Theorem follow from each other. As a result, Theorem 1 provides an alternative characterization of revealed preference.

Similarly, Theorem 5 is comparable to the work of McFadden (2005) on revealed stochastic preference. Let us follow the previous logic in the stochastic setting. We will think of "output" Y as a summary of uncertainty subsequent to the determination of an agent's type t. Thus, we now think of a random allocation $\mathbf{x} : T \times Y \to X$, where the randomness comes from Y. We appeal once more to the interpretation of $v(t, y) \in \mathbb{R}^X$ as a vector of "nonlinear prices," although we now allow it to be random. This is captured by its dependence on y. Similarly, we may think of $\mathbf{x}(t, y)$ as "random choices." Hence, y determines both prices v and choices $\mathbf{x}$. In particular, these two variables could be correlated given t. Intuitively, we are given a collection of price-choice distributions, and ask whether or not such a distribution may be generated by

[12] However, results in the quasi-linear context can be used to derive general rationalizability results, as Afriat (1967, p. 72) does in a neoclassical setting. See also Afriat (1963), Richter and Wong (2005).

a *population* of quasi-linear utility maximizers, indexed by y, whose members make choices given personalized nonlinear prices.

Following the argument for revealed preference, we seek to interpret $\xi'(t,y) = -\xi(t,y)$ as a (random, expected) utility function over choices by appealing to a suitable version of the taxation principle. Unfortunately, this principle is not available, since interim implementability does not require that $\xi(t,y) = \xi(s,y)$ whenever $\mathbf{x}(t,y) = \mathbf{x}(s,y)$. Therefore, Theorem 5 does not directly capture revealed stochastic preference. On the other hand, Theorem 5 may be extended by *imposing* such restrictions on ξ. The outcome of this exercise is documented in the next result.

Say that $\mathbf{x}$ is *Y-interim implementable as a Y-contingent menu* if it is Y-interim implementable with a scheme ξ such that $\xi(t,y) = \xi(s,y)$ whenever $\mathbf{x}(t,y) = \mathbf{x}(s,y)$. To characterize such version of implementability, we require further notation. Let $\mathscr{R} = \{(t,s,y) : \mathbf{x}(t,y) = \mathbf{x}(s,y)\}$ be the set that indexes restrictions on ξ, and write $\mathbf{1}_{\mathscr{R}}$ for the indicator function of $\mathscr{R}$, so $\mathbf{1}_{\mathscr{R}}(t,s,y) = 1$ if $(t,s,y) \in \mathscr{R}$ and 0 otherwise.

Theorem 10. *An allocation $\mathbf{x}$ is Y-interim implementable as a Y-contingent menu if and only if for any net $\{(\lambda_\delta, \mu_\delta)\}$ such that $\lambda_\delta \in \mathbb{R}_+^{(T\times T)}$ and $\mu_\delta \in M(Y)^{(T\times T)}$,*[13] *if*

$$\lim \sum_{s\in T} (\lambda_\delta(t,s)p(s) - \lambda_\delta(s,t)p(t)) - \sum_{s\in T} \int_Y [y]\mathbf{1}_{\mathscr{R}}(t,s,y)(\mu_\delta(dy|t,s) - \mu_\delta(dy|s,t)) \;=\; 0$$

for every t then $\lim \sum_{(t,s)} \lambda_\delta(t,s)w(t,s) \leq 0$, where $[y]$ stands for Dirac measure and the integral above is vector-valued in $M(Y)$.

This result follows similarly to previous results, so a proof is omitted. Theorem 10 generalizes previous results in several directions. To help describe them, think of y as a parameter for different types of decision maker in a heterogeneous population.[14]

First, Theorem 10 characterizes revealed stochastic preference of a population of decision makers under the following weaker assumptions: (i) it allows for "personalized budgets" because v may depend on y and therefore is compatible with correlation between prices, choices and utility, (ii) it allows for different populations in the sample of observed choices because p may depend on t, and (iii) it does not impose any

[13] The notation $M(Y)^{(T\times T)}$ stands for the set of functions whose domain is $T \times T$, whose range is the space of signed measures on Y, and such that $\mu(t,s) \neq 0$ for at most finitely many pairs (t,s).

[14] The interpretation of revealed stochastic preference as a population distribution of behavior rather than uncertain behavior by a single decision maker may be attributed to McFadden and Richter (1990, Footnote 25, pp. 174-5).

structure on T, so is compatible with any possibly infinite set of types. This contrasts with McFadden (2005), who in characterizing revealed stochastic preference in the infinite case confines attention to compact metric type spaces. On the other hand, Theorem 10 is restricted by the assumption of quasi-linearity, although this assumption may be dropped by applying Afriat's (1967, p. 72) trick.

Adding structure to the problem reveals important insights in Theorem 10. For instance, suppose that $p(t) = p$ does not depend on t, so the population does not change with observed behavior. Furthermore, suppose that $\mathbf{x}(t) \neq \mathbf{x}(s)$ with positive p-probability for every pair (t, s). Intuitively, this means that there are no duplicate observations. In this case, it is easy to see that Theorem 10 boils down to a version of Proposition 2, i.e., every profitable deviation is detectable.

Corollary 6. *Suppose that $p(t) = p$ does not depend on t and that $\mathbf{x}(t) \neq \mathbf{x}(s)$ with positive p-probability for every pair (t, s). An allocation $\mathbf{x}$ is Y-interim implementable as a Y-contingent menu if and only if every $\mathbf{x}$-pro table deviation is detectable.*

A slightly different version of Theorem 10 obtains by imposing $\xi(t, y) = \xi(s, y)$ for all y whenever $\mathbf{x}(t) = \mathbf{x}(s)$, instead of $\xi(t, y) = \xi(s, y)$ whenever $\mathbf{x}(t, y) = \mathbf{x}(s, y)$. This means that we may rewrite ξ as $\xi(\mathbf{x}(t), y)$ for every y. In other words, observed choices may be represented as coming from a population choosing an efficient Y-contingent allocation of goods amongst the population when individual utility functions are quasi-linear and exhibit consumption externalities. Indeed, the representation of ξ shows that individuals of type y care about the entire allocation $\mathbf{x}(t)$. This is an easy exercise given the techniques developed above, and therefore omitted.

4.4 Budget Balanced Interim Implementation

Let us return to the multiagent setting, where $I = \{1, \ldots, n\}$ is a finite set of agents, each T_i is a measurable space of types, $T = \prod_i T_i$ is the product space of type profiles with the product σ-algebra, and $v_i(t, \mathbf{x}(s_i, t_{-i})) \in \mathbb{R}$ is the utility to agent i under allocation $\mathbf{x} : T \to X$ when his type is t_i but he reported s_i.

Recall that an incentive scheme $\xi : I \times T \to \mathbb{R}$ is *budget balanced* if $\sum_i \xi_i(t) = 0$ for all $t \in T$. An allocation $\mathbf{x}$ is *T-interim implementable with budget balance* if there is a budget balanced scheme ξ such that $\mathbf{x}$ is T_{-i}-interim implementable for every i. For interim implementability to be well-defined we maintain Assumption 1. Notice

that this condition does *not* require v_i to be uniformly bounded.

A *strategy profile* is any family $\pi = \{\pi_i : i \in I\}$ of reporting strategies, where $\pi_i : S_i \to \Delta(S_i)$ for some finite subset $S_i \subset T_i$ for each i. Call π is **x**-*profitable* if

$$w \cdot \pi \;=\; \sum_{(i,t_i,s_i)} \int_{T_{-i}} \pi_i(s_i|t_i)[v_i(t, \mathbf{x}(s_i, t_{-i})) - v_i(t, \mathbf{x}(t))]p_i(dt_{-i}|t_i) \;>\; 0.$$

Say that *every infinitesimally* **x**-*profitable strategy profile is at most infinitesimally* T-*attributable* if $\mathscr{D} = \inf_\xi \sup_{\pi,\eta} w \cdot \pi / |D\pi(\xi)| < +\infty$, where $\eta \in \mathbb{R}^{(T)}$ and

$$D\pi(\xi|\eta) \;=\; \sum_{(i,t_i,s_i)} \int_{T_{-i}} \xi_i(s_i, t_{-i})\pi_i(s_i|t_i)[p_i(dt_{-i}|t_i) - p_i(dt_{-i}|s_i)] - \sum_{(i,t)} \xi_i(t)\eta(t).$$

Theorem 11. *A given allocation* **x** *is* T*-interim implementable with budget balance if and only if every infinitesimally* **x**-*profitable strategy profile is at most infinitesimally* T*-attributable.*

Theorem 11 can be proved by following that of Theorem 5 almost to the letter, so its proof is omitted. Note that the case of independent types is studied in Theorem 4.

Let us now characterize when ex post budget balance is not a binding constraint. To this end, say that *infinitesimal* T*-attribution implies at most infinitesimal* T*-detection* if $\inf_\xi \sup_{g,\eta} b \cdot \eta / |Dg(\xi|\eta)| < +\infty$ for every "budget" function $b : T \to \mathbb{R}$, where

$$Dg(\xi|\eta) \;=\; \sum_{(i,t_i,s_i)} \int_{T_{-i}} \xi_i(s_i, t_{-i})\lambda_i(t_i, s_i)[p_i(dt_{-i}|t_i) - p_i(dt_{-i}|s_i)] - \sum_{(i,t)} \xi_i(t)\eta(t).$$

We will say that *budget balance is not a binding constraint* if for any budget b there is an incentive scheme ξ such that $\sum_i \xi_i(t) = b(t)$ for all t and

$$\int_{T_{-i}} (\xi_i(s_i, t_{-i}) - \xi_i(t))p_i(dt_{-i}|t_i) \;\geq\; 0 \qquad \forall (i, t_i, s_i).$$

To understand this condition, suppose that an allocation **x** is implemented by the scheme ζ. If budget balance is not a binding constraint then there is an additional scheme ξ that absorbs any budgetary surpluses and deficits from ζ without disrupting any incentive compatibility constraints.

Proposition 4. *Budget balance is not a binding constraint if and only if infinitesimal* T*-attribution implies at most infinitesimal* T*-detection. With independent types, this holds if and only if detection implies attribution, i.e., for any strategy profile* π*, if* π *is unattributable then every* π_i *is undetectable.*

The proof of Proposition 4 is similar to previous ones, hence omitted.

Corollary 7. *With independent types, detection implies attribution. Hence, budget balance is not a binding constraint, so an allocation is T-interim implementable with budget balance if and only if it is T_{-i}-interim implementable for every agent i.*

Proof. By Proposition 4, it suffices to show that detection implies attribution with independent types. Suppose not. In this case, there exists an unattributable deviation profile π and an agent i such that π_i is detectable. Since π is unattributable, there exists η such that $\sum_{s_i}(\pi_i(s_i|t_i) - \pi_i(t_i|s_i)) = \eta(t)$ for each i. Hence, $\sum_t \eta(t) = 0$. By detectability, $\eta(t) > 0$ and $\eta(s) < 0$ for some pair (t, s). But then $\eta(s) = \eta(s_i, t_{-i}) < 0$ and $\eta(t) = \eta(t_j, s_i, t_{-ij}) > 0$, a contradiction. □

To illustrate, consider the special case of *private values*, where each agent's utility is independent of others' types, i.e., $v_i(t, x) = v_i(t_i, x)$ for all x, and an *ex post e cient* allocation, i.e., $\mathbf{x}^*$ such that $\mathbf{x}^*(t) \in \arg\max_x \sum_i v_i(t_i, x)$ for all $t \in T$. Below, we will prove that $\mathbf{x}^*$ is implementable with or without budget balance.

Corollary 8. *With private values, $\mathbf{x}^*$ is ex post implementable. Hence, $\mathbf{x}^*$ is interim implementable for any type space.*

Proof. By Corollary 4, we must show that every profitable deviation is detectable. Otherwise, suppose that π_i is a profitable, undetectable deviation and consider the welfare consequences of agent i reporting according to π_i instead of truthfully. Since π_i is undetectable and values are private, the expected utility to any agent $j \neq i$ is the same if i plays π_i or if he reports truthfully. On the other hand, agent i is strictly better off, therefore, welfare increases when agent i plays π_i instead of reporting truthfully. But this contradicts ex post efficiency. □

Corollary 9. *With independent private values, $\mathbf{x}^*$ is T-interim implementable with or without budget balance.*

4.5 Bargaining with Interdependent Values

A *bargaining problem* is the task of finding an incentive scheme that makes a given allocation interim implementable without violating budget balance or individual rationality, described below. The main goal of this section is to characterize existence of solutions for an arbitrary bargaining problem.

A mechanism $(\mathbf{x}, \xi)$ is called *individually rational* if

$$\int_{T_{-i}} [v_i(t, \mathbf{x}(t)) - \xi_i(t)] p_i(dt_{-i}|t_i) \;\geq\; \int_{T_{-i}} v_i(t, \mathbf{x}(0)) p_i(dt_{-i}|t_i) \qquad \forall (i, t_i),$$

where $\mathbf{x}(0)$ is the *disagreement outcome*, i.e., what happens when an agent decides to opt out of the mechanism. A *bargaining solution* for $\mathbf{x}$ is any incentive scheme ξ that T-interim implements $\mathbf{x}$ with budget balance and renders $(\mathbf{x}, \xi)$ individually rational.

In the dual problem, the multipliers on an agent's individual rationality constraint may be interpreted as the probability with which the agent deviates to opting out. Therefore, in this setting we redefine a strategy to be any map $\pi_i : S_i \to \Delta(S_i \cup \{0\})$, where $S_i \subset T_i$ is finite and $\pi_i(0|t_i)$ stands for the probability that agent i opts out when his type is t_i. Since $0 \notin T_i$ it is clear that every deviation where opting out has positive probability is detectable. A strategy profile π is called $\mathbf{x}$-profitable if

$$w \cdot \pi \;=\; \sum_{(i, t_i, s_i)} \int_{T_{-i}} \pi_i(s_i|t_i) [v_i(t, \mathbf{x}(s_i, t_{-i})) - v_i(t, \mathbf{x}(t))] p_i(dt_{-i}|t_i) \;>\; 0,$$

where the summation above is indexed by $i \in I$, $t_i \in S_i$ and $s_i \in S_i \cup \{0\}$, and $\mathbf{x}(0, t_{-i})$ is defined to equal $\mathbf{x}(0)$, i.e., the disagreement outcome. (Obviously, everything goes through even if disagreement outcomes depend on who opts out and others' types.) The definition of attribution and its infinitesimal counterpart is just the same as in the previous subsection, except for the caveat that s_i ranges across $S_i \cup \{0\}$.

Theorem 12. (1) *A bargaining solution for* $\mathbf{x}$ *exists if and only if every infinitesimally* $\mathbf{x}$*-profitable strategy profile is at most infinitesimally* T*-attributable.* (2) *When types are independent, a bargaining solution exists if and only if every* $\mathbf{x}$*-profitable strategy profile is attributable.*

Once again, since the proof of this result is similar to previous ones, it is omitted. This result may be contrasted with Segal and Whinston (2009) in that—using duality—Theorem 12 characterizes existence of bargaining solutions even when values are interdependent, types are correlated, the type space is arbitrary and utility functions are not necessarily uniformly bounded.

4.6 Optimal Mechanisms

We now turn to a characterization of optimal mechanisms. We focus on the case of one agent for simplicity, although the multi-agent case follows easily from this one. To

this end, we give the principal a linear objective u over random allocations, i.e., maps $\mu : T \to \Delta(X)$, where now X is a measurable space. We assume that the principal has some beliefs $q \in \Delta(T)$ over the agent's types, so T is a measurable space, too. For the principal's objective to be well defined, we require a further assumption.

Assumption 2. The information structure $p : T \to \Delta(Y)$ is a measurable map.[15] The functions u and ξ are integrable in each of their variables as well as jointly.

The *principal's problem* is given by:

$$\begin{aligned}
\sup_{\mu \geq 0, \xi} \int_T \int_Y \int_X u(t,x,y)\mu(dx|t) + \xi(t,y)p(dy|t)q(dt) \;\text{ s.t. }\; \mu(X|t) &= 1 \quad \forall t, \\
\int_Y \int_X v(t,x,y)(\mu(dx|s) - \mu(dx|t))p(dy|t) \;\leq\; \int_Y \xi(s,y) - \xi(t,y)p(dy|t) &\quad \forall (t,s), \\
\int_Y \int_X v(t,x,y)\mu(dx|t) - \xi(t,y)p(dy|t) \;\geq\; 0 &\quad \forall t.
\end{aligned}$$

We will now we will make further assumptions on the problem to guarantee that (i) the principal's problem has a value, and (ii) the value of the principal's problem may be characterized with an alternative problem that subsumes any reference to money. Afterwards, we will discuss briefly the case of costly reporting.

Theorem 13. *If u and v are both measurable and uniformly bounded then the value of the principal's problem equals*

$$\begin{aligned}
&\inf_{\lambda \geq 0} \sup_{\mu \geq 0} \int [u(t,x,y) + v(t,x,y)]\mu(dx|t)p(dy|t)q(dt) \\
&+ \int [v(t,x,y) - v(s,x,y)]\mu(dx|t)p(dy|s)\lambda(ds,dt) \;\; \textit{s.t.} \;\; \mu(X|t) = 1 \;\forall t, \\
&\qquad\qquad p(\cdot)q(\cdot) \;=\; \int_T p(\cdot)\lambda(\cdot,ds) - p(s)\lambda(ds,\cdot) + p(\cdot)\lambda_0(\cdot),
\end{aligned}$$

where $\mu \in B(T, M(X))$.

Theorem 13 characterizes the value of the principal's problem in terms of its dual problem and finds conditions under which there is no duality gap.

The dual chooses an allocation to maximize "virtual welfare" (see, e.g., Myerson, 1981) subject to an undetectability constraint on λ. Specifically, the constraint stipulates that the detectability of a feasible deviation must equal the prior probability

[15] Specifically, p is measurable with the σ-algebra generated by the weak topology on $\Delta(Y)$.

net of probability with which the agent opts out in the dual problem. Interestingly, the allocation may be chosen *after* the agent has chosen his deviation, λ.

To illustrate the usefulness of Theorem 13, in a finite version of Myerson's (1981) setting (i.e., $T \subset \mathbb{R}$ is finite, $X = \{0,1\}$, $u \equiv 0$ and $v(t,x,y) = 0$ if $x = 0$ and $v(t,x,y) = t$ if $x = 1$) the dual to the principal's problem becomes

$$\begin{aligned} \inf_{\lambda \geq 0} \sum_{t \in T} \max\{tq(t) + \sum_{s \in T} \lambda(s,t)(t-s), 0\} \quad \text{s.t.} \\ p(t)q(t) \;=\; \sum_{s \in T} \lambda(t,s)p(t) - \lambda(s,t)p(s) + \lambda_0(t)p(t) \quad \forall t. \end{aligned}$$

Of course, in the regular case (see Myerson, 1981, p. 66) this problem may be simplified further. On the other hand, if $p(t) \notin \text{conv}\,\{p(s) : s \neq t\}$ for some t then t cannot misrepresent his own type, so his incentive constraint will not bind.

The framework above applies also to the case of costly reporting. Let $v(t,s,x,y)$ be the utility of type t from reporting s, getting the choice x and y realizing. Now the dual to the principal's problem becomes:

$$\begin{aligned} \inf_{\lambda \geq 0} \sup_{\mu \geq 0} \int [u(t,x,y) + v(t,x,y)]\mu(dx|t)p(dy|t)q(dt) \\ + \int [v(t,t,x,y) - v(s,t,x,y)]\mu(dx|t)p(dy|s)\lambda(ds,dt) \quad \text{s.t.} \quad \mu(X|t) \;=\; 1 \;\forall t, \\ p(\cdot)q(\cdot) \;=\; \int_T p(\cdot)\lambda(\cdot,ds) - p(s)\lambda(ds,\cdot) + p(\cdot)\lambda_0(\cdot). \end{aligned}$$

4.7 Revenue Equivalence Revisited

In this subsection, we (i) recast the problem of revenue equivalence in interim terms, and (ii) characterize interim-implementing incentive schemes. This characterization differs from other characterizations of interim revenue equivalence in the literature, such as Heydenreich et al. (2009). Indeed, they not only focus on the case of independent types, but moreover they characterize when expected payments only differ by a constant, rather than when the entire payment schedule is unique up to a constant.[16]

[16] Of course, when types are independent, the most we can hope for in terms of revenue equivalence is that expected payments di er by a constant. On the other hand, when types are not independent or there are other additional constraints imposed on the payment schemes, it may become meaningful to consider revenue equivalence in terms of the entire schedule of payments.

Mathematically, Theorem 5 shows that an allocation is interim implementable if and only if the function V, defined below, is subdifferentiable at 0 (assume $\alpha_{\pm} \geq 0$):

$$V(\alpha_{\pm}) \;=\; \sup_{\lambda \geq 0} \sum_{(t,s)} \lambda(t,s) w(t,s) \;\text{ s.t. } -\alpha_{-}(t) \leq \sum_{s \in T} \lambda(t,s) p(s) - \lambda(s,t) p(t) \leq \alpha_{+}(t) \; \forall t.$$

One way to obtain interim revenue equivalence is to first view implementing payment schemes as shadow values of the above profit-maximization problem for the agent. By the proof of Theorem 5, the set of all interim implementing payment schemes coincides with the subdifferential of V at 0. Therefore, interim revenue equivalence obtains if and only if for any $\xi_{\pm}$ and $\zeta_{\pm}$ in this subdifferential, there exists $\alpha \in \mathbb{R}$ such that $\xi_{+} - \xi_{-} + \alpha \mathbf{1} = \zeta_{+} - \zeta_{-}$. A comparable interpretation to that supplied for Theorem 3 applies here, too.

5 Conclusion

In this paper I characterize implementability (Theorem 1) as well as interim implementability Theorem 5) by making use of the Minimax Theorem, emphasizing a strategic interpretation. I also suggest some generalizations of these results. All these results improve on Rochet's Theorem both in supplying a strategic interpretation and also generalizing its result.

Mathematically, a notable difference with Rochet's or Rockafellar's approach is that they derive cyclic monotonicity in some sense by "integrating" a subdifferential correspondence. They then obtain a convex function and a "fundamental theorem of calculus" for convex functions. To them, the payment scheme is obtained by "integrating." On the other hand, I take the alternative system of inequalities from incentive compatibility and think of payment schemes as multipliers on the dual constraints, i.e., I view them as (directional) derivatives. Hence, I obtain the payment schemes by differentiating a dual value function, rather than integrating a subdifferential correspondence.

As a final comment, although the approach used in this paper bears some resemblance to linear semi-infinite programming (LSIP, see, e.g., Goberna and López, 1998), please note that in this paper there may be both (a) infinitely many (incentive) constraints and (b) infinitely many unknowns. Therefore, this isn't strictly speaking LSIP.

A Preliminaries

This appendix presents ancillary results that are used in the main body of the paper. Let us begin with Clark's (2006) extension of The Theorem of the Alternative.

Let X and Y be ordered, locally convex real vector spaces, with positive cones X_+ and Y_+ and topological dual spaces X^* and Y^* such that $X^{**} = X$ and $Y^{**} = Y$. Let $A : X \to Y$ be a continuous linear operator with adjoint operator $A^* : Y^* \to X^*$ and x any $b \in Y$. Finally, for any set S let $\overline{S}$ denote its closure.

Lemma A.1 (Clark, 2006, page 479). *For any $b \in Y$, there exists $x \in X_+$ such that $A(x) = b$ if and only if $A^*(y_0^*) \in \overline{X_+^* - \{A^*(y^*) : y^*(b) = 0\}}$ implies that $y_0^*(b) \geq 0$.*

Now consider the characterization of strong duality by Gretsky et al. (2002). With the same notation as above, a *linear program* is any triple (A, b, c^*) such that A is as above, $b \in Y$ and $c^* \in X^*$. The *primal* is given by the linear optimization problem $\sup\{c^*(x) : A(x) \leq b, x \geq 0\}$, and the *dual* by $\inf\{y^*(b) : A^*(y^*) \leq c^*, y^* \geq 0\}$. Say that *there is no duality gap* if the value of the primal equals the value of the dual. Denote by $V(b)$ the value of the primal as a function of b. The *subdifferential* of a function V at b is the set $\partial V(b) = \{y^* : V(y) - V(b) \leq y^*(y - b)\ \forall y \in Y\}$. V is *subdifferentiable* at b if $\partial V(b) \neq \emptyset$.

Lemma A.2 (Gretsky et al., 2002, page 265). *Both the dual has a solution and there is no duality gap if and only if V is subdifferentiable at b.*

For the next lemma, we need some de nitions. Let $f : X \times Y \to \mathbb{R} \cup \{\pm\infty\}$ be any function. Let $\text{dom} f = \{(x, y) : |f(x, y)| < \infty\}$. Write $\text{dom}_1 f$ and $\text{dom}_2 f$ for the projections of $\text{dom} f$ on X and Y, respectively. Say that f is *closed* if both $\{x' : f(x', y) \geq c\}$ and $\{y' : f(x, y') \leq c\}$ are closed sets for every $c \in \mathbb{R}$, $x \in \text{dom}_1 f$ and $y \in \text{dom}_2 f$. The function f is *concave-convex* if it is concave with respect to x for all $y \in \text{dom}_2 f$ and convex with respect to y for all $x \in \text{dom}_1 f$.

Lemma A.3 (Io e and Tikhomirov, 1968, page 84). *Let $f : X \times Y \to \mathbb{R} \cup \{\pm\infty\}$ be a closed concave-convex function, and define the following functions on X^* and Y^*:*

$$h(z) = \inf_{y \in \text{dom}_2 f} \sup_x \{x \cdot z - f(x, y)\} \quad \text{and} \quad k(w) = \sup_y \inf_{x \in \text{dom}_1 f} \{f(x, y) + y \cdot w\}.$$

For f to have a saddle point it is necessary and sufficient that $\partial h(0) \neq \emptyset \neq \partial k(0)$. The set of saddle points coincides with the product $\partial h(0) \times \partial k(0)$.

B Proof of Theorem 1

First of all, let us prove the theorem under the restriction that T is a nite set.

Lemma 3. *If T is a finite set then an allocation* x *is implementable if and only if every* x*-profitable deviation is detectable.*

Proof. By the Theorem of the Alternative (see, e.g., Rockafellar, 1970, page 198), a scheme $\xi \in \mathbb{R}^T$ exists such that $v(t, \mathsf{x}(s)) - v(t, \mathsf{x}(t)) \leq \xi(s) - \xi(t)$ for every $t, s \in T$ if and only if there does not exist a vector $\lambda \geq 0$ satisfying (i) $\sum_s \lambda(s,t) = \sum_s \lambda(t,s)$ for all $t \in T$, and (ii) $\sum_{(t,s)} \lambda(s,t)[v(t, \mathsf{x}(s)) - v(t, \mathsf{x}(t))] > 0$. Each of these two conditions on λ is independent of $\lambda(t,t)$ for all $t \in T$, so assume without loss of generality that $\lambda(t,t) = \max\{\sum_{s \neq r} \lambda(s,r) : r \in T\} - \sum_{s \neq t} \lambda(s,t)$ for all $t \in T$. Now λ is proportional to a doubly stochastic matrix | in other words, a reporting strategy, call it π | which satis es (i) and (ii) if and only if λ satis es (i) and (ii). But (i) is just the requirement that π be undetectable, and (ii) states that π is x-pro table. □

Now suppose that T is not necessarily nite. We begin with some preliminaries.

For any set Z, let $\mathbb{R}^Z$ be the space of all real-valued functions on Z endowed with the product topology, and let $\mathbb{R}^Z_+ = \{f \in \mathbb{R}^Z : f(z) \geq 0 \ \forall z \in Z\}$ denote its positive cone. Let $\mathbb{R}^{(Z)}$ be the subspace of all real-valued functions g on Z with nite support, i.e., such that the set $\{z \in Z : g(z) \neq 0\}$ is nite. Any $g \in \mathbb{R}^{(Z)}$ is described by a nite set $\operatorname{supp} g = \{z_1, \ldots, z_m\}$ of elements in Z (the support of g) together with a nite-dimensional vector $(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$. Such a g acts on $\mathbb{R}^Z$ as follows:

$$g(f) = \sum_{k=1}^{m} \lambda_k f(z_k) \qquad \forall f \in \mathbb{R}^Z.$$

An important example is the *evaluation* functional $\mathsf{e}_z \in \mathbb{R}^{(Z)}$, de ned by $\mathsf{e}_z(f) = f(z)$. Clearly, any $g \in \mathbb{R}^{(Z)}$ can be written in terms of these evaluations as $g = \sum_k \lambda_k \mathsf{e}_{z_k}$. It is well known (e.g., Conway, 1990, p. 115) that $\mathbb{R}^{(Z)}$ is the topological dual of $\mathbb{R}^Z$, i.e., the space of continuous linear functionals on $\mathbb{R}^Z$. Let $\mathbb{R}^{(Z)}_+$ be its positive cone.

Any $g \in \mathbb{R}^{(Z \times Z)}$ is given by a nite support $\{(z_{11}, z_{21}), \ldots, (z_{1m}, z_{2m})\}$ and a vector $(\lambda_1, \ldots, \lambda_m)$. We will describe it instead by the subset $\{z : z = z_{ik} \text{ for some } i, k\}$ of Z with, say, n elements, denoted by $\operatorname{supp}_Z g = \{z_1, \ldots, z_n\}$ together with the $n \times n$ matrix $(\lambda_{11}, \ldots, \lambda_{1n}, \ldots, \lambda_{n1}, \ldots, \lambda_{nn})$ de ned by $\lambda_{k\ell} = \lambda_i$ if $(z_k, z_\ell) = (z_{1i}, z_{2i})$ and 0 if no such i exists. Clearly, both descriptions are equivalent.

Let $w \in \mathbb{R}^{T \times T}$ be the function de ned pointwise by $w(t,s) = v(t, \mathsf{x}(s)) - v(t, \mathsf{x}(t))$.

De ne pointwise the following operator $D : \mathbb{R}^{(T\times T)} \to \mathbb{R}^{(T)}$. Given $g \in \mathbb{R}^{(T\times T)}$, let $Dg = \sum_{(k,\ell)} \lambda_{k\ell}(\mathsf{e}_{t_k} - \mathsf{e}_{t_\ell})$. Hence, $Dg(f) = \sum_{(k,\ell)} \lambda_{k\ell}[f(t_k) - f(t_\ell)]$ for all $f \in \mathbb{R}^T$.

Lemma 4. *The following are equivalent:*
(i) *For every $g \in \mathbb{R}_+^{(T\times T)}$, $Dg = 0$ implies that $g(w) \leq 0$.*[17]
(ii) *There exists a net $\{\xi_\delta\}$ such that $w(t,s) \leq \liminf_\delta \xi_\delta(s) - \xi_\delta(t)$ for all (t,s).*

Proof. Let $X = \mathbb{R}^{(T\times T)}$ and $Y = \mathbb{R}^{(T)} \times \mathbb{R}$. Let $A : X \to Y$ be the operator de ned pointwise by $A(g) = (Dg, g(w))$. Since $\mathbb{R}_+^{(T\times T)}$ is a cone, (i) fails if and only if there exists $g \in \mathbb{R}_+^{(T\times T)}$ such that $Dg = 0$ and $g(w) = 1$, i.e., $A(g) = (0,1)$. The operator A is clearly linear and continuous, so by Lemma A.1, $A(g) = (0,1)$ if and only if given any number ε, incentive scheme ξ and net $\{(w_\delta, \xi_\delta) \in \mathbb{R}_+^{T\times T} \times \mathbb{R}^T\}$,

$$\xi(s) - \xi(t) + \varepsilon w(t,s) \;=\; \lim w_\delta(t,s) - [\xi_\delta(s) - \xi_\delta(t)] \quad \forall (t,s) \quad \Rightarrow \quad \varepsilon \geq 0.$$

Since $w_\delta \geq 0$, this condition is equivalent to

$$\xi(s) - \xi(t) + \varepsilon w(t,s) \;\geq\; \limsup -[\xi_\delta(s) - \xi_\delta(t)] \quad \forall (t,s) \quad \Rightarrow \quad \varepsilon \geq 0.$$

Rearranging, multiplying by -1 and replacing without any loss of generality ξ_δ with $\xi_\delta + \xi$ yields the equivalent condition

$$-\varepsilon w(t,s) \;\leq\; \liminf \xi_\delta(s) - \xi_\delta(t) \quad \forall (t,s) \quad \Rightarrow \quad \varepsilon \geq 0.$$

Therefore, (i) holds if and only if there exists a number $\varepsilon > 0$ and a net $\{\xi_\delta\}$ such that $\varepsilon w(t,s) \leq \liminf \xi_\delta(s) - \xi_\delta(t)$ for all (t,s). This last requirement is clearly equivalent to (ii) by dividing both sides by ε and replacing ξ_δ with $\varepsilon\xi_\delta$, as claimed. □

It is easy to see that (i) is necessary and su cient for every undetectable deviation to be x-unpro table. Indeed, given $t \in T$ let $f_t \in \mathbb{R}^T$ be the indicator function of t, i.e., $f_t(s) = 1$ if $s = t$ and 0 otherwise. For su ciency, if $g \in \mathbb{R}_+^{(T\times T)}$ satis es $\sum_\ell \lambda_{k\ell} = 1$ for all k then g is a reporting strategy. If $Dg = 0$ then $\sum_k \lambda_{k\ell} = 1$, too, since $Dg(f_{t_\cdot}) = \sum_k \lambda_{k\ell} - \lambda_{\ell k}$ for every $t_\ell \in \mathrm{supp}_T\, g$, so g is undetectable. Finally, it is clear that $g(w) \leq 0$ is equivalent to g being x-unpro table. For necessity, every $g \in \mathbb{R}_+^{(T\times T)}$ is proportional to a reporting strategy, and the value of Dg is determined by $Dg(f_{t_\cdot})$ for every $t_\ell \in \mathrm{supp}_T\, g$, so if it is doubly stochastic then $Dg = 0$.

The last step in our proof of Theorem 1 is to show that (ii) implies implementability.

Lemma 5. *The following statements are equivalent:*
(i) *There exists a net $\{\xi_\delta\}$ such that $w(t,s) \leq \liminf_\delta \xi_\delta(s) - \xi_\delta(t)$ for all (t,s).*
(ii) *There exists an incentive scheme ξ such that $w(t,s) \leq \xi(s) - \xi(t)$ for all (t,s).*

[17] $0 \in \mathbb{R}^{(T)}$ denotes the zero functional such that $0(f) = 0$ for all $f \in \mathbb{R}^T$.

Proof. That (ii) implies (i) is immediate. For the converse, without loss of generality we may x any $t_0 \in T$ and assume that $\xi_\delta(t_0) = 0$ for all δ in the net, since it will not a ect the any of the di erences $\xi_\delta(s) - \xi_\delta(t)$. By hypothesis,

$$w(t,t_0) \;\leq\; \liminf \xi_\delta(t) \;\leq\; \limsup \xi_\delta(t) \;=\; -\liminf -\xi_\delta(t) \;\leq\; -w(t_0,t) \qquad \forall t \in T.$$

Hence, $\xi(t) = \liminf \xi_\delta(t)$ is bounded. Since the lim inf function is superadditive, it follows that $\liminf \xi_\delta(s) - \xi_\delta(t) + \liminf \xi_\delta(t) \leq \liminf \xi_\delta(s)$ for every (t,s). Hence, $\liminf \xi_\delta(s) - \xi_\delta(t) \leq \xi(s) - \xi(t)$. By (i), $w(t,s) \leq \liminf \xi_\delta(s) - \xi_\delta(t)$. Collecting these last two inequalities nally yields $w(t,s) \leq \xi(s) - \xi(t)$, as required. □

C Proof of Theorem 4

If T is nite then the result follows by a similar argument to the one used to prove Lemma 3. Let $R = \{(i,s_i,t) : i \in I,\ s_i \in T_i \text{ and } t \in T\}$. By a similar argument to that of Lemma 4, there exists a net of incentive schemes $\{\xi^\delta\}$ such that both $v_i(t, \mathsf{x}(s_i,t_{-i})) - v_i(t,\mathsf{x}(t)) \leq \liminf_\delta \xi_i^\delta(s) - \xi_i^\delta(t)$ for every (i,t_i,s_i,t_{-i}) and $\lim_\delta \sum_i \xi_i^\delta(t) = 0$ for all t (call this condition $(*)$) if and only if for every $\lambda \in \mathbb{R}_+^{(R)}$ and $\eta \in \mathbb{R}^{(T)}$, the system of equations given by $\sum_{s_i}[\lambda_i(s_i,t) - \lambda_i(t_i,s_i,t_{-i})] = \eta(t)$ for every (i,t) implies $\sum_{(i,s_i,t)} \lambda_i(s_i,t)[v_i(t,\mathsf{x}(s_i,t_{-i})) - v_i(t,\mathsf{x}(t))] \leq 0$ (call this condition $(**)$). Clearly, $(**)$ is equivalent to (ii). To see this, just divide every $\lambda_i(s_i,t)$ by $\quad = \max_{(i,t)} \sum_{s_i} \lambda_i(s_i,t)$ (if this equals zero then there's nothing to prove), as well as $\eta(t)$, and replace $\lambda_i(t_i,t)$ with $\quad - \sum_{s_i} \lambda_i(s_i,t)$ for every (i,t). Now λ is proportional (with weight $\quad$) to an unattributable deviation pro le that is also unpro table. That (ii) implies $(**)$ is obvious. It remains to prove that $(*)$ is equivalent to (i). Again, that (i) implies $(*)$ is obvious. Conversely, let $\{\xi^\delta\}$ be a net that satis es $(*)$. Fix any $t^0 \in T$. Given (i,t,δ), de ne the net $\{\zeta^\delta\}$ by $\zeta_i^\delta(t) = \xi_i^\delta(t) + \sum_{j \neq i} \xi_j^\delta(t_i^0, t_{-i})$. By $(*)$, $\lim_\delta \zeta_i^\delta(t_i^0,t_{-i}) = 0$ for all t_{-i}, and $v_i(t,\mathsf{x}(s_i,t_{-i})) - v_i(t,\mathsf{x}(t)) \leq \liminf_\delta \zeta_i^\delta(s) - \zeta_i^\delta(t)$ for every (i,t_i,s_i,t_{-i}). Hence, following the proof of Lemma 5, the scheme ζ de ned by $\zeta_i(t) = \liminf_\delta \zeta_i^\delta(t) \in \mathbb{R}$ for every (i,t) ex post implements x. Let $\{\zeta^\gamma\}$ be a subnet of $\{\zeta^\delta\}$ such that $\lim_\gamma \zeta_i^\gamma(t) = \zeta_i(t)$ for all (i,t). One such subnet exists by de nition of lim inf and Lemma 1. Finally, for every $i_1 \in I$ and $t \in T$ let

$$\zeta_{i_1}^0(t) \;=\; \zeta_{i_1}(t) - \sum_{i_2 \neq i_1} \zeta_{i_2}(t_{i_1}^0, t_{-i_1}) + \sum_{i_3 \neq i_2} \zeta_{i_3}(t_{i_1}^0, t_{i_2}^0, t_{-i_1 i_2}) - \cdots + \sum_{i_n \neq i_{n-1}} \zeta_{i_n}(t_{-i_n}^0, t_{i_n}).$$

Clearly, ζ^0 ex post implements x because ζ does, too, since for all (i,t), $\zeta_i^0(t)$ equals $\zeta_i(t)$ plus something that does not depend on t_i. By construction, it is easy to see that the

scheme ζ^0 also satis es budget balance, since

$$\begin{aligned}\sum_{i_1\in I}\zeta^0_{i_1}(t) &= \sum_{i_1\in I}\zeta_{i_1}(t) - \sum_{i_2\neq i_1}\zeta_{i_2}(t^0_{i_1}, t_{-i_1}) + \sum_{i_3\neq i_2}\zeta_{i_3}(t^0_{i_1}, t^0_{i_2}, t_{-i_1 i_2}) - \cdots + \sum_{i_n\neq i_{n-1}}\zeta_{i_n}(t^0_{-i_n}, t_{i_n})\\ &= \lim \sum_{i_1\in I}\zeta^\gamma_{i_1}(t) - \sum_{i_2\neq i_1}\zeta^\gamma_{i_2}(t^0_{i_1}, t_{-i_1}) + \sum_{i_3\neq i_2}\zeta^\gamma_{i_3}(t^0_{i_1}, t^0_{i_2}, t_{-i_1 i_2}) - \cdots + \sum_{i_n\neq i_{n-1}}\zeta^\gamma_{i_n}(t^0_{-i_n}, t_{i_n})\\ &= \lim \sum_{i_1\in I}\xi^\gamma_{i_1}(t) + \sum_{i_2\neq i_1}\zeta^\gamma_{i_2}(t^0_{i_1}, t_{-i_1}) - \sum_{i_2\neq i_1}\zeta^\gamma_{i_2}(t^0_{i_1}, t_{-i_1}) + \cdots\\ &\quad - \sum_{i_n\neq i_{n-1}}\zeta^\gamma_{i_n}(t^0_{-i_n}, t_{i_n}) + \sum_{i_n\neq i_{n-1}}\zeta^\gamma_{i_n}(t^0_{-i_n}, t_{i_n}) + \sum_{j\neq i_n}\xi^\gamma_j(t^0) = 0.\end{aligned}$$

Therefore, ζ^0 ex post implements x with budget balance.

D Proofs of Theorem 5 and Proposition 2

This result extends Theorem 1 by admitting any measurable space Y instead of just the trivial one. However, the t091s95(o2[-39t091el41(Ho)2the)-d5(t091 ee)-not5(o2[-foll41(Ho)2the)-326(the)-sam

Proof. The proof is rather simple. Suppose that x is Y-interim implementable with scheme ξ. By de nition, (ξ, θ) is an equilibrium of F that pays 0 to the agent in the hypothetical zero-sum game. Conversely, if x is not Y-interim implementable then for any ξ there is a reporting strategy g that gives the agent positive pro t. Doubling g doubles the agent's pro t. Further doubling and doubling implies that there is no best response for the agent. Hence, equilibrium fails to exist. □

Let h and k be functions with respective domains $[B(Y)^*]^{(T)}$ and $\mathbb{R}^{T\times T}$ given by

$$h(z) = \inf_{g\in \mathrm{dom}_2 F} \sup_{\xi} \{\xi \cdot z - F(\xi, g)\} \quad \text{and} \quad k(v) = \sup_{g} \inf_{\xi\in \mathrm{dom}_1 F} \{F(\xi, g) + g \cdot v\}.$$

We may now characterize equilibrium in terms of subdi erentiability of h and k. For the de nition of subdi erential ∂f of a function f, see Appendix A.

Lemma 7. *An equilibrium of F exists if and only if both h and k are subdifferentiable at the origin. Moreover, the set of equilibria of F is the product $\partial h(0) \times \partial k(0)$.*

Proof. By Lemma A.3, it su ces to prove that F is a closed convex-concave function, but this follows immediately from the de nition of F. □

By Lemma 6, if an equilibrium of F exists without loss it involves truthful reporting, so $\theta \in \partial k(0)$. This observation is useful because it facilitates verifying equilibrium existence. Indeed, by Lemma 7, equilibrium exists if and only if there is a scheme ξ and a truthful reporting strategy θ such that

$$h(z) - h(0) \geq \xi \cdot z \quad \forall z \quad \text{and} \quad k(v) - k(0) \geq \theta \cdot v \quad \forall v. \tag{6}$$

By Lemma 7, an equilibrium exists only if both $h(0)$ and $k(0)$ equal zero. Since equilibrium involves truth-telling, if equilibrium fails to exist then either $k(0) = +\infty$, hence an x-pro table Y-undetectable deviation exists, or $k(0) = 0$ and $\partial k(0) \neq \emptyset$,[18] hence $\partial h(0) = \emptyset$. This helps to characterize equilibrium existence as follows.

Lemma 8. *An equilibrium of F exists if and only if every infinitesimally Y-detectable deviation is at most infinitesimally* x*-profitable.*

Proof. If equilibrium exists then $h(0) = 0$ and there exists $\xi^* \in \partial h(0)$ by Lemma 7. If $h(z)$ is nite then there exists g such that $\xi \cdot z = Dg(\xi)$ for all ξ. In this case, $h(z) = \inf_g\{-w \cdot g : Dg = z\}$. Dividing by $|\xi^* \cdot z|$ in the subdi erential inequality for h yields $\sup_g\{w \cdot g/|\xi^* \cdot z| : Dg = z\} \leq 1$ for all z (even if $h(z)$ is in nite), therefore we may also take the supremum of the left-hand side with respect to z. Subsituting for $\xi^* \cdot y =$

[18] To see this, notice that $k(v) \geq \theta \cdot v$ by replacing g with θ in the supremum that de nes k.

$Dg(\xi^*)$ and dividing both the numerator and denominator by $\sum_{(t,s)} \lambda(t,s)$, we obtain that $\sup_g w \cdot g / |Dg(\xi^*)| = \sup_\pi w \cdot \pi / |D\pi(\xi^*)| \leq 1$, so $\mathscr{D} = \inf_\xi \sup_\pi w \cdot \pi / |D\pi(\xi^*)| < +\infty$ and su ciency follows. For necessity, if no equilibrium exists then by Lemma 7 and the observation above, for every ξ there exists z such that $h(z) < \xi \cdot z$, so $h(z) < +\infty$ and $R(y|\xi) = \sup_g \{w \cdot g / |Dg(\xi)| : Dg = z\} > 1$. Hence, $R(\xi) = \sup_z R(z|\xi) \geq 1$. But now halving ξ makes $R(\xi/2) > 2$, and halving again and again implies that $\inf_\xi \sup_g w \cdot g / |Dg(\xi)| = +\infty$, as desired. □

This proves Theorem 5. Now, let us prove Proposition 2 with the next two results.

Lemma 9. *An equilibrium of F exists if and only if for every net $\{g_\delta\} \subset \mathbb{R}_+^{(T\times T)}$,*

$$\lim Dg_\delta(\xi) = 0 \quad \forall \xi \in B(Y)^T \quad \Rightarrow \quad \limsup w \cdot g_\delta \leq 0, \tag{7}$$

where $Dg_\delta(\xi) = \int_Y \sum_{(t,s)} \xi(t,y)[\lambda_\delta(t,s)p(dy|s) - \lambda_\delta(s,t)p(dy|t)]$ for every scheme ξ and $w \cdot g_\delta = \sum_{(t,s)} w(t,s)\lambda_\delta(t,s)$ for every $w \in \mathbb{R}^{T\times T}$.

Proof. By Lemma 7, an equilibrium exists if and only if (6) holds. If (6) holds then $h(0) = 0$, so there exists ξ^* such that $h(z) \geq \xi^* \cdot z$ for all z. For any $g \in \text{dom}_2\ F$, if $Dg(\xi) \neq \xi \cdot z$ for some ξ then $\sup_\xi \{\xi \cdot z - F(\xi, g)\} = +\infty$ by linearity. Therefore, $h(z) = -\sup_g \{w \cdot g : Dg = z, g \in \text{dom}_2\ F\}$, which implies that ξ^* exists such that $\sup_g w \cdot g - Dg(\xi^*) \leq 0$, where $g \in \text{dom}_2\ F$ satis es $Dg = z$. By letting $z \to 0$, su ciency now follows. Conversely, suppose that an equilibrium fails to exist. If there exists g such that $Dg(\xi) = 0$ for all ξ yet $w \cdot g > 0$, the consequent of the claim above fails and we are done, so suppose not. Hence, $h(0) = k(0) = 0$. Notice that the subdi erential inequality holds for k. Indeed, $k(v) \geq F(0,\theta) + \theta \cdot v = \theta \cdot v$, since $w \cdot \theta = 0$. Therefore, failure of subdi erentiability applies to h, so for every ξ there exists y such that $\sup_g \{w \cdot g : Dg = z, g \in \text{dom}_2\ F\} > \xi \cdot z$. Setting $\xi = 0$ and taking a sequence $\{g_n\}$ that achieves this sup establishes necessity. □

We now establish the last lemma we need to prove Proposition 2.

Lemma 10. *Let $g \in \mathbb{R}_+^{(T\times T)}$ and suppose that $p(t)$ does not depend on t. $Dg(\xi) = 0$ for all ξ implies that $w \cdot g \leq 0$ if and only if (7) holds for every net $\{g_\delta\}$.*

Proof. Necessity is evident. For su ciency, consider a net $\{g_\delta\}$ with $\lim Dg_\delta(\xi) = 0$ for all ξ but $w \cdot g_\delta = 1$. Without loss, we consider a convergent subsequence, $\{g_m\}$, by picking g_m such that $|w \cdot g_m - 1| < 1/m$. De ne the vector $\mu_m \in [B(Y)^*]^T$ pointwise by $\mu_m(t) = \sum_s \lambda_m(t,s)p(s) - \lambda_m(s,t)p(t)$ for every t. We consider three cases.

– *Case 1a*: Suppose that supp λ_m does not depend on m.

Let $S_m = \{t : \lambda_m(t,s) > 0 \text{ for some } s\} \cup \{t : \lambda_m(s,t) > 0 \text{ for some } s\}$ be the set of types to which λ_m gives positive weight. By hypothesis, and $S_m = S$ is independent of m and $|S| < \infty$. Consider the cone $C = \{g \in \mathbb{R}_+^{S\times S} : Dg = 0\}$. Since this cone is nitely generated, it is closed. Therefore, by the Theorem of the Alternative, there does not exist $g \geq 0$ such that $Dg = 0$ in $[B(Y)^*]^S$ and $w \cdot g > 0$ if and only if there is a scheme $\xi : S \to \mathbb{R}$ that Y-interim implements x assuming that the type space is S. Applying Lemma 9, the result now follows.

– *Case 1b*: Suppose that supp μ_m does not depend on m but supp λ_m does.

Let supp $\mu_m = T_0$, which, by hypothesis, does not depend on m. Clearly, $T_0 \subset S_m$. If $T_0 = S_m$ then we are in Case 1a, and we are done. Otherwise, $S_m \setminus T_0 \neq \emptyset$. Let $\lambda_m^0(t,s) = \lambda_m(t,s)$ if either t or s (or both) belong to $S_m \setminus T_0$ and 0 otherwise, and let $\lambda_m^1 = \lambda_m - \lambda_m^0$. Consider the following optimization problem:

$$V_m = \min_{\eta_m \geq 0} \left\|\lambda_m^1 - \eta_m\right\|_1 \quad \text{s.t.}$$
$$\sum_{s\in T_0} \eta_m(t,s)p(s) - \eta_m(s,t)p(t) + \sum_{s\in S_m\setminus T_0} \lambda_m(t,s)p(s) - \lambda_m(s,t)p(t) = 0 \qquad \forall t \in T_0.$$

If types are independent, i.e., $p(t)$ does not depend on t, then this problem has a feasible solution and we may avoid reference to y without any loss of generality. Indeed, for any vector $b \in \mathbb{R}^{T_0}$, by the Theorem of the Alternative $\eta \geq 0$ exists such that $\sum_s \eta(t,s) - \eta(s,t) = b(t)$ for all t if and only if $\sum_t b(t) = 0$, and clearly $\sum_{t\in T_0}\sum_{s\in S_m\setminus T_0} \lambda_m(t,s) - \lambda_m(s,t) = \sum_{(t,s)} \lambda_m^1(t,s) - \lambda_m^1(s,t) = 0$.[19] Taking the dual of this problem, manipulating it and applying strong duality yields

$$V_m = \max_{\zeta\in\mathbb{R}^{T_0}} \sum_{t\in T_0} \zeta(t) \sum_{s\in S_m} \lambda_m(t,s) - \lambda_m(s,t) \quad \text{s.t.} \quad -1 \leq \zeta(t) - \zeta(s) \leq 1 \quad \forall (t,s).$$

But this dual problem is easily solved, yielding $V_m = \frac{1}{2}\sum_t |\sum_s \lambda_m(t,s) - \lambda_m(s,t)|$. Since T_0 is nite, it follows that $\mu_m \to 0$ in norm, hence $V_m \to 0$. If a subsequence exists such that for every k there is (t,s) such that $w(t,s) = -\infty$ and $\lambda_{m_k}(t,s) > 0$ then we are done, so suppose not, i.e., for m su ciently large, $w(t,s) > -\infty$ for all (t,s) such that $\lambda_m(t,s) > 0$. Finally, if η_m is an optimal primal solution then $w \cdot \lambda_m = w \cdot (\lambda_m^0 + \eta_m + \lambda_m^1 - \eta_m) \leq w \cdot (\lambda_m^1 - \eta_m) \leq \|w\|_\infty \left\|\lambda_m^1 - \eta_m\right\|_1 \to 0$, where the rst inequality follows because by construction $\lambda_m^0 + \eta_m$ is undetectable, and we are assuming that $w \cdot g \leq 0$ for every undetectable strategy. But this contradicts the original hypothesis that $w \cdot \lambda_m \to 1$, and the claim is established for this case.

[19] Notice that the conclusion that the problem above is feasible does not necessarily follow if types fail to be independent. For instance, in the setting of Example 1, suppose that $\lambda_m(t,s) = 1$ if t and s both belong to $\{0, \frac{1}{2}, 1\}$ and $s = t \pm \frac{1}{2}$, otherwise $\lambda_m(t,s) = 0$.

– *Case 2*: Suppose that supp μ_m depends on m.

If supp $\mu_m = T_m$ has a subsequence $\{T_{m_k}\}$ that does not depend on k then we are back to Case 1, so suppose not. Hence, there is a subsequence $\{T_{m_k}\}$ such that $T_{m_k} \setminus \bigcup_{\ell<k} T_{m_\ell} \neq \emptyset$ for all k. Without loss, assume that this is the sequence with which we began. Construct $\{t_m\}$ according to innovations in T_m, i.e., such that $t_m \in T_m$ yet $t_m \notin T_k$ for all $k < m$.

By taking a subsequence if necessary, without loss $\{\lambda_m\}$ satisfies $\|\mu_m(t_k)\| < 2^{-m}$ for all $k \leq m$. Let $\xi(t_m, y) = [1_{P_m}(y) - 1_{N_m}(y)]/\|\mu_m(t_m)\|$ for all m and all y, where P_m and N_m are the positive and negative sets in a Hahn decomposition of Y relative to μ_m (see, e.g., Folland, 1999, for a definition of Hahn decomposition), and $1_X(y)$ is the indicator function of $X \subset Y$. (If $t \notin \{t_m\}$ then $\xi(t, y) = 0$.) For every $m \in \mathbb{N}$,

$$\begin{aligned} \xi \cdot \mu_m &= \sum_{k=1}^{m} \xi(t_k) \cdot \mu_m(t_k) = 1 + \sum_{k<m} \xi(t_k) \cdot \mu_m(t_k) \\ &> 1 - \sum_{k<m} 2^{-m} \cdot 2^k = 1 - \sum_{k<m} 2^{k-m} = 1 - (1 - 2^{-m}) \geq 1/2. \end{aligned}$$

Therefore, it is not the case that $Dg_m(\xi) \to 0$ for all ξ, and (7) follows vacuously. □

We end with a few remarks. With regard to Theorem 5, notice that the proof exploits the construction alluded to in the introduction of a zero-sum game between the principal and the agent and then solving for its equilibrium. This approach allows for greater generality[20] than that of Theorem 1 because it allows for w to take possibly infinite values. This may be due to conditionally infeasible reports, for instance.[21]

The proof of Proposition 2 is useful for two reasons. Firstly, Lemma 9 shows how the dual system of inequalities that are equivalent to implementability generalizes from the standard case to the interim case. Secondly, Lemma 10 reconciles the two theorems on implementation by showing that when types are independent we revert back to the original requirement of detecting profitable deviations to characterize implementability. Hence, the additional requirement of infinitesimally detectable deviations being at most infinitesimally profitable loses any bite in this setting.

It is interesting to note that the independence assumption is used to prove Lemma 10 only in Case 1b. In the other two cases, the assumption is not necessary, and in fact not used. This observation reveals the structure of the examples used to illustrate the differences between Theorems 1 and 5. There, the sequence of deviations constructed fit into Case 1b, i.e., the support of μ_m was independent of m but the support of π_m crucially was not. As a

[20] A technical improvement is that the proof of Theorem 1 uses a version of the Hahn-Banach Theorem, hence relies on (a weakening of) the axiom of choice. The proof of Theorem 5 does not.

[21] The literature sometimes refers to this as "evidence."

nal remark, note that the duality used to establish Case 1b can be used to provide a dual characterization of when detecting pro table deviations implies its in nitesimal counterpart. Namely, as long as the primal problem of Case 1b is feasible, or equivalently its dual is bounded, we obtain our desired result.

E Proofs of Theorems 7, 8, 9 and Proposition 3

Let us begin with the proof of Theorem 7. The proof proceeds in three steps. In the rst step, we describe full surplus extraction with a family of linear inequalities. In the second step, apply Lemma A.1 to obtain a necessary and su cient condition for full surplus extraction. Finally, in the last step we relate this dual condition to virtual convex independence when utility functions are bounded.

By de nition, all the surplus can be extracted if there exists a scheme ξ | called a *surplus-extracting scheme* | such that

$$v(t,t) = \int_Y \xi(t,y)p(dy|t) \qquad \forall t, \quad \text{and}$$
$$w(t,s) \leq \int_Y [\xi(s,y) - \xi(t,y)]p(dy|t) \qquad \forall (t,s),$$

where $v(t,s) = \int_Y v(t, \mathsf{x}(s,y), y)p(dy|t)$ and $w(t,s) = v(t,s) - v(t,t)$. Clearly, this is a system of linear inequalities with respect to ξ. Appealing to duality, we obtain the following characterization of existence of solutions to this linear system.

Lemma 11. *There exists a surplus-extracting scheme ξ if and only if for every net $\{(\lambda_\delta, \eta_\delta)\}$ such that $\lambda_\delta \in \mathbb{R}_+^{(T\times T)}$ and $\eta_\delta \in \mathbb{R}^{(T)}$,*

$$\eta_\delta(\cdot)p(\cdot) + \sum_{s\in T} \lambda_\delta(s,\cdot)p(s) - \lambda_\delta(\cdot,s)p(\cdot) \to 0 \quad \Rightarrow \quad \lim \lambda_\delta \cdot (w + W) \leq 0,$$

where $W(t,s) = v(s,s) - v(t,t)$.

Proof. B62 -31.114 Td [(wher)51(e)]TJ/F11 10.9091 Tf 17.843 0 Td [(0)]TJ/F14 10.9091 Tf4J/F11 10.909

By construction of $\{(\lambda_\delta, \eta_\delta)\}$ and (a), this is nally equivalent to

$$\lim \sum_{(s,t)} \lambda_\delta^+(t,s)[w(t,s) - W(t,s)] \ \leq \ 0,$$

and the claimed result follows. □

The proof of Theorem 7 is almost completed by the next easy lemma.

Lemma 12. *For every net $\{(\lambda_\delta, \eta_\delta)\}$ such that $\lambda_\delta \in \mathbb{R}_+^{(T\times T)}$ and $\eta_\delta \in \mathbb{R}^{(T)}$,*

$$\eta_\delta(\cdot)p(\cdot) + \sum_{s\in T} \lambda_\delta(s,\cdot)p(s) - \lambda_\delta(\cdot,s)p(\cdot) \ \to \ 0$$

implies that $\lim \lambda_\delta \cdot w = 0$ for all w such that $w(t,t) = 0$ given t if and only if p exhibits virtual convex independence.

Proof. Su ciency is immediate by letting $\eta_\delta(t) = \sum_s \lambda_\delta(t,s) - \lambda_\delta(s,t)$ for all (δ, t). For necessity, suppose that p exhibits virtual convex independence and that the above limiting condition holds for $\{(\lambda_\delta, \eta_\delta)\}$. We will show that $\lim \lambda_\delta \cdot w = 0$ for all requisite w. By integrating with respect to y, notice that $\eta_\delta(\cdot) - \sum_s \lambda_\delta(\cdot,s) - \lambda_\delta(s,\cdot) \to 0$ is necessary. Substituting, we obtain $\sum_s \lambda_\delta(s,\cdot)[p(s) - p(\cdot)] \to 0$. Hence, by virtual convex independence, $\lambda_\delta \cdot w = 0$ for all relevant w, as required. □

That virtual convex independence implies full surplus extraction now follows. For the converse, suppose that virtual convex independence fails, so there is a net $\{\lambda_\delta\}$ with $\lambda_\delta \geq 0$ such that $\sum_s \lambda_\delta(s,\cdot)[p(s) - p(\cdot)] \to 0$ yet $\lim \lambda_\delta \cdot w > 0$ for some w with $w(t,t) = 0$ given t. Now de ne $v(t,s) = w(t,s)$ for all (t,s). By Lemmata 11 and 12, there is no surplus-extracting scheme. The proof of Theorem 7 is now complete.

Next, let us prove Theorem 8. We will broadly follow the same steps as for the previous proof, but discuss in some detail the dual condition to surplus extraction given interim implementability before equating it to asymptotic convex independence.

Recall that by de nition all the surplus can be extracted from (v, x) conditional on interim implementability if there is a scheme ξ such that

$$\begin{aligned} v(t,t) \ &= \ \int_Y \xi(t,y)p(dy|t) \qquad \forall t, \quad \text{and} \\ 0 \leq \int_Y &[\xi(s,y) - \xi(t,y)]p(dy|t) \qquad \forall (t,s). \end{aligned}$$

Lemma 13. *All the surplus can be extracted from (v, x) assuming interim implementability if and only if for every net $\{(\lambda_\delta, \eta_\delta)\}$ with $\lambda_\delta \in \mathbb{R}_+^{(T\times T)}$ and $\eta_\delta \in \mathbb{R}^{(T)}$,*

$$\eta_\delta(\cdot)p(\cdot) + \sum_{s\in T} \lambda_\delta(s,\cdot)p(s) - \lambda_\delta(\cdot,s)p(\cdot) \ \to \ 0 \quad \Rightarrow \quad \lim \eta_\delta \cdot v \ \leq \ 0. \tag{11}$$

The proof of this result is almost identical to that of Lemma 11, hence omitted. Our next step in the proof of Theorem 8 is to show that the dual condition (11) above is equivalent to asymptotic convex independence. But before we take this step, let us manipulate and interpret the dual condition, to help understand it. As a useful preliminary step, let us temporarily assume that both T and Y are nite sets.

Claim 1. *Suppose that both T and Y are finite sets. All the surplus can be extracted from any given (v, x) assuming interim implementability if and only if p satisfies the following condition, called* convex dependence implies undetectability*: For any strategy π, if $p(t) = \sum_s \pi(s|t)p(s)$ for all t then $\sum_s \pi(t|s) = 1$ for all t.*

Proof. By the Alternative Theorem, for every (v, x) there exists ξ such that $v(t,t) = \sum_y \xi(t,y)p(y|t)$ for all t and $0 \leq \sum_y [\xi(s,y) - \xi(t,y)]p(y|t)$ for all (t,s) if and only if for every v and every pair (η, λ) with $\lambda \geq 0$, if $\eta(t)p(t) = \sum_s \lambda(s,t)p(s) - \lambda(t,s)p(t)$ for all t then $\sum_t \eta(t)v(t) \leq 0$. This latter condition is equivalent to the following: if $\eta(t)p(t) = \sum_s \lambda(s,t)p(s) - \lambda(t,s)p(t)$ for all t then $\eta \equiv 0$. Rearranging terms, the antecedent may be written as $[\eta(t) + \sum_s \lambda(t,s)]p(t) = \sum_s \lambda(s,t)p(s)$. Integrating out y, notice that $\eta(t) + \sum_s \lambda(t,s) = \sum_s \lambda(s,t)$ for every t. Without any loss of generality, we may assume that $\lambda(t,t) > 0$, and since $\lambda \geq 0$, it follows that $\eta(t) + \sum_s \lambda(t,s) = \sum_s \lambda(s,t) > 0$. By choosing $\lambda(t,t)$ appropriately, we may assume without loss that $\sum_s \lambda(s,t) =$ does not depend on t. Dividing both sides of the previous system of equations by , we nally obtain that if $p(t) = \sum_s \pi(s|t)p(s)$ then $\sum_s \pi(t|s) = 1$. □

To see how this condition works, consider the following example.

Example 8. Let $T = \{a, b, c\}$ and $Y = \{0, 1\}$. De ne $p(a) = p(b) =$ [0] and $p(c) =$ [1]. Here convex independence *does not* imply undetectability. To see this, consider the following strategy: $\pi(b|a) = \pi(b|b) = \pi(c|c) = 1$ and $\pi(s|t) = 0$ for all other (t,s). Clearly, $p(t) = \sum_s \pi(s|t)p(s)$ for all t, yet $\sum_s \pi(a|s) = 0$.

This example suggests that \convex dependence implies undetectability" is intimately related to convex independence. This intuition is correct, as the next result shows.

Claim 2. *The information structure p exhibits convex independence if and only if convex dependence implies undetectability.*

Proof. If convex independence fails then $p(\hat{t}) \in \text{conv}\{p(s) : s \neq \hat{t}\}$ for some type $\hat{t}$. Let $\hat{\pi}(t) = [t]$ if $t \neq \hat{t}$ and $\hat{\pi}(\hat{t})$ be any strategy that solves $p(\hat{t}) = \sum_s \pi(s|t)p(s)$. Now $\sum_s \hat{\pi}(\hat{t}|s) \neq 1$, so convex dependence does not imply undetectability. Conversely, assuming convex independence, if $p(t) = \sum_s \pi(s|t)p(s)$ for all t then $\pi(t) = [t]$ for all t, hence $p(t) = \sum_s \pi(t|s)p(s)$ for all t. Integrating out y, $\sum_s \pi(t|s) = 1$ for all t. □

Notice that Claim 2 holds regardless of the cardinality of T and Y. It follows from this last claim that when both T and Y are finite, all the surplus can be extracted assuming interim implementability if and only if p exhibits convex independence. By Cremer and McLean's result, it follows that convex independence characterizes both full surplus extraction and surplus extraction assuming interim implementability.

Let us now extend Claim 1 to the case where both T and Y may be infinite sets.

Lemma 14. *All the surplus can be extracted from any given* (v, x) *assuming interim implementability if and only if* p *exhibits asymptotic convex independence.*

Proof. By Lemma 13, all the surplus can be extracted from any given (v, x) assuming interim implementability if and only if condition (11) holds for all v. This is equivalent to requiring that for every net $\{(\lambda_\delta, \eta_\delta)\}$ with $\lambda_\delta \in \mathbb{R}_+^{(T\times T)}$ and $\eta_\delta \in \mathbb{R}^{(T)}$,

$$\eta_\delta(\cdot)p(\cdot) - \sum_{s\in T} \lambda_\delta(s,\cdot)p(s) - \lambda_\delta(\cdot,s)p(\cdot) \;\to\; 0 \quad\Rightarrow\quad \lim \eta_\delta \cdot v \;=\; 0 \quad \forall v.$$

In other words, $\eta_\delta \to 0$ weakly. Without loss, we may assume that $\lambda_\delta(t,t) \geq 1$ for all (t,δ) and that $\sum_s \lambda_\delta(s,t) = \;_\delta \geq 1$ for all (t,δ), so $\sum_s \lambda_\delta(s,t)$ does not depend on t. Integrating out y yields $\eta_\delta(\cdot) - \sum_s \lambda_\delta(s,\cdot) - \lambda_\delta(\cdot,s) \to 0$. Let $\pi_\delta(s|t) = \lambda_\delta(s,t)/\;_\delta$. Now, divide the antecedent above by $\;_\delta$ and rearrange to obtain the equivalent condition $[\eta_\delta(\cdot)/\;_\delta + \sum_s \pi_\delta(\cdot,s)]p(\cdot) - \sum_s \pi_\delta(s,\cdot)p(s) \to 0$. Since $\;_\delta$ is bounded below, $\eta_\delta(\cdot)/\;_\delta \to 0$, too. Hence, $\sum_s \pi_\delta(\cdot,s) \to 1$. Therefore, (11) above is equivalent to the following: $\sum_s \pi_\delta(s|\cdot)p(s) \to p(\cdot)$ implies that $\sum_s \pi_\delta(\cdot|s) \to 1$. □

Theorem 8 now follows from Lemma 14. We now present a result that explains the difference between asymptotic convex independence and condition $(*)$.

Lemma 15. *Asymptotic convex independence implies condition* $(*)$.

Proof. If condition $(*)$ fails then a net of strategies $\{\pi_\delta\}$ exists with $\sum_s \pi_\delta(s|t)p(s) \to p(t)$ for all t yet $\pi_\delta(\hat{t}) \not\to [\hat{t}]$ for some $\hat{t}$. Let $\hat{\pi}_\delta(t) = \pi_\delta(t)$ if $t = \hat{t}$ and $[t]$ otherwise. Now, $\sum_s \hat{\pi}_\delta(\hat{t}|s) = \hat{\pi}_\delta(\hat{t}|\hat{t}) \not\to 1$, so asymptotic convex independence fails. □

We now sketch a proof of Proposition 3. Consider the system of inequalities that describe surplus extraction assuming interim implementability. Given the restrictions of Proposition 3 and that ξ is continuous in t, by Lemma A.1 a feasible ξ exists for all v if and only if for every sequence $\{(\lambda_m, \eta_m)\}$ with $\lambda_m \in \mathbb{R}_+^{(T\times T)}$ and $\eta_m \in \mathbb{R}^{(T)}$,

$$\eta_m(\cdot)p(\cdot) - \sum_{s\in T} \lambda_m(s,\cdot)p(s) - \lambda_m(\cdot,s)p(\cdot) \;\to\; 0 \quad\Rightarrow\quad \lim \eta_m \cdot v \;=\; 0 \quad \forall v,$$

where now weak convergence in the above antecedent is with respect to all continuous functions on T by viewing $\eta_m(\cdot)p(\cdot) - \sum_s \lambda_m(s,\cdot)p(s) - \lambda_m(\cdot,s)p(\cdot)$ as a measure with nite support. Manipulating this implication as in the proof of Lemma 14, we obtain the equivalent condition that $\sum_s \pi_m(s|\cdot)p(s) \to p(\cdot)$ implies $\sum_s \pi_m(\cdot|s) \to 1$. Since T is a compact metric space, so is (T), and the nite measures are dense.

Consider an arbitrary function $\mu : T \to (T)$ such that $\int_T p(s)\mu(ds|t) = p(t)$. Let $\{\pi_m\}$ be any sequence of strategies such that $\pi_m(\cdot|t) \to \mu(\cdot|t)$ for all t. Therefore, $\sum_s \pi_m(s|t)p(s) - p(t) \to 0$ for all t. Since $\sum_s \pi_m(s|t)p(s) - p(t)$ has nite support as a function of t, it also converges in the weak$*$ topology, i.e., $\sum_t \xi(t)[\sum_s \pi_m(s|t)p(s) - p(t)]$ converges to the same limit for all continuous ξ, hence, this limit is zero.

By asymptotic convex independence, it follows that $\sum_s \pi_m(\cdot|s) \to 1$. But since pointwise convergence is implied by weak convergence, it follows that $\sum_s \pi_m(t|s) \to 1$ for all t, i.e., $\mu(\{t\}|t) = 1$. We have now established that condition (*) is implied by asymptotic convex independence, proving Proposition 3.

Finally, we turn to prove Theorem 9. We will just prove the rst statement here, as the same argument establishes the second one. First of all, one direction is immediate, since full surplus extraction implies virtually full surplus extraction. For the converse, recall that by de nition virtually all the surplus can be extracted from (v, x) if for every $\varepsilon > 0$ there is a scheme ξ such that

$$\begin{aligned} w(t,s) &\le \int_Y [\xi(s,y) - \xi(t,y)]p(dy|t) \qquad \forall (t,s), \quad \text{and} \\ 0 &\le \int_Y [v(t,\mathsf{x}(t,y),y) - \xi(t,y)]p(dy|t) \le \varepsilon \qquad \forall t. \end{aligned}$$

Our usual duality argument yields the following equivalence, whose proof is omitted.

Lemma 16. *Virtually all the surplus can be extracted from (v, x) if and only if for every $\varepsilon > 0$ and every net $\{\lambda_\delta\}$ such that $\lambda_\delta \ge 0$,*

$$\begin{aligned} [\lambda_\delta(1,\cdot) - \lambda_\delta(0,\cdot)]p(\cdot) + \sum_{s\in T} \lambda_\delta(s,\cdot)p(s) - \lambda_\delta(\cdot,s)p(\cdot) &\to 0 \\ \Rightarrow \quad \lim \lambda_\delta \cdot (w + W) - \varepsilon \sum_{(t,s)} \lambda_\delta(t,s) &\le 0, \end{aligned}$$

where $W(t,s) = v(s,s) - v(t,t)$.

By Lemma 16, virtual surplus extraction requires that for every net $\{\lambda_\delta\}$ satisfying the antecedent above and every $\varepsilon > 0$, $\lim \lambda_\delta \cdot (w + W) - \varepsilon \sum_{(t,s)} \lambda_\delta(t,s) \le 0$. But since $\varepsilon > 0$ is arbitrary, this implies that $\lim \lambda_\delta \cdot (w + W) \le 0$. However, this is precisely the requirement for full surplus extraction. This establishes Theorem 9.

F Proof of Theorem 13

Since the argument below is close to previous ones, we sketch the proof here. The *dual* of the principal's problem is given by the following linear program:

$$\inf_{\lambda\geq 0,\kappa} \kappa(T) \quad \text{s.t.} \quad p(\cdot)q(\cdot) = \int_T p(\cdot)\lambda(\cdot,ds) - p(s)\lambda(ds,\cdot) + p(\cdot)\lambda_0(\cdot),$$
$$\kappa(\cdot) \geq \int_Y u(\cdot,x,y)p(dy|\cdot)q(\cdot) + \int_T\int_Y v(\cdot,x,y)p(dy|\cdot)\lambda(\cdot,ds) - v(s,x,y)p(dy|s)\lambda(ds,\cdot)$$
$$+ \int_Y v(\cdot,x,y)p(dy|\cdot)\lambda_0(\cdot) \quad \forall x \in X,$$

where $\kappa, \lambda_0 \in M(T)$ and $\lambda \in M(T\times T)$.

Substituting the first constraint into the second yields the equivalent version below.

$$\inf_{\lambda\geq 0,\kappa} \kappa(T) \quad \text{s.t.} \quad p(\cdot)q(\cdot) = \int_T p(\cdot)\lambda(\cdot,ds) - p(s)\lambda(ds,\cdot) + p(\cdot)\lambda_0(\cdot),$$
$$\kappa(\cdot) \geq \int_Y [u(\cdot,x,y) + v(\cdot,x,y)]p(dy|\cdot)q(\cdot) + \int_T\int_Y [v(\cdot,x,y) - v(s,x,y)]p(dy|s)\lambda(ds,\cdot).$$

This is equivalent to the following saddle-point problem.

$$\inf_{\lambda\geq 0}\sup_{\mu\geq 0} \int [u(t,x,y) + v(t,x,y)]\mu(dx|t)p(dy|t)q(dt)$$
$$+ \int [v(t,x,y) - v(s,x,y)]\mu(dx|t)p(dy|s)\lambda(ds,dt) \quad \text{s.t.} \quad \mu(X|t) = 1 \ \forall t,$$
$$p(\cdot)q(\cdot) = \int_T p(\cdot)\lambda(\cdot,ds) - p(s)\lambda(ds,\cdot) + p(\cdot)\lambda_0(\cdot),$$

where $\mu \in B(T, M(X))$. By Io e and Tikhomirov (1968, Theorem 2.2, p. 83), it remains to show that the value function of the dual problem is lower semicontinuous at 0. Clearly, the dual is feasible since 0 is feasible, so the value of the dual is less than $+\infty$. If the dual is unbounded then by weak duality the primal is infeasible, so there is no duality gap and the theorem is proved. Assume now that the dual value is nite at 0. Let $(\alpha_+, \alpha_-) \in M(T\times Y)_+$ and $\beta \in B(X, M(T))_+$ be any perturbations to the dual right-hand side constraints. The perturbed problem looks like this:

$$\inf_{\lambda\geq 0,\kappa} \kappa(T) \quad \text{s.t.} \quad \alpha_-(\cdot) \leq \int_T p(\cdot)\lambda(\cdot,ds) - p(s)\lambda(ds,\cdot) + p(\cdot)\lambda_0(\cdot) - p(\cdot)q(\cdot) \leq \alpha_+(\cdot),$$
$$\kappa(\cdot) \geq \int [u(\cdot,x,y) + v(\cdot,x,y)]p(dy|\cdot)q(\cdot)$$
$$+ \int [v(\cdot,x,y) - v(s,x,y)]p(dy|s)\lambda(ds,\cdot) - \beta(\cdot|x).$$

Feasibility is maintained in the perturbed problem because the perturbations relax the constraints. For each $\varepsilon > 0$, consider a feasible solution (λ, κ) to the perturbed problem

such that $\kappa(T)$ is within ε of the value of this perturbed dual. Now, as the perturbations diminish, i.e., as $(\alpha_{\pm}, \beta) \to 0$, notice that

$$\kappa(T) \geq \sup_{\mu \geq 0} \int [u(t,x,y) + v(t,x,y)]\mu(dx|t)p(dy|t)q(dt)$$
$$+ \int [v(t,x,y) - v(s,x,y)]\mu(dx|t)p(dy|s)\lambda(ds,dt) - \int \mu(dx|t)\beta(dt|x) \geq$$
$$\sup_{\mu \geq 0} \int [u(t,x,y) + v(t,x,y)]\mu(dx|t)p(dy|t)q(dt)$$
$$+ \int [v(t,x,y) - v(s,x,y)]\mu(dx|t)p(dy|s)\hat{\lambda}(ds,dt) - 2\|v\|\,||\lambda - \hat{\lambda}|| - \|\beta\| \to$$
$$\sup_{\mu \geq 0} \int [u(t,x,y) + v(t,x,y)]\mu(dx|t)p(dy|t)q(dt)$$
$$+ \int [v(t,x,y) - v(s,x,y)]\mu(dx|t)p(dy|s)\hat{\lambda}(ds,dt),$$

where $V = \inf_{\eta} \|\eta - \lambda\|$ subject to η satisfying $p(\cdot)q(\cdot) = \int_T p(\cdot)\eta(\cdot, ds) - p(s)\eta(ds, \cdot) + p(\cdot)\lambda_0$ for some $\lambda_0 \geq 0$ and $\hat{\lambda}$ satis es $||\hat{\lambda} - \lambda|| - V < \|\alpha\|\,\varepsilon$. Intuitively, $\hat{\lambda}$ is arbitrarily close to a \projection" of λ onto the set of feasible solutions to the unperturbed equality constraints. Notice that for any λ, V is bounded above by $\|\alpha\|$, so the limit claimed above now follows. Therefore, the limit of value of the perturbed problems is greater than or equal to the value of the unperturbed problem. Hence, the value function is lower semicontinuous, as required.

References

Afriat, S. (1963): \The system of inequalities $a_{rs} > X_r - X_s$," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 59, 125. 23

| | | (1967): \The construction of utility functions from expenditure data," *International Economic Review*, 67{77. 4, 23, 25

Aliprantis, C. and K. Border (2006): *Infinite dimensional analysis: a hitchhiker's guide*, Springer Verlag. 21

Clark, S. A. (2006): \Necessary and Su cient Conditions for In nite-Dimensional Linear Inequalities," *Positivity*, 10, 475{489. 32

Conway, J. B. (1990): *A Course in Functional Analysis*, Graduate Texts in Mathematics, New York: Springer, second ed. 33

Cremer, J. and R. McLean (1988): \Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247{1257. 4, 17, 18, 19, 44

FOLLAND, G. (1999): *Real Analysis: Modern Techniques and their Applications*, John Wiley & Sons, second ed. 40

GOBERNA, M. AND M. LÓPEZ (1998): *Linear semi-infinite optimization*, John Wiley & Son Ltd. 31

GRETSKY, N., J. OSTROY, AND W. ZAME (2002): \Subdi erentiability and the Duality Gap," *Positivity*, 6, 1{16. 32

HEYDENREICH, B., R. MÜLLER, M. UETZ, AND R. VOHRA (2009): \Characterization of Revenue Equivalence," *Econometrica*, 77, 307{316. 3, 4, 10, 30

IOFFE, A. D. AND V. M. TIKHOMIROV (1968): \Duality of Convex Functions and Extremum Problems," *Russian Mathematical Surveys*, 23, 53{124. 32, 46

KOS, N. AND M. MESSNER (2009): \Incentive Compatible Transfers," Mimeo. 4

MCAFEE, R. AND P. RENY (1992a): \Correlated Information and Mecanism Design," *Econometrica*, 60, 395{421. 4, 17, 19, 20, 21, 22

| | | (1992b): \A Stone-Weierstrass theorem without closure under suprema," *Proceedings of the American Mathematical Society*, 114, 61{67. 17, 19

MCFADDEN, D. (2005): \Revealed stochastic preference: a synthesis," *Economic Theory*, 26, 245{264. 4, 23, 25

MCFADDEN, D. AND M. RICHTER (1990): \Stochastic rationality and revealed stochastic preference," *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz*, 161{186. 4, 24

MYERSON, R. (1981): \Optimal Auction Design," *Mathematics of Operations Research*, 6, 58{73. 29, 30

RICHTER, M. AND K. WONG (2005): \In nite inequality systems and cardinal revelations," *Economic Theory*, 26, 947{971. 23

ROCHET, J. C. (1987): \A Necessary and Su cient Condition for Rationalizability in a Quasi-Linear Context," *Journal of Mathematical Economics*, 16, 191{200. 3, 6, 10, 23, 31, 36

ROCKAFELLAR, R. T. (1970): *Convex Analysis*, Princeton, New Jersey: Princeton University Press. 6, 31, 33

SEGAL, I. AND M. D. WHINSTON (2009): \An Expected-E cient Status Quo Allows E cient Bargaining," *Theoretical Economics*, forthcoming. 28