



University of  
Pittsburgh

Department of Economics  
Dietrich School of Arts and Sciences

## WORKING PAPER SERIES

20/006

Detectability, Duality, and Surplus Extraction

Giuseppe Lopomo, Luca Rigotti and Chris Shannon

May, 2020

# Detectability, Duality, and Surplus Extraction\*

Giuseppe Lopomo  
Duke University

Luca Rigotti  
University of Pittsburgh

Chris Shannon  
UC Berkeley

this version: May 2020

## Abstract

We study surplus extraction in the general environment of McAfee and Reny (1992), and provide two alternative proofs of their main theorem. The first is an analogue of the classic argument of Crémer and McLean (1985, 1988), using geometric features of the set of agents' beliefs to construct a menu of contracts extracting the desired surplus. This argument, which requires a finite state space, also leads to a counterexample showing that full extraction is not possible without further significant conditions on agents' beliefs or surplus, even if the designer offers an infinite menu of contracts. The second argument uses duality and applies for an infinite state space, thus yielding the general result of McAfee and Reny (1992). Both arguments suggest methods for studying surplus extraction in settings beyond the standard model.

---

\*Financial support from the NSF under grant SES 1227707 is gratefully acknowledged.  
Contact: giuseppe.lopomo@duke.edu, luca@pitt.edu, cshannon@econ.berkeley.edu

# 1 Introduction

In most settings with asymmetric information, private information generates rents for the agents who hold it. This underlies many central results, such as the inefficiency of outcomes in many mechanisms. A series of important results by Crémer and McLean (1985, 1988) and McAfee and Reny (1992) proved that such rents can be fragile, however, and depend crucially on the assumption that agents' private information is independent. If instead agents' information is correlated, even only to an arbitrarily small degree, then appropriately designed mechanisms can typically leverage this correlation to extract all, or virtually all, information rents.

Crémer and McLean started this important strand of work. They considered the problem of surplus extraction in two particular settings, the monopolist screening problem (1985), and private values auctions (1988), each with agents whose private information is summarized by finitely many types. McAfee and Reny (1992) instead consider a very general environment, and allow for infinitely many types. McAfee and Reny (1992) show that a natural analogue of Crémer and McLean's convex independence condition on beliefs is necessary and sufficient for virtual surplus extraction in this class of environments, meaning that for every  $\varepsilon > 0$  a designer can offer a finite menu of contracts leaving each agent with no more than  $\varepsilon$  surplus.

As McAfee and Reny argue, considering the infinite type case is not merely a technical exercise in mathematical completeness, but is of central importance for understanding the explanatory power of the finite model and its ability to approximate the infinite case. Although McAfee and Reny's conditions are natural analogues of Crémer and McLean's for the infinite case, their result is in no sense a limit of Crémer and McLean's. In addition, they establish their result by a significantly different argument, from which a connection to the finite case is not clear. Crémer and McLean's result that full extraction holds in the finite case can be proven constructively using the separating hyperplane theorem, while McAfee and Reny's proof that virtual extraction holds in the infinite case relies on their elegant generalization of the classic Stone-Weierstrass approximation theorem. More importantly, from their work and subsequent work, it is unclear whether the gap between virtual and full extraction is due to the restriction to a finite menu of con-

tracts, or is instead an integral consequence of allowing for infinitely many types.

In this paper we provide two alternative proofs of the main theorem of McAfee and Reny (1992). Each provides insight into connections between the finite and infinite types models, including the extent to which the finite model approximates the infinite one, as well as suggesting novel techniques and results. The first is a natural analogue of the argument of Crémer and McLean, using geometric features of the set of agents' beliefs to explicitly construct a menu of contracts extracting the desired surplus. This argument requires a finite set of states over which agents are uncertain. By highlighting the connection to the geometry of the set of beliefs, this argument also leads to a counterexample showing that full extraction is not possible without further significant conditions on agents' beliefs or surplus, even if the designer offers an infinite menu of contracts.

The second argument is based on duality, using the characterization of extraction as the existence of a solution to a particular family of inequalities. This argument, while not constructive, applies for an infinite state space, and thus yields the general result of McAfee and Reny (1992). By providing a connection to duality, this argument suggests a method for studying surplus extraction in other models in which agents or the designer might have objectives other than risk-neutral value maximization.

Our use of duality in the surplus extraction problem is inspired by and builds on the work of Rahman (2012), which introduced duality arguments to study surplus extraction in arbitrary type spaces. Rahman (2012) argues that full surplus extraction holds under an analogue of convex independence in a setting with general type spaces allowing the designer to offer an infinite menu of contracts, and that this setting includes the environment of McAfee and Reny (1992) as a special case. We show by example that full extraction can fail under the assumptions of McAfee and Reny (1992), even allowing for an infinite menu of contracts. The two proofs we provide shed some light on why virtual extraction holds while full extraction can fail, and the extent to which duality arguments can be used to study surplus extraction in environments more general than the standard model. Other recent papers have also emphasized the importance of duality for different mechanism design questions. This includes work on multidimensional screening using

optimal transport methods by Daskalakis, Deckelbaum, and Tzamos (2017), and work on optimal auction design under robustness concerns by Carroll and Segal (2018) and Bergemann, Brooks, and Morris (2017a, b).

The paper proceeds as follows. In section 2 we give some preliminary definitions and results from convex analysis that will be used throughout the paper. In section 3 we set up the basic model and definitions, including notions of surplus extraction. In section 4 we consider the case in which the state space is finite. We give a necessary condition for full extraction, and use this to derive a counterexample showing that full extraction can fail under the assumptions of McAfee and Reny (1992), even if the designer can offer an infinite menu of contracts. We then give a constructive proof that virtual extraction holds in this setting. In section 5 we consider the case in which the state space can be infinite, and provide a proof of the general virtual extraction result based on duality. In section 6 we give an example illustrating how these methods can be used to study surplus extraction in models beyond the standard case. Additional results, some of which might be of independent interest, are collected in the appendix.

## 2 Preliminaries

We recall and collect here some preliminary definitions and results from convex analysis.

**Definition 1.** Let  $C \subseteq \mathbf{R}^n$  be a convex set. An *extreme point* of  $C$  is a point  $x \in C$  with the property that if  $x = \alpha y + (1 - \alpha)z$  for some  $y, z \in C$  and some  $\alpha \in [0, 1]$ , then  $x = y$  or  $x = z$ .

An *exposed point* of  $C$  is a point  $x \in C$  such that there is some real linear functional  $f$  on  $\mathbf{R}^n$  such that  $f(y) < f(x)$  for all  $y \in C$  with  $y \neq x$ .

**Note:** Every exposed point is an extreme point, but the converse does not hold. That is, extreme points need not be exposed. If  $C$  has only finitely many extreme points, however, then every extreme point of  $C$  is exposed.

**Definition 2.** A nonempty subset  $F \subseteq C$  of a convex set  $C$  is a *face* of  $C$  if whenever  $x, y \in C$  and  $\alpha x + (1 - \alpha)y \in F$  for some  $\alpha \in (0, 1)$ , then  $x, y \in F$ . A face  $F$  of  $C$  is a *proper face* if it is a proper subset of  $C$ .

**Definition 3.** A nonempty subset  $E \subseteq C$  of a convex set  $C \subseteq \mathbf{R}^n$  is an *exposed set* of  $C$  if there is a real linear functional  $f$  on  $\mathbf{R}^n$  such that  $f(y) \leq f(x)$  for all  $x, y \in C$  with  $x \in E$ , and  $f(y) < f(x)$  if  $y \notin E$ .

**Note:** An exposed set is a face, but a face need not be exposed.

**Definition 4.** Let  $C \subseteq \mathbf{R}^n$  be a convex set and let  $W \subseteq \mathbf{R}^n$  be the unique affine subspace of  $\mathbf{R}^n$  such that  $C \subseteq W$  and such that  $C$  has a nonempty relative interior in  $W$ . The *dimension* of  $C$ , denoted  $\dim C$ , is the dimension of  $W$ .

We record a useful result that connects these concepts next.

**Theorem 1.** *Let  $C \subseteq \mathbf{R}^n$  be a compact convex set, and  $F \subseteq C$  be a proper face of  $C$ . If  $\dim C = \ell$ , then  $\dim F < \ell$ .*

For example, see Simon (2011, Proposition 8.10).

Finally, note that if  $C \subseteq \mathbf{R}^n$  is a compact, convex set with  $\dim C = 1$ , then  $C$  has finitely many extreme points (at most 2), and thus all extreme points of  $C$  are exposed.

### 3 Set-up and Extraction Notions

In this section we lay out the basic set-up and notation used throughout the paper, and give the definitions of surplus extraction underlying the main results.

We use the following standard notation. For a compact metric space  $B$ ,  $C(B)$  is the space of continuous real-valued functions on  $B$ , and  $\mathcal{M}(B)$  is

the space of finite signed Borel measures on  $B$ . Similarly,  $\Delta(B)$  is the space of Borel probability measures on  $B$ .

For  $x \in C(B)$  and  $\eta \in \mathcal{M}(B)$ , we write  $x \cdot \eta = \eta \cdot x$  for the bilinear form  $\langle x, \eta \rangle = \langle \eta, x \rangle$ , that is,

$$\eta \cdot x = x \cdot \eta = \langle x, \eta \rangle = \langle \eta, x \rangle = \int x(b) \eta(db)$$

We follow McAfee and Reny (1992) in giving a reduced form description of the surplus extraction problem. In a prior, unmodeled stage, agents play a game that leaves them with some information rents as a function of their private information. Private information is summarized by the type  $t \in T$ , where  $T$  denotes the set of possible types. Unless specified otherwise, we let  $T = [0, 1]$  be the set of types (for all of the results it suffices that  $T$  is a compact, convex metric space). The current stage also has an exogenous source of uncertainty, summarized by a set of states  $S$ , on which contract payments can depend. For some applications, it is natural to take  $S = T$ , although we follow McAfee and Reny (1992) in allowing  $S$  to be arbitrary. We assume throughout that  $S$  is a compact metric space. To each type  $t \in T$  is then associated a value  $v(t) \in \mathbf{R}$ , representing the rents from the prior stage, and beliefs  $\pi(t) \in \Delta(S)$ . Throughout we maintain the assumption that  $v : T \rightarrow \mathbf{R}$  is continuous and that  $\pi : T \rightarrow \Delta(S)$  is norm continuous.

Let

$$C := \overline{\text{co}}\{\pi(t) \in \Delta(S) : t \in T\}$$

where for a subset  $A$  of a topological vector space  $X$ ,  $\text{co}(A)$  denotes the convex hull of  $A$ , and  $\overline{\text{co}}(A)$  denotes the closed convex hull of  $A$ .

Following our general notation, let  $\Delta(T)$  be the set of Borel probability measures on  $T$ . For  $t \in T$ ,  $\delta_t \in \Delta(T)$  denotes the Dirac measure concentrated on  $t$ . For  $x \in C(T \times S)$  and  $t \in T$ , we write  $x(t) \in C(S)$  for the function such that  $x(t)(s) = x(t, s)$  for each  $s \in S$ .

Next we give definitions for the main notions of surplus extraction in this setting, full extraction and virtual extraction. Both reflect the idea that the designer offers agents a menu of stochastic contracts from which they choose, based on minimizing their expected costs. Exploiting correlation between

types and beliefs might allow the designer to construct such a menu that leaves every agent with zero expected surplus, in the case of full extraction, or no more than  $\varepsilon$  expected surplus for any  $\varepsilon > 0$ , in the case of virtual extraction.

**Definition 5.** *Full extraction* holds if, for each given  $v : T \rightarrow \mathbf{R}$ , there exists a collection  $\{c(t) \in C(S) : t \in T\}$  such that for each  $t \in T$ :

$$v(t) - \pi(t) \cdot c(t) = 0$$

and

$$v(t) - \pi(t) \cdot c(s) \leq 0 \quad \forall s \neq t$$

*Virtual extraction* holds if, for each given  $v : T \rightarrow \mathbf{R}$ , for each  $\varepsilon > 0$  there exists a collection  $\{c_\varepsilon(t) \in C(S) : t \in T\}$  such that for each  $t \in T$ :

$$0 \leq v(t) - \pi(t) \cdot c_\varepsilon(t) \leq \varepsilon$$

and

$$v(t) - \pi(t) \cdot c_\varepsilon(s) \leq \varepsilon \quad \forall s \neq t$$

As defined, full extraction or virtual extraction might require the designer to offer an infinite menu of contracts when  $T$  is infinite. In the case of virtual extraction, such a menu need not have an expected cost minimizing element for all agents. By allowing for an infinite menu of contracts, this might also appear to be a weaker notion of virtual extraction than considered by McAfee and Reny (1992), which instead shows that for each  $\varepsilon > 0$ , there is a finite menu  $\{c_1, \dots, c_n\}$  such that for each  $t \in T$ ,

$$0 \leq \max_{j=1, \dots, n} \{v(t) - \pi(t) \cdot c_j\} \leq \varepsilon$$

We note, however, that whenever virtual extraction holds (using the definition above), then it is always possible to find a finite menu of contracts that would achieve the same bounds on surplus. We record this observation and its proof below.

**Theorem 2.** *If virtual extraction holds, then virtual extraction can be achieved with a finite menu of contracts. That is, given  $v : T \rightarrow \mathbf{R}$ , for each  $\varepsilon > 0$  there exists a finite menu  $\{c_1, \dots, c_n\} \subseteq C(S)$  such that for each  $t \in T$ ,*

$$0 \leq \max_{j=1, \dots, n} \{v(t) - \pi(t) \cdot c_j\} \leq \varepsilon$$



*Proof.* Let  $v : T \rightarrow \mathbf{R}$  be given and fix  $\varepsilon > 0$ . Choose  $\{c_\varepsilon(t) \in C(S) : t \in T\}$  such that for each  $t \in T$ :

$$0 \leq v(t) - \pi(t) \cdot c_\varepsilon(t) \leq \varepsilon$$

and

$$0 \leq \sup_{s \in T} \{v(t) - \pi(t) \cdot c_\varepsilon(s)\} \leq \varepsilon$$

Then for each  $t \in T$  there exists  $\delta_t > 0$  such that  $s \in B_{\delta_t}(t) \Rightarrow$

$$-\frac{\varepsilon}{2} \leq v(s) - \pi(s) \cdot c_\varepsilon(t) \leq \varepsilon$$

where  $B_{\delta_t}(t) = \{s \in T : \|s - t\| < \delta_t\}$ . Since  $T$  is compact and  $\{B_{\delta_t}(t) : t \in T\}$  is an open cover of  $T$ , there exists  $t_1, \dots, t_n$  such that  $T \subseteq \cup_i B_{\delta_{t_i}}(t_i)$ . Then for each  $i$ , set  $c_i = c_\varepsilon(t_i) + \varepsilon$ . For each  $t \in T$  there exists  $i$  such that  $t \in B_{\delta_{t_i}}(t_i)$ . Thus

$$0 \leq v(t) - \pi(t) \cdot c_i \leq 2\varepsilon$$

and for each  $j = 1, \dots, n$ ,

$$v(t) - \pi(t) \cdot c_j \leq 2\varepsilon$$

Thus

$$0 \leq \max_{j=1, \dots, n} \{v(t) - \pi(t) \cdot c_j\} \leq 2\varepsilon$$

The result follows. □

As with the original formulation of McAfee and Reny (1992), these notions of full and virtual extraction do not explicitly address incentive compatibility. When full extraction holds, incentive compatibility will follow. Incentive compatibility does not immediately follow from the definition of virtual extraction, however; this would require that the menu of contracts satisfies the additional incentive constraints  $v(t) - \pi(t) \cdot c(s) \leq v(t) - \pi(t) \cdot c(t)$  for all  $s, t \in T$ . But if virtual extraction holds, then there is always some menu achieving virtual extraction that is also incentive compatible, by virtue of Theorem 2. If virtual extraction holds, then it can be achieved with some finite menu of contracts  $\{c_1, \dots, c_n\}$ . Choosing  $c(t) \in \arg \max\{v(t) - \pi(t) \cdot c_i : i = 1, \dots, n\}$  for each  $t \in T$  then yields a menu that achieves virtual extraction and is also

incentive compatible. More generally, an infinite menu of contracts might allow for virtual extraction while also satisfying incentive compatibility, but need not. Our results do not address this question directly, although they suggest that probabilistic independence is not sufficient to guarantee incentive compatibility for every infinite menu.

Following Crémer and McLean (1988) and McAfee and Reny (1992), we consider conditions on beliefs under which full extraction or virtual extraction holds. McAfee and Reny (1992) show that virtual extraction is possible whenever beliefs satisfy the following condition.

**Definition 6.** Types satisfy *probabilistic independence* if for all  $t \in T$ :

$$\pi(t) = \int \pi(s)\mu(ds) \text{ for some } \mu \in \Delta(T) \Rightarrow \mu = \delta_t$$

**Note:** If types satisfy probabilistic independence, then  $\pi(t)$  is an extreme point of  $C$  for each  $t \in T$ .

**Note:** If  $T$  is finite, then probabilistic independence reduces to the standard *convex independence* condition of Crémer-McLean, that is

$$\pi(t) = \sum_{s \in T} \mu_s \pi(s) \text{ for some } \mu \in \Delta(T) \Rightarrow \mu_t = 1$$

Crémer and McLean (1988) show that when  $T$  and  $S$  are finite, then full extraction holds whenever beliefs satisfy convex independence. This is no longer true when  $T$  is infinite, as the example in the next section illustrates. We sketch a standard argument for the classic result when  $T$  and  $S$  are both finite next, to motivate the main ideas we develop in the following sections.

Suppose  $T$  and  $S$  are finite, and types satisfy convex independence. Fix a type  $t \in T$ . Since types satisfy convex independence,  $\pi(t) \notin \overline{\text{co}}\{\pi(s) : s \in T, s \neq t\}$ . Thus there exists  $z(t) \in \mathbf{R}^S$  such that

$$\pi(t) \cdot z(t) = 0$$

and

$$\pi(s) \cdot z(t) > 0 \quad \forall s \in T, s \neq t$$

That is,  $\pi(t)$  is an exposed point of  $\{\pi(s) : s \in T\}$ , and of  $C = \overline{\text{co}}\{\pi(s) : s \in T\}$ . Alternatively, by convex independence,  $\{\pi(s) : s \in T\}$  is the set of extreme points of  $C$ ; since  $T$  is finite this set is finite, so each element  $\pi(t)$  must also be an exposed point of  $C$ . Now consider a contract of the form  $c(t) = v(t) + \alpha(t)z(t)$  where  $\alpha(t) \in \mathbf{R}_+$ , which requires the constant payment  $v(t)$  and a stochastic payment that is a scaled version of  $z(t)$  (throughout we use a constant  $r \in \mathbf{R}$  interchangeably with the function  $r\mathbf{1}(S)$ , where  $\mathbf{1}(S)$  denotes the identity on  $S$ ). For type  $t$ , this contract has expected cost  $v(t)$ , as

$$\pi(t) \cdot c(t) = v(t) + \alpha(t)(\pi(t) \cdot z(t)) = v(t)$$

while for types  $s \neq t$ , the expected cost is

$$\pi(s) \cdot c(t) = v(t) + \alpha(t)(\pi(s) \cdot z(t))$$

Since  $\pi(s) \cdot z(t) > 0$  for all types  $s \neq t$ , the designer can take advantage of this difference in beliefs to set  $\alpha(t)$  sufficiently large to make the resulting contract unattractive to all types  $s \neq t$  while keeping the expected cost constant for type  $t$ . To that end, set  $\alpha(t) > 0$  sufficiently large so that

$$\alpha(t) > \max_{s \neq t} \frac{v(s) - v(t)}{\pi(s) \cdot z(t)}$$

Note that since  $\pi(s) \cdot z(t) > 0$  for all  $s \neq t$  and  $T$  is finite, the term on the right above is well-defined (the ratio is bounded), and thus  $\alpha(t)$  is well-defined. Then for type  $t$ ,

$$v(t) - \pi(t) \cdot c(t) = v(t) - v(t) = 0$$

while for types  $s \neq t$ ,

$$v(s) - \pi(s) \cdot c(t) = v(s) - v(t) - \alpha(t)(\pi(s) \cdot z(t)) < 0$$

by choice of  $\alpha(t)$ . Repeating this construction for each type  $t \in T$  yields a menu  $\{c(t) : t \in T\}$  that achieves full extraction. Finally, if in addition  $|S| \geq |T|$ , then convex independence is satisfied for almost all elements of  $\Delta(S)^T$ .

## 4 Finite State Space

In this section, we consider the case in which the state space  $S$  is finite. With abuse of notation, we use the symbol  $S$  interchangeably for the state space and its cardinality throughout this section.

Focusing on the simpler case in which  $S$  is finite leads to several important observations. First, we can connect the extraction problem to the underlying convex geometry, as in the setting with finitely many types. This lets us establish a more direct connection between the extraction problem with finitely many types and with infinitely many types. From this connection we identify a simple necessary condition for full extraction, based on the geometry of the set of beliefs  $\{\pi(t) : t \in T\}$ . This necessary condition sheds light on why probabilistic independence is no longer sufficient for full extraction with infinitely many types, and allows us to give an example to illustrate this breakdown. The example also sheds light on why virtual extraction holds nonetheless, and on the nature of types to whom some surplus might need to be left. Finally, these observations lead to a constructive proof that virtual extraction holds under probabilistic independence, analogous to classic arguments in the finite type case, in which we will explicitly construct a menu of contracts to achieve extraction of all but at most  $\varepsilon$  surplus for each  $\varepsilon > 0$ .

We start with the observation that if full extraction holds, then for each  $t \in T$ ,  $\pi(t)$  must be an exposed point of  $\{\pi(s) : s \in T\}$  and of  $C = \overline{\text{co}}\{\pi(s) : s \in T\}$ .

**Theorem 3.** *Let  $S$  be finite. If full extraction holds, then for each  $t \in T$ ,  $\pi(t)$  is an exposed point of  $\{\pi(s) : s \in T\}$  and of  $C = \overline{\text{co}}\{\pi(s) : s \in T\}$ .*

*Proof.* Fix  $t \in T$ . Choose  $v : T \rightarrow \mathbf{R}$  such that  $v(s) > v(t)$  for all  $s \neq t$ . Since full extraction holds, there exists a collection  $\{c(s) \in \mathbf{R}^S : s \in T\}$  such that

$$v(t) - \pi(t) \cdot c(t) = 0$$

and

$$v(s) - \pi(s) \cdot c(t) \leq 0 \quad \forall s \neq t$$

Write  $c(t) = v(t) + z(t)$ , where  $z(t) \in \mathbf{R}^S$  and with abuse of notation  $v(t) =$

$v(t)\mathbf{1}_S$ . Since  $\pi(t) \cdot c(t) = v(t)$ ,

$$\pi(t) \cdot z(t) = 0$$

Then for  $s \neq t$ ,

$$\pi(s) \cdot c(t) = v(t) + \pi(s) \cdot z(t) \geq v(s)$$

Thus

$$\pi(s) \cdot z(t) \geq v(s) - v(t) > 0 \quad \forall s \neq t$$

That is,  $\pi(t)$  is an exposed point of  $\{\pi(s) : s \in T\}$ .

Now take  $\pi \in C = \overline{\text{co}}\{\pi(s) : s \in T\} \subseteq \mathbf{R}^S$  with  $\pi \neq \pi(t)$ . Since  $\{\pi(s) : s \in T\} \subseteq \mathbf{R}^S$  is compact, its convex hull is closed, thus  $C = \text{co}\{\pi(s) : s \in T\}$ . Then by definition, there exists  $\{t_1, \dots, t_n\} \subseteq T$  and  $\alpha_1, \dots, \alpha_n > 0$  with  $\sum_i \alpha_i = 1$  such that  $\pi = \sum_i \alpha_i \pi(t_i)$ . Since  $\pi \neq \pi(t)$ , there exists  $i$  such that  $t_i \neq t$ . Then

$$\pi \cdot z(t) = \sum_i \alpha_i \pi(t_i) \cdot z(t) > 0$$

since  $\exists i$  such that  $t_i \neq t$  and  $\pi(t_i) \cdot z(t) > 0$  for all such  $t_i$ .

Thus  $\pi(t)$  is an exposed point of  $C$ . Since  $t \in T$  was arbitrary,  $\pi(t)$  is an exposed point of  $\{\pi(s) : s \in T\}$ , and of  $C = \overline{\text{co}}\{\pi(s) : s \in T\}$  for each  $t \in T$ .  $\square$

This observation motivates the following definition, reframing the necessary condition above in terms of agents' beliefs.

**Definition 7.** Type  $t \in T$  is *detectable* if  $\exists z \in \mathbf{R}^S$  such that

$$\begin{aligned} \pi(t) \cdot z &= 0 \\ \pi(s) \cdot z &> 0 \quad \forall s \neq t \end{aligned}$$

**Note:** A type  $t$  is detectable if and only if  $\pi(t)$  is an exposed point of  $C$ . Using this terminology, we can then restate Theorem 3 as follows: when  $S$  is finite, full extraction requires that all types are detectable.

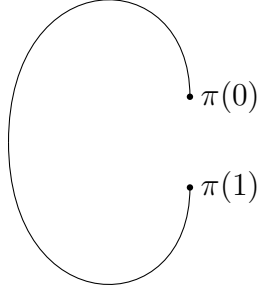


Figure 1: Graph of  $\{\pi(t) : t \in T\}$

Next we give an example to illustrate the failure of full extraction when  $T$  is infinite. The example illustrates why full extraction can fail and also suggests why virtual extraction can hold despite the failure of full extraction. In the example, probabilistic independence is satisfied, but some types are not detectable.

**Example 1:** Let  $\pi : T \rightarrow \Delta(S)$  be as in Figure 1, and suppose  $v(t) > v(0)$  for all  $t \neq 0$ . Note that for every  $t \neq 0, 1$ ,  $\pi(t)$  is an exposed point of  $C = \overline{\text{co}}\{\pi(s) : s \in T\}$ , while  $\pi(0)$  and  $\pi(1)$  are extreme points of  $C$  that are not exposed. That is,  $t$  is detectable in  $T$  for all  $t \neq 0, 1$ , while types  $t = 0$  and  $t = 1$  are not detectable in  $T$ . See Figures 2 and 3.

To see that these beliefs satisfy probabilistic independence, suppose

$$\pi(t) = \int \pi(s) \mu(ds) \quad \text{for some } t \in T \text{ and } \mu \in \Delta(T)$$

First suppose  $t \notin \{0, 1\}$ . Then since  $\pi(t)$  is an exposed point of  $C$ , there exists  $z(t) \in \mathbf{R}^S$  such that  $\pi(t) \cdot z(t) = 0$  and  $\pi(s) \cdot z(t) > 0 \quad \forall s \neq t$ . Then

$$0 = \pi(t) \cdot z(t) = \int \pi(s) \cdot z(t) \mu(ds)$$

Since  $\pi(s) \cdot z(t) > 0$  for all  $s \neq t$  and  $\pi(t) \cdot z(t) = 0$ , this implies  $\mu = \delta_t$ .

Now suppose  $t \in \{0, 1\}$ . In this case, there exists  $z(t) \in \mathbf{R}^S$  such that  $\pi(t) \cdot z(t) = \pi(0) \cdot z(t) = \pi(1) \cdot z(t) = 0$  and  $\pi(s) \cdot z(t) > 0 \quad \forall s \neq 0, 1$ . Then

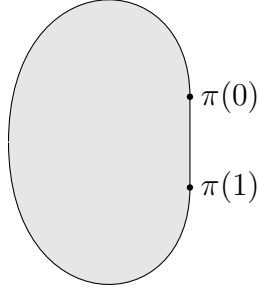


Figure 2: Graph of  $C = \overline{\text{co}}\{\pi(t) : t \in T\}$ ;  $\pi(0)$  and  $\pi(1)$  are extreme points of  $C$  that are not exposed.

as above,

$$0 = \pi(t) \cdot z(t) = \int \pi(s) \cdot z(t) \mu(ds)$$

Since  $\pi(s) \cdot z(t) > 0$  for all  $s \neq 0, 1$  and  $\pi(0) \cdot z(t) = \pi(1) \cdot z(t) = 0$ , this implies  $\text{supp } \mu \subseteq \{0, 1\}$ . Thus  $\pi(t) = \alpha\pi(0) + (1-\alpha)\pi(1)$  for some  $\alpha \in [0, 1]$ . But since both  $\pi(0)$  and  $\pi(1)$  are extreme points of  $C$ , this implies  $\mu = \delta_t$ .

Thus beliefs satisfy probabilistic independence. Full extraction does not hold, however, by Theorem 3, since  $\pi(0)$  and  $\pi(1)$  are not exposed points in  $C$ , that is, types  $t = 0$  and  $t = 1$  are not detectable in  $T$ .

In this case, it is not difficult to see directly why full extraction fails. As above, suppose  $v(t) > v(0)$  for all  $t \neq 0$ . Notice in particular this implies  $v(1) > v(0)$ , so the surplus of type 1 is greater than the surplus of type 0. Now if full extraction were possible, there must exist a contract  $c(0) = v(0) + z(0)$  for some  $z(0) \in \mathbf{R}^S$  such that

$$v(0) = \pi(0) \cdot c(0) = v(0) + \pi(0) \cdot z(0) \iff \pi(0) \cdot z(0) = 0$$

and

$$v(s) \leq \pi(s) \cdot c(0) = v(0) + \pi(s) \cdot z(0) \iff \pi(s) \cdot z(0) \geq v(s) - v(0) > 0 \quad \forall s \neq 0$$

Here, however, if  $\pi(0) \cdot z = 0$  and  $\pi(s) \cdot z \geq 0$  for all  $s \in T$ , it must be the case that  $\pi(1) \cdot z = 0$  as well (see Figure 3). But then the contract  $c(0)$

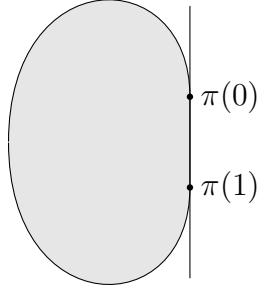


Figure 3: Full extraction is impossible when  $v(t) > v(0)$  for all  $t \neq 0$ .

must leave type 1 with strictly positive surplus, as

$$v(1) - \pi(1) \cdot c(0) = v(1) - v(0) - \pi(1) \cdot z(0) = v(1) - v(0) > 0$$

So any contract that extracts full surplus from type 0 and does not provide surplus to other types  $t \in (0, 1)$  must leave strictly positive surplus for type 1.  $\diamond$

Although full extraction is not possible in this example, virtual extraction is. Figures 4 and 5 illustrate the idea, and also illustrate that it might be necessary to give small rents to some types close to a type that is not detectable in order to extract surplus from other types from which that type cannot be distinguished. We sketch the argument here, which then serves as the template for the general constructive proof we give below.

First, consider the case of  $t \notin \{0, 1\}$ , so  $t$  is detectable in  $T$  and  $\pi(t)$  is an exposed point of  $C = \overline{\text{co}}\{\pi(s) : s \in T\}$ . Then there exists  $z(t) \in \mathbf{R}^S$  such that

$$\pi(t) \cdot z(t) = 0$$

and

$$\pi(s) \cdot z(t) > 0 \quad \forall s \neq t$$

See Figure 4. Following the idea of the proof for the finite types case, consider a contract of the form  $c(t) = v(t) + \alpha(t)z(t)$ . For type  $t$ , this contract has expected cost  $v(t)$ :

$$\pi(t) \cdot c(t) = v(t) + \alpha(t)(\pi(t) \cdot z(t)) = v(t)$$



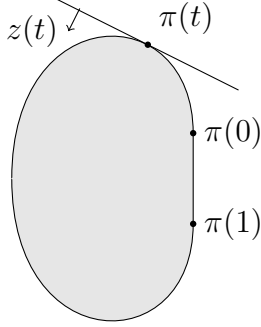


Figure 4: Construction of  $c(t)$  when  $\pi(t)$  is an exposed point of  $C$ .

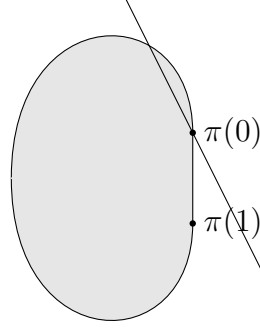


Figure 5: Leaving some surplus for types close to  $t = 0$  might be necessary to extract surplus from type  $t = 1$ .

while for types  $s \neq t$ , the expected cost is

$$\pi(s) \cdot c(t) = v(t) + \alpha(t)(\pi(s) \cdot z(t))$$

As in the finite case, to make this contract unattractive to types  $s \neq t$ , we would like to set  $\alpha(t) > 0$  sufficiently large so that

$$\alpha(t) > \sup_{s \neq t} \frac{v(s) - v(t)}{\pi(s) \cdot z(t)}$$

The term on the right need not be finite with infinitely many types, however (indeed, both the denominator and the numerator go to 0 as  $s \rightarrow t$ ), so this scaling term need not be defined. For types sufficiently close to  $t$ , however, the difference in values  $v(s) - v(t)$  is small, so the surplus such types could gain by choosing the contract  $c(t)$  is small, as

$$v(s) - \pi(s) \cdot c(t) = v(s) - v(t) - \alpha(t)(\pi(s) \cdot z(t)) < v(s) - v(t)$$

Given  $\varepsilon > 0$ , we can find  $\delta > 0$  such that  $\|s - t\| < \delta \Rightarrow v(s) - v(t) < \varepsilon$ . Since the set of types in  $T$  with  $\|s - t\| \geq \delta$  is compact, we can now choose  $\alpha(t) > 0$  sufficiently large so that

$$\alpha(t) > \max_{\|s-t\| \geq \delta} \frac{v(s) - v(t)}{\pi(s) \cdot z(t)}$$

For the resulting contract  $c(t) = v(t) + \alpha(t)z(t)$ ,

$$v(t) - \pi(t) \cdot c(t) = v(t) - v(t) = 0$$

and

$$v(s) - \pi(s) \cdot c(t) \leq \varepsilon \quad \forall s \neq t$$

Now consider  $t \in \{0, 1\}$ , and without loss of generality take  $t = 0$ . First note that, as depicted in Figure 3, there exists  $z(0) \in \mathbf{R}^S$  such that

$$\pi(0) \cdot z(0) = \pi(1) \cdot z(0) = 0$$

and

$$\pi(t) \cdot z(0) > 0 \quad \forall t \notin \{0, 1\}$$

Then consider  $T_1 = \{0, 1\}$  and

$$C_1 = \{\alpha\pi(0) + (1 - \alpha)\pi(1) : \alpha \in [0, 1]\} = \text{co}\{\pi(0), \pi(1)\}$$

In  $C_1$ ,  $\pi(0)$  is an exposed point, thus there exists  $z_1(0) \in \mathbf{R}^S$  such that

$$\pi(0) \cdot z_1(0) = 0 \quad \text{and} \quad \pi \cdot z_1(0) > 0 \quad \forall \pi \in C_1 \setminus \{\pi(0)\}$$

In particular,  $\pi(1) \cdot z_1(0) > 0$ . See Figures 6 and 7. Mimicking the construction in the previous case for a detectable type, choose  $\alpha_1(0) > 0$  sufficiently large so that

$$\alpha_1(0) > \frac{v(1) - v(0)}{\pi(1) \cdot z_1(0)}$$

Then set  $c_1(0) = v(0) + \alpha_1(0)z_1(0)$ .

By construction,

$$v(0) - \pi(0) \cdot c_1(0) = v(0) - v(0) - \alpha_1(0)(\pi(0) \cdot z_1(0)) = 0$$

and

$$v(1) - \pi(1) \cdot c_1(0) = v(1) - v(0) - \alpha_1(0)(\pi(1) \cdot z_1(0)) < 0$$

The contract  $c_1(0)$  might leave surplus to some types  $t \notin \{0, 1\}$ , however. To take care of this, we can use the additional stochastic payment  $z(0)$ , which type 0 and type 1 believe has expected value of zero. Since all other types believe  $z(0)$  has positive expected value, we can appropriately scale  $z(0)$  to

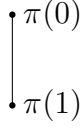


Figure 6: Graph of  $C_1 = \{\alpha\pi(0) + (1 - \alpha)\pi(1) : \alpha \in [0, 1]\}$ .

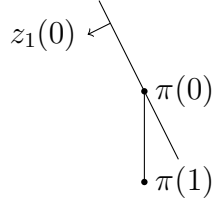


Figure 7: Construction of  $c_1(0)$ .

ensure that no more than  $\varepsilon$  expected surplus is left for all such types, as follows.

Note that the surplus  $c_1(0)$  leaves to types arbitrarily close to 0 must be arbitrarily small, since  $c_1(0)$  leaves zero surplus for type 0. Similarly, the surplus  $c_1(0)$  leaves to types arbitrarily close to 1 must be negative, since  $c_1(0)$  leaves negative surplus for type 1. Then choose  $\delta > 0$  such that for  $\|s - 0\| < \delta$  or  $\|s - 1\| < \delta$ ,

$$v(s) - \pi(s) \cdot c_1(0) = v(s) - v(0) - \alpha_1(0)(\pi(s) \cdot z_1(0)) < \varepsilon$$

This is possible, by the continuity of  $v$  and  $\pi$ , and the fact that  $v(0) - \pi(0) \cdot c_1(0) = 0$  and  $v(1) - \pi(1) \cdot c_1(0) < 0$ . Then set  $\alpha(0) > 0$  such that

$$\alpha(0) > \max_{\substack{\|s-0\| \geq \delta \\ \|s-1\| \geq \delta}} \frac{v(s) - \pi(s) \cdot c_1(0)}{\pi(s) \cdot z(0)}$$

Then set

$$c(0) = c_1(0) + \alpha(0)z(0) = v(0) + \alpha_1(0)z_1(0) + \alpha(0)z(0)$$

By construction,  $\pi(0) \cdot c(0) = \pi(0) \cdot c_1(0)$  and  $\pi(1) \cdot c(0) = \pi(1) \cdot c_1(0)$ . Thus  $v(0) - \pi(0) \cdot c(0) = 0$  and  $v(1) - \pi(1) \cdot c(0) < 0$ .

For  $t \notin \{0, 1\}$ , again by construction,

$$v(t) - \pi(t) \cdot c(0) = v(t) - \pi(t) \cdot c_1(0) - \alpha(0)(\pi(t) \cdot z(0)) < \varepsilon$$

The collection  $\{c(t) : t \in T\}$  thus constructed achieves extraction of all but at most  $\varepsilon$  surplus.

Next we will show that an analogous construction works in general whenever probabilistic independence is satisfied to yield a collection of contracts that achieves virtual extraction. Three key aspects of the construction in the example lead to the general result. First, for any detectable type there is always a contract that leaves that type with zero expected surplus while leaving at most  $\varepsilon$  expected surplus for all other types. Second, although some types might not be detectable, such as types 0 and 1 in the example, probabilistic independence implies that every type is eventually detectable within some (non-singleton) subset of types, in a sense we make precise below. Finally, this weaker property is sufficient to construct a suitable analogous contract for a type that is not detectable. We formalize this next, using additional results about convex sets and their extreme points that help illuminate the connection between the extraction problems with finitely and infinitely many types.

We start with a weaker notion of exposed point for a convex set, motivated by the above construction of contracts for the case when types are not detectable.

**Definition 8.** Let  $C$  be a convex set. A point  $x \in C$  is *eventually exposed* if there exists a sequence  $\{F_0, F_1, \dots, F_n\}$  of subsets of  $C$  such that

- (i)  $F_0 = C, F_n = \{x\}$
- (ii)  $\dim F_i \geq 1$  for  $i = 0, 1, \dots, n-1$
- (iii) for each  $i, F_{i+1} \subseteq F_i$  and  $F_{i+1}$  is an exposed face of  $F_i$

**Note:** If  $C$  is a convex set and  $x \in C$  is an exposed point of  $C$ , then  $x$  is also eventually exposed, using the trivial sequence  $F_0 = C$  and  $F_1 = \{x\}$ .

**Note:** If  $C$  is compact and convex and  $x \in C$  is eventually exposed, then each set  $F_i$  is compact and convex. In addition,  $x$  is an exposed point in the set  $F_{n-1}$ .

Next, we show that while a convex set can have extreme points that are not exposed, every extreme point in a compact, convex subset of  $\mathbf{R}^S$  is eventually exposed.

**Theorem 4.** *Let  $C$  be a compact, convex subset of  $\mathbf{R}^S$ . If  $x \in C$  is an extreme point then it must be eventually exposed.*

*Proof.* Let  $x \in C$  be an extreme point. If  $x$  is an exposed point of  $C$ , then it is also eventually exposed. So suppose  $x$  is not exposed. Since  $x$  is an extreme point of  $C$ , there exists  $z_1 \in \mathbf{R}^S$  and  $b_1 \in \mathbf{R}$  such that

$$\begin{aligned} x \cdot z_1 &= b_1 \\ y \cdot z_1 &\geq b_1 \quad \forall y \in C \\ y \cdot z_1 &> b_1 \quad \text{for some } y \in C \end{aligned}$$

Set

$$F_1 = \{y \in C : y \cdot z_1 = b_1\}$$

Since  $x$  is not an exposed point of  $C$ ,  $F_1$  is a proper face of  $C$ , so  $\dim F_1 < \dim C$ , and  $\dim F_1 \geq 1$ .

If  $x$  is an exposed point of  $F_1$ , we are done, setting  $F_2 = \{x\}$ . Else,  $x$  must be an extreme point of  $F_1$  (since  $F_1 \subseteq C$ ) that is not exposed. Repeating the above argument, choose  $z_2 \in \mathbf{R}^S$  and  $b_2 \in \mathbf{R}$  such that

$$\begin{aligned} x \cdot z_2 &= b_2 \\ y \cdot z_2 &\geq b_2 \quad \forall y \in F_1 \\ y \cdot z_2 &> b_2 \quad \text{for some } y \in F_1 \end{aligned}$$

Set

$$F_2 = \{y \in F_1 : y \cdot z_2 = b_2\}$$

If  $x$  is an exposed point of  $F_2$ , we are done. Else,  $F_2$  is a proper face of  $F_1$  and

$$1 \leq \dim F_2 < \dim F_1 \leq \dim C - 1$$

Repeating this argument, since  $\dim F_i \leq \dim F_{i-1} - 1$  for each  $i$ , eventually must have  $\dim F_n = 1$ , and then because  $x$  is an extreme point of  $F_n$ ,  $x$  must also be an exposed point of  $F_n$ .  $\square$

Returning to the extraction problem, these definitions and results suggest analogous conditions connecting types to beliefs, and the geometry of the set of beliefs, as with the connection between exposed points and detectable types. First, we extend the notion of a detectable type to a set of types in the natural way.

**Definition 9.** A set of types  $T^* \subseteq T$  is *detectable* in  $T$  if there exists  $z \in \mathbf{R}^S$  such that

$$\begin{aligned}\pi(t) \cdot z &= 0 \quad \forall t \in T^* \\ \pi(t) \cdot z &> 0 \quad \forall t \in T \setminus T^*\end{aligned}$$

Next, in analogy with the connection between exposed points and eventually exposed points, there is a natural notion of eventually detectable types.

**Definition 10.** A type  $t^* \in T$  is *eventually detectable* if there is a nested sequence of compact sets  $\{T_0, T_1, \dots, T_n\}$  with  $T_0 = T$ ,  $T_n = \{t^*\}$ , and  $T_{i+1} \subseteq T_i$  for each  $i$ , and a corresponding sequence  $\{z_1, \dots, z_n\} \subseteq \mathbf{R}^S$  such that for each  $i = 1, \dots, n$ ,

$$\begin{aligned}\pi(t) \cdot z_i &= 0 \quad \forall t \in T_i \\ \pi(t) \cdot z_i &> 0 \quad \forall t \in T_{i-1} \setminus T_i\end{aligned}$$

**Note:** A type  $t^* \in T$  is eventually detectable if there is a nested sequence of compact sets  $\{T_0, T_1, \dots, T_n\}$ , with  $T_0 = T$ ,  $T_n = \{t^*\}$ , and  $T_{i+1} \subseteq T_i$  for each  $i$  such that  $T_{i+1}$  is detectable in  $T_i$  for each  $i$ .

**Note:** If  $t^* \in T$  is detectable, then it is eventually detectable, using the trivial sequence  $T_0 = T$ ,  $T_1 = \{t^*\}$ .

The following lemma provides two key results that connect these ideas to the extraction problem. First, when a type  $t^*$  is detectable in some compact subset of types  $T^*$ , then for any  $\varepsilon > 0$  a contract can be constructed that leaves type  $t^*$  zero expected surplus while leaving the other types in  $T^*$  expected surplus at most  $\varepsilon$ . Second, building on this observation, if a type is eventually detectable, such a contract can be constructed to leave expected surplus at most  $\varepsilon$  for all types in  $T$ .

**Lemma 1.** *Let  $S$  be finite and  $v : T \rightarrow \mathbf{R}$  be given.*

- (i) Let  $T^* \subseteq T$  be compact and  $t^* \in T^*$ . If  $t^*$  is detectable in  $T^*$ , then for every  $\varepsilon > 0$  there exists  $c(t^*) \in \mathbf{R}^S$  such that  $v(t^*) - \pi(t^*) \cdot c(t^*) = 0$  and  $v(t) - \pi(t) \cdot c(t^*) < \varepsilon$  for all  $t \in T^* \setminus \{t^*\}$ .
- (ii) If  $t^* \in T$  is eventually detectable, then for every  $\varepsilon > 0$  there exists  $c(t^*) \in \mathbf{R}^S$  such that  $v(t^*) - \pi(t^*) \cdot c(t^*) = 0$  and  $v(t) - \pi(t) \cdot c(t^*) < \varepsilon$  for all  $t \in T \setminus \{t^*\}$ .

*Proof.* For (i), since  $t^*$  is detectable in  $T^*$ , choose  $z(t^*) \in \mathbf{R}^S$  such that

$$\begin{aligned}\pi(t^*) \cdot z(t^*) &= 0 \\ \pi(t) \cdot z(t^*) &> 0 \quad \forall t \in T^* \setminus \{t^*\}\end{aligned}$$

Fix  $\varepsilon > 0$ . Choose  $\delta > 0$  such that  $\|t - t^*\| < \delta \Rightarrow |v(t) - v(t^*)| < \varepsilon$ . Set  $\alpha(t^*) > 0$  such that

$$\alpha(t^*) > \max_{\{t \in T^* : \|t - t^*\| \geq \delta\}} \frac{v(t) - v(t^*)}{\pi(t) \cdot z(t^*)}$$

Note that  $\alpha(t^*)$  is well-defined (i.e., finite) since  $\{t \in T^* : \|t - t^*\| \geq \delta\}$  is compact,  $v$  and  $\pi$  are continuous, and  $\pi(t) \cdot z(t^*) > 0$  for all  $t \in T^* \setminus \{t^*\}$ .

Then define a contract  $c(t^*) \in \mathbf{R}^S$  as follows (here and in what follows, we use the constant  $r \in \mathbf{R}$  interchangeably with the constant vector  $(r, \dots, r) \in \mathbf{R}^S$ ):

$$c(t^*) = v(t^*) + \alpha(t^*)z(t^*)$$

Now for type  $t^*$ :

$$v(t^*) - \pi(t^*) \cdot c(t^*) = v(t^*) - v(t^*) - \alpha(t^*)(\pi(t^*) \cdot z(t^*)) = 0$$

For types  $t \in T^* \setminus \{t^*\}$ , first suppose  $\|t - t^*\| \geq \delta$ . Then

$$v(t) - \pi(t) \cdot c(t^*) = v(t) - v(t^*) - \alpha(t^*)(\pi(t) \cdot z(t^*)) < 0$$

by definition of  $\alpha(t^*)$ . If instead  $\|t - t^*\| < \delta$ , then

$$\begin{aligned}v(t) - \pi(t) \cdot c(t^*) &= v(t) - v(t^*) - \alpha(t^*)(\pi(t) \cdot z(t^*)) \\ &< v(t) - v(t^*) \quad \text{since } \pi(t) \cdot z(t^*) > 0 \text{ and } \alpha(t^*) > 0 \\ &< \varepsilon \quad \text{since } \|t - t^*\| < \delta\end{aligned}$$

Thus for all types  $t \in T^* \setminus \{t^*\}$ ,  $v(t) - \pi(t) \cdot c(t^*) < \varepsilon$ . This establishes the claim in (i).

For (ii), since  $t^*$  is eventually detectable in  $T$ , there exists a nested sequence of compact sets  $\{T_0, T_1, \dots, T_n\}$  with  $T_0 = T$ ,  $T_n = \{t^*\}$ , and  $T_{i+1} \subseteq T_i$  for each  $i$ , and a corresponding sequence  $\{z_1, \dots, z_n\} \subseteq \mathbf{R}^S$  such that for each  $i = 1, \dots, n$ ,

$$\begin{aligned}\pi(t) \cdot z_i &= 0 \quad \forall t \in T_i \\ \pi(t) \cdot z_i &> 0 \quad \forall t \in T_{i-1} \setminus T_i\end{aligned}$$

Fix  $\varepsilon > 0$ . Now consider  $T_{n-1}$  and  $t^*$ . Since  $t^*$  is detectable in  $T_{n-1}$ , by (i) there is a contract  $c_{n-1}(t^*)$  such that

$$v(t^*) - \pi(t^*) \cdot c_{n-1}(t^*) = 0$$

and

$$v(t) - \pi(t) \cdot c_{n-1}(t^*) < \frac{1}{n}\varepsilon \quad \forall t \in T_{n-1}$$

Now we claim, by induction, that for each  $k = 0, \dots, n-1$ , there is a contract  $c_k(t^*)$  such that

$$v(t^*) - \pi(t^*) \cdot c_k(t^*) = 0$$

and

$$v(t) - \pi(t) \cdot c_k(t^*) < \frac{n-k}{n}\varepsilon \quad \forall t \in T_k$$

To see this, fix  $i \geq 0$  and suppose there exists  $c_i(t^*)$  such that

$$v(t^*) - \pi(t^*) \cdot c_i(t^*) = 0$$

and

$$v(t) - \pi(t) \cdot c_i(t^*) < \frac{n-i}{n}\varepsilon \quad \forall t \in T_i$$

Then we claim that there exists  $c_{i-1}(t^*)$  such that

$$v(t^*) - \pi(t^*) \cdot c_{i-1}(t^*) = 0$$

and

$$v(t) - \pi(t) \cdot c_{i-1}(t^*) < \frac{n-(i-1)}{n}\varepsilon \quad \forall t \in T_{i-1}$$



To show this, first note that for each  $t \in T_i$ , there exists  $\delta(t) > 0$  such that  $\|s - t\| < \delta(t) \Rightarrow$

$$|v(t) - \pi(t) \cdot c_i(t^*) - [v(s) - \pi(s) \cdot c_i(t^*)]| < \frac{1}{n}\varepsilon$$

For each  $t \in T_i$ , let  $B_{\delta(t)}(t) = \{s \in T : \|s - t\| < \delta(t)\}$ . The collection  $\{B_{\delta(t)}(t) : t \in T_i\}$  is an open cover of the compact set  $T_i$ , so there exists  $\{t^1, \dots, t^m\} \subseteq T_i$  such that  $T_i \subseteq \cup_j B_{\delta(t^j)}(t^j)$ . Moreover,  $\cup_j B_{\delta(t^j)}(t^j)$  is open, so  $T_{i-1} \setminus \cup_j B_{\delta(t^j)}(t^j)$  is compact. Choose  $R_i > 0$  such that

$$R_i > \max_{t \in T_{i-1} \setminus \cup_j B_{\delta(t^j)}(t^j)} v(t) - \pi(t) \cdot c_i(t^*)$$

and choose  $\alpha_i(t^*) > 0$  such that

$$\alpha_i(t^*) > \max_{t \in T_{i-1} \setminus \cup_j B_{\delta(t^j)}(t^j)} \frac{R_i}{\pi(t) \cdot z_i(t^*)}$$

Then set

$$c_{i-1}(t^*) = c_i(t^*) + \alpha_i(t^*) \cdot z_i(t^*)$$

For each  $t \in T_i$ :

$$\begin{aligned} v(t) - \pi(t) \cdot c_{i-1}(t^*) &= v(t) - \pi(t) \cdot c_i(t^*) \\ &< \frac{n-i}{n}\varepsilon \\ &< \frac{n-(i-1)}{n}\varepsilon \end{aligned}$$

and

$$v(t^*) - \pi(t^*) \cdot c_{i-1}(t^*) = v(t^*) - \pi(t^*) \cdot c_i(t^*) = 0$$

For  $t \in T_{i-1} \setminus T_i$ :

- if  $t \in \cup_j B_{\delta(t^j)}(t^j)$ :

$$\begin{aligned} v(t) - \pi(t) \cdot c_{i-1}(t^*) &= v(t) - \pi(t) \cdot c_i(t^*) - \alpha_i(t^*)(\pi(t) \cdot z_i(t^*)) \\ &< v(t) - \pi(t) \cdot c_i(t^*) \quad \text{since } \pi(t) \cdot z_i(t^*) > 0 \\ &< \frac{n-i}{n}\varepsilon + \frac{1}{n}\varepsilon \quad \text{by construction} \\ &= \frac{n-(i-1)}{n}\varepsilon \end{aligned}$$

- if  $t \in T_{i-1} \setminus \cup_j B_{\delta(t^j)}(t^j)$ :

$$\begin{aligned} v(t) - \pi(t) \cdot c_{i-1}(t^*) &= v(t) - \pi(t) \cdot c_i(t^*) - \alpha_i(t^*)(\pi(t) \cdot z_i(t^*)) \\ &< 0 \quad \text{by choice of } \alpha_i(t^*) \end{aligned}$$

Thus for all  $t \in T_{i-1} \setminus \{t^*\}$ ,

$$v(t) - \pi(t) \cdot c_{i-1}(t^*) < \frac{n - (i - 1)}{n} \varepsilon$$

and

$$v(t^*) - \pi(t^*) \cdot c_{i-1}(t^*) = 0$$

Thus by induction, for each  $k$  there exists a contract  $c_k(t^*)$  such that for all  $t \in T_k \setminus \{t^*\}$ ,

$$v(t) - \pi(t) \cdot c_k(t^*) < \frac{n - k}{n} \varepsilon$$

and

$$v(t^*) - \pi(t^*) \cdot c_k(t^*) = 0$$

In particular, consider  $k = 0$ : there exists  $c(t^*) = c_0(t^*)$  such that for all  $t \in T_0 \setminus \{t^*\} = T \setminus \{t^*\}$ ,

$$v(t) - \pi(t) \cdot c(t^*) < \frac{n}{n} \varepsilon = \varepsilon$$

and

$$v(t^*) - \pi(t^*) \cdot c(t^*) = 0$$

This establishes the claim.  $\square$

From Lemma 1, it follows immediately that virtual extraction holds whenever all types are eventually detectable. We show next that when  $S$  is finite and types satisfy probabilistic independence, then while types need not be detectable, every type is eventually detectable, which in turn guarantees that virtual extraction holds.

**Theorem 5.** *Let  $S$  be finite.*

- (i) *If every type is eventually detectable, then virtual extraction holds.*

(ii) If types satisfy probabilistic independence, then every type is eventually detectable.

*Proof.* Part (i) follows from part (ii) of Lemma 1. For (ii), let  $t^* \in T$ . By probabilistic independence,  $\pi(t^*)$  is an extreme point of  $C$ . If  $t^*$  is not detectable, then  $\pi(t^*)$  is an extreme point of  $C$  that is not exposed. Hence by Theorem 4,  $\pi(t^*)$  is eventually exposed in  $C$ . Thus there is a sequence  $\{F_0, F_1, \dots, F_n\}$  with  $F_0 = C$ ,  $F_n = \{\pi(t^*)\}$ , and a corresponding sequence  $\{z_1, \dots, z_n\}$  such that for each  $i$ ,  $\pi(t^*) \in F_i$  and

$$\begin{aligned}\pi \cdot z_i &= 0 \quad \forall \pi \in F_i \\ \pi \cdot z_i &> 0 \quad \forall \pi \in F_{i-1} \setminus F_i\end{aligned}$$

Set  $T_0 = T$ , and

$$T_1 = \{t \in T : \pi(t) \in F_1\}$$

Note that since  $T$  and  $F_1$  are compact and the map  $t \mapsto \pi(t)$  is continuous,  $T_1$  is compact. Then for each  $i \geq 2$ , set

$$T_i = \{t \in T_{i-1} : \pi(t) \in F_i\}$$

By induction,  $T_i$  is compact for each  $i$ , and by construction  $T_i \subseteq T_{i-1}$  for each  $i$ . Also by construction,

$$T_n = \{t \in T_{n-1} : \pi(t) \in F_n\} = \{\pi(t^*)\} = \{t^*\}$$

Then let  $m$  be the minimum index  $i$  for which  $T_i = \{t^*\}$ . Consider the sequence  $\{T_0, T_1, \dots, T_m\}$  and the corresponding sequence  $\{z_1, \dots, z_m\}$ . Then for each  $i = 1, \dots, m$ ,

$$\begin{aligned}\pi(t) \cdot z_i &= 0 \quad \forall t \in T_i \\ \pi(t) \cdot z_i &> 0 \quad \forall t \in T_{i-1} \setminus T_i\end{aligned}$$

Thus  $t^*$  is eventually detectable.  $\square$

As Example 1 illustrated, full extraction requires that all types are detectable. With infinitely many types however, this is not sufficient for full

extraction. We close this section by noting a stronger condition which is sufficient for full extraction in general, and then see probabilistic independence implies this stronger condition with finitely many types.<sup>1</sup>

**Definition 11.** A type  $t^* \in T$  is *strongly detectable* if there exists  $z \in \mathbf{R}^S$  such that

$$\pi(t^*) \cdot z = 0$$

and

$$\inf_{t \neq t^*} \pi(t) \cdot z > 0$$

Note that if all types are strongly detectable, then full extraction is possible for any  $t \mapsto v(t)$ , and any type set  $T$ . This follows from the basic argument for the finite case, as sketched and adapted above. Also note that when  $T$  is finite and types satisfy probabilistic independence, which is equivalent to convex independence because  $T$  is finite, then all types are strongly detectable. We record these observations below.

**Theorem 6.** *Let  $S$  be finite.*

- (i) *If every type is strongly detectable, then full extraction holds for any  $v : T \rightarrow \mathbf{R}$ .*
- (ii) *If  $T$  is finite and types satisfy probabilistic independence, then every type is strongly detectable.*

*Proof.* Part (i) is a straightforward adaptation of arguments above; we omit the details. For (ii), fix  $t \in T$ . Since types satisfy probabilistic independence,  $\pi(t) \notin \overline{\text{co}}\{\pi(s) : s \in T, s \neq t\}$ . Thus there exists  $z(t) \in \mathbf{R}^S$  such that  $\pi(t) \cdot z(t) = 0$  and  $\pi(s) \cdot z(t) > 0$  for all  $s \in T, s \neq t$ . Since  $T$  is finite, this implies  $\inf_{s \neq t} \pi(s) \cdot z(t) > 0$ . Thus  $t$  is strongly detectable by definition. Repeating for each  $t \in T$  yields the result.  $\square$

---

<sup>1</sup>See McAfee and Reny (1992) for a discussion of several other sufficient conditions for full extraction when  $T$  is infinite.

## 5 General State Space

In this section we consider the general case in which the state space  $S$  is an arbitrary compact metric space. For example, this allows the model to accommodate the case in which  $S = T$ , or more generally when  $|S| \geq |T|$ .

We start by again considering the problem of full extraction, and cast the problem in slightly stronger terms (we will see that while this gives a stronger condition, probabilistic independence guarantees that virtual extraction holds under this stronger condition). Rather than looking for a collection of contracts  $\{c(t) \in C(S) : t \in T\}$ , we add the requirement that the contracts also be jointly continuous in types and states, and thus consider the existence of a schedule of contracts  $c \in C(T \times S)$  such that for each  $t \in T$ :

$$\begin{aligned} v(t) - \pi(t) \cdot c(t) &= 0 \\ v(t) - \pi(t) \cdot c(s) &\leq 0 \quad \forall s \neq t \end{aligned}$$

First, we note that full extraction is equivalent to the seemingly weaker condition *weak full extraction*

$$\begin{aligned} v(t) - \pi(t) \cdot c(t) &\geq 0 \\ v(t) - \pi(t) \cdot c(s) &\leq 0 \quad \forall s \neq t \end{aligned}$$

We establish this equivalence in the lemma below.

**Lemma 2.** *For each  $v : T \rightarrow \mathbf{R}$ ,  $c \in C(T \times S)$  satisfies full extraction if and only if  $c$  satisfies weak full extraction.*

*Proof.* Full extraction clearly implies weak full extraction. To see that these are equivalent, fix  $v : T \rightarrow \mathbf{R}$  and  $t \in T$ . Then choose a sequence  $s_n \rightarrow t$  with  $s_n \neq t$  for each  $n$ ; this is possible by the connectedness of  $T$ . Suppose  $c \in C(T \times S)$  satisfies weak full extraction. Then for each  $n$ , since  $s_n \neq t$ ,

$$v(s_n) - \pi(s_n) \cdot c(t) \leq 0$$

and

$$v(t) - \pi(t) \cdot c(t) \geq 0$$

Since  $v(s_n) \rightarrow v(t)$  and  $\pi(s_n) \rightarrow \pi(t)$ ,

$$v(t) - \pi(t) \cdot c(t) \leq 0$$

Thus

$$v(t) - \pi(t) \cdot c(t) = 0$$

Since  $t \in T$  was arbitrary, the equivalence follows.  $\square$

From Lemma 2, it is enough to consider the relaxed problem of weak full extraction. Now write

$$c(t) = v(t) + z(t)$$

where  $z \in C(T \times S)$ , and we use  $v(t) \in \mathbf{R}$  interchangeably with  $v(t)\mathbf{1}(S)$ , where  $\mathbf{1}(S)$  denotes the identity on  $S$ . Note that any  $c \in C(T \times S)$  can be written this way for appropriate choice of  $z$ . Then  $c$  satisfies full extraction if and only if for each  $t \in T$ ,

$$\begin{aligned} \pi(t) \cdot z(t) &\leq 0 \\ \pi(t) \cdot z(s) &\geq v(t) - v(s) \quad \forall s \neq t \end{aligned}$$

This follows from observing that for each  $t$ ,

$$v(t) - \pi(t) \cdot c(t) = v(t) - v(t) - \pi(t) \cdot z(t) = -\pi(t) \cdot z(t)$$

and

$$v(t) - \pi(t) \cdot c(s) = v(t) - v(s) - \pi(t) \cdot z(s)$$

Thus we will consider the existence of  $z \in C(T \times S)$  such that for each  $t \in T$ ,

$$\begin{aligned} \pi(t) \cdot z(t) &\leq 0 \\ \pi(t) \cdot z(s) &\geq v(t) - v(s) \quad \forall s \neq t \end{aligned}$$

To that end, let  $f : T \times C(T \times S) \rightarrow \mathbf{R}$  be given by

$$f(t, z) = \pi(t) \cdot z(t)$$

and for each  $t \in T$ , let  $f_t : C(T \times S) \rightarrow \mathbf{R}$  be given by  $f_t(z) = f(t, z)$ . Similarly, let  $g : T \times T \times C(T \times S) \rightarrow \mathbf{R}$  be given by

$$g_{(s,t)}(z) := g(s, t, z) = \pi(s) \cdot z(t)$$

Then to show that full extraction is possible, it suffices to show that there exists  $z \in C(T \times S)$  such that

$$\begin{aligned} f(t, z) &\leq 0 \quad \forall t \in T \\ g(t, s, z) &\geq v(t) - v(s) \quad \forall s \neq t, \forall t \in T \end{aligned}$$

**Note:** For each  $t \in T$ ,  $f_t$  is convex, and for every  $t, s \in T$ ,  $g_{(t,s)}$  is concave. (Both are in fact linear.) In addition,  $f(0) = g(0) = 0$ .

Now consider the problem

$$\begin{aligned} p^* := & \inf_{c \in \mathbf{R}, z \in C(T \times S)} c \\ & \text{subject to} \quad f(t, z) \leq c \quad \forall t \in T \\ & \quad \quad \quad v(t) - v(s) - g(t, s, z) \leq c \quad \forall t, s \in T \end{aligned} \tag{vse}$$

Note that if the optimal value  $p^*$  of this problem is less than or equal to zero, then at least virtual surplus extraction is possible. We establish this in the next lemma.

**Lemma 3.** *If  $p^* \leq 0$ , then virtual surplus extraction holds. If  $p^* < 0$ , or if  $p^* = 0$  and is obtained in (vse), then full extraction holds.*

*Proof.* To see this, first suppose  $p^* < 0$ . Then there must exist  $z \in C(T \times S)$  such that

$$f(t, z) \leq 0 \quad \forall t \in T$$

and

$$v(t) - v(s) - g(t, s, z) \leq 0 \quad \forall t, s \in T$$

In this case, full extraction holds, and thus a fortiori, virtual extraction holds as well. Similarly, if  $p^* = 0$ , then either  $p^*$  is realized, in which case again there must exist such a  $z \in C(T \times S)$  as above so that full extraction holds, or if  $p^*$  is not realized, then for each  $\varepsilon > 0$  there exists  $z_\varepsilon \in C(T \times S)$  such that

$$f(t, z_\varepsilon) \leq \varepsilon \quad \forall t \in T$$

and

$$v(t) - v(s) - g(t, s, z_\varepsilon) \leq \varepsilon \quad \forall t, s \in T$$

Now note that using continuity of  $\pi$  and  $v$ , this implies

$$\pi(t) \cdot z_\varepsilon(t) \geq -\varepsilon \quad \forall t \in T$$

Thus

$$-\varepsilon \leq \pi(t) \cdot z_\varepsilon(t) \leq \varepsilon \quad \forall t \in T$$

Now set

$$z := z_\varepsilon - \varepsilon$$

and for each  $t \in T$ , set the contract  $c(t)$  to be

$$c(t) = v(t) + z(t)$$

Then  $c \in C(T \times S)$ , and for each  $t \in T$ ,

$$\begin{aligned} v(t) - \pi(t) \cdot c(t) &= v(t) - v(t) - \pi(t) \cdot z(t) \\ &= -\pi(t) \cdot z(t) \\ &= \varepsilon - \pi(t) \cdot z_\varepsilon(t) \end{aligned}$$

and by the preceding argument,

$$0 \leq \varepsilon - \pi(t) \cdot z_\varepsilon(t) \leq 2\varepsilon$$

Thus for each  $t \in T$ ,

$$0 \leq v(t) - \pi(t) \cdot c(t) \leq 2\varepsilon$$

Then fix  $t \in T$ , and consider  $s \neq t$ .

$$\begin{aligned} v(t) - \pi(t) \cdot c(s) &= v(t) - v(s) - \pi(t) \cdot z(s) \\ &= v(t) - v(s) - \pi(t) \cdot z_\varepsilon(s) + \varepsilon \\ &\leq v(t) - v(s) - g(t, s, z_\varepsilon) + \varepsilon \\ &\leq 2\varepsilon \end{aligned}$$

Thus for all  $t \in T$ ,

$$v(t) - \pi(t) \cdot c(s) \leq 2\varepsilon \quad \forall s \neq t$$

So for all  $t \in T$ ,

$$0 \leq \sup_{s \in T} v(t) - \pi(t) \cdot c(s) \leq 2\varepsilon$$

The result follows. □



Thus to show that virtual surplus extraction is possible, it suffices to show that  $p^* \not\asymp 0$ . We establish this below by considering the dual of the optimization problem (vse), and making use of duality to argue that these problems have the same value. The heart of the proof is then to show that this common value cannot be positive under probabilistic independence.

**Theorem 7.** *Let  $S$  be a compact metric space. If types satisfy probabilistic independence, then virtual extraction holds.*

*Proof.* By Lemma 3, to show that virtual surplus extraction is possible it suffices to show that  $p^* \not\asymp 0$ . To that end, note that the Lagrange dual function for the problem (vse) is

$$\mathcal{L}(\lambda, \nu) = \inf_{\substack{c \in \mathbf{R} \\ z \in C(T \times S)}} \{c + \lambda \cdot (f(t, z) - c) + \nu \cdot (v(t) - v(s) - g(t, s, z) - c)\}$$

where  $\lambda \in \mathcal{M}(T)$  and  $\nu \in \mathcal{M}(T \times T)$ . Let  $d \in C(T \times T)$  be given by

$$d(t, s) = v(t) - v(s)$$

Note that  $d(t, t) = 0$  for all  $t \in T$ .

Define

$$h(\lambda, \nu) = \inf_{z \in C(T \times S)} \{\lambda \cdot f(z) + \nu \cdot (d - g(z))\}$$

Using this notation, we can rewrite the Lagrange dual function for (vse) as follows:

$$\mathcal{L}(\lambda, \nu) = \begin{cases} h(\lambda, \nu) & \text{if } \int \lambda(dt) + \iint \nu(ds dt) = 1 \\ -\infty & \text{otherwise} \end{cases}$$

Thus the dual problem of (vse) is

$$\begin{aligned} d^* &:= \sup_{\lambda \in \mathcal{M}(T), \nu \in \mathcal{M}(T \times T)} h(\lambda, \nu) \\ &\text{subject to} \quad (\lambda, \nu) \geq 0 && \text{(d-vse)} \\ &\quad \int \lambda(dt) + \iint \nu(ds dt) = 1 \end{aligned}$$

Then note that Slater's condition holds for the original problem (vse). To see this, set  $z = 0$ , so

$$f(t, z) = g(t, s, z) = 0 \quad \forall t, s \in T$$

Then choose

$$\bar{c} > \sup_{t,s \in T} v(t) - v(s) \geq 0$$

For  $(z, c) = (0, \bar{c})$ ,

$$\sup_{t \in T} f(t, z) - c = -\bar{c} < 0$$

and

$$\sup_{t,s \in T} v(t) - v(s) - g(t, s, z) - c = \sup_{t,s \in T} v(t) - v(s) - \bar{c} < 0$$

Thus  $p^* = d^*$  and in addition  $d^*$  is obtained, where  $p^*$  is the optimal value of (vse) and  $d^*$  is the optimal value of (d-vse).

Now it suffices to show that  $p^* = d^* \not\geq 0$ . To show this, suppose by way of contradiction that  $p^* = d^* > 0$ . Since  $d^*$  is obtained in (d-vse), there exists  $(\lambda, \nu) \geq 0$  such that

$$d^* = h(\lambda, \nu) > 0 \text{ and } \int \lambda(dt) + \iint \nu(ds dt) = 1$$

Recall that, by definition,

$$h(\lambda, \nu) = \inf_{z \in C(T \times S)} (\lambda \cdot f(z) + \nu \cdot (d - g(z)))$$

and  $f(0) = g(0) = 0$ , which implies

$$h(\lambda, \nu) \leq \nu \cdot d$$

Since  $h(\lambda, \nu) > 0$ , this implies  $\nu \cdot d > 0$ . Thus  $\nu \neq 0$ . Since  $\lambda, \nu \geq 0$ , this implies  $\nu > 0$ .

Let  $F : C(T \times S) \rightarrow \mathbf{R}$  be given by

$$F(z) = \lambda \cdot f(z) + \nu \cdot (d - g(z))$$

Note that  $F$  is convex and continuous, and by definition,

$$h(\lambda, \nu) = \inf_{z \in C(T \times S)} F(z) > 0$$

In particular, this implies  $\inf_{z \in C(T \times S)} F(z) \in \mathbf{R}$ . By Ekeland's Variational Principle (see Lemma 4 in the Appendix), there exists a sequence  $\{z_n\}$  and a sequence  $\{\gamma_n\}$  with  $\gamma_n \in \partial F(z_n)$  for each  $n$  such that

$$F(z_n) = \lambda \cdot f(z_n) + \nu \cdot (d - g(z_n)) \rightarrow \inf_{z \in C(T \times S)} F(z) = h(\lambda, \nu) > 0$$

and

$$\|\gamma_n\| \rightarrow 0$$

By Lemmas 5, 6 and 7 (see the Appendix), since  $\gamma_n \in \partial F(z_n)$  for each  $n$ ,  $\gamma_n$  is the measure for which

$$\gamma_n \cdot y = \int \pi(t) \cdot y(t) \lambda(dt) - \iint \pi(s) \cdot y(t) \nu(ds dt)$$

for any measurable function  $y$ . But then note that  $\gamma_n$  is constant for each  $n$ ; let this constant measure be denoted  $\gamma$ . Since  $\|\gamma_n\| \rightarrow 0$ , this implies  $\|\gamma\| = 0$ , and that  $\gamma \cdot y = 0$  for any such  $y$ .

Now fix  $A \subseteq T$  and let  $y$  be given by

$$y(t) = \begin{cases} 0 & \text{if } t \notin A \\ \mathbf{1}(S) & \text{if } t \in A \end{cases}$$

where  $\mathbf{1}(S)$  is the indicator of  $S$ . Then

$$\begin{aligned} \gamma \cdot y &= \int_A \pi(t) \cdot y(t) \lambda(dt) - \iint_{T \times A} \pi(s) \cdot y(t) \nu(ds dt) \\ &= \int_A \lambda(dt) - \iint_{T \times A} \nu(ds dt) \\ &= \lambda(A) - \nu(T \times A) \end{aligned}$$

And  $\gamma \cdot y = 0$ , which implies  $\lambda(A) - \nu(T \times A) = 0$ , that is,  $\lambda(A) = \nu(T \times A)$ . Since  $A$  was arbitrary,  $\lambda(A) = \nu(T \times A)$  for each  $A \subseteq T$ . From this it follows first that  $\lambda(T) = \nu(T \times T)$ , and since  $\nu > 0$ , this implies  $\lambda(T) = \nu(T \times T) > 0$ . Then without loss of generality, rescaling if necessary, take  $\lambda(T) = \nu(T \times T) = 1$ . Second, this implies that, using disintegration of measures, we can write

$$\nu = \int \nu_t(ds) \lambda(dt)$$

where  $\nu_t$  is a measure on  $T$ ,  $\nu_t \geq 0$  and  $\nu_t(T) = 1$  for each  $t$  in the support of  $\lambda$ .

For each  $t \in T$ , let

$$\gamma(t) = \pi(t) - \int \pi(s) \nu_t(ds)$$

Then  $\gamma$  is the measure given by

$$\gamma(E) = \int \gamma(t)(E_t) \lambda(dt) \quad \forall E \subseteq T \times S$$

where for  $E \subseteq T \times S$ ,  $E_t := \{s \in S : (t, s) \in E\}$ .

Then note that

$$\begin{aligned} \|\gamma\| &= \sup_E |\gamma(E)| \quad \text{by definition} \\ &= \gamma^+(T \times S) + \gamma^-(T \times S) \quad \text{by definition} \\ &= \int [\gamma^+(t)(S) + \gamma^-(t)(S)] \lambda(dt) \\ &= \int \|\gamma(t)\| \lambda(dt) \end{aligned}$$

Recall from above

$$\|\gamma\| = \int \|\gamma(t)\| \lambda(dt) = 0$$

By definition,  $\|\gamma(t)\| \geq 0$  for each  $t \in T$ , hence  $\|\gamma(t)\| = 0$  for  $\lambda$ -a.e  $t \in T$ .

Thus for  $\lambda$ -a.e  $t \in T$ ,

$$\gamma(t) = \pi(t) - \int \pi(s) \nu_t(ds) = 0$$

where  $\nu_t \in \Delta(T)$ . Thus by probabilistic independence,  $\nu_t = \delta_t$  for  $\lambda$ -a.e  $t \in T$ .

But then

$$\begin{aligned} \nu \cdot d &= \iint d(s, t) \nu(ds dt) \\ &= \iint d(s, t) \nu_t(ds) \lambda(dt) \\ &= \int_{\text{supp } \lambda} d(t, t) \lambda(dt) \\ &= 0 \quad \text{since } d(t, t) = 0 \text{ for all } t \in T \end{aligned}$$

This is a contradiction, as  $\nu \cdot d > 0$ . Thus  $p^* \leq 0$ . □

## 6 Discussion

The techniques developed in sections 4 and 5 provide new insights into the foundational result of McAfee and Reny (1992), and the surplus extraction problem more generally. Surplus extraction theorems are central results in mechanism design. Their conclusions that designers can extract all, or virtually all, information rents under standard assumptions yield strong and important predictions. These predictions are arguably implausible in a variety of practical settings, motivating significant work reconsidering many standard assumptions in mechanism design. New methods that help to understand what drives these results and their limitations are useful as a consequence. The constructive methods developed in section 4 highlight the connection between surplus extraction and the geometry of the set of beliefs, while the methods developed in section 5 instead highlight the connection between surplus extraction and duality by using the natural characterization of full surplus extraction contracts as solutions to families of inequalities. Both techniques are useful for understanding when surplus extraction holds, when it might fail, and limits on designers in settings where virtual or full extraction fails. Both techniques also suggest methods for studying surplus extraction in models beyond the standard case.

To indicate how the methods we developed can be used to study surplus extraction in other settings, we close by considering a modification of the standard model in which each type is associated with a set of beliefs rather than a single belief. This set could stem from the designer's objective to have mechanisms that are robust to misspecifying agents' beliefs, or from agents' perceptions of ambiguity. We use the constructive methods from section 4 in a simple example in this framework, building on Example 1, to illustrate.

**Example 2:** Suppose  $S$  is finite. Let  $\pi_1, \pi_2 : T \rightarrow \Delta(S)$  where  $\pi_1$  is as in Example 1, and  $\pi_2$  is a translation of  $\pi_1$  as depicted in Figure 8. For each  $t \in T$ , set  $\Pi(t) = \{a\pi_1(t) + (1-a)\pi_2(t) : a \in [0, 1]\}$ . See Figure 8. Then  $\Pi(t) \subseteq \Delta(S)$  is compact and convex for each  $t \in T$ , and  $\Pi : T \rightarrow 2^{\Delta(S)}$  is a continuous correspondence.

Given  $v : T \rightarrow \mathbf{R}$ , for each  $t \in T$  define  $V_t : \mathbf{R}^S \rightarrow \mathbf{R}$  by

$$V_t(x) = \min_{\pi \in \Pi(t)} \pi \cdot (v(t) - x) = v(t) - \max_{\pi \in \Pi(t)} \pi \cdot x$$

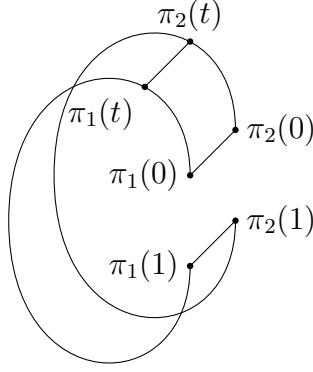


Figure 8: Graph of  $\{\Pi(t) : t \in T\}$

For each  $t \in T$ ,  $V_t(x)$  gives the minimum expected surplus for type  $t$  from the contract  $x \in \mathbf{R}^S$ , computed with respect to beliefs in  $\Pi(t)$ . This is a version of maxmin expected utility, as in Gilboa and Schmeidler (1989).

We consider the designer's surplus extraction problem, given these utilities for agents. As we noted above, this can be motivated either by robustness concerns of the designer, or the perception of ambiguity by agents. *Virtual extraction* holds here if for each  $v : T \rightarrow \mathbf{R}$  and for each  $\varepsilon > 0$ , there exists a menu  $\{c(t) \in \mathbf{R}^S : t \in T\}$  such that for each type  $t \in T$ ,

$$0 \leq V_t(c(t)) \leq \varepsilon \quad \text{and} \quad V_t(c(s)) \leq \varepsilon \quad \forall s \neq t$$

We adapt the constructive argument in section 4 to show that virtual extraction holds in this setting.<sup>2</sup>

First, let  $t \in T \setminus \{0, 1\}$ . In this case, there exists  $z(t) \in \mathbf{R}^S$  such that

$$\pi \cdot z(t) = 0 \quad \forall \pi \in \Pi(t) \quad \text{and} \quad \pi \cdot z(t) > 0 \quad \forall \pi \in \Pi(s), \quad \forall s \neq t$$

See Figure 9. As in Example 1, we construct a contract of the form  $c(t) = v(t) + \alpha(t)z(t)$  for appropriate choice of  $\alpha(t)$ .

Let  $\varepsilon > 0$  be given. Choose  $\delta > 0$  such that  $\|s - t\| < \delta \Rightarrow v(s) - v(t) < \varepsilon$ .

---

<sup>2</sup>As in the standard case, it is straightforward to show that virtual extraction can also be achieved with a finite menu; we omit this extension of Theorem 2.

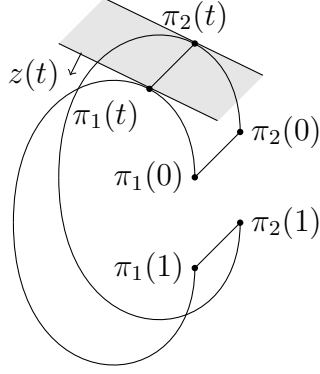


Figure 9: Construction of  $c(t)$ ,  $t \in T \setminus \{0, 1\}$ .

Then choose  $\alpha(t) > 0$  sufficiently large so that

$$\alpha(t) > \max_{\substack{\|s-t\| \geq \delta \\ \pi \in \Pi(s)}} \frac{v(s) - v(t)}{\pi \cdot z(t)}$$

For the resulting contract  $c(t) = v(t) + \alpha(t)z(t)$ ,

$$V_t(c(t)) = v(t) - \max_{\pi \in \Pi(t)} \pi \cdot c(t) = v(t) - v(t) - \alpha(t) \max_{\pi \in \Pi(t)} \pi \cdot z(t) = 0$$

and for all  $s \neq t$ ,

$$V_s(c(t)) = v(s) - \max_{\pi \in \Pi(s)} \pi \cdot c(t) = v(s) - v(t) - \alpha(t) \max_{\pi \in \Pi(s)} \pi \cdot z(t) \leq \varepsilon$$

Now consider  $t \in \{0, 1\}$ , and without loss of generality take  $t = 0$ . First note that, as depicted in the left side of Figure 10, there exists  $z(0) \in \mathbf{R}^S$  such that

$$\pi \cdot z(0) = 0 \quad \forall \pi \in \Pi(t), \quad t = 0, 1 \quad \text{and} \quad \pi \cdot z(0) > 0 \quad \forall \pi \in \Pi(s), \quad \forall s \notin \{0, 1\}$$

Then note that there exists  $z_1(0) \in \mathbf{R}^S$  such that

$$\pi \cdot z_1(0) = 0 \quad \forall \pi \in \Pi(0) \quad \text{and} \quad \pi \cdot z_1(0) > 0 \quad \forall \pi \in \Pi(1)$$

See the right side of Figure 10. Choose  $\alpha_1(0) > 0$  sufficiently large so that

$$\alpha_1(0) > \max_{\pi \in \Pi(1)} \frac{v(1) - v(0)}{\pi \cdot z_1(0)}$$

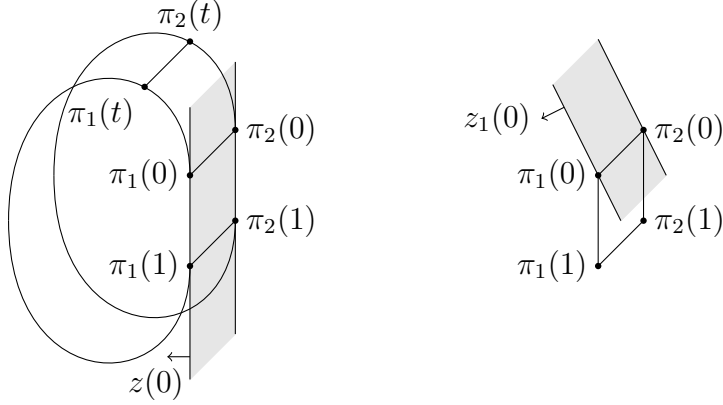


Figure 10: Construction of  $c_1(0)$  and  $c(0)$ .

Then set  $c_1(0) = v(0) + \alpha_1(0)z_1(0)$ . By construction,

$$V_0(c_1(0)) = v(0) - \max_{\pi \in \Pi(0)} \pi \cdot c_1(0) = v(0) - v(0) - \alpha_1(0) \max_{\pi \in \Pi(0)} \pi \cdot z_1(0) = 0$$

and

$$V_1(c_1(0)) = v(1) - \max_{\pi \in \Pi(1)} \pi \cdot c_1(0) = v(1) - v(0) - \alpha_1(0) \max_{\pi \in \Pi(1)} \pi \cdot z_1(0) < 0$$

The contract  $c_1(0)$  might leave surplus for other types  $t \notin \{0, 1\}$ , however. As in Example 1, we use the additional stochastic payment  $z(0)$  to account for this.

To that end, for each  $t \in T$  set

$$\bar{V}_t(c_1(0)) = v(t) - v(0) - \min_{\pi \in \Pi(t)} \pi \cdot (\alpha_1(0)z_1(0))$$

For type  $t$ ,  $\bar{V}_t(c_1(0))$  is the maximum expected surplus from the contract  $c_1(0)$  over all beliefs in  $\Pi(t)$ . Note that by construction,  $\bar{V}_0(c_1(0)) = 0$  and  $\bar{V}_1(c_1(0)) < 0$ . Then choose  $\delta > 0$  such that for  $\|s - 0\| < \delta$  or  $\|s - 1\| < \delta$ ,  $\bar{V}_s(c_1(0)) < \varepsilon$ . This is possible, by the continuity of  $v$  and  $\Pi$ , and the fact that  $\bar{V}_0(c_1(0)) = 0$  and  $\bar{V}_1(c_1(0)) < 0$ .

Now set  $\alpha(0) > 0$  such that

$$\alpha(0) > \max_{\substack{\|s-0\| \geq \delta \\ \|s-1\| \geq \delta \\ \pi \in \Pi(s)}} \frac{\bar{V}_s(c_1(0))}{\pi \cdot z(0)}$$



and set

$$c(0) = c_1(0) + \alpha(0)z(0) = v(0) + \alpha_1(0)z_1(0) + \alpha(0)z(0)$$

By construction, for types  $t \in \{0, 1\}$ ,  $\pi \cdot c(0) = \pi \cdot c_1(0)$  for all  $\pi \in \Pi(t)$ . Thus  $V_0(c(0)) = 0$  and  $V_1(c(0)) < 0$ . For  $t \notin \{0, 1\}$ , again by construction,

$$V_t(c(0)) = v(t) - v(0) - \max_{\pi \in \Pi(t)} \pi \cdot (\alpha_1(0)z_1(0) + \alpha(0)z(0)) < \varepsilon$$

The collection  $\{c(t) : t \in T\}$  thus constructed achieves extraction of all but at most  $\varepsilon$  surplus.  $\diamond$

## 7 Appendix

Because these results might be of independent interest, we include here the derivation of the version of Ekeland's Variational Principle that we used in the proof of Theorem 7.

Before giving the main result of the appendix, we start with some preliminary definitions and results, including the classic version of Ekeland's Variational Principle, and an extension due to Borwein, from which the main result follows quickly.

**Definition 12.** For  $x \in \text{dom } f$  and  $\varepsilon > 0$ , the  $\varepsilon$ -subdifferential of  $f$  at  $x$ , denoted  $\partial_\varepsilon f(x)$  is

$$\partial_\varepsilon f(x) = \{x^* \in X^* : f(y) \geq f(x) + \langle x^*, y - x \rangle - \varepsilon \quad \forall y \in X\}$$

**Note:** For any  $x \in \text{dom } f$  and any  $\varepsilon > 0$ ,

- $\partial_\varepsilon f(x)$  is closed and convex for every  $x$
- $\inf f \leq f(x_\varepsilon) \leq \inf f + \varepsilon \iff 0 \in \partial_\varepsilon f(x_\varepsilon)$
- for  $\varepsilon = 0$ ,  $\partial_0 f(x) = \partial f(x)$
- if  $f$  is convex and lsc, then  $\partial_\varepsilon f(x) \neq \emptyset$  for every  $x \in \text{dom } f$

Next we state the classic version of Ekeland's Variational Principle (Ekeland, 1974).

**Theorem 8. (Ekeland's Variational Principle)** *Let  $V$  be a complete metric space and  $F : V \rightarrow \mathbf{R} \cup \{+\infty\}$  be a proper, lsc function such that  $\inf F > -\infty$ . Let  $\varepsilon > 0$  and  $\beta > 0$ . For every  $u \in V$  such that*

$$\inf F \leq F(u) \leq \inf F + \varepsilon$$

*there exists  $v \in V$  such that*

$$(i) \quad F(v) \leq F(u)$$

$$(ii) \quad d(u, v) \leq \beta$$

$$(iii) \quad F(u) \geq F(v) - \frac{\varepsilon}{\beta}d(v, w) \quad \forall w \neq v$$

*If in addition  $F$  is convex, then*

$$(iv) \quad v \text{ can be chosen such that there exists } g \in \partial F(v) \text{ such that } \|g\| \leq \frac{\varepsilon}{\beta}$$

More precise approximations can be given for convex functions, as shown by Borwein (1982).

**Theorem 9. (Borwein, 1982, Theorem 1)** *Let  $X$  be a Banach space and  $f : X \rightarrow \mathbf{R} \cup \{+\infty\}$  be a proper, convex, lsc function. Let  $\varepsilon > 0$  and  $k \geq 0$  be given. Let*

$$x_0^* \in \partial_\varepsilon f(x_0)$$

*Then there exist  $x_\varepsilon$  and  $x_\varepsilon^*$  such that*

$$x_\varepsilon^* \in \partial f(x_\varepsilon)$$

*and such that*

$$\begin{aligned} \|x_\varepsilon - x_0\| &\leq \sqrt{\varepsilon} \\ |f(x_\varepsilon) - f(x_0)| &\leq \sqrt{\varepsilon}(\sqrt{\varepsilon} + \frac{1}{k}) \\ \|x_\varepsilon^* - x_0^*\| &\leq \sqrt{\varepsilon}(1 + k\|x_0^*\|) \\ |x_\varepsilon^*(h) - x_0^*(h)| &\leq \sqrt{\varepsilon}(\|h\| + k|x_0^*(h)|) \\ x_\varepsilon^* &\in \partial_{2\varepsilon} f(x_0) \end{aligned}$$

Putting these two results together yields the following.

**Lemma 4.** *Let  $X$  be a Banach space and  $f : X \rightarrow \mathbf{R}$  be a proper, convex, lsc function such that  $\inf f > -\infty$ . Then there exists a sequence  $\{x_n\}$  in  $X$  such that  $f(x_n) \rightarrow \inf f$  and  $d(0, \partial f(x_n)) \rightarrow 0$ , i.e., there exists  $\{g_n\}$  such that  $g_n \in \partial f(x_n)$  for each  $n$  and  $\|g_n\| \rightarrow 0$ .*

*Proof.* For each  $n \in \mathbf{N}$  there exists  $x_n$  such that

$$\inf f \leq f(x_n) \leq \inf f + \frac{1}{4n^2}$$

Then  $0 \in \partial_{\frac{1}{4n^2}} f(x_n)$  for each  $n$ .

By Borwein (1982, Theorem 1), for each  $n$  there exist  $\bar{x}_n$  and  $\bar{x}_n^*$  such that (with  $k = 1$  here)

$$\begin{aligned} \bar{x}_n^* &\in \partial f(\bar{x}_n) \\ \|\bar{x}_n - x_n\| &\leq \frac{1}{2n} \\ |f(\bar{x}_n) - f(x_n)| &\leq \frac{1}{2n} \left( \frac{1}{2n} + 1 \right) \\ &\leq \frac{1}{4n^2} + \frac{1}{2n} \leq \frac{1}{n} \\ \|\bar{x}_n^* - 0\| = \|\bar{x}_n^*\| &\leq \frac{1}{2n} (1 + 0) = \frac{1}{2n} \\ |\bar{x}_n^*(h)| &\leq \frac{1}{2n} \|h\| \end{aligned}$$

Then for each  $n$ ,

$$\begin{aligned} \inf f \leq f(\bar{x}_n) &\leq f(x_n) + \frac{1}{n} \\ &\leq \inf f + \frac{1}{4n^2} + \frac{1}{n} \\ &\leq \inf f + \frac{2}{n} \end{aligned}$$

and

$$\bar{x}_n^* \in \partial f(\bar{x}_n) \text{ with } \|\bar{x}_n^*\| \leq \frac{1}{2n}$$

Thus  $f(\bar{x}_n) \rightarrow \inf f$  and  $\|\bar{x}_n^*\| \rightarrow 0$ . □

**Lemma 5.** Let  $B$  be a compact metric space and  $X = C(B)$ . Let  $f : B \times X \rightarrow \mathbf{R}$  be continuous, and for each  $b \in B$ , let  $f_b : X \rightarrow \mathbf{R}$  be given by  $f_b(x) = f(b, x)$ . Suppose  $f_b$  is convex for each  $b \in B$ . Let  $h : X \rightarrow \mathbf{R}$  be given by

$$h(x) = \int f_b(x) \mu(db)$$

where  $\mu \in \mathcal{M}(B)$  and  $\mu \geq 0$ . Then  $h$  is convex, and

$$\partial h(x) = \int \partial f_b(x) \mu(db)$$

That is, for each  $\gamma \in \partial h(x)$  there is a mapping  $b \mapsto \gamma_b$  such that  $\gamma_b \in \partial f_b(x)$  for  $\mu$ -a.e  $b \in B$  and

$$\gamma \cdot y = \int \gamma_b \cdot y \mu(db)$$

for any measurable  $y$ .

*Proof.* Since  $B$  is a compact metric space,  $X = C(B)$  is separable. The result then follows from Ioffe and Levin (1972); see also Clarke (1990) Theorem 2.7.2 and discussion on pp. 76-77.  $\square$

**Lemma 6.** Let  $f : T \times C(T \times S) \rightarrow \mathbf{R}$  be given by

$$f(t, z) = \pi(t) \cdot z(t)$$

and for each  $t \in T$ , let  $f_t : C(T \times S) \rightarrow \mathbf{R}$  be given by  $f_t(z) = f(t, z)$ . Then  $f$  is continuous and  $f_t$  is convex for each  $t \in T$ . For each  $t \in T$ , if  $\gamma \in \partial f_t(z)$  then  $\gamma \in \mathcal{M}(T \times S)$  is a measure such that for any measurable function  $y$ ,

$$\gamma \cdot y = \pi(t) \cdot y(t)$$

*Proof.* First,  $f$  is continuous, by construction. To see this, let  $(s_n, z_n) \rightarrow (t, z)$ . Then

$$\begin{aligned} \|\pi(s_n) \cdot z_n(s_n) - \pi(t) \cdot z(t)\| &= \|(\pi(s_n) - \pi(t)) \cdot z_n(s_n) + \pi(t) \cdot (z_n(s_n) - z(t))\| \\ &\leq \|(\pi(s_n) - \pi(t)) \cdot z_n(s_n)\| + \|\pi(t) \cdot (z_n(s_n) - z(t))\| \\ &\leq \|\pi(s_n) - \pi(t)\| \|z_n(s_n)\| + \|\pi(t)\| \|z_n(s_n) - z(t)\| \end{aligned}$$

Since  $z_n \rightarrow z$ ,  $\{z_n(s_n)\}$  and  $\{z_n(s_n) - z(t)\}$  are bounded, and  $z_n(s_n) \rightarrow z(t)$  pointwise. Then  $\|\pi(s_n) - \pi(t)\| \|z_n(s_n)\| \rightarrow 0$ , since  $\pi(s_n) \rightarrow \pi(t)$  in norm, and  $\pi(t) \cdot (z_n(s_n) - z(t)) \rightarrow 0$  by the bounded convergence theorem. Thus  $f(s_n, z_n) \rightarrow f(t, z)$ .

By construction,  $f_t$  is linear and continuous for each  $t \in T$ , and  $\partial f_t(z)$  is the measure  $\gamma_t \in \mathcal{M}(T \times S)$  such that

$$\gamma_t \cdot y = \pi(t) \cdot y(t) \quad \text{for } y \text{ measurable function}$$

□

**Lemma 7.** *Let  $g : T \times T \times C(T \times S) \rightarrow \mathbf{R}$  be given by*

$$g_{(s,t)}(z) := g(s, t, z) = \pi(s) \cdot z(t)$$

*Then  $g$  is continuous, and  $g_{(s,t)}$  is concave for each  $t, s \in T$ . For each  $t, s \in T$ , if  $\gamma \in \partial g_{(s,t)}(z)$  then  $\gamma \in \mathcal{M}(T \times S)$  is a measure such that for any measurable function  $y$ ,*

$$\gamma \cdot y = \pi(s) \cdot y(t)$$

*Proof.* This follows from arguments analogous to those used in the proof of Lemma 6. □

## References

- [1] Bergemann, D., B. Brooks, and S. Morris (2017a), “Selling to Intermediaries: Optimal Auction Design in a Common Value Model,” working paper.
- [2] Bergemann, D., B. Brooks, and S. Morris (2017b), “Informationally Robust Optimal Auction Design,” working paper.
- [3] Borwein, J. (1982), “A Note on  $\varepsilon$ -subgradients and Maximal Monotonicity,” *Pacific Journal of Mathematics*, 103(2): 307-314.
- [4] Carroll, G., and I. Segal (2018), “Robustly Optimal Auctions with Unknown Resale Opportunities,” forthcoming *Review of Economic Studies*.

- [5] Clarke, F.H. (1990), *Optimization and Nonsmooth Analysis*, New York: Wiley.
- [6] Crémer, J. and R. McLean (1985), “Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent,” *Econometrica*, 53(2): 345-61.
- [7] Crémer, J. and R. McLean (1988), “Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions,” *Econometrica*, 56(6): 1247-57.
- [8] Daskalakis, C., A. Deckelbaum, and C. Tzamos (2017), “Strong Duality for a Multiple-Good Monopolist,” *Econometrica*, 85(3): 735-767.
- [9] Ekeland, I. (1974), “On the Variational Principle,” *Journal of Mathematical Analysis and Applications*, 47: 324-353.
- [10] Gilboa, I. and D. Schmeidler (1989), “Maxmin Expected Utility with Non-Unique Prior,” *Journal of Mathematical Economics*, 18(2): 141-153.
- [11] Ioffe, A. and V. Levin (1972), “Subdifferentials of Convex Functions,” *Transactions of the Moscow Mathematical Society*, 26: 1-72.
- [12] McAfee, P. and P. Reny (1992), “Correlated Information and Mechanism Design,” *Econometrica*, 60(2): 395-421.
- [13] Rahman, D. (2012), “Surplus Extraction on Arbitrary Type Spaces,” working paper.
- [14] Simon, B. (2011), *Convexity: An Analytic Viewpoint*, Cambridge: Cambridge University Press, 2011.