Kernel-based Type Spaces*

Konrad Grabiszewski†

2010

Abstract

Type space is of fundamental importance in epistemic game theory. This paper shows how to build type space if players approach the game in a way advocated by Bernheim's justification procedure. If an agent fixes a strategy profile of her opponents and ponders which of their beliefs about her set of strategies make this profile optimal, such an analysis is represented by kernels and yields disintegrable beliefs. Our construction requires that underlying space is Polish.

Keywords: Universal Type Spaces; Rationalizability; Epistemic Game Theory

JEL classification: C72; D80

*This paper would not exist without the enormous support and guidance of Adam Brandenburger. I thank Amanda Friedenberg for our talks about epistemic game theory and her valuable insight and help. I also thank the associate editor, an anonymous referee, Jean-Pierre Benoît, Janko Hernández Cortés, and David Pearce for their comments. Support from the Asociación Mexicana de Cultura A.C. and the Stern School of Business is gratefully acknowledged.

†Address: Department of Business Administration, ITAM, Av. Camino a Sta. Teresa No. 930, Col. Magdalena Contreras, C.P. 10700 Mexico D.F., Mexico, konrad.grabiszewski@gmail.com

1 Introduction

Fix a game played by Ann and Bob with their strategy sets, S^a and S^b, respectively. Ann's first-order belief is her conjecture over Bob's choices. It is natural to assume that Bob ponders Ann's strategies, as well, and that Ann knows this. Hence, she tries to link Bob's alternatives with his first-order beliefs. Ann fixes Bob's strategy, s^b, and selects his conjectures that make s^b optimal. Bob conducts the same analysis and, in consequence, we obtain infinite structures representing the players' thinking about the game. This way of interactive reasoning lies behind the concept of rationalizability introduced by Bernheim (1984) and Pearce (1984). According to the former:

> *Since the state of the world, as perceived by A, is uncertain, he must construct some assessment of B's action and optimize accordingly. (...) A knows that B has an assessment of what A will do for which B's strategy is a best response. (...) A must not only have an assessment of what B will do subject to which A's choice is a best response, but for every forecast of B's strategy to which A ascribes positive probability, A must also be able to construct some conjecture of B's assessment of A's action, for which this forecast of B's strategy is a best response. Since conformity with Savage's axioms is common knowledge, this reasoning can be extended indefinitely. If it is possible to justify the choice of a particular strategy by constructing infinite sequence of self-justifying conjectured assessments in this way, then I call the strategy "rationalizable."*

The infinite hierarchy of beliefs and type space were introduced by Harsanyi (1967/8) (see Myerson (2004) for a non-technical review). Type space is an essential tool of epistemic game theory (see a recent three-article survey by Brandenburger Brandenburger (2008), Heifetz (2008), and Siniscalchi (2008)). In particular, we want to know whether the collection of all of Ann's hierarchies, T^a, and the space of her beliefs over $S^b \times T^b$, $P^a(S^b \times T^b)$, are homeomorphic. Proving this establishes the existence of the universal type space (see Siniscalchi (2008) and Friedenberg (2010) for discussion of universal, terminal, and complete type spaces). Coherency of agents' conjectures is a minimal condition. That is, we require the higher and lower order beliefs to agree on appropriate spaces. If this is not true, then it is impossible for an infinite hierarchy to induce a unique belief over $S^b \times T^b$. However, coherency is not enough, as is shown by Heifetz and Samet (1999). Their result is based on violating topological assumptions in the Kolmogorov Extension Theorem. In order to obtain a positive answer, we need to introduce restriction on either underlying space or on agents' hierarchies. Mertens and Zamir (1985) and Brandenburger and Dekel (1993) focus on topological constraints. The former assumes the space of uncertainty to be compact, while the latter considers a Polish space. In Heifetz (1993), that space is Hausdorff, and agents' beliefs are defined as regular probability measures.

Our construction of the universal type space is based on reconceptualizing the idea of an agent's belief. We want to capture Bernheim's justification procedure leading to rationalizability. Ann's first-order belief is defined as a probability, λ^a, over the set of Bob's strategies, S^b. However, instead of defining higher-order beliefs in a standard way directly on product spaces, we use the notion of a kernel. A second level of Ann's hierarchy consists of a family of kernels between S^b and $P^b(S^a)$. For each of Bob's strategies, s^b, a kernel, ν, assigns Ann's conjecture, $\nu(s^b)$, over the set of Bob's first-order beliefs, $P^b(S^a)$. As Bernheim noted, "*A* must also be able to construct some conjecture of *B*'s assessment of *A*'s action, for which this forecast of *B*'s strategy is a best response." Ann's belief, λ^a, and kernel, ν, generate a unique belief over $S^b \times P^b(S^a)$. The collection of kernels that constitutes the second-order belief is determined by the notion of equivalency. We say that ν and $\widetilde{v}$ are equivalent with respect to λ^a, if they are the same except for the set of measure zero. Our belief is called a **kernel-based belief**, since kernel constitutes the basis of our construction.

In the standard formulation of type spaces, the belief is a probability directly defined on a product space, $X \times Y$. If such a standard belief can be reconstructed from some kernel between X and Y and a measure on X, then we say that it is a disintegrable probability. In our construction, kernels are the primitives of the model. In consequence, the kernel-based belief generates probability on $X \times Y$ which satisfies the disintegrability condition by default. This restricted definition of belief is the price we must pay in order to capture Bernheim's procedure.

The model we propose also applies to a more general case in which the parameter set is common to both agents. In fact, we conduct our analysis and proofs in this setup. As usual, we begin with Ann's first-order belief, $\lambda^a \in P(S)$. Next, Ann ponders the relationship between the parameter set, S, and the set of Bob's first-order conjectures, $P^b(S)$. She builds a kernel, v, between S and $P^b(S)$ by assigning conjecture, $\nu(s)$, over $P^b(S)$ for each s. Naturally, if we presume that the parameters and agents' beliefs are independent, and that this fact is commonly known to the agents, then our construction and the constructions offered in previously mentioned papers are overly complex and even unnecessary. In such a case, Ann's conjectures are directly defined on spaces S, $P^b(S)$, $P^b(P^a(S))$, etc. However, independence is only a special case of the general type space formulation.

In order to prove the existence of the universal type space, we assume that the underlying space of uncertainty is Polish. The key property we use in the proof of existence is **coherency**. This says that the family of kernels at the nth level of hierarchy is equivalent with respect to the measure induced by the first $n-1$ levels of that hierarchy.

In Section 2, we review the concept of kernel and indicate some of its properties. In Section 3, we discuss the idea of a kernel-based belief, a key element of our analysis. We also propose a topology associated with the set of these beliefs. In Section 4, we prove the existence of canonical homeomorphism. In Section 5, we compare the standard construction of type space with our construction of kernel-based type space. Appendix A contains proofs of all results presented in Section 3.

2 Kernels

Throughout this and the next sections, we assume that X and Y are Polish spaces endowed with Borel σ-algebras, $\mathcal{E}$ and $\mathcal{F}$, respectively. E and F are generic elements of $\mathcal{E}$ and $\mathcal{F}$, respectively. We endow $X \times Y$ with the product σ-algebra, of which B is its generic element, and $B(x)$ is a section of B at x. $P(X)$ denotes a collection of probability measures on X with the weak-* topology assigned to it.

$\nu : X \times \mathcal{F} \to \mathbb{R}$ is a **kernel**[1] between X and Y, if for each x, $\nu\ (x;.)$ is probability measure on Y, and for each $F \in \mathcal{F}$, $\nu(.;F)$ is a measurable function. If λ^X is a probability measure on X, then for a given ν, there is the unique probability, $\lambda^{X\times Y}$ on $X \times Y$, such that for each B

$$\lambda^{X\times Y}(B) = \int_X \nu(x; B(x)) d\lambda^X. \tag{1}$$

Any measure on $X \times Y$ that can be represented as in (1) is called a **disintegrable measure** with respect to (its marginal) λ^X. Under our topological assumption, every probability measure on $X \times Y$ is disintegrable (see, for instance, Chapter 21 in Fristedt and Gray (1997)).

We say that ν and $\widetilde{\nu}$ are **equivalent kernels** with respect to λ^X if for any F, $\nu(x;F) = \widetilde{\nu}(x;F)$ λ^X-a.s. Since equivalent kernels differ only on the set of measure zero, it is easy to show the following result.

[1]Kernels are also called transition probabilities, Markov kernels, or transition kernels.

Lemma 2.1

Let ν and $\widetilde{\nu}$ be equivalent with respect to λ^X. For each measurable $B \subset X \times Y$,

$$\lambda^{X\times Y}(B) = \int_X \nu(x; B(x) d\lambda^X = \int_X \widetilde{\nu}(x; B(x)) d\lambda^X = \widetilde{\lambda}^{X\times Y}(B). \tag{2}$$

For a product of more than two Polish spaces, $\nu^{0,..,n-1;n}$ will denote a kernel between $X_1 \times ... \times X_{n-1}$ and X_n. Let $P(X_0 \times ... \times X_n)$ denote the collection of (fully disintegrable) measures on $X_1 \times ... \times X_n$. If $\lambda^{0,...,n} \in P(X_0 \times ... \times X_n)$, then there is a probability, λ^0, on X_0 and a collection of kernels, $(\nu^{0;1}, \nu^{0,1;2}, ..., \nu^{0,...,n-1;n})$, such that for any measurable $B \subset X_0 \times ... \times X_n$, we have:

$$\lambda^{0,...,n}(B) = \int_{X_0} d\lambda^0 \int_{X_1} \nu^{0;1}(x_0; dx_1) ... \int_{X_n} 1_B \nu^{0,...,n-1;n}(x_0, ..., x_{n-1}; dx_n) \tag{3}$$

where 1_B is an indicator function.

Construction of type space, which we demonstrate in Section 4, is based on infinite sequences of the form, $(\lambda^0, \nu^{0;1}, \nu^{0,1;2}, ..., \nu^{0,...,n-1;n}, ...)$. The Ionescu-Tulcea Theorem (see Chapter II.9 in Shiryaev (1996)), which is the basis of our main results, indicates that such a sequence generates the unique measure, λ^∞, on $X_0 \times ... \times X_n \times ....$ In addition, λ^∞ agrees with $\lambda^{0,...,n}$ – constructed in (3) – on cylinders where cylinder C^n is defined as $B^{0,...,n} \times X_{n+1} \times ....$ The base of cylinder, $B^{0,...,n}$, is a measurable subset of $X_0 \times ... \times X_n$.

Lemma 2.2 contains auxiliary results related to measures on product spaces, which will help us in the forthcoming sections. We skip the proofs, as they are based on standard methods.

Lemma 2.2

1a. *Take $\lambda^{0,1}$ on $X_0 \times X_1$ and λ^0. B^0 is a λ^0-continuity set*[2] *if and only if $B^0 \times X_1$ is a $\lambda^{0,1}$-continuity set.*

1b. *Take λ^∞ and its marginal $\lambda^{0,...,n}$. $B^{0,...,n}$ is a $\lambda^{0,...,n}$-continuity set if and only if $C^n = B^{0,...,n} \times X_{n+1} \times ...$ is a λ^∞-continuity set.*

2a. *If $\{\lambda_t^{0,1}\}$ weak-* converges to $\lambda^{0,1}$, then the sequence of marginals generated by $\{\lambda_k^{0,1}\}$, $\{\lambda_k^0\}$, weak-* converges to the marginal of $\lambda^{0,1}$, λ^0.*

[2]Set E is a λ-continuity set if $\lambda(\partial E) = 0$, where ∂E denotes the boundary of E.

2b. If $\{\lambda_k^\infty\}$ weak- converges to λ^∞, then for each n, the sequence of marginals generated by $\{\lambda_k^\infty\}$, $\{\lambda_k^{0,\dots,n}\}$, weak-* converges to the marginal of λ^∞, $\lambda^{0,\dots,n}$.*

3. λ^∞ is uniquely determined by cylinders.

So far, we have discussed a general construction of kernels. However, our main interest lies in continuous kernels.

Definition 2.1 ***Continuous kernel***

Kernel, ν, between X and Y is continuous if $\nu(x_k;.)$ weak- converges to $\nu(x;.)$ whenever $\{x_k\}$ converges to x.*

Continuous kernels are also called Feller kernels and they satisfy a very useful property (see Theorem 2.5.53 in Denkowski et al. (2003)), which we will use in construction of topology for the families of kernels.

Lemma 2.3

If a kernel, ν, between X and Y is continuous, then the function

$$g(x) := \int_Y f(x,y)\nu(x;dy) \tag{4}$$

is bounded continuous for a bounded continuous f on $X \times Y$.

Let $P_C(X_0 \times ... \times X_n)$ be collection of measures on $X_0 \times ... \times X_n$ that are disintegrable with respect to continuous kernels.

3 Kernel-based Beliefs

A kernel, ν, between X and Y is a natural probabilistic representation of Bernheim's justification procedure. Ann starts with a conjecture over X, the set of Bob's strategies. Such a conjecture is called a first-order belief. Bob conducts the same reasoning, and Y denotes the set of his first-order beliefs. In the next step, Ann fixes some strategy of Bob, x, and searches for Bob's conjectures that make x optimal. Kernel $\nu(x,.)$ represents that search. However, ν is not unique from the probabilistic perspective. Any kernel equivalent to ν with respect to Ann's first-order belief would represent the same reasoning of Ann. Hence, instead of defining Ann's second-order belief as a kernel, we should consider a family of equivalent kernels.

In fact, we restrict our attention to continuous kernels. There are two arguments supporting this additional requirement. First, continuity of kernel implies some form of consistency in Ann's reasoning. She does not "jump" in her pondering about Bob's choices – if two strategies, x and y, are close to each other, then we may expect that Bob's beliefs supporting them are not very different. In consequence, $\nu(x,.)$, Ann's conjecture given x, and $\nu(y,.)$, Ann's conjecture given y, should be close to each other. The second reason to impose continuity of kernels is more technical and will become clear when we define the convergence of kernel-based beliefs. Our discussion implies the following definition of kernel-based belief.

Definition 3.1 ***Kernel-based Belief***

Kernel-based belief between X and Y is a family of continuous kernels equivalent to continuous kernel ν with respect to some probability measure on X, λ^X.

Let $K(X;Y)$ denote the collection of kernel-based beliefs on $X \times Y$, with its generic element K. In the above definition, we say that the pair, (λ^X, ν), **generates** K and that λ^X is **associated with** K. The following result is a consequence of Lemma 2.1.

Corollary 3.1

Kernel-based belief, K, and its associated measure, λ^X, generate the unique and continuously disintegrable measure, $\lambda^{X\times Y}$, on X such that $\lambda^{X\times Y}$ and λ^X agree on X.

Corollary 3.1 suggests a natural definition of **coherency**.

Definition 3.2 ***Coherency***

1. *(λ^X, K) is coherent if λ^X is associated with K.*
2. *$(K^{0,\ldots,n-2;n-1}, K^{0,\ldots,n-1;n})$ is coherent if $K^{0,\ldots,n-2;n-1}$ together with its associated probability measure, $\lambda^{0,\ldots,n-2}$, generates $\lambda^{0,\ldots,n-1}$ that is associated with $K^{0,\ldots,n-1;n}$.*

Our definition naturally extends to hierarchies. $(\lambda^0, K^{0;1}, \ldots, K^{0,\ldots,n-1;n}, \ldots)$ is coherent if every pair of its adjacent elements is coherent. Such a hierarchy is characterized by a useful property.

Lemma 3.1

For a coherent $(\lambda^0, K^{0;1}, ..., K^{0,...,n-1;n}, ...)$ *there exists a unique sequence,* $(\lambda^0, \lambda^{0,1}, ...)$*, where* $\lambda^{0,...,n}$ *is a measure on* $X_0 \times ... \times X_n$*, and a unique* λ^∞ *on* $X_0 \times ... \times X_n \times ...$ *that, for each* n*, agrees with* $\lambda^{0,...,n}$ *on cylinders.*

The equivalency of kernels forces us to define a kernel-based belief as a family of kernels. Such a belief together with its associated measure yields unique probability on the product. However, associated measure is not unique. In the next lemma, we discover the relationship between kernel-based beliefs and absolute continuity of measures.[3] This result will be important in our construction of topology on $K(X;Y)$.

Lemma 3.2

λ^X *and* $\widetilde{\lambda}^X$ *are both associated with* K *if and only if* λ^X *and* $\widetilde{\lambda}^X$ *are absolutely continuous with respect to each other.*

Observe that a kernel-based belief, together with its associated measure, is a linear bounded operator on the space of real-valued, continuous bounded functions on $X \times Y$, $CB(X \times Y)$. We assign the weak-* topology to $K(X;Y)$. That is, $\{K_k\}$ converges to K if and only if there exists a sequence, $\{\lambda_k^X\}$, and λ^X such that each λ_k^X is associated with K_k, and λ^X is associated with K and

$$\int_X d\lambda_k^X \int_Y f\nu_k(x;dy) \to \int_X d\lambda^X \int_Y f\nu(x;dy), \text{ for every } f \in CB(X \times Y). \tag{5}$$

From the (generalized) Fubini Theorem, we know that $\int_X d\lambda_k^X \int_Y f\nu_k(x;dy) = \int_{X\times Y} fd\lambda_k^{X\times Y}$, where $\lambda_k^{X\times Y}$ is the unique measure constructed from (λ_k^X, K_k). Thus, the convergence of kernel-based beliefs is a weak-* convergence of the measures on $X \times Y$ generated by these beliefs and their associated probabilities.

However, as we discovered in Lemma 3.2, kernel-based belief does not have the unique associated measure. This observation raises a natural concern about our construction of convergence in $K(X;Y)$. What if, in (5), instead of $\{\lambda_k^X\}$ and λ^X, we have $\{\widetilde{\lambda}_k^X\}$ and $\widetilde{\lambda}^X$ such that λ_k^X and $\widetilde{\lambda}_k^X$ are absolutely continuous with respect to each other, and λ^X and $\widetilde{\lambda}^X$

[3]Let λ^X and $\widetilde{\lambda}^X$ be two probabilities on X. We say that λ^X is absolutely continuous with respect to $\widetilde{\lambda}^X$, denoted by $\lambda^X \ll \widetilde{\lambda}^X$, if $\lambda^X(E) = 0$ whenever $\widetilde{\lambda}^X(E) = 0$.

are absolutely continuous with respect to each other? It is possible that the convergence described in (5) does not hold for the replaced measures. Of course, this does not immediately mean that our definition is not appropriate. Suppose that $\{\widetilde{\lambda}_k^X\}$ does not weak-* converge to $\widetilde{\lambda}^X$. In this case, there is no reason to expect the convergence in (5). However, if $\{\widetilde{\lambda}_k^X\}$ does weak-* converge to $\widetilde{\lambda}^X$ and, nevertheless, convergence in (5) does not hold, then, clearly, we need to change the definition of convergence of kernel-based beliefs. Fortunately, as we show in the next lemma, our construction does not fail the consistency check.

Lemma 3.3

Suppose that $\{K_k\}$ converges to K for $\{\lambda_k^X\}$ and λ^X, as in (5). Then, the following are true:

1. *$\{\lambda_k^X\}$ weak-* converges to λ^X.*
2. *Suppose that $\{\widetilde{\lambda}_k^X\}$ weak-* converges to $\widetilde{\lambda}^X$. If $\widetilde{\lambda}_k^X$ is associated with K_k for each k, and $\widetilde{\lambda}^X$ is associated with K, then*

$$\int_X d\widetilde{\lambda}_k^X \int_Y f\nu_k(x;dy) \to \int_X d\widetilde{\lambda}^X \int_Y f\nu(x;dy), \text{ for every } f \in CB(X \times Y). \tag{6}$$

Lemma 3.3 supports our choice of topology assigned to $K(X;Y)$. We can rewrite the convergence condition in a way that appears more natural, as it does not rely on the specific sequence, $\{\lambda_k^X\}$, that we used in (5).

> *A sequence of kernel-based beliefs, $\{K_k\}$, converges to kernel-based belief, K, if and only if for any sequence, $\{\lambda_k^X\}$, converging to λ^X such that each λ_k^X is associated with K_k, and λ^X is associated with K, we have*

$$\int_X d\lambda_k^X \int_Y f\nu_k(x;dy) \to \int_X d\lambda^X \int_Y f\nu(x;dy), \text{ for every } f \in CB(X \times Y). \tag{7}$$

Analysis of the proof of Lemma 3.3 indicates why we restrict our attention to continuous kernels. With a discontinuous kernel, we cannot expect the function, $g_k := \int_Y f\nu_k(x;dy)$, to be continuous. This, in turn, prevents us from using the Portmanteau Theorem, which was essential in proving Lemma 3.3.

Lemma 3.3 not only serves as our argument for focusing only on continuous kernels, but also helps us to establish the equivalency result for the convergence of sequences of the form, $\{\lambda_k^0, K_k^{0;1}, ...\}$. Such a convergence is crucial for the construction of canonical homeomorphism in the next section.

Lemma 3.4

Let $\{(\lambda_k^0, K_k^{0;1})\}$ be a sequence of coherent pairs and let $(\lambda^0, K^{0;1})$ be a coherent pair. Let $\{\lambda_k^{0,1}\}$ be the sequence of unique measures induced by $\{(\lambda_k^0, K_k^{0;1})\}$ and let $\lambda^{0,1}$ be the unique measure induced by $(\lambda^0, K^{0;1})$. Then, $\{(\lambda_k^0, K_k^{0;1})\}$ converges to $(\lambda^0, K^{0;1})$ if and only if $\{\lambda_k^{0,1}\}$ weak- converges to $\lambda^{0,1}$.*

Our choice of topology for $K(X;Y)$ turns out to be very useful as it implies that $K(X;Y)$ is metrizable and separable space, a result that we will use in Lemma 4.1.

Lemma 3.5

If X and Y are Polish, then $K(X;Y)$ is metric separable.

4 Kernel-based Type Spaces

Let S be a Polish space endowed with Borel σ-algebra. This is an uncertainty space faced by the players. A **kernel-based type** is an infinite collection of kernel-based beliefs. In order to construct it, we inductively define spaces:

$\Omega_0 := S$

$\Omega_1 := P(S)$

$\Omega_2 := K(S; P(S)) = K(\Omega_0; \Omega_1)$

$\vdots$

$\Omega_n := K(\Omega_0 \times ... \times \Omega_{n-2}; \Omega_{n-1})$.

Let $W_0 := \underset{i=1}{\times} \Omega_i$ be the (canonical) space of kernel-based types, with generic element, $w := (\lambda^0, K^{0;1}, K^{0,1;2}, ...)$. Each $K^{0,...,n-1;n}$ is a family of equivalent kernels between $\Omega_0 \times ... \times \Omega_{n-1}$ and Ω_n. We defined coherency of a sequence, $(\lambda^0, K^{0;1}, K^{0,1;2}, ...)$, in the previous section. Let W_1 be the set of coherent types. Our construction of canonical homeomorphism uses the approach based on Proposition 2 in Brandenburger and Dekel (1993). The first step consists of showing that coherent types are homeomorphically mapped to the subset of beliefs on $S \times W_0$. Recall that kernel-based beliefs are constructed from continuous kernels. This implies that that we should not consider $P(S \times W_0)$, but rather $P_C(S \times W_0)$.

Lemma 4.1

If S is Polish, then there exists a homeomorphism, $p : W_1 \to P_C(S \times W_0)$.

Proof of Lemma 4.1.

1. *Existence, injectivity, and surjectivity of p.* Take $w \in W_1$. From Lemma 3.1, we know that w yields a unique probability, λ^∞, on $S \times W_0$, which we denote as $p(w)$. Note that existence of p does not require S to be even a topological space. This is distinct from the previously mentioned literature based on the Kolmogorov Extension Theorem, where topological assumptions are necessary for the existence of such a map. To show injectivity, take distinct w and $\widetilde{w}$. They generate not only unique $\lambda^\infty = p(w)$ and $\widetilde{\lambda}^\infty = p(\widetilde{w})$, respectively, but also unique sequences of measures $(\lambda^0, \lambda^{0,1}, ...)$ and $(\widetilde{\lambda}^0, \widetilde{\lambda}^{0,1}, ...)$, respectively. The difference between w and $\widetilde{w}$ occurs at some level of hierarchy. It means that either $\lambda^0 \neq \widetilde{\lambda}^0$ or $K^{0,...,n-1;n} \neq \widetilde{K}^{0,...,n-1;n}$ for some n. If the former is true, then $p(w) \neq p(\widetilde{w})$, as they do not agree on Ω_0. If the latter is true, then $\lambda^{0,...,n}$ and $\widetilde{\lambda}^{0,...,n}$ differ. Since w and $p(w)$ agree on cylinders, $p(w) \neq p(\widetilde{w})$. Surjectivity is a consequence of taking $P_C(S \times W_0)$, instead of $P(S \times W_0)$, as a range of p.

2. *Continuity of p.* Take a sequence, $\{w_k\}$, converging to w. That is, $(\lambda_k^0, K_k^{0;1}, K_k^{0,1;2}, ...) \to (\lambda_k^0, K_k^{0;1}, K_k^{0,1;2}, ...)$. Let $\lambda_k^{0,...,n}$ and $\lambda^{0,...,n}$ be probability measures on $\Omega_0 \times ... \times \Omega_n$ generated by the first n levels of $\{w_k\}$ and w, respectively. Let $\lambda_k^\infty = p(w_k)$ and $\lambda^\infty = p(w)$. Assuming that $\{w_k\} \to w$ implies, due to Lemma 3.4, that $\lambda_k^{0,...,n} \to \lambda^{0,...,n}$ for each n. To show that $\{p(w_k)\}$ weak-* converges to $p(w)$, we use the convergence-determining-class technique (see Chapters 1.2 and 1.3, especially Problem 7 on p. 22, in Billingsley (1968)): For a countable product of separable metric spaces, if $\lambda_k^\infty(C^n) \to \lambda^\infty(C^n)$ for each λ^∞-continuity cylinder set C^n, then $\{\lambda_k^\infty\}$ weak-* converges to λ^∞. In our case, separability is satisfied (see Lemma 3.5). Fix λ^∞-continuity cylinder, $C^n = B^{0,...,n} \times \Omega_{n+1} \times ...$, where $B^{0,...,n}$ is a measurable subset of $\Omega_0 \times ... \times \Omega_n$. Note that $\lambda^\infty(C^n) = \lambda^{0,...,n}(B^{0,...,n})$. Similarly, $\lambda_k^\infty(C^n) = \lambda_k^{0,...,n}(B^{0,...,n})$. Note that C^n is a λ^∞-continuity set if and only if $B^{0,...,n}$ is a $\lambda^{0,...,n}$-continuity set (Lemma 2.2.1b). Since $\{\lambda_k\}$ weak-* converges to λ_k, by the Portmanteau Theorem, $\lim_{k\to\infty} \lambda_k^{0,...,n}(B^{0,...,n}) = \lambda^{0,...,n}(B^{0,...,n})$. Thus, $\lim_{k\to\infty} \lambda_k^\infty(C^n) = \lambda^\infty(C^n)$, which implies that $\{p(w_k)\}$ weak-* converges to $p(w)$.

3. *Continuity of p^{-1}.* Take $\{\lambda_k^\infty\}$ weak-* convergent to λ^∞. $p^{-1}(\lambda_k^\infty)$ is a hierarchy $(\lambda_k^0, K_k^{0;1}, K_k^{0,1;2}, ...)$ that yields a unique hierarchy of marginals, $(\lambda_k^0, \lambda_k^{0,1}, \lambda_k^{0,1,2}, ...)$. From $p^{-1}(\lambda^\infty)$ we obtain unique $(\lambda^0, K^{0;1}, K^{0,1;2}, ...)$ and $(\lambda^0, \lambda^{0,1}, \lambda^{0,1,2}, ...)$. We want to show that weak-* convergence of $\{\lambda_k^\infty\}$ to λ^∞ implies the convergence of marginals.

From Lemma 3.4, we know that convergence of marginals is enough to establish the convergence of hierarchies, $(\lambda_k^0, K_k^{0;1}, K_k^{0,1;2}, ...) \to (\lambda^0, K^{0;1}, K^{0,1;2}, ...)$. Let $\lambda_k^{0,...,n}$ and $\lambda^{0,...,n}$ be these marginals on $\Omega_0 \times ... \times \Omega_n$. Take $B^{0,...,n}$, a measurable subset of $\Omega_0 \times ... \times \Omega_n$, such that it is a $\lambda^{0,...,n}$-continuity set. Hence, as we argued above, $C^n = B^{0,...,n} \times \Omega_{n+1} \times ...$ is a λ^∞-continuity set. By the Portmanteau Theorem, weak-* convergence $\lambda_k^\infty \to \lambda^\infty$ is equivalent with the convergence on λ^∞-continuity set: $\lim_{k\to\infty} \lambda_k^\infty(C^n) = \lambda^\infty(C^n)$. But $\lambda^\infty(C^n) = \lambda^{0,...,n}(B^{0,...,n})$ and for each k, $\lambda_k^\infty(C^n) = \lambda_k^{0,...,n}(B^{0,...,n})$. Hence, $\lim_{k\to\infty} \lambda_k^{0,...,n}(B^{0,...,n}) = \lambda^{0,...,n}(B^{0,...,n})$. Invoking again the Portmanteau Theorem, we establish that $\{\lambda_k^{0,...,n}\}$ weak-* converges to $\lambda^{0,...,n}$. Since marginals converge at each level n, it means that $(\lambda_k^0, K_k^{0;1}, ..., K_k^{0,1,...,n-1;n}) \to (\lambda^0, K^{0;1}, ..., K^{0,1,...,n-1;n})$ for each n. In consequence, $\{p^{-1}(\lambda_k^\infty)\}$ weak-* converges to $p^{-1}(\lambda^\infty)$. ■

Let $W_2 := \{w \in W_1 : p(w)(S \times W_1) = 1\}$ be the set of types that are not only coherent but also believe in coherency (i.e., these types assign measure one to the fact that coherency is satisfied). Inductively, we define $W_n := \{w \in W_{n-1} : p(w)(S \times W_{n-1}) = 1\}$ and $W := \cap W_n$ as the set of types that satisfy both coherency and the common belief of coherency. The relationship among sets W_n is captured in the following lemma.

Lemma 4.2

For each $n = 1, ...,$ W_n is a closed subset of W_{n-1}.

Proof of Lemma 4.2. First, we prove that W_1 is closed in W_0. Take a sequence of coherent types, $\{w_k\}$, converging to w. Each w_k generates hierarchy $(\lambda_k^0, \lambda_k^{0,1} ...)$. That is, from Lemma 3.4, we have convergence $(\lambda_k^0, \lambda_k^{0,1} ...) \to (\lambda^0, \lambda^{0,1} ...)$. Note that if $\{\lambda_k^{0,...,n}\}$ converges to $\lambda^{0,...,n}$, then the sequence of marginals generated by $\{\lambda_k^{0,...,n}\}$ converges to the marginal of $\lambda^{0,...,n}$. Hence, w is coherent, and it follows that W_1 is closed, which indicates that it is a measurable subset of W_0. This means that W_2 is well-defined; that is, requiring $p(w)(S \times W_1) = 1$ is a valid operation. Note that continuity of p implies that W_2 is closed in W_0. Consequently, by induction, each W_n is a closed subset of W_0. We prove that $\{W_n\}$ is a decreasing sequence by induction. We already established that $W_0 \supset W_1$. Suppose that $W_0 \supset W_1 \supset ... \supset W_n$ is true. To verify for $n+1$, take $w \in W_{n+1}$. That is, $p(w)(S \times W_n) = 1$. Since $W_n \supset W_{n-1}$, it implies that $p(w)(S \times W_{n-1}) = 1$, which makes w belong to W_n. Hence, $W_n \supset W_{n+1}$, as required. ■

In the next proposition, we demonstrate the existence of canonical homeomorphism.

Proposition 1

If S is Polish, then there exists a homeomorphism, $q : W \to P_C(S \times W)$.

Proof of Proposition 1. First, we must show that $W = \{w \in W_1 : p(w)(S \times W) = 1\}$. Take $w \in W$. In Lemma 4.2 we showed that $\{W_n\}$ is a decreasing sequence. Using that result and the continuity of probability measure, we deduce that $p(w)(S \times W) = p(w)(\bigcap_n (S \times W_n)) = \lim_{n\to\infty} p(w)(S \times W_n) = 1$. Hence, $W \subset \{w \in W_1 : p(w)(S \times W) = 1\}$. Next, take $w \in \{w \in W_1 : p(w)(S \times W) = 1\}$. Suppose there is m such that $p(w)(S \times W_m) < 1$. Since $\{W_n\}$ is a decreasing sequence, this implies that $p(w)(S \times W_k) < 1$ for all $k \geq m$. Using again the continuity of probability measure, we deduce that $p(w)(S \times W) < 1$, which is a contradiction. Hence, $p(w)(S \times W_n) = 1$ for all n, and $\{w \in W_1 : p(w)(S \times W) = 1\} \subset W$. From the fact that $W = \{w \in W_1 : p(w)(S \times W) = 1\}$, we derive that $p(W) = \{\lambda^\infty \in \widetilde{P}_C(S \times W_0) : \lambda^\infty(S \times W) = 1\}$. Since $\{\lambda^\infty \in \widetilde{P}_C(S \times W_0) : \lambda^\infty(S \times W) = 1\}$ and $\widetilde{P}_C(S \times W)$ are homeomorphic, and $p(W)$ and W are also homeomorphic, we deduce the desired relation. ■

5 Kernel-based and Standard Type Spaces

We inductively define the standard type:

$X_0 := S$

$X_1 := X_0 \times P(X_0)$

$\vdots$

$X_n := X_{n-1} \times P(X_{n-1})$.

We assume that S is Polish, which implies, due to Theorems 3.1 and 3.5 in Varadarajan (1958), that each X_n is Polish.

$T_0 := \underset{i=0}{\times} P(X_i)$ is the space of standard types with a generic element, $t = (\mu^0, \mu^{0,1}, ...)$. Each $\mu^{0,\ldots,n}$ is defined on a product of Polish spaces and, as such, is disintegrable, as in (3). However, such a type does not to have to be coherent, as in Brandenburger and Dekel (1993). Hence, we say that a type is **coherent** if disintegration of the n-level conjecture is conducted with respect to the $n-1$-level belief. If, in addition, the disintegration procedure yields continuous kernel, then we say that a type is **coherent and continuous**. Let T_1 be the set of standard types that satisfy coherency and continuity. As in the previous section, we inductively define $T_n := \{t \in T_1 : f(t)(S \times T_{n-1}) = 1\}$. Let $T := \cap T_n$ be the set of types that satisfy both (a) coherency and continuity, and (b) the common belief of coherency and continuity. We obtain the existence of homeomorphisms, $f : T_1 \to P_C(S \times T_0)$ and $g : T \to P_C(S \times T)$. The proofs are omitted since, under assumptions of coherency and continuity, they are virtually identical to the proofs of Lemma 4.1 and Proposition 1.

Lemma 5.1

If S is Polish, then there exists a homeomorphism, $f : T_1 \to P_C(S \times T_0)$.

Proposition 2

If S is Polish, then there exists a homeomorphism, $g : T \to P_C(S \times T)$.

The following commutative diagram describes the relationship between the kernel-based and standard types.

$$\begin{array}{ccc} W & \xrightarrow{\Psi} & T \\ {\scriptstyle q}\downarrow & & \downarrow{\scriptstyle g} \\ P_C(S \times W) & \xrightarrow[\phi]{} & P_C(S \times T) \end{array} \tag{8}$$

In Propositions 1 and 2, we established the existence of homeomorphisms, q and g. In Proposition 3, which concludes our paper, we establish that W and T are homeomorphic. Function ϕ is defined as $\phi := q^{-1} \circ \Psi \circ g$ which, given that it is a combination of homeomorphisms, is a homeomorphism.

Proposition 3

There exists a homeomorphism, $\Psi : W \to T$.

Proof of Proposition 3. First, we build a homeomorphic function, $\Psi : W_1 \to T_1$. Take $w \in w_1$. Let ψ_0 be a projection of w on $P(S)$. In other words, $\psi_0(w)$ is the first-level belief. Let $\psi_{0,1}(w)$ be the unique probability on $S \times P(S) = P(X_1)$ induced by the first two levels of w. We know that $\psi_{0,1}(w) \in \widetilde{P}_C(S \times P(S))$ is unique, since w is coherent. In general, $\psi_{0,\ldots,n}$ assigns to w unique probability on X_n. Since ψ_0 is the identity function, it is homeomorphic. According to Lemma 3.4, each $\psi_{0,\ldots,n}$ is homeomorphic as well. Next, define $\Psi := (\psi_0(w), \psi_{0,1}(w), ..., \psi_{0,...,n}(w), ...)$. Each component of Ψ is homeomorphic, which implies that Ψ is homeomorphic as well. Next, we want to show that $\Psi(W_n) = T_n$ for each n. We already established that this claim is true for $n = 1$ and we assume that this claim is also true for n. In order to verify the claim for $n + 1$, take $w \in W_{n+1}$. By assumption, w assigns measure one to $S \times W_n$. Since Ψ is homeomorphic, a measure associated with $\Psi(w)$ on $S \times T_1$ assigns one to $S \times \Psi(W_n)$. But this means that $S \times T_n$ is assigned probability one by $\Psi(w)$. Hence, $\Psi(w) \in T_{n+1}$. In order to prove that t, which belongs to T_{n+1}, belongs to $\Psi(W_{n+1})$, we again follow reasoning based on the fact that W_n and T_n are homeomorphic. Finally, given that $\Psi(W_n) = T_n$ for each n and that Ψ is a homeomorphism, we note that $T = \cap T_n = \cap \Psi(W_n) = \Psi(\cap W_n) = \Psi(W)$. Thus, Ψ carries W homeomorphically to T. ■

6 Appendix

Proof of Lemma 3.1. Take coherent $(\lambda^0, K^{0;1}, ..., K^{0,...,n-1;n}, ...)$. First, consider $(\lambda^0, K^{0;1})$. By assumption λ^0 is associated with $K^{0;1}$. Hence, by Corollary 3.1, we know that $(\lambda^0, K^{0;1})$ generates a unique $\lambda^{0,1}$. By assumption, $\lambda^{0,1}$ is associated with $K^{0,1;2}$, and together they generate a unique $\lambda^{0,1,3}$. The reasoning continues ad infinitum. To show the uniqueness of λ^∞, take $(\lambda^0, \nu^{0;1}, \nu^{0,1;2}, ..., \nu^{0,...,n-1;n}, ...)$ where $\nu^{0,...,n-1;n}$ is some kernel that belongs to $K^{0,...,n-1;n}$. By the already mentioned Ionescu-Tulcea Theorem, such a sequence generates a unique λ^∞ on $X_0 \times ... \times X_n \times ....$ Next, take different $(\lambda^0, \widetilde{\nu}^{0;1}, \widetilde{\nu}^{0,1;2}, ..., \widetilde{\nu}^{0,...,n-1;n}, ...)$ where $\widetilde{\nu}^{0,...,n-1;n}$ belongs to $K^{0,...,n-1;n}$. Let $\widetilde{\lambda}^\infty$ denote the unique measure on $X_0 \times ... \times X_n \times ...$ generated by this sequence. Let $\lambda^{0,...,n}$ be the unique measure generated by $(\lambda^0, \nu^{0;1}, ..., \nu^{0,...,n-1;n})$. By what we just discussed, $(\lambda^0, \widetilde{\nu}^{0;1}, ..., \widetilde{\nu}^{0,...,n-1;n})$ generates the same $\lambda^{0,...,n}$. Since the measure on $X_0 \times ... \times X_n \times ...$ is uniquely determined by cylinders (Lemma 2.2.3), we deduce that $\lambda^\infty = \widetilde{\lambda}^\infty$. ■

Proof of Lemma 3.2. First, suppose that λ^X and $\widetilde{\lambda}^X$ are both associated with K. Assume also that there is E such that $\lambda^X(E) = 0$ while $\widetilde{\lambda}^X(E) > 0$. Take $\nu \in K$. Let $\widetilde{\nu}$ be a kernel that is identical with ν except for subset E. Hence, $\widetilde{\nu}$ and ν are equivalent with respect to λ^X while they are not equivalent with respect to $\widetilde{\lambda}^X$. This is a contradiction. Next, suppose that λ^X and $\widetilde{\lambda}^X$ are absolutely continuous with respect to each other. Take ν and $\widetilde{\nu}$, which are equivalent with respect to λ^X. That is, these kernels are identical except for set E such that $\lambda^X(E)$. Since $\widetilde{\lambda}^X(E) = 0$ as well, we conclude that ν and $\widetilde{\nu}$ are equivalent with respect to $\widetilde{\lambda}^X$ as well. ■

Proof of Lemma 3.3. The first claim is a direct consequence of combining Lemma

2.2.1a with the fact that a kernel and its associated measure generate the unique measure on the product. In order to prove the second claim, take $\int_X d\widetilde{\lambda}_k^X \int_Y f\nu_k(x;dy)$. Denote $g_k := \int_Y f\nu_k(x;dy)$. According to Lemma 2.3, g_k is bounded continuous. Similarly, $g := \int_Y f\nu(x;dy)$, which also belongs to $CB(X)$. We rewrite $\int_X d\widetilde{\lambda}_k^X \int_Y f\nu_k(x;dy) = \int_X g_k d\widetilde{\lambda}_k^X$ and $\int_X d\widetilde{\lambda}^X \int_Y f\nu(x;dy) = \int_X g d\widetilde{\lambda}^X$ and obtain:

$$\begin{aligned}\int_X g_k d\widetilde{\lambda}_k^X - \int_X g d\widetilde{\lambda}^X &= \left(\int_X g_k d\widetilde{\lambda}_k^X - \int_X g_k d\widetilde{\lambda}^X\right) + \left(\int_X g_k d\widetilde{\lambda}^X - \int_X g d\widetilde{\lambda}^X\right)\\ &= \left(\int_X g_k d\widetilde{\lambda}_k^X - \int_X g_k d\widetilde{\lambda}^X\right) + \int_X (g_k - g)\, d\widetilde{\lambda}^X.\end{aligned}$$

Since $\{\lambda_k^X\}$ weak-* converges to λ^X, we know that, due to the Portmanteau Theorem, $\int_X h d\widetilde{\lambda}_k^X \to \int_X h d\widetilde{\lambda}^X$ for all $h \in CB(X)$. Since each g_k and g are in $CB(X)$, we deduce that $\int_X g_k d\widetilde{\lambda}_k^X - \int_X g_k d\widetilde{\lambda}^X$ converges to zero. Second, we want to show that $\int_X (g_k - g)\, d\widetilde{\lambda}^X$ converges to zero as well. By assumption, $\int_X g_k d\lambda_k^X \to \int_X g d\lambda^X$. We manipulate the terms as follows:

$$\begin{aligned}\int_X g_k d\lambda_k^X - \int_X g d\lambda^X &= \left(\int_X g_k d\lambda_k^X - \int_X g_k d\lambda^X\right) + \left(\int_X g_k d\lambda^X - \int_X g d\lambda^X\right)\\ &= \left(\int_X g_k d\lambda_k^X - \int_X g_k d\lambda^X\right) + \int_X (g_k - g)\, d\lambda^X.\end{aligned}$$

As we argued above, $\{\lambda_k^X\}$ weak-* converges to λ^X. This implies that $\int_X g_k d\lambda_k^X - \int_X g_k d\lambda^X$ converges to zero. Since $\int_X g_k d\lambda_k^X \to \int_X g d\lambda^X$, it must be true that $\int_X (g_k - g)\, d\lambda^X$ converges to zero as well. Hence, for large enough k, it is true that $\int_X (g_k - g)\, d\lambda^X \le \int_X \frac{1}{k} d\lambda^X = \frac{1}{k}$. Now, we go back to $\int_X (g_k - g)\, d\widetilde{\lambda}^X$. Due to the Radon-Nikodyn Theorem, for each k, there exists a function, ρ_k, such that (a) $\widetilde{\lambda}_k^X(E) = \int_E \rho_k d\lambda_k^X$ for each measurable E, and (b) $\int_X \gamma d\widetilde{\lambda}_k^X = \int_X \gamma\rho_k d\lambda_k^X$ for each measurable γ. In consequence, we know that $\int_X (g_k - g)\, d\widetilde{\lambda}^X = \int_X (g_k - g)\, \rho_k d\lambda^X$. But this implies that for large k, $\int_X (g_k - g)\, \rho_k d\lambda^X \le \int_X \frac{1}{k}\rho_k d\lambda^X = \frac{1}{k}\int_X \rho_k d\lambda^X = \frac{1}{k}$ as $\int_X \rho_k d\lambda^X = \widetilde{\lambda}^X(X) = 1$. That is, $\int_X (g_k - g)\, \rho_k d\lambda^X$ converges to zero. ■

Proof of Lemma 3.4. First, suppose that $\{\lambda_k^0, K_k^{0;1}\}$ converges to $\{\lambda^0, K^{0;1}\}$. This means that $\{\lambda_k^0\}$ weak-* converges to λ^0, and $\{K_k^{0;1}\}$ converges to $K^{0;1}$. By Lemma 3.3, the latter means that

$$\int_X d\lambda_k^X \int_Y f\nu_k(x;dy) \to \int_X d\lambda^X \int_Y f\nu(x;dy), \text{ for every } f \in CB(X \times Y).$$

In other words, $\{\lambda_k^{0,1}\}$ weak-* converges to $\lambda^{0,1}$. Next, suppose that $\{\lambda_k^{0,1}\}$ weak-* converges to $\lambda^{0,1}$. This indicates that $\{\lambda_k^0\}$ weak-* converges to λ^0 (see Lemma 2.2.2a). Take sequence $\{K_k^{0;1}\}$ and belief $K^{0;1}$. According to the definition of convergence of kernel-based belief (equation (5)), convergence of $\{\lambda_k^{0,1}\}$ implies that $\{K_k^{0;1}\}$ converges to $K^{0;1}$. ■

Proof of Lemma 3.5. First, we note that $P(X \times Y)$ is metric separable (see Theorem 3.1 in Varadarajan (1958)). Hence, $P_C(X \times Y)$, a subset of $P(X \times Y)$, is also metric separable. We need to find a continuous and surjective map between a subset of $P_C(X \times Y)$ and $K(X;Y)$. We are interested in the subset of $P_C(X \times Y)$, since each K, together with its associated measure, generates continuously disintegrable probability. We build this subset in the following way. To each K we assign one of its associated measures, λ^X. Together, a pair, (K, λ^X), generates the unique measure, $\lambda^{X \times Y}$. Collection of all such measures constitutes subset $\hat{P}_C(X \times Y)$ of $P_C(X \times Y)$. Let $\theta : \hat{P}_C(X \times Y) \to K(X;Y)$ be a natural function that assigns to each $\lambda^{X \times Y} \in \hat{P}(X \times Y)$ a kernel K that generated it. Such a function is surjective. To prove continuity of θ, suppose that $\lambda_k^{X \times Y} \to \lambda^{X \times Y}$. Hence, for every $f \in CB(X \times Y)$, $\int_{X \times Y} f d\lambda_k^{X \times Y} \to \int_{X \times Y} f d\lambda^{X \times Y}$. Since all measures in $\hat{P}_C(X \times Y)$ are continuously disintegrable, we can rewrite the convergence as $\int_X d\lambda_k^X \int_Y f\nu_k(x;dy) \to \int_X d\lambda^X \int_Y f\nu(x;dy)$, where ν_k belongs to the kernel-based belief that formed $\lambda_k^{X \times Y}$, $\theta(\lambda_k^{X \times Y})$. The same holds for ν, which is a member of $\theta(\lambda^{X \times Y})$. Hence, by definition of convergence of kernel-based beliefs, we deduce that $\theta(\lambda_k^{X \times Y}) \to \theta(\lambda^{X \times Y})$. ■

References

Bernheim, B. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–1028.

Billingsley, P. (1968): *Convergence of Probability Measures*, New York, NY: John Wiley & Sons.

Brandenburger, A. (2008): "Epistemic Game Theory: Complete Information," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Palgrave Macmillan, second ed.

Brandenburger, A. and E. Dekel (1993): "Hierarchies of Beliefs and Common Knowledge," *Journal of Economic Theory*, 59, 189–198.

Denkowski, Z., S. Migórski, and N. Papageorgiou (2003): *An Introduction to Nonlinear Analysis: Theory*, New York, NY: Kluwer Academic/Plenum Publishers.

Friedenberg, A. (2010): "When Do Type Structures Contain All Hierarchies of Beliefs?" *Games and Economic Behavior*, 68, 108–129.

Fristedt, B. E. and L. F. Gray (1997): *A Modern Approach to Probability Theory*, Cambridge, MA: Birkhauser Boston.

Harsanyi, J. C. (1967/8): "Games with Incomplete Information Played by "Bayesian" Players, I-III." *Management Science*, 14, 159–182, 320–334, 320–334.

Heifetz, A. (1993): "The Bayesian Formulation of Incomplete Information — The Non-Compact Case," *International Journal of Game Theory*, 21, 329–338.

——— (2008): "Epistemic Game Theory: Incomplete Information," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Palgrave Macmillan, second ed.

Heifetz, A. and D. Samet (1999): "Coherent Beliefs Are Not Always Types," *Journal of Mathematical Economics*, 32, 475–488.

Mertens, J.-F. and S. Zamir (1985): "Formulation of Bayesian Analysis for Games with Incomplete Information," *International Journal of Game Theory*, 14, 1–29.

Myerson, R. B. (2004): "Comments on "Games with Incomplete Information Played by "Bayesian" Players, I-III"," *Management Science*, 50, 1818–1824.

Pearce, D. (1984): "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029–1050.

Shiryaev, A. (1996): *Probability*, New York, NY: Springer-Verlag.

Siniscalchi, M. (2008): "Epistemic Game Theory: Beliefs and Types," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, New York, NY: Palgrave Macmillan, second ed.

Varadarajan, V. (1958): "Weak Convergence of Measures on Separable Metric Spaces," *Sankhya*, 19, 15–22.