The Genericity of the McAfee-Reny Condition for Full Surplus Extraction in Models with a Continuum of Types

Alia Gizatulina
Martin Hellwig

MAX PLANCK SOCIETY

The Genericity of the McAfee-Reny Condition for Full Surplus Extraction in Models with a Continuum of Types

Alia Gizatulina / Martin Hellwig

July 2015

The Genericity of the McAfee-Reny Condition for Full Surplus Extraction in Models with a Continuum of Types [*]

Alia Gizatulina[†] and Martin Hellwig[‡]

This version: July 5, 2015

Abstract

McAfee and Reny (1992) have given a necessary and sufficient condition for full surplus extraction in models with a continuum of types. We interpret their condition as significantly stronger version of the requirement of injectiveness of the function mapping abstract types into beliefs and prove that their condition is satisfied by a generic set of model specifications. Our analysis involves an extension of the classical embedding theorem for continuous functions. Our proof does not rely on finite approximations and the topology on beliefs that we use is compatible with strategic continuity.

JEL Classification: D40, D44, D80, D82

Keywords: surplus extraction, mechanism design, BDP, correlated information, strategic continuity

1 Introduction

A central theme of the economics of information concerns the ability of agents to earn rents because they have private information. For example, the buyer of a good may be

1

able to obtain a surplus because the seller does not know how much the good is worth to the buyer.

However, Crémer and McLean (1988) have shown that, when there are multiple potential buyers for a good and these buyers have correlated private values, then, under certain conditions, in a Bayesian setting, a seller can extract all the surplus from the sale of his good, i.e. all information rents can be made to disappear. Specifically, if the potential buyers have only finitely many types, a Bayesian incentive mechanism that extracts all the potentially available surplus from buyers can be designed if and only if, for each agent i and each type t_i of this agent, the vector of beliefs that agent i has about the other agents' types when his own type is t_i cannot be represented as a convex combination of the vectors of beliefs about the other agents' types that he has at types other than t_i.

McAfee and Reny (1992) have extended the analysis of Crémer and McLean (1988) to the case where each agent's type set is the unit interval and where each agent's beliefs about other agents' types are given by a probability distribution with a continuous density function. They showed that (approximately) full surplus extraction can be obtained if and only if the density functions that represent agents' beliefs satisfy a function space analogue of the Crémer-McLean condition.

With finitely many types, the Crémer-McLean condition for full surplus extraction is necessarily fulfilled if the belief vectors that are associated with the different types of an agent are linearly independent. If there are at least as many constellations of other agents' types as there are possible types of agent i, linear independence of belief vectors holds for an open and dense set of specifications of beliefs in the Crémer-McLean model. For any given finite type space, therefore, the Crémer-McLean condition for full surplus extraction is generic.[1]

This paper establishes an analogous result for the McAfee-Reny condition for full surplus extraction in models with a continuum of types. Specifically, we will show that if the different agents' type spaces are compact metric spaces, then, for any agent i, the McAfee-Reny condition for full surplus extraction holds for a residual set of belief mappings of this agent, i.e., for a countable intersection of open and dense sets of functions mapping the agent's types to his beliefs about the other agents' types. With finitely many agents,

---

[1]Throughout the paper we use the word "model" in a wide sense, like Bergemann and Morris (2005) or Heifetz and Neeman (2006). In contrast, Barelli (2009), as well as Chen and Xiong (2011), Chen and Xiong (2013) use the word more specifically for a common prior model with a minimal consistent belief subspace of the universal type space.

the vectors of belief mappings of the different agents that all satisfy the McAfee-Reny condition form a residual set in the set of all vectors of belief mappings for the different agents. Moreover, this latter statement remains valid if attention is restricted to those vectors of belief mappings that are consistent with the existence of a common prior on the space of vectors of the different agents' types.

The McAfee-Reny condition requires that, for each agent i and each type t_i of this agent, the density function that indicates the agent's beliefs when his type is t_i cannot be represented as a convex combination of the beliefs that he has at types other than t_i. Formally, for any t_i and any measure μ on the agent's type space T_i, unless μ is the degenerate measure that assigns all mass to the singleton $\{t_i\}$, it must be the case that

$$\int f_i(\cdot|t_i')d\mu(t_i') \neq f_i(\cdot|t_i). \tag{1}$$

The right-hand side of (1) represents the density function $f_i(\cdot|t_i)$ of agent i's beliefs about the other agents' types when his own type is t_i. The left-hand side represents an average with respect to μ of the density functions $f_i(\cdot|t_i')$ of the beliefs that the agent has about the other agents' types when his own type is t_i'.

For an interpretation of this condition, observe that the average on the left-hand side corresponds to the beliefs that agent i would have about the other agents' types if he thought that his own type was distributed as μ. For example, if the agent believes that his type is $t_i' \neq t_i$, μ is the degenerate measure concentrated at t_i' and the left-hand side is just $f_i(\cdot|t_i')$. More generally, the measure μ in (2) might be nondegenerate, reflecting uncertainty of the agent about his own type. The McAfee-Reny condition requires that, in either case, if μ is a degenerate measure concentrated at some $t_i' \neq t_i$ and if μ is nondegenerate, the agent's beliefs when he thinks that his own type is distributed as μ must differ from the beliefs that the agent forms when he knows that his own type is t_i.

The notion of agent i's forming beliefs about the other agents' types on the basis that his own type is distributed as μ is at odds with the standard assumption that each agent knows his own type. This notion is natural though if we replace the type space T_i by the space $\mathcal{M}(T_i)$ of probability measures on T_i and if we think about an *extended type* of agent i as measure μ on T_i. If the measure μ is nondegenerate, the agent knows his extended type but has only probabilistic beliefs about his own type (and by implication, his payoff type and his belief type). In this formulation, the McAfee-Reny condition is equivalent to the requirement that the beliefs that are associated with any degenerate extended type, i.e. any measure on T_i that is concentrated at a singleton, must be distinguishable from

the beliefs that are associated with any other extended type. This requirement is stronger than injectiveness of the belief function on the original type space T_i but weaker than injectiveness of the belief function on the extended type space $\mathcal{M}(T_i)$; it does not exclude the possibility that the beliefs that are associated with any two nondegenerate extended types might be the same.[2]

The McAfee-Reny condition for full surplus extraction can thus be interpreted as a much stronger version of the requirement that agents' belief mappings should be *injective*. If an agent's belief function is injective, a mechanism designer who knows the agent's beliefs can infer his type. Moreover, regardless of how the agent's payoff parameters depend on his type, the mechanism designer can also infer the agent's payoff parameters from his beliefs. This latter property is sometimes referred to as the BDP property (short for *beliefs determine preferences*). Neeman (2004) and Heifetz and Neeman (2006) showed that this property is necessary for full surplus extraction, and they suggested that, with arbitrary non-finite type spaces, this property is only obtained in exceptional cases.[3] In their view, the genericity of full surplus extraction in Crémer and McLean (1988) is due to their assumption of fixed finite type spaces.

Our analysis contradicts this suggestion. In Gizatulina and Hellwig (2014), we had already shown that the BDP property is generic in the space of continuous functions from agents' types to their beliefs. Our proof of this result relied on the classical embedding theorem for continuous functions. This theorem asserts that, if X is a compact n-dimensional metric space and Y is a metric space with dimension at least $2n + 1$, then the set of embeddings, i.e., of continuous injective functions, is residual in the space of continuous functions from X to Y, endowed with the uniform topology. Thus, if the range of the functions is much richer than the domain, injectiveness is the rule and a failure of injectiveness the

---

[2]McAfee and Reny (1992) themselves give an interpretation of (1) that is based on the notion that μ can be thought of as a prior on T_i. In this case, a violation of (1) would indicate that, relative to the prior μ, agent i's learning that his own type is t_i does not provide him with any new information about the other agents' types. This explanation however is problematic if the measure μ does not have t_i in its support, for example, if μ is a degenerate measure that assigns all mass to $t_i' \neq t_i$. Moreover, the interpretation of μ as a prior raises the question of how to interpret the requirement that condition (1) hold for all measures μ on T_i except for the degenerate measure with unit mass at t_i.

[3]Their formal analysis showed that, if a family of common priors on a given type space is convex, and if this family contains at least one non-BDP prior (and hence at least one non-FSE prior), then, in a geometric and a measure theoretic sense, the set of BDP priors is sparse in the given convex family. Their analysis does not consider the genericity properties of the set convex families of common priors that contain at least one non-BDP prior in the set of all convex families of common priors on the given type space.

exception. Genericity of the BDP property follows directly because any injective belief function has this property, regardless of how payoff characteristics depend on types.

However, because the BDP property is only *necessary* and not *sufficient* for full surplus extraction, that result does not permit any inference about the genericity of the latter. As explained above, the McAfee-Reny condition, which is sufficient as well as necessary for full surplus extraction, is much stronger than the BDP property or injectivity. Therefore, we cannot rely on the classical embedding theorem for continuous functions. Indeed, if the range of an agent's belief functions were finite-dimensional, the set of belief functions satisfying the McAfee-Reny condition would not be dense no matter how large the dimension of the range might be relative to the dimension of the domain.

To establish the genericity of the McAfee-Reny condition, we actually need a new mathematical result, which makes use of the fact that beliefs are given by probability measures and that the space of probability measures on any set with more than finitely many elements is an infinite-dimensional space. The arguments we use involve similar ideas as the proof of the classical embedding theorem. However, whereas the classical embedding theorem involves functions whose range and domain have finitely many dimensions, our theorem involves functions whose range is infinite-dimensional (while the domain can be an arbitrary finite- or infinite-dimensional compact metric space).

In the context of belief mappings, the dimensionality requirement on the range is unproblematic. If each agent has more than finitely many types, beliefs, i.e. probability measures on type spaces, are necessarily infinite-dimensional. For example, in McAfee and Reny (1992), where each agent's type space is the unit interval, an agent's belief function maps the unit interval into the space of continuous density functions on the product of the other agents' type spaces, an infinite-dimensional space.

Genericity of full surplus extraction has also been studied in Chen and Xiong (2013). They also suggest that full surplus extraction is generically possible. However, their approach and their result are very different from ours. Most importantly, Chen and Xiong (2013) focus on beliefs that are derived from common priors and they treat the surplus extraction property as a property of these priors.[4] By contrast, we treat the McAfee-Reny condition as a property of each individual agent's belief function. We also have a genericity result for common priors under which all agents' belief functions satisfy the McAfee-Reny condition, but this result is derived from our genericity result for belief functions,

---

[4]More precisely, they treat agents' types as elements of the universal type space and consider belief-closed subsets of this space in which the different agents' belief functions admit common priors.

which does not require the existence of a common prior.

The McAfee-Reny condition itself is a condition on a belief function of each single agent which does not presume the existence of a common prior. For the analysis of Chen and Xiong (2013), however, the common-prior assumption is essential because they treat agents' types as elements of the universal type space. In the universal type space it does not even make sense to ask for which belief functions the McAfee-Reny condition is satisfied. In this space, there is single belief function, each "type" is defined by a vector consisting of the agent's payoff parameters and the agent's beliefs about the other agents' payoff parameters, the agent's beliefs about the other agents' beliefs about other agents' payoff parameters, etc. The belief function is given by the projection from the space of universal types to the space of belief hierarchies and by the observation that each belief hierarchy defines a unique measure on the space of the other agents' universal types. Because the universal type space approach has no room for considering different belief functions, Chen and Xiong (2013) do not look at the scope for full surplus extraction as a property of belief functions but as a property of belief-closed subsets of the universal type space and of common priors on such subsets. The fact that the McAfee-Reny condition for surplus extraction refers to each individual agent's belief function, without reference to mutually consistent belief hierarchies, is thereby lost.[5]

For models with continuous type spaces, Chen and Xiong (2013) do not actually consider the feasibility of surplus extraction, or the McAfee-Reny condition, directly. Instead they rely on approximations of such models by sequences of models with finite type spaces. For their denseness result, they observe that, if the space of common priors is given the topology of weak convergence, i.e., the weak* topology, then the set of common priors with finite supports is dense in the set of all common priors and that, as mentioned above, the set of priors for which full surplus extraction is possible is itself dense in the set of priors with finite supports. This argument is unsatisfactory because, as is well known, the weak* topology on priors is not well suited to capture the continuity properties of strategic behaviour under incomplete information.[6]

Any genericity result depends on the chosen topology. In a context involving strategic behaviour, the topology on the space of exogenous data should be chosen in such a way that, with this topology, the map from exogenous data to strategic behaviour exhibits nat-

---

[5]This comments also apply to the analysis of the BDP property in Chen and Xiong (2011). For an extensive discussion, see Gizatulina and Hellwig (2014).

[6]See Dekel, Fudenberg, and Morris (2006), Chen, DiTillio, Faingold, and Xiong (2010).

ural continuity properties. For type spaces and belief functions in incomplete-information models, this requirement is met if "similar" types and "similar" belief functions - in the sense of the given topology - induce similar behaviours in strategic games. Along the lines of Kajii and Morris (1994) or Dekel, Fudenberg, and Morris (2006), we interpret "similar behaviours" in terms of what Kajii and Morris (1994) refer to as *Approximate Lower Hemicontinuity of Actions*: In any strategic game, for any $\varepsilon > 0$ and any type of an agent, an action that is optimal for the agent with this type under some belief function, when the other agents' strategies are taken as given, will also be ε-optimal under other belief function that is sufficiently close. We obtain this property by modelling beliefs as measures with continuous density functions and specifying belief functions as continuous functions from types to beliefs, where the space of beliefs is given the topology of uniform convergence of density functions and the space of belief functions is given the topology of uniform convergence.

The assumption that beliefs have continuous density functions and that the topology on beliefs is given by the uniform topology on density functions is not actually necessary for our analysis. To be sure, McAfee and Reny (1992) assumed that beliefs have continuous density functions. However, if beliefs are specified as arbitrary measures and belief functions as continuous functions from types to measures, a suitable generalization of the McAfee-Reny condition can still be shown to be necessary and sufficient for full surplus extraction. Moreover, our genericity result can be extended to this setting. However, we suspect that, in this more general setting, *Approximate Lower Hemicontinuity of Actions* would fail.

Our analysis should not be interpreted as saying that full surplus extraction is to be considered very likely. One may have serious doubts whether, in any concrete situation, a mechanism designer may be presumed to know enough to implement the mechanism that is needed to fully exploit the dependence of beliefs on types for surplus extraction. However, our result suggests that such doubts ought to be articulated by explicitly modelling the mechanism designer's lack of information about the environment or by imposing a requirement of robustness along the lines of Bergemann and Morris (2005). The notion that "beliefs determine preferences" may seem implausible, and one may be tempted to formalize this implausibility at the level of incomplete-information models, but our analysis shows that in settings with infinite type spaces, and therefore infinite-dimensional belief spaces, such notions of plausibility must be treated with caution. In such settings, neither the McAfee-Reny condition for full surplus extraction nor the weaker condition of

injectiveness, is very special.

In the following, Section 2 lays out the basic model and reviews the result of McAfee and Reny (1992). Section 3 states our main theorem on the genericity of belief functions satisfying the McAfee-Reny condition and compares it to the result of Gizatulina and Hellwig (2014) on the genericity of the BDP property. In particular, this section provides some geometric intuition for the similarities and differences between the McAfee-Reny condition and injectiveness and for the difference between our theorem and the classical embedding theorem for continuous functions. The proof of our main theorem is given in Section 4. In Section 5, we use our main theorem to provide a genericity result for common priors under which the belief functions, i.e., the regular conditional distributions of all agents, satisfy the McAfee-Reny condition. In the concluding discussion in Section 6, we relate our analysis to the work of Heifetz and Neeman (2006), Gizatulina and Hellwig (2014) and Chen and Xiong (2013). In that section, we also discuss the role of the topology we use, as well as the relation between the abstract type space approach, which we use, and the universal type space approach. A generalization of our analysis to the case where belief functions do not have continuous densities is discussed in the appendix.

2 The McAfee-Reny Result

McAfee and Reny (1992) consider the following problem: Suppose that a game of incomplete information between agents $i = 1, ..., I$ had an equilibrium in dominant strategies and that the equilibrium payoff of agent i is

$$\pi_i(t_1, ..., t_I), \tag{2}$$

where $t_1, ..., t_I$ are the different agents' types. Is it possible to design incentive-compatible systems of participation fees that extract this surplus from the agents? For example, we may think of the incomplete-information game as a second-price auction for a given object and of the payoff $\pi_i(t_1, ..., t_I)$ as the surplus that agent i obtains if he has the highest valuation for the object and pays a price equal to the second-highest valuation, i.e.,

$$\pi_i(t_1, ..., t_I) = \theta_i(t_i) - \max_{j \neq i} \theta_j(t_j)$$

if $\theta_i(t_i) \geq \max_{j \neq i} \theta_j(t_j)$ and

$$\pi_i(t_1, ..., t_I) = 0$$

8

if $\theta_i(t_i) \leq \max_{j \neq i} \theta_j(t_j)$, where, for $j = 1, ..., I$, $\theta_j(t_j)$ is the value that agent j with type t_j assigns to the object. Instead of a second-price auction, one might also have the message game that is induced by a Groves mechanism for public-good provision, or any other game with a dominant-strategy equilibrium.

For $i = 1, ..., I$, let T_i be the set of possible types of agent i, and let $T_{-i} := \prod_{j \neq i} T_j$ be the set of vectors of the other agents' types. The type t_i of agent i determines not only the agent's payoff characteristics but also his beliefs $b_i(t_i)$ about the other agents' types; formally, $b_i(t_i)$ is a probability measure on T_{-i}.

Given the belief $b_i(t_i)$ and given the equilibrium function π_i determined by (2), the agent's expected payoff from participating in the game is given as

$$\bar{\pi}_i(t_i) := \int_{T_{-i}} \pi_i(t_1, ..., t_I) \, b_i(dt_{-i}|t_i). \tag{3}$$

The question is whether there exists an incentive-compatible system of participation fees that make the agent pay (approximately) $\bar{\pi}_i(t_i)$, for each $t_i \in T_i$, to the mechanism designer/organizer of the game.

McAfee and Reny (1992) consider a system of participation fees with the following structure. Knowing his type t_i, agent i can choose one out of N_i schedules $z_1^i, ..., z_{N_i}^i$ where each schedule z_n^i, $n = 1, ... N_i$ makes the actual fee depend on the other agents' types. Thus, if agent i chooses the schedule z_n^i, his payment will be $z_n^i(t_{-i})$. Given his belief $b_i(t_i)$, his expected payment under the fee schedule z_n is

$$\bar{z}_n^i(t_i) := \int_{T_{-i}} z_n^i(t_{-i}) \, b_i(dt_{-i}|t_i). \tag{4}$$

The dependence of agent i' payment on the other agents' types obviously does not give rise to any incentive concerns. If agent i can choose one out of the N_i schedules $z_1^i, ..., z_{N_i}^i$, agent i will choose the one that entails the smallest expected payment $\bar{z}_n(t_i)$. Given this choice, his expected payment is equal to

$$\bar{Z}_i(t_i) := \min(\bar{z}_1^i(t_i), ..., \bar{z}_{N_i}^i(t_i)). \tag{5}$$

Let $b_i : T_i \to \mathcal{M}(T_{-i})$ be a belief function of agent i where $\mathcal{M}(T_{-i})$ is the space of probability measures over T_{-i}. The belief function b_i is said to *admit full surplus extraction* if and only if, for every continuous function $\bar{\pi}_i : T_i \to \mathbb{R}_+$ and every $\varepsilon > 0$, there exists a system $z_1^i, ..., z_{N_i}^i$ of participation fee schedules for agent i such that the induced expected

payment $\bar{Z}_i(t_i)$ as given by (4) and (5) satisfies

$$\bar{\pi}_i(t_i) - \varepsilon \leq \bar{Z}_i(t_i) \leq \bar{\pi}_i(t_i) \tag{6}$$

for all $t_i \in [0,1]$. Whereas the ability to extract the surplus of agent i depends on the belief function b_i and the expected-payoff function $\bar{\pi}_i$ jointly, McAfee and Reny (1992) ask which belief functions admit full surplus extraction for all expected-payoff function $\bar{\pi}_i$ within the class of continuous functions. For brevity, belief functions admitting full surplus extraction will sometimes be referred to as FSE belief functions.

Condition (6) provides for approximate rather than full surplus extraction. As explained by McAfee and Reny, exact surplus extraction is not to be expected. For example, if the belief function b_i has a continuous density, one can find an expected-payoff function $\bar{\pi}_i$ for which exact surplus extraction is not possible, i.e. there is no system of fee schedules $z_1^i, ..., z_{N_i}^i$ such that $\min(\bar{z}_1^i(t_i), ..., \bar{z}_{N_i}^i(t_i)) = \bar{\pi}_i(t_i)$ for all t_i. However, since the choice of ε in (6) is arbitrary, the divergence from full surplus extractions can be made arbitrarily small.

In their formal analysis, McAfee and Reny (1992) equate the type spaces $T_i, i = 1, ..., I$, with the unit interval and restrict the analysis to belief functions b_i such that, for each $t_i \in [0,1]$, the measure $b_i(t_i)$ has a density function $f_{b_i}(\cdot|t_i)$ and, moreover, $f_{b_i}(t_{-i}, t_i)$ is jointly continuous in t_{-i} and t_i: For this specification, they establish

THEOREM 2.1 (MCAFEE AND RENY) *A belief function $b_i : T_i \rightarrow \mathcal{M}(T_{-i})$ with a density function $f_{b_i} : T_i \times T_{-i} \rightarrow \mathbb{R}_+$ that is jointly continuous in t_i and t_{-i} admits full surplus extraction if and only if, for every $\bar{t}_i \in T_i$ and every probability measure μ on T_i,*

$$\mu(\{\bar{t}_i\}) \neq 1 \ \text{implies} \ f_{b_i}(\cdot|\bar{t}_i) \neq \int_{T_i} f_{b_i}(\cdot|t_i)\mu(dt_i). \tag{*}$$

In other words, there must not exist any type $\bar{t}_i$ such that the density function $f_{b_i}(.|\bar{t}_i)$ associated with the belief $b_i(\bar{t}_i)$ belongs to the convex hull of the density functions $f_{b_i}(.|t_i)$ associated with the beliefs $b_i(t_i)$, $t_i \in [0,1]\backslash\{\bar{t}_i\}$. This is the condition we discussed in the introduction. To interpret it, McAfee and Reny (1992) point to the fact that, if the map $t_i \rightarrow f_{b_i}(\cdot|t_i)$ is a conditional density induced by some prior on $T_i \times T_{-i}$, then $\int_{T_i} f_{b_i}(.|t_i)\mu(dt_i)$ is a conditional density for agent i's beliefs about the other agents' types following the observation of a possibly noisy signal that induces μ as the agent's conditional distribution of his own type. Condition (*) requires that there must not exist any signal about the

agent's own type such that, given the observation of this signal, his conditional beliefs about the other agents' types are the same as when he knows that his type is $\bar{t}_i$.

3 Genericity of FSE Belief Functions

As noted by McAfee and Reny (1992), Theorem 2.1 can be extended to the case where the type spaces are compact metric spaces. We therefore assume that, for $i = 1, ..., I$ the type set T_i is a compact metric space. For any $t_i \in T_i$, the belief $b_i(t_i)$ is an element of the space $\mathcal{M}(T_{-i})$ of probability measures on T_{-i}. In fact, we impose the stronger assumption that there exists some arbitrary, but fixed measure λ on T_{-i}, such that, for any $t_i \in T_i$, $b_i(t_i)$ belongs to the subset $\mathcal{M}^\lambda(T_{-i})$ of those measures in $\mathcal{M}(T_{-i})$ that have continuous densities with respect to λ. An example would be the case where, for some natural number n, each of the sets T_i is a compact subset of $\mathbb{R}^n$ and, for each $t_i \in T_i$, the measure $b_i(t_i)$ that is induced by the belief function b_i has a continuous density function $f_{b_i}(.|t_i)$ with respect to Lebesgue measure on $\mathbb{R}^{(I-1)n}$. However, we are not assuming that type spaces are finite-dimensional.[7]

In this setting, the assumption that $f_{b_i}(t_{-i}|t_i)$ be continuous in t_{-i} and t_i translates into the requirement that the belief function b_i map the type space T_i continuously into the space $\mathcal{M}^\lambda(T_{-i})$, where $\mathcal{M}^\lambda(T_{-i})$ has the topology that is induced by the uniform topology on the space of density functions. In other words, b_i belongs to the space $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ of continuous functions from T_i to $\mathcal{M}^\lambda(T_{-i})$.[8]

---

[7]The assumption that $b_i(t_i)$ belongs to $\mathcal{M}^\lambda(T_{-i})$ is not actually necessary for our analysis. In the Appendix, we show that the result of McAfee and Reny (1992), i.e., Theorem 2.1, as well as our own result, Theorem 3.1, can be extended to the case where belief functions are simply continuous functions from T_i to the space $\mathcal{M}(T_{-i})$ of probability measures on T_{-i}, when this latter space has the topology of weak convergence, i.e. the weak* topology. In the text, we focus on the case where beliefs functions take values in $\mathcal{M}^\lambda(T_{-i})$ because, as we discuss in Section 6, this specification automatically generates natural continuity properties of strategic behaviour.

[8]The joint continuity assumption on $f_{b_i}(t_{-i}|t_i)$ is equivalent to the assumption that $b_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$. To see this, suppose first that $b_i \notin \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$. Then there exists $\varepsilon > 0$, and there exist sequences $\{t_{-i}^k\}, \{t_i^k\}$ such that $\{t_i^k\}$ converges to a limit t_i^∞ and $|f_{b_i}(t_{-i}^k|t_i^k) - f_{b_i}(t_{-i}^k|t_i^\infty)| \geq \varepsilon$ for all k. Since T_{-i} is compact, we may assume without loss of generality that $\{t_{-i}^k\}$ converges to a limit t_{-i}^∞. By the triangle inequality it follows that either $|f_{b_i}(t_{-i}^k|t_i^k) - f_{b_i}(t_{-i}^\infty|t_i^\infty)| \geq \frac{\varepsilon}{2}$ for all k, or $|f_{b_i}(t_{-i}^\infty|t_i^\infty) - f_{b_i}(t_{-i}^k|t_i^\infty)| \geq \frac{\varepsilon}{2}$ for all k. In either case, continuity of f_{b_i} is violated. Hence, if f_{b_i} is continuous, $b_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$. Conversely, if f_{b_i} is not continuous, there exists $\varepsilon > 0$ and there exist sequences $\{t_{-i}^k\}, \{t_i^k\}$ converging to $t_{-i}^\infty, t_i^\infty$

If the set T_{-i} is finite, but T_{-i} has at least as many elements as T_i, the analysis of Crémer and McLean (1988) implies that the set $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ is actually an open and dense subset of $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$.[9] [10] With infinite type spaces, we only obtain the weaker property of residualness and denseness. Formally, we obtain:

THEOREM 3.1 *Let $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ be the set of continuous functions from T_i to $\mathcal{M}^\lambda(T_{-i})$ that satisfy condition (*). If T_{-i} has more than finitely many elements, then $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ is a residual and dense subset of the space $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ with the uniform topology, i.e., $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ contains a countable intersection of open and dense subsets of $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ and is itself dense.*

Theorem 3.1 is the main result of this paper. We prove this result in the next section. Before going there, we explain the relation between Theorem 3.1 and the result in Gizatulina and Hellwig (2014) on the genericity of injective belief function. For this purpose, we discuss the relation between condition (*) and injectiveness and we give some geometric intuition for the genericity of these properties, showing in particular, why the classical embedding theorem, on which we relied in Gizatulina and Hellwig (2014), cannot be used to prove Theorem 3.1.

We begin our discussion with a reformulation of condition (*). With any $b_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$, we can associate a map φ_{b_i} from the space $\mathcal{M}(T_i)$ of probability measures on T_i into

---

such that $|f_{b_i}(t^k_{-i}|t^k_i) - f_{b_i}(t^\infty_{-i}|t^\infty_i)| \geq \varepsilon$ for all k. Then either, $|f_{b_i}(t^k_{-i}|t^k_i) - f_{b_i}(t^k_{-i}|t^\infty_i)| \geq \frac{\varepsilon}{2}$ for all k or $|f_{b_i}(t^k_{-i}|t^\infty_i) - f_{b_i}(t^\infty_{-i}|t^\infty_i)| \geq \frac{\varepsilon}{2}$ for all k. The second inequality contradicts the assumption that the measure $b_i(t^\infty_i)$ has a continuous density, i.e. that $b_i(t^\infty_i) \in \mathcal{M}^\lambda(T_{-i})$. The first inequality contradicts the assumption that the map $t_i \to f_{b_i}(\cdot|t_i)$ is continuous when the range of this map has the uniform topology. In either case, it follows that $b_i \notin \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$.

[9]**Proof.** If T_i is a finite set, with n_i elements, any measure $\mu \in \mathcal{M}(T_i)$ is represented by an n_i-dimensional vector $(\mu_1, ..., \mu_{n_i})$; the density of the measure $\varphi_{b_i}(\cdot|\mu)$ takes the form

$$f_{b_i}(t_{-i}|\mu) = \sum_{i' \in T_i} f_{b_i}(t_{-i}|t'_i)\mu_{t'_i}.$$

Condition (**) takes the form

$$f_{b_i}(t_{-i}|t_i) = \sum_{i' \in T_i} f_{b_i}(t_{-i}|t'_i)\mu_{t'_i} \quad \text{implies} \quad \mu_{t_i} = 1 \text{ and } \mu_{t'_i} = 0 \text{ for } t'_i \neq t_i.$$

This condition is satisfied whenever the vectors $(f_{b_i}(t_{-i}|t_i))_{t_{-i} \in T_{-i}}$, $t_i \in T_i$, are linearly independent. If the cardinality N_{-i} of T_{-i} is no less than that of T_i, the result follows because, with $N_{-i} \geq n_i$, generic n-tuples of N_{-i}-vectors are linearly independent. ∎

[10]For the case of finite type sets, Crémer and McLean also showed that full surplus extraction through a dominant-strategy mechanism can be achieved if and only if the matrix of posterior beliefs of all types of each agent has the rank n_i where n_i is the cardinality of the type space of agent i.

$\mathcal{M}^\lambda(T_{-i})$ by setting

$$
\begin{aligned}
\varphi_{b_i}(B|\mu) \quad : \quad &= \int_{T_i} b_i(B|t_i)d\mu(t_i) = \int_{T_i} \int_B f_{b_i}(t_{-i}|t_i)dt_{-i}d\mu(t_i) \\
&= \int_B \int_{T_i} f_{b_i}(t_{-i}|t_i)d\mu(t_i)dt_{-i}
\end{aligned}
\tag{7}
$$

for any $\mu \in \mathcal{M}(T_i)$ and any measurable $B \subset T_{-i}$. By inspection of (7), for any $\mu \in \mathcal{M}(T_i)$, the measure $\varphi_{b_i}(\cdot|\mu)$ has the density function

$$
f_{b_i}(\cdot|\mu) := \int_{T_i} f_{b_i}(\cdot|t_i)d\mu(t_i).
\tag{8}
$$

With this notation, condition (*) can be reformulated as

$$
f_{b_i}(\cdot|\mu) = f_{b_i}(\cdot|\bar{t}_i) \text{ implies } \mu = \delta_{\bar{t}_i},
$$

or

$$
\varphi_{b_i}(\mu) = \varphi_{b_i}(\delta_{\bar{t}_i}) \text{ implies } \mu = \delta_{\bar{t}_i},
\tag{**}
$$

where $\delta_{\bar{t}_i}$ is the degenerate measure which assigns unit mass to the singleton $\bar{t}_i$.

If the measure μ itself is degenerate, i.e., if $\mu = \delta_{t_i}$ for some $t_i \in T_i$, condition (**) specializes to the requirement that

$$
\varphi_{b_i}(\delta_{t_i}) = \varphi_{b_i}(\delta_{\bar{t}_i}) \text{ implies } \delta_{t_i} = \delta_{\bar{t}_i},
$$

or

$$
f_{b_i}(\cdot|t_i) = f_{b_i}(\cdot|\bar{t}_i) \text{ implies } t_i = \bar{t}_i,
\tag{9}
$$

which means that the belief mapping b_i is *injective*.

The set of continuous injective functions, i.e. the set of embeddings, from T_i to $\mathcal{M}^\lambda(T_{-i})$ is thus a subset of the set $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ of continuous functions from T_i to $\mathcal{M}^\lambda(T_{-i})$ that satisfy condition (*). From Theorem 3.1, therefore, one immediately obtains:

COROLLARY 3.2 *Let $\mathcal{E}(T_i, \mathcal{M}^\lambda(T_{-i}))$ be the set of embeddings from T_i to $\mathcal{M}^\lambda(T_{-i})$. If T_{-i} has more than finitely many elements, then $\mathcal{E}(T_i, \mathcal{M}^\lambda(T_{-i}))$ is a residual and dense subset of the space $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ with the uniform topology, i.e., $\mathcal{E}(T_i, \mathcal{M}^\lambda(T_{-i}))$ contains a countable intersection of open and dense subsets of $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ and is itself dense.*

Corollary 3.2 is closely related to the condition that Neeman (2004) and Heifetz and Neeman (2006) refer to as the BDP property - *"beliefs determine preferences"*. If the belief mapping is injective, no type $\bar{t}_i \neq t_i$ has the belief $b_i(t_i)$, so a mechanism designer who knows that agent i has the belief $b_i(t_i)$ can infer that the agent's type is t_i, rather than some $\bar{t}_i \neq t_i$. As a result, he can also infer the agent's payoff characteristics - his preferences - no matter how these characteristics may depend on the agent's type. As discussed by Neeman (2004) and Heifetz and Neeman (2006), the ability to infer the agent's preferences from his beliefs, the BDP property, is *necessary* for full surplus extraction.

However, neither the BDP property nor the injectiveness of an agent's belief mapping are sufficient for full surplus extraction. A necessary and sufficient condition is given by McAfee and Reny's condition (*) or our equivalent reformulation (**). From (**) and (9), it is obvious that (**) is much stronger than merely injectiveness or the BDP property.

In Gizatulina and Hellwig (2014), we proved a weaker version of Corollary 3.2 in order to show that belief functions exhibiting the BDP property are generic. The proof made use of the classical embedding theorem for continuous functions. For *finite-dimensional* compact metric spaces X and Y, this theorem asserts that, if the dimension of Y is strictly greater than twice the dimension of X, then the set of embeddings, i.e. of continuous injective functions, is a residual subset of the set of continuous functions from X to Y when the space of these functions has the topology of uniform convergence.[11] Relying on this theorem, the result in Gizatulina and Hellwig (2014) shows that, if the type sets $T_1, ..., T_I$ are finite-dimensional compact metric spaces, the set $\mathcal{E}(T_i, \mathcal{M}^\lambda(T_{-i}))$ of embeddings of T_i into $\mathcal{M}^\lambda(T_{-i})$ is a residual subset of the set $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ of continuous functions from T_i into $\mathcal{M}^\lambda(T_{-i})$ where $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ is endowed with the uniform topology. Because $\mathcal{M}^\lambda(T_{-i})$ is an infinite-dimensional space, the dimensionality condition of the embedding theorem is automatically fulfilled whenever the dimension of T_i is finite.

By contrast, the McAfee-Reny condition for full surplus extraction is much stronger than injectivity. Therefore, the proof of Theorem 3.1 cannot just cite the embedding theorem but needs a new argument. The arguments involve similar ideas as the proof of the embedding theorem, but the analysis makes essential use of the fact that the space $\mathcal{M}^\lambda(T_{-i})$ is infinite-dimensional. Even if the domain T_i of the belief functions of agent i is a finite-dimensional set, the conclusion of the theorem can only be obtained if the range is infinite-dimensional.

We illustrate the distinction between injectivity and (**) by means of some diagrams.

---

[11]See, e.g., Hurewicz and Wallman (1941)

(a) Image of $[0,1]$ in $\mathbb{R}^2$ (b) Image of $[0,1]$ in $\mathbb{R}^3$
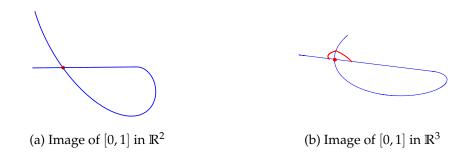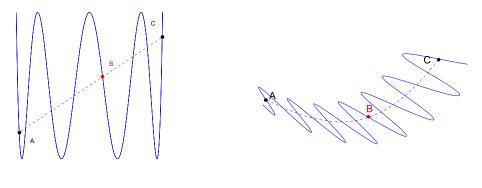
Figure 1: Denseness of injectivity

The functions in these diagrams should be interpreted as abstract examples for the two properties, without reference to any belief functions. Each diagram shows the image of some function that has domain $[0,1]$ and either a two- or a three-dimensional range. Because the unit interval has dimension one, the embedding theorem asserts that any continuous function from the unit interval to $\mathbb{R}^3$ can be approximated by a sequence of continuous *injective* functions from the unit interval to $\mathbb{R}^3$. Thus, the function which generates the solid blue loop in the plane in Figure 1 is *not* injective because the point of self-intersection corresponds to the value of the function under two distinct points in its domain. However, as indicated by the red "detour", the third dimension provides ample room for perturbations of this function that are injective.

(a) Image of $[0,1]$ in $\mathbb{R}^2$ (b) Image of $[0,1]$ in $\mathbb{R}^3$

Figure 2: Denseness of (**) ?

In Figure 2, the solid blue line in the left half of the figure is the image of a sinusoid function taking values in a plane. The sinusoid is injective, but fails to satisfy condition

15

(**) because its value at point B is a convex combination of points A and C on the red dashed line (and of course any other point on the crossing of the red line and blue line is a convex combination of "extreme" points A and C). The right half of the figure shows that, in this case, too, an additional dimension can provide room for perturbations of the sinusoid that satisfy condition (**). The "curved" sinusoid in part (b) of Figure 2 can be seen as an element of a sequence of functions approximating the "flat" sinusoid" in part (a). Each element of the sequence of "curved" sinusoids satisfies condition (**) even though the "flat" sinusoid does not.

However, there is a difference. In Figure 1, approximation of the non-injective function by injective functions is obtained by *local* perturbations. In Figure 2, the approximation by functions satisfying condition (**) is based on *global* perturbations. Condition (**) hinges on comparing the value of $\varphi_{b_i}(\delta_{\bar{t}_i})$ with the value of $\varphi_{b_i}(\mu)$ for any μ, a comparison which depends on types $\hat{t}_i$ with $\varphi_{b_i}(\delta_{\hat{t}_i})$ far away from $\varphi_{b_i}(\delta_{\bar{t}_i})$.

For injectivity, the local nature of the required perturbations explains why embeddings are dense if the dimension of the range is greater than twice the dimension of the domain. If the domain is the unit interval, as was assumed for Figure 1, then at any point of self-intersection, the range behaves locally like a plane, which leaves another dimension to provide room for perturbations that are injections.

For condition (**), this kind of reasoning is unavailable. Figure 3 gives an example with a function on $[0,1]$ whose image is the union of a vertical line and an expanding spiral around the vertical line so that the vertical line belongs to the convex hull of the
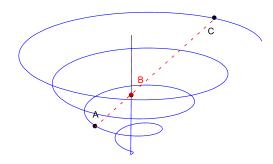

Figure 3: Image of $[0,1]$ in $\mathbb{R}^3$: Failure of denseness of (**)

16

expanding spiral, which is a non-negligible subset of $\mathbb{R}^3$. In this case there is no sequence of functions from $[0,1]$ to $\mathbb{R}^3$ that satisfy condition (**) and approximate the given function – for any such function, the part that approximates the vertical line inside the spiral must belong to the convex hull of the part that approximates the spiral.

In Appendix A, we show more generally that, for any n, there is a function from $[0,1]$ to $\mathbb{R}^n$ that violates condition (**) and that cannot be approximated by functions from $[0,1]$ to $\mathbb{R}^n$ that satisfy condition (**). If the convex hull of the range of a function from $[0,1]$ to $\mathbb{R}^n$ is a non-negligible subset of $\mathbb{R}^n$, one needs yet another dimension (at least) for the perturbations that satisfy condition (**) and approximate the given function. This is why the proof of Theorem 3.1 makes essential use of the fact that $M^\lambda(T_{-i})$ is infinite-dimensional. By contrast, in the argument of Gizatulina and Hellwig (2014), this fact matters only because, with an infinite-dimensional range, the dimensionality condition of the embedding theorem is satisfied no matter what the (finite) dimension of the domain may be.

4   Proof of the Main Theorem

The argument is closely related to that given by Hurewicz and Wallman (1941) for the classical embedding theorem for continuous functions. Let d be the metric on T_i. Given this metric, let $\mathcal{M}(T_i)$, the space of probability measures on T_i, be endowed with the associated Prohorov metric; this metric induces the topology of weak convergence, or weak* topology, on $\mathcal{M}(T_i)$. In an abuse of notation, we write $d(\cdot,\cdot)$ for this metric as well. We also endow $\mathcal{M}^\lambda(T_{-i})$ with the metric ρ such that, for any β and $\hat{\beta}$ in $\mathcal{M}^\lambda(T_{-i})$,

$$\rho(\beta,\hat{\beta}) = \max_{t_{-i} \in T_{-i}} |f_\beta(t_{-i}) - f_{\hat{\beta}}(t_{-i})|; \tag{10}$$

the maximum in (10) is well defined for all β and $\hat{\beta}$ because T_{-i} is compact and the density functions f_β and $f_{\hat{\beta}}$ that are associated with β and $\hat{\beta}$ in $\mathcal{M}^\lambda(T_{-i})$ are continuous.

Proof. For any $\varepsilon > 0$, we now define G_ε as the subset of $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ that contains all mappings b_i with the property that, for any $t_i \in T_i$ and all $\mu \in \mathcal{M}(T_i)$,

$$\varphi_{b_i}(\delta_{t_i}) = \varphi_{b_i}(\mu) \tag{11}$$

implies

$$d(\mu, \delta_{t_i}) < \varepsilon \tag{12}$$

17

where $d(\mu, \delta_{t_i})$ is the Prohorov distance between the measures μ and δ_{t_i}.

Claim 1: $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i})) = \cap_{k=1}^\infty G_{1/k}$.

Proof: If $b_i \in G_{1/k}$ for all k, then for any $t_i \in T_i$ and all $\mu \in \mathcal{M}(T_i)$, $\varphi_{b_i}(\delta_{t_i}) = \varphi_{b_i}(\mu)$ implies $0 \leq d(\mu, \delta_{t_i}) \leq 1/k$, hence $d(\mu, \delta_{t_i}) = 0$. Conversely, if $b_i \in \mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$, then, by (**), $\varphi_{b_i}(\delta_{t_i}) = \varphi_{b_i}(\mu)$ implies $d(\mu, \delta_{t_i}) = 0 < 1/k$ for all $t_i \in T_i$, all $\mu \in \mathcal{M}(T_i)$, and all k.

Claim 2: For any $\varepsilon > 0$, the set G_ε is an open subset of $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$.

Proof: Fix any $b_i \in G_\varepsilon$, and let

$$\Gamma := \{(t_i, \mu) \in T_i \times \mathcal{M}(T_i) | d(\mu, \delta_{t_i}) \geq \varepsilon\}$$

and let

$$\eta = \min_{(t_i, \mu) \in \Gamma} \rho(\varphi_{b_i}(\delta_{t_i}), \varphi_{b_i}(\mu)). \tag{13}$$

Because T_i and therefore also $\mathcal{M}(T_i)$ are compact, the minimum in (13) is well defined, i.e., there exists $(t_i, \mu) \in \Gamma$ such that $\rho(\varphi_{b_i}(\delta_{t_i}), \varphi_{b_i}(\mu)) = \eta$. Since $b_i \in G_\varepsilon$ and, for $(t_i, \mu) \in \Gamma$, $d(\mu, \delta_{t_i}) \geq \varepsilon$, it follows that $\eta > 0$.

Now let $\bar{b}_i$ be any mapping in $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ such that

$$\rho(\bar{b}_i(t_i), b_i(t_i)) < \frac{\eta}{2} \tag{14}$$

for all $t_i \in T_i$. Since

$$
\begin{aligned}
\rho(\varphi_{\bar{b}_i}(\mu), \varphi_{b_i}(\mu)) &= \max_{t_{-i} \in T_{-i}} |f_{\bar{b}_i}(t_{-i}|\mu) - f_{b_i}(t_{-i}|\mu)| \\
&= \max_{t_{-i} \in T_{-i}} \left| \int_{T_i} f_{\bar{b}_i}(t_{-i}|t_i) d\mu(t_i) - \int_{T_i} f_{b_i}(t_{-i}|t_i) d\mu(t_i) \right| \\
&\leq \int_{T_i} \max_{t_{-i} \in T_{-i}} \left| f_{\bar{b}_i}(t_{-i}|t_i) - f_{b_i}(t_{-i}|t_i) \right| d\mu(t_i) \\
&= \int_{T_i} \rho(\bar{b}_i(t_i), b_i(t_i)) d\mu(t_i),
\end{aligned}
$$

by (10) and (8), we infer that

$$\rho(\varphi_{\bar{b}_i}(\mu), \varphi_{b_i}(\mu)) < \frac{\eta}{2} \tag{15}$$

for all $\mu \in \mathcal{M}(T_i)$. Thus, if t_i and μ are such that $\varphi_{\bar{b}_i}(\delta_{t_i}) = \varphi_{\bar{b}_i}(\mu)$, it must be the case that $\rho(\varphi_{b_i}(\delta_{t_i}), \varphi_{b_i}(\mu)) < \eta$. By the definition of η, it follows that $d(\mu, \delta_{t_i}) < \varepsilon$, which proves that $\bar{b}_i$ is also an element of G_ε. Thus, along with b_i, any element of the open $\frac{\eta}{2}$–ball around b_i is an element of G_ε. The claim follows immediately.

18

Claim 3: For any $\varepsilon > 0$, the set G_ε is a dense subset of $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$.

Proof: Fix any $b_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ and any $\eta > 0$. We will show that there exists a function $\bar{b}_i \in G_\varepsilon$ such that

$$\rho(\bar{b}_i(t_i), b_i(t_i)) < \eta \tag{16}$$

for all $t_i \in T_i$.

Relying on the fact that the continuous function b_i is uniformly continuous on the compact set T_i, let $\zeta \in (0, \varepsilon)$ be such that, for any t_i and $\bar{t}_i$ in T_i, $d(\delta_{t_i}, \delta_{\bar{t}_i}) < \zeta$ implies $\rho(b_i(t_i), b_i(t'_i)) < \frac{\eta}{2}$. Because T_i is a metric space, there exists a covering $U_1, U_2, \ldots$ of T_i such that

$$d(U_k) < \zeta \tag{17}$$

and

$$\rho(b_i(U_k)) < \frac{\eta}{2} \tag{18}$$

for all k. where $d(U_k)$ and $\rho(b_i(U_k))$ are defined as the suprema of $d(t_i, \bar{t}_i)$ and of $\rho(b_i(t_i), b_i(\bar{t}_i))$, respectively, over $t_i, \bar{t}_i$ in U_{ki}.

Because T_i is compact, the covering $U_1, U_2, \ldots$ of T_i has a finite subcovering, which we denote as $U_1, U_2, ..U_K$. Given the sets $U_1, U_2, ..U_K$, we may select measures $\beta_1, \ldots, \beta_K$ in $\mathcal{M}^\lambda(T_{-i})$ such that, for $k = 1, \ldots, K$, $\rho(\beta_k, b_i(\bar{t}_i)) < \frac{\eta}{2}$ for some $\bar{t}_i \in U_k$ and, moreover, the associated density functions $f_{\beta_1}, \ldots, f_{\beta_K}$ are linearly independent; linear independence of the density functions $f_{\beta_1}, \ldots, f_{\beta_K}$ may be assumed because T_{-i} has infinitely many elements.

For each $t_i \in T_i$ and $k = 1, 2, \ldots, K$, set

$$w_k(t_i) = \min_{\bar{t}'_i \in T_i \setminus U_k} d(t_i, \bar{t}_i)$$

and

$$\alpha_k(t_i) = \frac{w_k(t_i)}{\sum_{\ell=1}^K w_\ell(t_i)}.$$

This is well defined because for each $t_i \in T_i$, there exists at least one ℓ such that $t_i \in U_\ell$ and therefore $w_\ell(t_i) > 0$ and $\sum_{\ell=1}^K w_\ell(t_i) > 0$.

We now define $\bar{b}_i : T_i \to \mathcal{M}^\lambda(T_{-i})$ by setting

$$\bar{b}_i(t_i) = \sum_{k=1}^K \alpha_k(t_i)\beta_k. \tag{19}$$

19

Because the functions w_k and α_k are obviously continuous, $\bar{b}_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$. Moreover, for any $t_i \in T_i$ and $k = 1, ..., K$, $\alpha_k(t_i) > 0$ implies $t_i \in U_k$ and therefore, $\rho(\beta_k, b_i(t_i)) < \frac{\eta}{2}$. Therefore

$$
\begin{aligned}
\rho(\bar{b}_i(t_i), b_i(t_i)) &= \max_{t_{-i} \in T_{-i}} \left| f_{\bar{b}_i}(t_{-i}|t_i) - f_{b_i}(t_{-i}|t_i) \right| \\
&= \max_{t_{-i} \in T_{-i}} \left| \sum_{k=1}^{K} \alpha_k(t_i) f_{\beta_k}(t_{-i}) - f_{b_i}(t_{-i}|t_i) \right| \\
&\leq \sum_{k=1}^{K} \alpha_k(t_i) \max_{t_{-i} \in T_{-i}} \left| f_{\beta_k}(t_{-i}) - f_{b_i}(t_{-i}|t_i) \right| \\
&= \sum_{k=1}^{K} \alpha_k(t_i) \rho(\beta_k, b_i(t_i)) \\
&< \sum_{k=1}^{K} \alpha_k(t_i) \left[\inf_{\bar{t}_i \in U_k} \rho(\beta_k, b_i(\bar{t}_i)) + \frac{\eta}{2} \right] \\
&< \eta,
\end{aligned}
\tag{20}
$$

which establishes (16). In the derivation for (20), the last inequality follows because β_k had been chosen so that $\rho(\beta_k, b_i(\bar{t}_i)) < \frac{\eta}{2}$ for some $\bar{t}_i \in U_k$; the last inequality but one follows from (18) and the triangle inequality.

It remains to be shown that $\bar{b}_i \in G_\varepsilon$. For this purpose, consider any $t_i \in T_i$ and $\mu \in \mathcal{M}(T_i)$, and suppose that

$$
\varphi_{\bar{b}_i}(\delta_{t_i}) = \varphi_{\bar{b}_i}(\mu).
$$

By construction,

$$
\varphi_{\bar{b}_i}(\delta_{t_i}) = \bar{b}_i(t_i) = \sum_{k=1}^{K} \alpha_k(t_i) \, \beta_k
$$

and

$$
\varphi_{\bar{b}_i}(\mu) = \int \bar{b}_i(\bar{t}_i) d\mu(\bar{t}_i) = \sum_{k=1}^{K} \int \alpha_k(\bar{t}_i) d\mu(\bar{t}_i) \, \beta_k.
$$

Thus, $\varphi_{\bar{b}_i}(\delta_{t_i}) = \varphi_{\bar{b}_i}(\mu)$ implies

$$
\sum_{k=1}^{K} \alpha_k(t_i) \, \beta_k = \sum_{k=1}^{K} \int \alpha_k(\bar{t}_i) d\mu(\bar{t}_i) \, \beta_k.
$$

The associated densities satisfy

$$
\sum_{k=1}^{K} \alpha_k(t_i) \, f_{\beta_k} = \sum_{k=1}^{K} \int \alpha_k(\bar{t}_i) d\mu(\bar{t}_i) \, f_{\beta_k},
$$

20

or

$$\sum_{k=1}^{K} \left[\alpha_k(t_i) - \int \alpha_k(\bar{t}_i) d\mu(\bar{t}_i) \right] f_{\beta_k} = 0. \tag{21}$$

Because the densities $f_{\beta_1}, ..., f_{\beta_K}$ are linearly independent, equation (21) implies that

$$\alpha_k(t_i) - \int \alpha_k(\bar{t}_i) d\mu(\bar{t}_i) = 0 \tag{22}$$

for all k. For any k, therefore, $\alpha_k(t_i) = 0$ implies $\int \alpha_k(\bar{t}_i) d\mu(\bar{t}_i) = 0$. Since $\alpha_k(t_i) = 0$ if $t_i \notin U_k$ and $\alpha_k(\bar{t}_i) > 0$ if $\bar{t}_i \in U_k$, it follows that, if $t_i \notin U_k$, then $\mu(U_k) = 0$. The measure μ is concentrated on the union of the sets U_k that contain t_i. By (17), any one of these sets, and therefore their union, is contained in the open ζ-ball $B_\zeta(t_i)$ around t_i. Thus, $\mu(B_\zeta(t_i)) = 1$, and therefore $d(\delta_{t_i}, \mu) < \zeta$. Since $\zeta < \varepsilon$, it follows that $\bar{b}_i \in G_\varepsilon$. Claim 3 is thereby established.

To complete the proof of Theorem 3.1, we note that $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ is a complete metric space[12] and therefore a Baire space. The residual set $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i})) \subset \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ is therefore itself dense in $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$. ∎

5   Common Priors With the FSE Property

Whereas the preceding analysis has focussed on a single agent, we now consider the scope for surplus extraction from all participants together. We restrict our analysis to the case where the belief functions b_i, $i = 1, ..., I$, can be interpreted as regular conditional distributions, derived from a common prior on the space $T := \prod_{i=1}^{I} T_i$ of vectors of all agents' types. We continue to assume that, for $i = 1, ... I$, T_i is a compact metric space. For some fixed measure λ on T, let $\mathcal{M}^\lambda(T)$ be the space of probability measures on T that have continuous density functions with respect to λ. For simplicity, we assume that λ is a product measure, i.e., $\lambda = \prod_{i=1}^{I} \lambda_i$, where, for $i = 1, ..., I$, λ_i is a measure on T_i.[13] (Again, the leading example would be the case $T_i \subset \mathbb{R}^n$, with λ as Lebesgue measure on $\mathbb{R}^{In}$.) We restrict our

---

[12]Completeness follows from the fact that T_{-i} is compact and hence the space of densities over T_{-i} is the space of bounded functions. So is $\mathcal{M}^\lambda(T_{-i})$. Finally as T_i is compact as well, the same argument implies that $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$, with uniform metric, is complete as well.

[13]This assumption means, roughly, that, if agent i's observation of his type t_i contains any information about the other agents' types, this is captured by the density of the prior, rather than the reference measure λ.

analysis to the subspace $\mathcal{M}_+^\lambda(T) \subset \mathcal{M}^\lambda(T)$ consisting of those measures ν for which the marginal densities

$$\bar{f}_i^\nu(t_i) := \int_{T_{-i}} f^\nu(t_i, t_{-i}) d\lambda_{-i}(t_{-i}) \tag{23}$$

are everywhere strictly positive, where $\lambda_{-i} := \prod_{j \neq i} \lambda_j$ is the marginal distribution on T_{-i} that is induced by λ.[14] The strict positivity of $\bar{f}_i^\nu(t_i)$ implies that the expression

$$f_i^\nu(t_{-i}|t_i) = \frac{f^\nu(t_i, t_{-i})}{\bar{f}_i^\nu(t_i)} \tag{24}$$

is well defined for all $t_{-i} \in T_{-i}$, and so is the expression

$$b_i(B|t_i, \nu) = \int_B f_i^\nu(t_{-i}|t_i) d\lambda_{-i}(t_{-i}), \tag{25}$$

for any measurable set $B \subset T_{-i}$. The function $t_i \to b_i(t_i, \nu)$ is a regular conditional distribution for t_{-i} given t_i under the measure ν.

For $\nu \in \mathcal{M}_+^\lambda(T)$, Lebesgue's bounded-convergence theorem implies that the marginal density function $\bar{f}_i^\nu$ that is defined by (23) is continuous. Because T_i is compact, it follows that $\bar{f}_i^\nu(t_i)$ is actually bounded away from zero. By inspection of (24), therefore, one sees that the conditional density $f_i^\nu(t_{-i}|t_i)$ is jointly continuous in t_i and t_{-i}. Thus, $b_i(\cdot, \nu)$ is an element of the space $\mathcal{C}(T_i, \mathcal{M}^{\lambda_{-i}}(T_{-i}))$ of continuous functions from T_i to the space $\mathcal{M}^{\lambda_{-i}}(T_{-i})$ of measures on T_{-i} that have continuous density functions with respect to λ_{-i}, where the topology on $\mathcal{M}^{\lambda_{-i}}(T_{-i})$ is induced by the uniform topology on the space of the associated density functions.

We will say that the prior ν *admits full surplus extraction* if and only if each of the belief functions $b_1(\cdot, \nu), ..., b_I(\cdot, \nu)$ admits full surplus extraction in the sense of the definition of McAfee and Reny (1992). We denote the set of priors admitting full surplus extraction as $\mathcal{N}^{**}$. We will sometimes refer to the priors in $\mathcal{N}^{**}$ as *FSE priors*.

In the remainder of this section, we will use the genericity property of FSE belief functions that was established in Theorem 3.1 in order to establish an analogous genericity property for FSE priors. We begin with a lemma on the topological properties of the mapping relating belief functions to priors.

LEMMA 5.1 *Let $\mathcal{M}_+^\lambda(T)$ have the topology that is induced by the uniform topology on the space of density functions on T. For any i, let $\mathcal{C}(T_i, \mathcal{M}^{\lambda_{-i}}(T_{-i}))$ have the uniform topology. Then the mapping $\nu \to b_i(\cdot, \nu)$ from $\mathcal{M}_+^\lambda(T)$ into $\mathcal{C}(T_i, \mathcal{M}^{\lambda_{-i}}(T_{-i}))$ is continuous and open.*

---

[14]Because f^ν is continuous and the set T_{-i} is compact, the integral in (23) is well defined for all $t_i \in T_i$.

Proof. We first prove continuity. If $\{v^k\}$ is a sequence in $\mathcal{M}_+^\lambda(T)$ that converges to a limit $v \in \mathcal{M}_+^\lambda(T)$, the associated densities satisfy

$$\lim_{k\to\infty} f^{v^k}(t_i, t_{-i}) = f^v(t_i, t_{-i}), \tag{26}$$

uniformly on T (by the definition of the topology on $\mathcal{M}_+^\lambda(T)$) By (26), it follows that the densities f^{v^k} are uniformly bounded. For any i, therefore, Lebesgue's bounded convergence theorem implies that

$$\lim_{k\to\infty} \bar{f}_i^{v^k}(t_i) = \lim_{k\to\infty} \int_{T_{-i}} f^{v^k}(t_i, t_{-i}) d\lambda_{-i}(t_{-i}) = \int_{T_{-i}} f^v(t_i, t_{-i}) d\lambda_{-i}(t_{-i}) = \bar{f}_i^v(t_i), \tag{27}$$

uniformly on T_i.

Because $v \in \mathcal{M}_+^\lambda(T)$, we have $\bar{f}_i^v(t_i) > 0$ for all $t_i \in T_i$; indeed, because $\bar{f}_i^v(\cdot)$ is continuous, $\bar{f}_i^v(t_i)$ is bounded away from zero on T_i. Because the marginal densities $\bar{f}_i^{v^k}$ converge uniformly to $\bar{f}_i^v$, it follows that they are uniformly bounded away from zero. If we combine (26) and (27) with (24), we obtain

$$\lim_{k\to\infty} f_i^{v^k}(t_{-i}|t_i) = f_i^v(t_{-i}|t_i), \tag{28}$$

uniformly on T. For any $t_i \in T_i$, therefore, the sequence $\{b_i(\cdot|t_i, v^k)\}$ converges to $b_i(\cdot|t_i, v)$, uniformly on T_i. Continuity of the map $v \to b_i(\cdot, v))$ follows immediately.

We next prove openness. For this purpose, we show that the mapping $v \to (\bar{v}_i(v), b_i(\cdot, v))$, which assigns to each $v \in \mathcal{M}_+^\lambda(T)$ the marginal distribution $\bar{v}_i(v)$ on T_i and the belief function, $b_i(\cdot, v)$, is open, where $\bar{v}_i(v)$ is an element of the space $\mathcal{M}_+^{\lambda_i}(T_i)$ of measures on T_i that have continuous densities, with the topology that is induced by the uniform topology for density functions. The mapping $v \to b_i(\cdot, v))$ is the composition of the mapping $v \to (\bar{v}_i(v), b_i(\cdot, v))$ with the projection from $\mathcal{M}_+^{\lambda_i}(T_i) \times \mathcal{C}(T_i, \mathcal{M}^{\lambda_{-i}}(T_{-i}))$ to $\mathcal{C}(T_i, \mathcal{M}^{\lambda_{-i}}(T_{-i}))$. Because the projection mapping is open and the composition of open mappings is open, openness of the mapping $v \to (\bar{v}_i(v), b_i(\cdot, v))$ automatically yields openness of the mapping $v \to b_i(\cdot, v))$.

To prove that the mapping $v \to (\bar{v}_i(v), b_i(\cdot, v))$ is open, consider any $v \in \mathcal{M}_+^\lambda(T)$ and any sequence $\{v^k\}$ such that

$$\lim_k (\bar{v}_i(v^k), b_i(\cdot, v^k)) = (\bar{v}_i(v), b_i(\cdot, v)) \tag{29}$$

(29) implies

$$\lim_{k\to\infty} \bar{f}_i^{v^k}(t_i) = \bar{f}_i^v(t_i) \tag{30}$$

23

and

$$\lim_{k \to \infty} f_i^{v^k}(t_{-i}|t_i) = f_i^v(t_{-i}|t_i), \tag{31}$$

uniformly, for all $t_i \in T_i$ and all $t_{-i} \in T_{-i}$. From (24), (30), and (31), we therefore obtain

$$
\begin{aligned}
\lim_{k \to \infty} f^{v^k}(t_i, t_{-i}) &= \lim_{k \to \infty} \bar{f}_i^{v^k}(t_i) \lim_{k \to \infty} f_i^{v^k}(t_{-i}|t_i) \\
&= \bar{f}_i^v(t_i) f_i^v(t_{-i}|t_i) = f^v(t_i, t_{-i}),
\end{aligned} \tag{32}
$$

uniformly, for all $t_i \in T_i$ and all $t_{-i} \in T_{-i}$. Thus (29) implies $\lim_{k\infty} v^k = v$, which proves that the mapping $v \to (\bar{v}_i(v), b_i(\cdot, v))$ is open. This completes the proof of Lemma 5.1. ∎

Because continuous and open mappings preserve open and dense sets, we immediately obtain:

THEOREM 5.2 *The set $\mathcal{N}^{**}$ of FSE priors is a residual subset of the set $\mathcal{M}_+^\lambda(T)$ of priors under consideration, i.e., $\mathcal{N}^{**}$ contains a countable intersection of open and dense subsets of $\mathcal{M}_+^\lambda(T)$.*

Proof. For $i = 1, ..., I$, let $\mathcal{N}_i^* \subset \mathcal{M}_+^\lambda(T)$ be the set of priors v for which the belief function $b_i(\cdot, v)$ belongs to the set $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ of functions in $\mathcal{C}(T_i, \mathcal{M}^{\lambda-i}(T_{-i})$ that satisfy condition (*). By the extension of Theorem 3.1,

$$\mathcal{N}^{**} = \bigcap_{i=1}^I \mathcal{N}_i^*. \tag{33}$$

To prove the theorem, it is therefore enough to show that each of the sets $\mathcal{N}_i^*, i = 1, ..., I$, contains a countable intersection of open and dense subsets of $\mathcal{M}_+^\lambda(T)$.

For any i, Theorem 3.1 implies that the set $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ contains a sequence $\{V_k^i\}_{k=1}^\infty$ of open and dense subsets of $\mathcal{C}(T_i, \mathcal{M}^{\lambda-i}(T_{-i})$. For any i and k, define

$$U_k^i := \{v \in \mathcal{M}_+^\lambda(T) | b_i(\cdot, v) \in V_k^i\}. \tag{34}$$

By the definition of $\mathcal{N}_i^*$, $U_k^i \subset \mathcal{N}_i^*$ for all k. Hence, $\cap_{k=1}^\infty U_k^i \subset \mathcal{N}_i^*$. Because V_k^i is open and the function $v \to b_i(\cdot, v)$ is continuous, U_k^i is also open for each k. Because the function $v \to b_i(\cdot, v)$ is continuous, the set

$$\{b_i(\cdot, v) \in \mathcal{C}(T_i, \mathcal{M}^{\lambda-i}(T_{-i}) | v \in U^*\}$$

is open whenever U^* is an open subset of $\mathcal{M}_+^\lambda(T)$. Since V_k^i is dense, it follows that the intersection

$$V_k^i \cap \{b_i(\cdot, v) \in \mathcal{C}(T_i, \mathcal{M}^{\lambda-i}(T_{-i}) | v \in U^*\}$$

24

is nonempty, and therefore also the intersection $U_k^i \cap U^*$ is nonempty whenever U^* is open. Thus U_k^i is dense as well as open. $\mathcal{N}_i^*$ contains a countable intersection of open and dense subsets of $\mathcal{M}_+^\lambda(T)$. This completes the proof of Theorem 5.2. ∎

As in the case of Theorem 3.1, the Baire property of the space $\mathcal{M}_+^\lambda(T)$ implies that $\mathcal{N}^{**}$ itself is dense in $\mathcal{M}_+^\lambda(T)$.

6 Discussion

6.1 Relation to the Literature

The thrust of our results runs counter to that of Heifetz and Neeman (2006) and parallels Chen and Xiong (2013). Heifetz and Neeman (2006) consider families $\{\mathcal{T}^k\}_{k \in K}$ of incomplete-information models of the form

$$\mathcal{T}^k = \{T_i^k, \hat{\theta}_i^k, b_i^k\}_{i=1}^I,$$

where, for each i, T_i^k is a type space for agent i, and $\hat{\theta}_i^k$ and b_i^k are functions indicating how agent i's payoff parameters and beliefs depend on his type. Restricting their attention to a family of models consistent with common priors, they study the genericity of the full surplus extraction property within the set $\mathcal{P}$ of common priors for the models T^k in a given family, which is taken to be fixed. Under the additional assumption that the family $\{\mathcal{T}^k\}_{k \in K}$ is what they call "closed under finite unions", they show that $\mathcal{P}$ is a convex set and that any prior of the form $F = \sum_{j=1}^J \alpha_j F^j$ with $\alpha_j > 0$ and $F^j \in \mathcal{P}$ for all j has the BDP property if and only if *each* of the distributions F^j has the BDP property. Because the BDP property is necessary for full surplus extraction, they conclude that, unless *all* incomplete-information models $\mathcal{T}^k$, $k \in K$, have BDP priors, the set of priors that do not admit full surplus extraction is geometrically and measure-theoretically generic in $\mathcal{P}$.

The difference between our results and those of Heifetz and Neeman (2006) is not only a matter of topological versus geometric or measure-theoretic genericity but also one of genericity in the full space versus genericity in the given set $\mathcal{P}$ of priors for the models in a given family. For a *fixed* family $\{T^k\}_{k \in K}$ of incomplete-information models that is closed under finite unions, Heifetz and Neeman (2006) compare the set of priors for models in this family that admit full surplus extraction to the set of all priors for models in this family. Their procedure begs the question what can be said about the set of families for

which they obtain non-genericity of full surplus extraction relative to the set of all families of incomplete-information models that are closed under finite unions.

The requirement that the family $\{\mathcal{T}^k\}_{k \in K}$ be closed under finite unions can actually be quite restrictive. In related work (Gizatulina and Hellwig (in progress)), we show that, if, for any i, the type spaces $T_i^k, k \in K$, are subsets of a complete separable metric space and if they have non-empty interiors, then the family $\{\mathcal{T}^k\}_{k \in K}$ cannot be closed under finite unions unless it is at most countable. As a consequence of Theorem 5.2, we then find that, for a given sequence $\{T^k\}_{k \in K}$, of such type spaces, full surplus extraction can be obtained for all models in a residual set of families. The set of families with at least one member for which full surplus extraction cannot be obtained, i.e., the set of families to which the Heifetz and Neeman (2006) analysis applies, is itself sparse in the set of all families of models with the given family of type spaces.

Topological genericity of models allowing full surplus extraction is also studied by Chen and Xiong (2013). They specify the type space $T = \prod_{i=1}^{I} T_i$ as the Θ-based universal type space, i.e., the space of payoff parameters and belief hierarchies that is generated by the payoff type space $\Theta = \prod_{i=1}^{I} \Theta_i$. Chen and Xiong (2013) endow the universal type space with the product topology and the set of common priors on this space with the associated weak* topology. For this topology, they show that priors allowing for full surplus extraction form a residual set in the set of all common priors. Their analysis relies on a result of Mertens, Sorin, and Zamir (1994), which implies that, with this topology, the set of common priors with finite supports is dense in the set of all common priors. Because the Crémer-McLean condition for full surplus extraction holds for an open and dense set of models with finitely many types for each agent and a dense subset of a dense set is itself dense in the ambient space, Chen and Xiong (2013) conclude that the set of priors admitting full surplus extraction is dense in the space of all priors endowed with the weak* topology.

6.2 The Choice of Topology

The use of finite approximations and of the weak* topology on the space of priors associated with the product topology on the universal type space is problematic. The product topology on the universal type space has been criticized on the grounds that, in this topology, the correspondence of strictly interim ε-rationalizable strategies is not generally

26

lower hemi-continuous.[15] Dekel, Fudenberg, and Morris (2006), as well as Chen, DiTillio, Faingold, and Xiong (2010) have therefore suggested that a stronger topology should be imposed, under which the correspondence of strictly interim ε-rationalizable strategies would be lower hemi-continuous for any $\varepsilon \geq 0$. In such a stronger topology, however, models with finitely many types need not be dense, in which case the argument of Chen and Xiong (2013) is not available.[16]

Our approach is immune to this criticism. To see this, consider the set of incomplete-information models $\{T_i, \hat{\theta}_i, b_i\}_{i=1}^I$ such that, for each i, the type space T_i is a compact metric space, the payoff function $\hat{\theta}_i$ maps T_i continuously into some topological space Θ_i of payoff parameters, and, for some measure $\lambda_{-i} \in \mathcal{M}(T_{-i})$, the belief function b_i maps T_i into the space $\mathcal{M}^\lambda(T_{-i})$ of measures that have continuous densities with respect to λ. As before, endow $\mathcal{M}^\lambda(T_{-i})$ with the topology of uniform convergence of density functions, and let b_i be an element of the space $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ of continuous functions from T_i to $\mathcal{M}^\lambda(T_{-i})$, endowed with the uniform topology. Define a game of incomplete information by specifying finite action sets $S_1, ..., S_I$ and continuous and bounded payoff functions $u_i : \Theta_i \times \prod_{j=1}^I S_j \to \mathbb{R}$, $i = 1, ..., I$ so that, for each i, $u_i(\hat{\theta}_i(t_i), s_i, s_{-i})$ is the payoff that agent i receives if his own type is t_i, he chooses the action $s_i \in S_i$, and the other agents' actions are given by $s_{-i} \in S_{-i} := \prod_{j \neq i} S_j$. A strategy for agent i is a map $\sigma_i : T_i \to S_i$ that indicates how the action chosen by the agent depends on his type. Given the vector $\sigma_{-i} := \{\sigma_j\}_{j \neq i}$ of strategies of the other agents, the strategy σ_i is said to be strictly interim ε-optimal for agent i, if it satisfies

$$\int_{T_{-i}} [u_i(\hat{\theta}_i(t_i), \sigma_i(t_i), \sigma_{-i}(t_{-i})) - u_i(\hat{\theta}_i(t_i), s_i, \sigma_{-i}(t_{-i}))] f_{b_i}(t_{-i}|t_i) d\lambda_{-i}(t_{-i}) > -\varepsilon \quad (35)$$

for all $s_i \in S_i$ and all $t_i \in T_i$. By adapting an argument from Kajii and Morris (1994) and Engl (1995) we obtain the following lower hemi-continuity result:

PROPOSITION 6.1 *Let* $\{T_i, \hat{\theta}_i, b_i\}_{i=1}^I$ *be an incomplete-information model such that, for any i,* $b_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ *for some* $\lambda_{-i} \in \mathcal{M}(T_{-i})$. *Let* $\sigma_1, ..., \sigma_I$ *be strategies for agents $1, ..., I$ in the game with finite action sets $S_1, ..., S_I$ and continuous and bounded payoff functions $u_1, ..., u_I$*

---

[15]See, e.g., Dekel, Fudenberg, and Morris (2006).

[16]In particular, finite models are not dense in the uniform strategic topology proposed by Dekel, Fudenberg, and Morris (2006) and the uniform weak topology proposed by Chen, DiTillio, Faingold, and Xiong (2010). Chen, DiTillio, Faingold, and Xiong (2010) provide an extensive discussion of the conceptual issues involved in choosing between the different topologies.

in the incomplete-information model $\{T_i, \hat{\theta}_i, b_i\}_{i=1}^I$. *If* σ_i *is strictly interim* ε-*optimal for agent* *i in this game when the other agents choose* σ_{-i}, *then there exists* $\delta > 0$ *such that, if the belief function* $\bar{b}_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ *is* δ-*close to the belief function* b_i, *then* σ_i *is also strictly interim* ε-*optimal for agent i in the game with action sets* $S_1, ..., S_I$ *and payoff functions* $u_1, ..., u_I$ *in the incomplete-information model* $\{T_i, \hat{\theta}_i, \bar{b}_i\}_{i=1}^I$ *when the other agents choose* σ_{-i}.

Proof. Consider any game with action sets $S_1, ..., S_I$ and payoff functions $u_1, ..., u_I$ in the incomplete-information model $\{T_i, \hat{\theta}_i, b_i\}_{i=1}^I$ and let U_i be a bound such that $|u_i(\theta_i, s_i, s_{-i})| \leq U_i$ for all θ_i, s_i, s_{-i}. Fix strategies σ_i, σ_{-i} and suppose that σ_i is strictly interim ε-optimal for agent i in the game with action sets $S_1, ..., S_I$ and payoff functions $u_1, ..., u_I$ in the incomplete-information model $\{T_i, \hat{\theta}_i, b_i\}_{i=1}^I$ when the other agents choose σ_{-i}. Then, for any $t_i \in T_i$ and $s_i \in S_i$,

$$\eta_i(t_i, s_i) > 0,$$

where

$$\eta_i(t_i, s_i) := \int_{T_{-i}} \Delta_i(t_i, t_{-i}, s_i) f_{b_i}(t_{-i}|t_i) d\lambda_{-i}(t_{-i}) + \varepsilon, \tag{36}$$

with

$$\Delta_i(t_i, t_{-i}, s_i) := u_i(\hat{\theta}_i(t_i), \sigma_i(t_i), \sigma_{-i}(t_{-i})) - u_i(\hat{\theta}_i(t_i), s_i, \sigma_{-i}(t_{-i})). \tag{37}$$

for any t_{-i}. The mapping $(t_i, s_i) \to \eta_i(t_i, s_i)$ is obviously continuous. Since T_i is compact and S_i is finite, it follows that there exists some $\hat{\eta} > 0$ such that, for all t_i,

$$\eta_i(t_i, s_i) > \hat{\eta}$$

for all $s_i \in S_i$. From (36), we then obtain

$$\int_{T_{-i}} \Delta_i(t_i, t_{-i}, s_i) f_{b_i}(t_{-i}|t_i) d\lambda_{-i}(t_{-i}) > -(\varepsilon - \hat{\eta}), \tag{38}$$

implying that σ_i is actually strictly interim $(\varepsilon - \hat{\eta})$-optimal for agent i.

For any other belief function $\bar{b}_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$, we have

$$\begin{aligned}
&\int_{T_{-i}} \Delta_i(t_i, t_{-i}, s_i) f_{\bar{b}_i}(t_{-i}|t_i) d\lambda_{-i}(t_{-i}) \\
\geq \ &\int_{T_{-i}} \Delta_i(t_i, t_{-i}, s_i) f_{b_i}(t_{-i}|t_i) d\lambda_{-i}(t_{-i}) \\
&- \int_{T_{-i}} |\Delta_i(t_i, t_{-i}, s_i)| \ |f_{b_i}(t_{-i}|t_i) - f_{\bar{b}_i}(t_{-i}|t_i)| d\lambda_{-i}(t_{-i}),
\end{aligned} \tag{39}$$

28

By (36), the first term on the right-hand side of (39) is nonnegative. Because $|u_i(\theta_i, s_i, s_{-i})| \le U_i$ for all i, we also have

$$|\Delta_i(t_i, t_{-i}, s_i)| \le 2U_i \tag{40}$$

for all t_i, t_{-i}, and s_i. Thus, (39) and (38) imply

$$\int_{T_{-i}} \Delta_i(t_i, t_{-i}, s_i) f_{\bar{b}_i}(t_{-i}|t_i) d\lambda_{-i}(t_{-i}) > -\varepsilon + \hat{\eta} - 2U_i \sup_{(t_i, t_{-i}) \in T} |f_{b_i}(t_{-i}|t_i) - f_{\bar{b}_i}(t_{-i}|t_i)|$$

and hence

$$\int_{T_{-i}} \Delta_i(t_i, t_{-i}, s_i) f_{\bar{b}_i}(t_{-i}|t_i) d\lambda_{-i}(t_{-i}) > -\varepsilon \tag{41}$$

whenever $\sup_{(t_i, t_{-i}) \in T} |f_{b_i}(t_{-i}|t_i) - f_{\bar{b}_i}(t_{-i}|t_i)| \le \frac{\hat{\eta}}{2U_i}$. Upon setting $\delta = \frac{\varepsilon}{2U_i}$, we see that, if $\bar{b}_i$ is δ-close to b_i, then σ_i is strictly interim ε-optimal for agent i in the game with action sets $S_1, ..., S_I$ and payoff functions $u_1, ..., u_I$ in the incomplete-information model $\{T_i, \hat{\theta}_i, \bar{b}_i\}_{i=1}^I$ when the other agents choose σ_{-i}. ∎

Using Proposition 6.1, one can easily show that, in our abstract type space setting, the correspondence of interim ε-rationalizable strategies as well as the correspondence of interim ε-Nash equilibria are lower hemi-continuous in types. The key to these results lies in the assumption that all beliefs of agent i are absolutely continuous with respect to some fixed measure λ_{-i}, and that their density functions are continuous and depend continuously on belief functions and types.[17] This assumption eliminates examples like Rubinstein's e-mail game, where lower hemi-continuity of strictly interim rationalizable strategies is known to fail.[18]

In the appendix, we show that suitably modified versions of Theorems 2.1 and 3.1 hold even if we drop the assumptions that the beliefs of agent i are absolutely continuous

---

[17]Engl (1995) uses setwise convergence of the measures that represent beliefs. Convergence of density functions of course implies setwise convergence; see, e.g. Royden (1988).

[18]To put Rubinstein's e-mail into an abstract type space setting of the kind we use in this paper, let $I = 2$, and set $T_1 = T_2 = \{0, \frac{1}{2}, \frac{2}{3}, ..., 1\}$. Specify a belief function b_1 for agent 1 so that, for some $\alpha \in (0, 1)$ and $n = 1, 2, ..., b_1(\frac{n}{n+1}) = \alpha \delta_{(n-1)/n} + (1-\alpha)\delta_{n/(n+1)}$ and $b_1(1) = \delta_1$, where for any $t \in [0, 1]$, δ_t is the degenerate measure that assigns all probability mass to the singleton $\{t\}$. Similarly, specify a belief function b_2 for agent 2 so that, for some $\beta \in (0, 1)$ and $n = 1, 2, ..., b_2(\frac{n}{n+1}) = \beta \delta_{n/(n+1)} + (1-\beta)\delta_{(n+1)/(n+2)}$ and $b_2(1) = \delta_1$.

If the measures $b_1(t_1), t_1 \in T_1$, are absolutely continuous with respect to some fixed measure λ on T_2, then, for $n = 1, 2, ...$, the density of $b_1(\frac{n}{n+1})$ with respect to λ satisfies $f_{b_1}(\frac{n}{n+1}|\frac{n}{n+1}) = (1-\alpha)/\lambda(\{\frac{n}{n+1}\})$. Because λ is a probability measure, it must be the case that $\lambda(\{\frac{n}{n+t}\})$ goes to zero as n goes out of bounds. Hence $f_{b_1}(\frac{n}{n+1}|\frac{n}{n+1})$ fails to converge to $f_{b_1}(1|1) = 1/\lambda(\{1\})$ as n goes out of bounds and $\frac{n}{n+t}$ converges to one.

with respect to some fixed measure λ_{-i}. In particular, if belief functions are treated as continuous functions from T_i to the space $\mathcal{M}(T_{-i})$ of probability measures on T_{-i}, where $\mathcal{M}(T_{-i})$ has the weak* topology, full surplus extraction is still possible for a residual set of belief functions. However, we do not have an analogue of Proposition 6.1 for this specification of belief functions.

6.3   Universal versus Abstract Type Spaces

There is also a deeper issue with the universal type space approach to studying the genericity of the McAfee-Reny condition for surplus extraction. Like the BDP property, the McAfee-Reny condition is a condition on the belief function of an individual agent which is given without reference to any notion to cross-agent consistency of beliefs, such as belief-closed subsets of the universal type space or to a common prior. In the universal type space, however, it does not even make sense to ask for which belief functions the McAfee-Reny condition is satisfied. In this space, there is single belief function, which is fixed and given by the formalism. Each "type" is defined by a vector consisting of the agent's payoff parameters and the agent's beliefs about the other agents' payoff parameters, the agent's beliefs about the other agents' beliefs about other agents' payoff parameters, et... The belief function is given by the projection from the space of universal types to the space of belief hierarchies, in combination with the observation that each belief hierarchy defines a unique measure on the space of the other agents' universal types.

Because the universal type space approach has no room for considering different belief functions, Chen and Xiong (2011) and Chen and Xiong (2013) do not actually look at the BDP property or the possibility of full surplus extraction as properties of belief functions but as properties of interactive belief systems, more precisely, as properties of belief-closed subsets of the universal type space and of common priors on such subsets. The fact that these properties refer to belief functions without any reference to interactive belief systems is thereby lost.[19]

The notion of an abstract type space approach is sometimes criticized for its lack of economic or game theoretic meaning.[20] This criticism is akin to criticizing the notion of a

---

[19]For a more extensive discussion of these issues in the context of the BDP property, see Gizatulina and Hellwig (2014).

[20]In this context, we have encountered the objection that, in view of the arbitrariness of the specification abstract types, our genericity result might be due to a double counting of belief functions that allow full

probability space on the grounds that it has no meaning in terms of observables. The notion of a probability space provides the grammatical infrastructure for thinking and talking sensibly about random variables, which then leads to statements about observables. In the same vein, Harsanyi's notion of an abstract type space provides the grammatical infrastructure for thinking and talking about incomplete information in a multi-agent setting, which then leads to statements about strategic behaviour in such situations.

From this perspective, it is important to note that the McAfee-Reny condition for full surplus extraction, as well as the weaker condition of injectiveness, appear as conditions on belief functions whose validity does not depend on the specification of the functions that determine payoffs, beliefs about payoffs, beliefs about beliefs about payoff, and so on. This observation is at odds with the Heifetz-Neeman phrase "beliefs determine preferences", which suggests that the scope for surplus extraction depends on how beliefs co-vary with payoff parameters.

Our genericity result suggests that, as a feature of an incomplete-information model, validity of the McAfee-Reny condition is to be expected because the space of beliefs is very rich. In models with infinite type spaces, beliefs are typically infinite-dimensional. The mathematical reasoning is similar to that given for the classical embedding theorem used in Gizatulina and Hellwig (2014), but in contrast to the embedding theorem, the fact that beliefs are infinite-dimensional plays an essential role in the argument establishing the genericity of the McAfee-Reny condition.

---

surplus extraction. Specifically, if a belief function $b_i : T_i \to \mathcal{M}^\lambda(T_{-i})$ satisfies the McAfee-Reny condition, then for any two homeomorphisms $h_i : T_i \to T_i$ and $h_{-i} : T_{-i} \to T_{-i}$, a new belief function $\bar{b}_i$ that also satisfies the McAfee-Reny condition is obtained by setting

$$\bar{b}_i(t_i) = b_i(h_i(t_i)) \circ h_{-i}^{-1}. \tag{42}$$

From a mechanism design perspective, the belief functions b_i and $\bar{b}_i$ are equivalent but in the context of Theorem 3.1, they are treated as distinct. Such "double counting", however, applies to belief functions violating the McAfee-Reny condition as well as those that satisfy it. To make this point formally, let $\mathcal{Q}(T_i, \mathcal{M}^\lambda(T_{-i}))$ be the quotient space of $\mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$ that is defined by treating belief functions b_i and $\bar{b}_i$ as equivalent if they satisfy (42) for some homeomorphisms h_i on T_i and h_{-i} on T_{-i}. For each $b_i \in \mathcal{C}(T_i, \mathcal{M}^\lambda(T_{-i}))$, let $\omega(b_i)$ be the element of $\mathcal{Q}(T_i, \mathcal{M}^\lambda(T_{-i}))$ to which b_i belongs. One easily verifies that $b_i \in \mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ if and only if $\bar{b}_i \in \mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$ for all $\bar{b}_i \in \omega(b_i)$. Hence we can write $\mathcal{E}^*_{\mathcal{Q}}(T_i, \mathcal{M}^\lambda(T_{-i}))$ for the subset of $\mathcal{Q}(T_i, \mathcal{M}^\lambda(T_{-i}))$ consisting of those equivalence classes of functions whose elements belong to $\mathcal{E}^*(T_i, \mathcal{M}^\lambda(T_{-i}))$. It is also easy to verify that, if $\mathcal{Q}(T_i, \mathcal{M}^\lambda(T_{-i}))$ has the quotient topology, then the mapping $b_i \to \omega(b_i)$ is open as well as continuous. By Theorem 3.1, it follows that the set $\mathcal{E}^*_{\mathcal{Q}}(T_i, \mathcal{M}^\lambda(T_{-i}))$ is residual in $\mathcal{Q}(T_i, \mathcal{M}^\lambda(T_{-i}))$.

As mentioned in the introduction, we do not interpret our result as saying that full surplus extraction is to be considered very likely. Indeed it seems unlikely that, in a given situation a mechanism designer should know the participants' belief functions as well as he needs to in order to exploit the dependence of beliefs on types for surplus extraction. However, this concern is not properly addressed by genericity considerations in terms of the space of incomplete-information models $\{T_i, \hat{\theta}_i, b_i\}_{i=1}^{I}$. The McAfee-Reny condition or even the BDP property may seem implausibly strong, but our analysis shows that, if beliefs are infinite-dimensional, such notions of plausibility must be treated with caution. In settings with infinite type spaces, and therefore infinite-dimensional spaces of beliefs, neither the McAfee-Reny condition for full surplus extraction nor the weaker condition of injectiveness, is very special. To deal with concerns about the mechanism designer's lack of information, one needs either an explicit model of the information available to him or a robustness requirement along the lines of Bergemann and Morris (2005), which would preclude his using very specific information about about beliefs and belief functions.

A   Necessity of infinite-dimensionality of the range

In Section 3, we asserted that the genericity claim made in Theorem 3.1 is false if the functions under considerations have a finite-dimensional range. This contrasts with the genericity of embeddings, which by the classical embedding theorem holds whenever the dimension of the range is more than twice the dimension of the domain of the functions under consideration. The following result provides a formal statement.

PROPOSITION A.1 *Let X be a metric space with $|X| \geq N$, and let Y be an N-dimensional, compact, convex, metric space. Let $\mathcal{C}(X,Y)$ be the space of continuous functions from X to Y, endowed with the uniform topology. There exists an open subset $\mathcal{U}$ of $\mathcal{C}(X,Y)$ such that, for every $f \in \mathcal{U}$, there exists $x \in X$ such that*

$$f(x) \in co\{f(X\backslash\{x\})\}, \tag{43}$$

i.e., the value of the function at x belongs to the convex hull of its values at points other than x.

COROLLARY A.2 *For X and Y as specified in Proposition A.1, let $\mathcal{E}^*(X,Y)$ be the subset of those functions $f \in \mathcal{C}(X,Y)$ for which there exists no $x \in X$ for which (43) holds. Then the set $\mathcal{E}^*(X,Y)$ is not dense in $\mathcal{C}(X,Y)$.*

Proof of Proposition A.1.   Using the fact that Y is N-dimensional and has a non-empty interior, choose $p_1, ..., p_{N+2}$ in Y so that $p_1, ..., p_{N+1}$ are in general position and

$$p_{N+2} = \sum_{i=1}^{N+1} \alpha_i p_i, \tag{44}$$

where $\sum_{i=1}^{N+1} \alpha_i = 1$ and $\alpha_i > 0$ for all i.

Fix a function $g \in C(X,Y)$ such that, for $i = 1, ..., N+2$,

$$g(x_i) = p_i. \tag{45}$$

Then, obviously,

$$g(x_{N+2}) \in co\{g(x_1), ..., g(x_{N+1})\}; \tag{46}$$

In fact, since $\alpha_i > 0$ for all i, $g(x_{N+2}) = p_{N+2}$ belongs to the interior of $co\{g(x_1), ..., g(x_{N+1})\}$, i.e., there exists $\varepsilon > 0$ such that the ε-ball $B^\varepsilon(p_{N+2})$ around $g(x_{N+2}) = p_{N+2}$ is fully contained in the convex hull of $g(x_1) = p_1, ..., g(x_{N+1}) = p_{N+1}$.

Now, fix $\eta = \frac{\varepsilon}{2}$ and let $\mathcal{U}$ be the open η-neighbourhood of g, i.e. the set of all functions f such that $d(f(x), g(x)) < \eta$ for all $x \in X$, where d is the metric on Y. We claim that, for any $f \in \mathcal{U}$,

$$f(x_{N+2}) \in co\{f(x_1), ..., f(x_{N+1})\}. \tag{47}$$

To prove this claim, we will show that the $\frac{\varepsilon}{2}$-ball $B^{\frac{\varepsilon}{2}}(p_{N+2})$ around $g(x_{N+2}) = p_{N+2}$ is a subset of the convex hull of $f(x_1), ..., f(x_{N+1})$, i.e., that

$$B^{\frac{\varepsilon}{2}}(p_{N+2}) \subset co\{f(x_1), ..., f(x_{N+1})\}. \tag{48}$$

(48) implies (47) because, for $f \in \mathcal{U}$, $d(f(x_{N+2}), g(x_{N+2})) < \eta = \frac{\varepsilon}{2}$, and therefore, $f(x_{N+2}) \in B^{\frac{\varepsilon}{2}}(p_{N+2})$.

To prove (48), we first note that

$$co\{p_1, ..., p_{N+1}\} \subset B^{\eta}(co\{f(x_1), ..., f(x_{N+1})\}). \tag{49}$$

To see this, observe that, for any $q \in co\{p_1, ..., p_{N+1}\}$, there exist $\alpha_i^q, i = 1, ..., N+1$, such that

$$q = \sum_{i=1}^{N+1} \alpha_i^q p_i.$$

For $f \in \mathcal{U}$, the distance between q and the element $\sum_{i=1}^{N+1} \alpha_i^q f(x_i)$ of $co\{f(x_1), ..., f(x_{N+1})\}$ is less than η, i.e. $q \in B^{\eta}(\{co\{f(x_1), ..., f(x_{N+1})\}\})$.

Since $B^{\varepsilon}(p_{N+2}) \subset co\{p_1, ..., p_{N+1}\}$, it follows that

$$B^{\varepsilon}(p_{N+2}) \subset B^{\eta}(co\{f(x_1), ..., f(x_{N+1})\}). \tag{50}$$

Now (48) follows because $\eta = \frac{\varepsilon}{2}$. For suppose that (48) fails so that there exists some $v \in B^{\frac{\varepsilon}{2}}(p_{N+2}) \backslash co\{f(x_1), ..., f(x_{N+1})\}$. Let y be the element of $co\{f(x_1), ..., f(x_{N+1})\}$ that is closest to v, and let

$$z = v + \delta(v - y) \tag{51}$$

where δ is chosen so that $d(z, v)$, the distance between z and v, is exactly η. By the triangle inequality,

$$d(p_{N+2}, z) \leq d(p_{N+2}, v) + d(v, z).$$

By construction, $d(v, z) = \eta = \frac{\varepsilon}{2}$ and $d(p_{N+2}, v) \leq \frac{\varepsilon}{2}$. Thus, $z \in B^{\varepsilon}(p_{N+2})$. By (50), it follows that there exists $u \in co\{f(x_1), ..., fx_{N+1})\}$ such that

$$d(z, u) \leq \eta. \tag{52}$$

34

Using (51), we obtain

$$d(z, u) = d(v + \delta(v - y), u) = (1 + \delta)d(v, \lambda u + (1 - \lambda)y),$$

where $\lambda := \frac{1}{1+\delta}$. By the definition of y and another application of (51), it follows that

$$d(z, u) \geq (1 + \delta)d(v, y) > \delta d(v, y) = d(z, v) \tag{53}$$

Upon combining (52) and (53), we find that $d(z, v) < \eta$, contrary to the assumption that $d(z, v) = \eta$. The assumption that (48) fails has thus led to a contradiction and must be false. ∎

B Belief Functions in $\mathcal{C}(T_i, \mathcal{M}(T_{-i}))$

In this appendix, we extend Theorems 2.1 and 3.1 to the case where the belief function b_i is treated as an element of $\mathcal{C}(T_i, \mathcal{M}(T_{-i}))$, where measures in $\mathcal{M}(T_{-i})$ need not have densities and the topology on $\mathcal{M}(T_{-i})$ is the topology of weak convergence of probability measures, i.e. the weak' topology. As in the main text, the type spaces $T_1, ..., T_I$ are assumed to be compact metric spaces with more than finitely many elements.

B.1 Generalizing the McAfee-Reny Result

From the main text, we recall that a belief function $b_i : T_i \to \mathcal{M}(T_{-i})$ is said to *admit full surplus extraction* if and only if, for *every* continuous function $\bar{\pi}_i : T_i \to \mathbb{R}$ and every $\varepsilon > 0$, there exists a system $z_1^i, ..., z_{N_i}^i$ of participation fee schedules for agent i such that the induced expected payment $\bar{Z}_i(t_i)$ as given by equations

$$\bar{Z}_i(t_i) := \min(\bar{z}_1^i(t_i), ..., \bar{z}_{N_i}^i(t_i)). \tag{54}$$

and

$$\bar{z}_n^i(t_i) := \int_{T_{-i}} z_n^i(t_{-i}) \, b_i(dt_{-i}|t_i). \tag{55}$$

(4) and (5) satisfies

$$\bar{\pi}_i(t_i) - \varepsilon \leq \bar{Z}_i(t_i) \leq \bar{\pi}_i(t_i) \tag{56}$$

for all $t_i \in [0, 1]$. Following McAfee and Reny (1992), we reformulate this condition as follows. Let $\mathcal{C}(T_i)$ be the set of continuous functions from T_i to $\mathbb{R}$, endowed with the

topology of uniform convergence. For any $b_i \in \mathcal{C}(T_i, \mathcal{M}(T_{-i}))$, let $R(b_i)$ be the set of functions $\bar{z} \in \mathcal{C}(T_i)$ such that, for some continuous function $z : T_{-i} \to \mathbb{R}$,

$$\bar{z}(t_i) := \int_{T_{-i}} z(t_{-i}) \, b_i(dt_{-i}|t_i)$$

for all $t_i \in T_i$; further let $r(b_i)$ be the set of functions $\bar{Z} \in \mathcal{C}(T_i)$ such that, for some N and some functions $\bar{z}_1, ..., \bar{z}_N$ in $R(b_i)$,

$$\bar{Z}(t_i) = \min(\bar{z}_1^i(t_i), ..., \bar{z}_{N_i}^i(t_i))$$

for all $t_i \in [0,1]$. Then b_i admits full surplus extraction if and only if

$$\bar{r}(b_i) = \mathcal{C}(T_i), \tag{57}$$

where $\bar{r}(b_i)$ is the closure of $r(b_i)$ in $\mathcal{C}(T_i)$.

We also extend the map φ_{b_i} that was defined in Section 3 by setting

$$\varphi_{b_i}(B|\mu) := \int_{T_i} b_i(B|t_i) d\mu(t_i) \tag{58}$$

for any measurable $B \subset T_{-i}$ and any $\mu \in \mathcal{M}(T_i)$. In this setting, we obtain the following generalization of the result of McAfee and Reny (1992).

THEOREM B.1 *A belief function $b_i : T_i \to \mathcal{M}(T_{-i})$ admits full surplus extraction if and only if, for all $\bar{t}_i \in T_i$ and all $\mu \in \mathcal{M}(T_i)$,*

$$\varphi_{b_i}(\mu) = \varphi_{b_i}(\delta_{\bar{t}_i}) \ \text{implies} \ \mu = \delta_{\bar{t}_i}, \tag{59}$$

where $\varphi_{b_i} : \mathcal{M}(T_i) \to \mathcal{M}(T_{-i})$ is the map that is defined by (58).

The proof of Theorem B.1 is by and large the same as the proof of Theorem 2 in McAfee and Reny (1992), with due changes to account for the fact that T_i is an arbitrary compact metric space, rather than the unit interval, and for the fact that beliefs need not have densities. Therefore we will not go into all the details but merely indicate where and how the argument must be adapted.

In the analysis of McAfee and Reny (1992), a special role is played by what they call the set of "(ε, δ)-u-shaped functions at t_{i0}". In the present, more general setting, an (ε, δ)-u-shaped function at t_{i0} is a function $u \in \mathcal{C}(T_i)$ such that

    (i) $u(t_i) \geq 0$ for all $t_i \in T_i$,

(ii) $u(t_{i0}) \leq \varepsilon$, and

(iii) $u(t_i) \geq 1$ for all $t_i \in T_i \backslash B^\delta(t_{i0})$, where $B^\delta(t_{i0})$ is the closed δ-ball around t_{i0}.

The set of such (ε, δ)-u-shaped function at t_{i0} is denoted as $U(\varepsilon, \delta, t_{i0})$. This set plays a key role in the following auxiliary result, which extends Theorem 1 of McAfee and Reny (1992).

PROPOSITION B.2 *Suppose that a set $A \subset \mathcal{C}(T_i)$ satisfies:*

(i) A is closed under addition and under multiplication by a positive scalar.

(ii) For any $N, x_1, ..., x_N \in A$ implies $y \in A$, where y is defined by setting $y(t_i) = \min(x_1(t_i), ..., x_N(t_i))$ for any $t_i \in T_i$.

(iii) $1, -1 \in A$

(iv) For all $\varepsilon > 0, \delta > 0$, and every $t_{i0} \in T_i$, $U(\varepsilon, \delta, t_{i0}) \cap \bar{A} \neq \varnothing$.

Then $\bar{A} = \mathcal{C}(T_i)$.

The proof of Proposition B.2 is step by step the same as the proof of Theorem 1 in McAfee and Reny (1992), except that the unit interval as the domain of functions must be replaced by T_i and intervals of the form $[t_{i0} - \delta, t_{i0} + \delta]$ must be replaced by the closed δ-balls $B^\delta(t_{i0})$ around t_{i0}.

Similarly, the proof of the necessity of condition (59) for full surplus extraction is step by step the same as the proof of the necessity statement in Theorem 2 of McAfee and Reny, again with the proviso that $[0, 1]$ be replaced by T_i and intervals of the form $[t_{i0} - \delta, t_{i0} + \delta]$ by the closed δ-balls $B^\delta(t_{i0})$ around t_{i0}.

Proof of the sufficiency statement in Theorem B.1. As in McAfee and Reny (1992), the proof is indirect. Suppose condition (59) is not sufficient for full surplus extraction. Then there exists $b_i \in \mathcal{C}(T_i, \mathcal{M}(T_{-i}))$ such that (59) holds for all $\bar{t}_i \in T_i$ and all $\mu \in \mathcal{M}(T_i)$ and $\mathcal{C}(T_i) \backslash \bar{r}(b_i) \neq \varnothing$. By Proposition B.2, $\mathcal{C}(T_i) \backslash \bar{r}(b_i) \neq \varnothing$ implies that the set $r(b_i)$ violates one of the conditions in that proposition; the only candidate is condition (iv). Thus there exist $\varepsilon_0 > 0, \delta_0 > 0$, and $t_{i0} \in T_i$ such that $U(\varepsilon_0, \delta_0, t_{i0}) \cap \bar{r}(b_i) = \varnothing$. Since $R(b_i) \subset r(b_i)$, it follows that $U(\varepsilon_0, \delta_0, t_{i0}) \cap \bar{R}(b_i) = \varnothing$.

As discussed by McAfee and Reny (1992), one can now use the separating hyperplane theorem and the Riesz representation theorem to assert the existence of a constant c and a regular, countably additive, signed measure $\mu \neq 0$ on T_i such that

$$\int_{T_i} x(t_i) d\mu(t_i) \leq c \quad \text{for all} \ x \in \bar{R}(b_i), \ \text{and} \tag{60}$$

37

$$\int_{T_i} x(t_i) d\mu(t_i) \geq c \quad \text{for all} \ \ x \in U(\varepsilon_0, \delta_0, t_{i0}). \tag{61}$$

Moreover, since $\bar{R}(b_i)$ is a linear subspace of $\mathcal{C}(T_i)$, it must be the case that

$$\int_{T_i} x(t_i) d\mu(t_i) = 0 \ \text{ for all} \ \ x \in \bar{R}(b_i), \tag{62}$$

and the constant c can be taken to be zero. By the definition of $R(b_i)$, it follows that

$$\int_{T_i} \int_{T_{-i}} z(t_{-i}) db_i(t_{-i}|t_i) d\mu(t_i) = 0 \tag{63}$$

for all $z \in \mathcal{C}(T_{-i})$.

By the Jordan decomposition theorem, we may write μ as the difference between two positive measures μ^+ and μ^-, at least one of which is finite. Thus, (63) can be rewritten in the form

$$\int_{T_i} \int_{T_{-i}} z(t_{-i}) db_i(t_{-i}|t_i) d\mu^+(t_i) = \int_{T_i} \int_{T_{-i}} z(t_{-i}) db_i(t_{-i}|t_i) d\mu^-(t_i). \tag{64}$$

If $z \in \mathcal{C}(T_{-i})$ is the constant function with value one, (64) specializes to the equation

$$\mu^+(T_i) = \mu^-(T_i),$$

so μ^+ and μ^- are both finite. Moreover, there is no loss of generality in setting $\mu^+(T_i) = \mu^-(T_i) = 1$, so both μ^+ and μ^- belong to $\mathcal{M}(T_i)$. (64) can therefore be rewritten in the form

$$\int_{T_{-i}} z(t_{-i}) d\varphi_{b_i}(t_{-i}|\mu^+) = \int_{T_{-i}} z(t_{-i}) d\varphi_{b_i}(t_{-i}|\mu^-). \tag{65}$$

If (65) is to hold for all $z \in \mathcal{C}(T_{-i})$, it must be the case that

$$\varphi_{b_i}(\mu^+) = \varphi_{b_i}(\mu^-). \tag{66}$$

By (59), it follows that neither μ^+ nor μ^- can be the point measure $\delta_{t_{i0}}$ at t_{i0}. For suppose that $\mu^+ = \delta_{t_{i0}}$. Then (66) implies $\varphi_{b_i}(\mu^-) = \varphi_{b_i}(\delta_{t_{i0}})$ and, by (59), $\mu^- = \delta_{t_{i0}}$, which is incompatible with $\mu = \mu^+ - \mu^- \neq 0$. By a precisely symmetric argument, $\mu^- = \delta_{t_{i0}}$ is also impossible.

Since μ^- is regular, it follows that there exists a closed set $A \subset T_i$, such that $\mu^-(A) > 0, \mu^+(A) = 0$, and for some $\delta \in (0, \delta_0]$, the intersection of A with the δ-ball $B^\delta(t_{i0})$ around t_{i0} is empty. Fixing $K > 1/\mu^-(A)$, we can define a step function x^K by setting

$$
\begin{aligned}
x^K(t_i) &= & 0 \text{ if } t_i \in B^\delta(t_{i0}), \\
x^K(t_i) &= & K \text{ if } t_i \in A, \\
x^K(t_i) &= & 1 \text{ otherwise.}
\end{aligned}
$$

For this step function, we find $\int x^K(t_i)d\mu(t_i) \leq 1 - K\mu^-(A) < 0$. Now x^K itself is not continuous, but can be approximated by a sequence $\{x_n\}_{n=1}^{\infty}$ of continuous functions so that the integrals $\int x_n(t_i)d\mu(t_i)$ converge to $\int x^K(t_i)d\mu(t_i)$ as n becomes large. For any sufficiently large n, therefore, $\int x_n(t_i)d\mu(t_i) < 0$.

However, as discussed by McAfee and Reny (1992), the sequence $\{x_n\}_{n=1}^{\infty}$ can be chosen so that $x_n \in U(\varepsilon_0, \delta_0, t_{i0})$. By (61) and the fact that $c = 0$, it follows that $\int x_n(t_i)d\mu(t_i) \geq 0$ for all n. The assumption that condition (59) is not sufficient for full surplus extraction has thus led to a contradiction and must be false. ∎

B.2 A Version of Theorem 3.1 for Belief Functions in $\mathcal{C}(T_i, \mathcal{M}(T_{-i}))$

We next provide a version of Theorem 3.1 for belief functions in $\mathcal{C}(T_i, \mathcal{M}(T_{-i}))$.

THEOREM B.3 *Let $\mathcal{E}^{**}(T_i, \mathcal{M}(T_{-i}))$ be the set of continuous functions from T_i to $\mathcal{M}(T_{-i})$ that satisfy the condition for full surplus extraction. If T_{-i} has more than finitely many elements, then $\mathcal{E}^{**}(T_i, \mathcal{M}(T_{-i}))$ is a residual and dense subset of the space $\mathcal{C}(T_i, \mathcal{M}(T_{-i}))$ with the uniform topology, i.e., $\mathcal{E}^{**}(T_i, \mathcal{M}(T_{-i}))$ contains a countable intersection of open and dense subsets of $\mathcal{C}(T_i, \mathcal{M}^{\lambda}(T_{-i}))$ and is itself dense.*

Proof Sketch. As before, we only indicate how the proof in the main text must be changed to accommodate the change in the range of the belief functions under consideration. The main change concerns the metric on the range. Whereas equation (10) in the main text provided for a metric in terms of density functions, this metric now is not available. Following Parthasarathy (1967), we fix a countable dense set of continuous functions $g^n : T_{-i} \to [0, 1]$ and define the distance between two measures β and $\hat{\beta}$ in $\mathcal{M}(T_{-i})$ to be given as

$$\rho^*(\beta, \hat{\beta}) = \sum_{k=1}^{\infty} \xi^n \left| \int_{T_{-i}} g^n(t_{-i})d\beta(t_{-i}) - \int_{T_{-i}} g^n(t_{-i})d\hat{\beta}(t_{-i}) \right|, \tag{67}$$

where $\xi \in (0, 1)$ is arbitrary but fixed. By Theorem 6.6 , p. 47, in Parthasarathy (1967), the metric ρ^* induces the topology of weak convergence of probability measures on T_{-i}.

In going through the proof of Theorem 3.1 in the main text, one observes that the argument for Claim 1 does not refer to the range of the belief functions at all. The argument for Claim 2 does refer to the range, but one easily verifies that every step remains valid if the metric ρ from equation (10) is replaced by the metric ρ^* that is defined by (67).

In the proof of Claim 3, we can proceed similarly, replacing ρ by ρ^* throughout. When it comes to selecting measures $\beta_1, ..., \beta_K$ such that, for $k = 1, ..., K$, the distance (now under ρ^*, rather than μ) between β_k and $b_i(\bar{t}_i)$ for some $\bar{t}_i$ in the open set U_k in the finite subcovering of T_i is less than $\frac{\eta}{2}$, we now replace the requirement that the associated density function be linearly independent by the requirement that the associated sequences $\{\int_{T_{-i}} g^n(t_{-i})d\beta_k(t_{-i})\}_{n=1}^{\infty}$, $k = 1, ..., K$, be linearly independent. Given the definition (19) of the belief mapping $\bar{b}_i$, the distance between $\bar{b}_i$ and the originally given belief mapping b_i now given by the metric ρ^* is again found to be no greater than η, as in (23) in the previous analysis.

The key step in the proof of Claim 3 concerns the assertion that $\bar{b}_i \in G_\varepsilon$. By construction,

$$\varphi_{\bar{b}_i}(\delta_{t_i}) = \bar{b}_i(t_i) = \sum_{k=1}^{K} \alpha_k(t_i)\beta_k$$

and

$$\varphi_{\bar{b}_i}(\mu) = \int \bar{b}_i(\bar{t}_i)d\mu(\bar{t}_i) = \sum_{k=1}^{K} \int \alpha_k(\bar{t}_i)d\mu(\bar{t}_i) \, \beta_k,$$

so, for any $t_i \in T_i$ and $\mu \in \mathcal{M}(T_i)$, $\varphi_{\bar{b}_i}(\delta_{t_i}) = \varphi_{\bar{b}_i}(\mu)$ implies

$$\sum_{k=1}^{K} \alpha_k(t_i)\beta_k = \sum_{k=1}^{K} \int \alpha_k(\bar{t}_i)d\mu(\bar{t}_i) \, \beta_k.$$

For any n, therefore,

$$\sum_{k=1}^{K} \alpha_k(t_i) \int g^n(t_{-i})d\beta_k(t_{-i}) = \sum_{k=1}^{K} \int \alpha_k(\bar{t}_i) \int g^n(\bar{t}_{-i})d\beta_k(\bar{t}_{-i})d\mu(\bar{t}_i).$$

Equivalently,

$$\sum_{k=1}^{K} \left[\alpha_k(t_i) - \int \alpha_k(\bar{t}_i)d\mu(\bar{t}_i) \right] \int g^n(t_{-i})d\beta_k(t_{-i}) = 0. \tag{68}$$

Because the sequences $\{\int_{T_{-i}} g^n(t_{-i})d\beta_k(t_{-i})\}_{n=1}^{\infty}$, $k = 1, ..., K$, are linearly independent, equation (68) implies that

$$\alpha_k(t_i) - \int \alpha_k(\bar{t}_i)d\mu(\bar{t}_i) = 0 \tag{69}$$

for all k. By the same argument as in the main text, it follows that the measure μ is concentrated on the union of the sets U_k that contain t_i. Claim 3 then follows by the same argument as in the proof of Theorem 3.1.

Finally, by Theorem 6.5 in Parthasarathy (1967) $\mathcal{M}(T_{-i})$ is a complete metric space (as T_{-i} is complete). Therefore, $\mathcal{C}(T_i, \mathcal{M}(T_{-i}))$ is also a complete metric space and hence a Baire space. The residual set $\mathcal{E}^{**}(T_i, \mathcal{M}(T_{-i}))$ is therefore itself dense in $\mathcal{C}(T_i, \mathcal{M}(T_{-i}))$. $\blacksquare$

In contrast to Theorem B.1 above, which generalizes Theorem 2.1, i.e., the result of McAfee and Reny (1992), strictly speaking, Theorem B.3 is *not* simply a generalization of Theorem 3.1. Besides enlarging the range of the belief functions under consideration, from $\mathcal{M}^\lambda(T_{-i})$ to $\mathcal{M}(T_{-i})$, we are also endowing this range with a different topology. In Theorem 3.1, $\mathcal{M}^\lambda(T_{-i})$ was endowed with the topology of uniform convergence of density functions, but now we endow the larger space $\mathcal{M}(T_{-i})$ with the topology of weak convergence of probability measures. The induced subspace topology on $\mathcal{M}^\lambda(T_{-i})$ is weaker than the topology of uniform convergence of density functions, so the range of belief functions has fewer open sets, and the set of continuous functions from T_i to $\mathcal{M}^\lambda(T_{-i})$ is larger.

References

BARELLI, P. (2009): "On the genericity of full surplus extraction in mechanism design," *J. of Econ. Theory*, 144, 1320 – 1333.

BERGEMANN, D. AND S. MORRIS (2005): "Robust mechanism design," *Econometrica*, 73, 1521–1534.

CHEN, Y., A. DITILLIO, E. FAINGOLD, AND S. XIONG (2010): "Uniform Topologies on Types," *Theoretical Economics*, 5, 445 – 478.

CHEN, Y. AND S. XIONG (2011): "The Genericity of Beliefs-Determine-Preferences Models Revisited," *Journal of Economic Theory*, 146, 751– 761.

——— (2013): "Genericity and Robustness of Full Surplus Extraction," *Econometrica*, 81, 825–847.

CRÉMER, J. AND R. MCLEAN (1988): "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica*, 56, 1247–1257.

DEKEL, E., D. FUDENBERG, AND S. MORRIS (2006): "Topologies on Types," *Theoretical Economics*, 1, 275 – 309.

ENGL, G. (1995): "Lower Hemicontinuity of the Nash Equilibirum Correspondance," *Games and Economic Behaviour*, 9, 151–160.

GIZATULINA, A. AND M. HELLWIG (2014): "Beliefs, Payoffs, Information: On the Robustness of the BDP Property in Models with Endogenous Beliefs," *Journal of Mathematical Economics*, 51, 136–153.

——— (in progress): "On the Designer Uncertainty and the Robustness of Full Surplus Extraction in Families of Incomplete-Information Models," Work in progress, Max Planck Institute for Research on Collective Goods.

HEIFETZ, A. AND Z. NEEMAN (2006): "On the generic (im)possibility of full surplus extraction in mechanism design," *Econometrica*, 117, 213–233.

HUREWICZ, W. AND H. WALLMAN (1941): *Dimension Theory*, Princeton, NJ: Princeton Mathematical Series, no. 4.

KAJII, A. AND S. MORRIS (1994): "Payoff Continuity in Incomplete Information Games and Almost Uniform Convergence of Beliefs," Caress working paper no 94-17, University of Pennsylvania.

MCAFEE, P. AND P. RENY (1992): "Correlated information and mechanism design," *Econometrica*, 60, 395–421.

MERTENS, J.-F., S. SORIN, AND S. ZAMIR (1994): "Repeated Games: Part A: Background Material," Working paper, CORE Université de Louvain.

NEEMAN, Z. (2004): "The relevance of private information in mechanism design," *J. Econ. Theory*, 117, 55–77.

PARTHASARATHY, K. (1967): *Probability Measures on Metric Spaces*, New York and London: Academic Press.

ROYDEN, H. (1988): *Real Analysis, 2nd Edition*, London: The Macmillian Company.