Epistemic Game Theory: Beliefs and Types

Marciano Siniscalchi*

March 28, 2007

1 Introduction

John Harsanyi [19] introduced the formalism of type spaces to provide a simple and parsimonious representation of belief hierarchies. He explicitly noted that his formalism was not limited to modeling a player's beliefs about payoff-relevant variables: rather, its strength was precisely the ease with wich Ann's beliefs about Bob's beliefs about payoff variables, Ann's beliefs about Bob's beliefs about Ann's beliefs about payoff variables, etc. could be represented.

This feature plays a prominent role in the epistemic analysis of solution concepts (see the article by Adam Brandenburger elsewhere in this volume), as well as in the literature on global games (Morris and Shin [25]) and on robust mechanism design (Bergemann and Morris [7]). All these applications place particular emphasis on the expressiveness of the type-space formalism. Thus, a natural question arises: just how expressive is Harsanyi's approach?

For instance, solution concepts such as Nash equilibrium or rationalizability can be characterized by means of restrictions on the players' mutual beliefs. In principle, these assumptions could be formulated directly as restrictions on players' hierarchies of beliefs; but, in practice, the analysis is mostly carried out in the context of a type space à la Harsanyi. This is without loss of generality only if Harsanyi type spaces do not themselves impose restrictions on the belief hierarchies that can be represented. Similar considerations apply in the context of robust mechanism design.

A rich literature addresses this issue from different angles, and for a variety of basic representations of beliefs. This article focuses on hierarchies of probabilistic beliefs; however, some extensions are also mentioned. For simplicity, attention is restricted to two players, denoted "1" and "2" or "i" and "$-i$."

*Economics Department, Northwestern University, Evanston, IL 60208-2600. Email: marciano@northwestern.edu. I thank Pierpaolo Battigalli and Adam Brandenburger for helpful discussion.

2 Probabilistic Type Spaces and Belief Hierarchies

Begin with some mathematical preliminaries. A topology on a space X is deemed Polish if it is separable and completely metrizable; in this case, X is itself deemed a Polish space. Examples include finite sets, Euclidean space $\mathbb{R}^n$ and closed subsets thereof. A countable product of Polish spaces, endowed with the product topology, is itself Polish. For any topological space X, the notation $\Delta(X)$ indicates the set of Borel probability measures on X. If the topology on X is Polish, then the weak* topology on $\Delta(X)$ is also Polish (e.g. Aliprantis and Border [4, Theorem 14.15]). A sequence $\{\mu^k\}_{k\geq 1}$ in $\Delta(X)$ converges in the weak* sense to a measure $\mu \in \Delta(X)$, written $\mu^k \xrightarrow{w^*} \mu$, if and only if, for every bounded, continuous function $\psi : X \to \mathbb{R}$, $\int_X \psi d\mu^k \to \int_X \psi d\mu$. The weak* topology on $\Delta(X)$ is especially meaningful and convenient when X is a Polish space: see [4, Chap. 14] for an overview of its properties. Finally, if μ is a measure on some product space $X \times Y$, the marginal of μ on X is denoted $\operatorname{marg}_X \mu$.

The basic ingredient of the players' hierarchical beliefs is a description of payoff-relevant or fundamental uncertainty. Fix two sets S_1 and S_2, hereinafter called the *uncertainty domains*; the intended interpretation is that S_{-i} describes aspects of the strategic situation that Player i is uncertain about. For example, in an independent private-values auction, each set S_i could represent bidder i's possible valuations of the object being sold, which is not known to bidder $-i$. In the context of interactive epistemology, S_i is usually taken to be Player i's strategy space. It is sometimes convenient to let $S_1 = S_2 \equiv S$; in this case, the formalism introduced below enables one to formalize the assumption that each player observes different aspects of the common uncertainty domain S (for instance, different signals correlated with the common, unknown value of an object offered for sale).

An (S_1,S_2)-*based type space* is a tuple $\mathcal{T} = (T_i, g_i)_{i=1,2}$ such that, for each $i = 1,2$, T_i is a Polish space and $g_i : T_i \to \Delta(S_{-i} \times T_{-i})$ is continuous. As noted above, type spaces can represent hierarchies of beliefs; it is useful to begin with an example. Let $S_1 = S_2 = \{a,b\}$ and consider the type space defined in Tab. 1. To interpret, for every $i = 1,2$, the entry in the row corresponding to t_i and (s_{-i}, t_{-i}) is $g_i(t_i)(\{(s_{-i}, t_{-i})\})$. Thus, for instance, $g_1(t_1)(\{(a, t_2')\}) = 0$; $g_2(t_2)(\{b\} \times T_1) = 0.5$.

T_1	a, t_2	a, t_2'	b, t_2	b, t_2'
t_1	1	0	0	0
t_1'	0	0.3	0	0.7

T_2	a, t_1	a, t_1'	b, t_1	b, t_1'
t_2	0	0.5	0.5	0
t_2'	0	0	0	1

Table 1: A type space

Consider type t_1 of Player 1. She is certain that $s_2 = a$; furthermore, she is certain that Player 2 believes that $s_1 = a$ and $s_1 = b$ are equally likely. Taking this one step further, type t_1 is certain that Player 2 assigns probability 0.5 to the event that Player 1 believes that $s_2 = b$ with probability 0.7.

These intuitive calculations can be formalized as follows. Fix an (S_1,S_2)-based type space $\mathcal{T} = (T_i, g_i)_{i=1,2}$;

for every $i = 1,2$, define the set X^0_{-i} and the function $h^1_i : T_i \to \Delta(X_0^{-i})$ by

$$X^0_{-i} = S_{-i} \quad \text{and} \quad \forall t_i \in T_i,\ h^1_i(t_i) = \text{marg}_{S_{-i}} g_i(t_i). \tag{1}$$

Thus, $h^1_i(t_i)$ represents the *first-order beliefs* of type t_i in type space $\mathscr{T}$—her beliefs about the uncertainty domain S_{-i}. Note that each $X^0_{-i} = S_{-i}$ is Polish. Proceeding inductively, assuming that $X^0_{-i},\ldots,X^{k-1}_{-i}$ and $h^1_i,\ldots,h^k_i$ have been defined up to some $k>0$ for $i=1,2$, and that all sets X^ℓ_{-i}, $\ell = 0,\ldots,k-1$ are Polish, define the set X^k_{-i} and the functions $h^{k+1}_i : T_i \to \Delta(X^k_{-i})$ for $i=1,2$ by

$$X^k_{-i} = X^{k-1}_{-i} \times \Delta(X^{k-1}_i) \quad \text{and} \quad \forall t_i \in T_i,\ h^{k+1}_i(t_i)(E) = g_i(t_i)\Big(\big\{(s_{-i},t_{-i}) \in S_{-i} \times T_{-i} : (s_{-i}, h^k_{-i}(t_{-i})) \in E\big\}\Big) \tag{2}$$

for every Borel subset E of X^k_{-i}. Thus, $h^2_1(t_1)$ represents the *second-order beliefs* of type t_1—her beliefs about *both* the uncertainty domain $S_2 = X^0_2$ and Player 2's beliefs about S_1, which by definition belong to the set $\Delta(X^0_1) = \Delta(S_1)$. Similarly, $h^{k+1}_i(t_i)$ represents type t_i's $(k+1)$-th order beliefs.

Observe that type t_1's second-order beliefs are defined over $X^0_2 \times \Delta(X^0_1) = S_2 \times \Delta(S_1)$, rather than just over $\Delta(X^0_1) = \Delta(S_1)$; a similar statement holds for her $(k+1)$-th order beliefs. This is crucial in many applications. For instance, a typical assumption in the literature on epistemic foundations of solution concepts is that Player 1 believes that Player 2 is rational. Letting S_i be the set of actions or strategies of Player i in the game under consideration, this can be modeled by assuming that the support of $h^2_1(t_1)$ consists of pairs $(s_2, \mu_1) \in S_2 \times \Delta(S_1)$ wherein s_2 is a best response to μ_1. Clearly, such an assumption could not be formalized if $h^2_1(t_1)$ only conveyed information about type t_1's beliefs on Player 2's first-order beliefs: even though type t_1's beliefs about the action played by Player 2 could be retrieved from $h^1_1(t_1)$, it would be impossible to tell whether each action that type t_1 expects to be played is matched with a belief that rationalizes it.

Note that, since X^{k-1}_i and X^{k-1}_{-i} are assumed Polish, so are $\Delta(X^{k-1}_i)$ and X^k_{-i}. Also, each function h^k_i is continuous.

Finally, it is convenient to define a function that associates to each type $t_i \in T_i$ an entire *belief hierarchy*: to do so, define the set H_i and, for $i=1,2$, the function $h_i : T_i \to H_i$ by

$$H_i = \prod_{k \geq 0} \Delta(X^k_{-i}) \quad \text{and} \quad \forall t_i \in T_i,\ h_i(t_i) = \big(h^1_i(t_i),\ldots,h^{k+1}_i(t_i),\ldots\big). \tag{3}$$

Thus, H_i is the set of all hierarchies of beliefs; notice that, since each X^k_{-i} is Polish, so is H_i.

3 Rich Type Spaces

The preceding construction suggests a rather direct way to ask how expressive Harsanyi's notion of a type space is: can one construct a type space that generates *all* hierarchies in H_i?

A moment's reflection shows that this question must be refined. Fix a type space $(T_i, g_i)_{i=1,2}$ and a type $t_i \in T_i$; recall that, for reasons described above, the first- and second-order beliefs of type t_i satisfy $h_i^1(t_i) \in \Delta(S_{-i})$ and $h_i^2(t_i) \in \Delta(X_{-i}^0 \times \Delta(X_i^0)) = \Delta(S_{-i} \times \Delta(S_i))$ respectively. This, however, creates the potential for redundancy or even contradiction, because both $h_i^1(t_i)$ and $\text{marg}_{S_{-i}} h_i^2(t_i)$ can be viewed as "type t_i's beliefs about S_{-i}." A similar observation applies to higher-order beliefs. Fortunately, it is easy to verify that, for every type space $(T_i, g_i)_{i=1,2}$ and type $t_i \in T_i$, the following *coherency* condition holds:

$$\forall k > 1, \quad \text{marg}_{X_{-i}^{k-2}} h_i^k(t_i) = h_i^{k-1}(t_i); \tag{4}$$

to interpret, recall that $h_i^k(t_i) \in \Delta(X_{-i}^{k-1}) = \Delta(X_{-i}^{k-2} \times \Delta(X_{-i}^{k-2}))$. Thus, in particular, $\text{marg}_{S_{-i}} h_i^2(t_i) = h_i^1(t_i)$.

Since H_i is defined as the set of *all* hierarchies of beliefs for Player i, some (in fact, "most") of its elements are not coherent. As noted above, no type space can generate incoherent hierarchies; more importantly, coherency can be viewed as an integral part of the interpretation of interactive beliefs. How could an individual simultaneously hold (infinitely) many distinct first-order beliefs? Which of these should be used, say, to verify whether she is rational? This motivates restricting attention to coherent hierarchies, defined as follows:

$$H_i^c = \left\{ (\mu_i^1, \mu_i^2, \ldots) \in H_i \, : \, \forall k > 1, \, \text{marg}_{X_{-i}^{k-2}} \mu_i^k = \mu_i^{k-1} \right\}. \tag{5}$$

Since $\text{marg}_{X_{-i}^{k-2}} : \Delta(X_{-i}^{k-1}) \to \Delta(X_{-i}^{k-2})$ is continuous, H_i^c is a closed, hence Polish subspace of H_i.

Brandenburger and Dekel [10, Proposition 1] show that there exist homeomorphisms $g_i^c : H_i^c \to \Delta(S_{-i} \times H_{-i})$: that is, *every coherent hierarchy corresponds to a distinct belief over the uncertainty domain and the hierarchies of the opponent, and conversely.* Furthermore, this homeomorphism is canonical, in the following sense. Note that $S_{-i} \times H_{-i} = S_{-i} \times \prod_{k \geq 0} \Delta(X_i^k) = X_{-i}^k \times \prod_{\ell > k} \Delta(X_i^\ell)$. Then it can be shown that, if $\mu_i = (\mu_i^1, \mu_i^2, \ldots) \in H_i^c$, then $\text{marg}_{X_{-i}^k} g_i^c(\mu_i) = \mu_i^{k+1}$. Intuitively, the marginal belief associated with μ_i over the first k orders of the opponent's beliefs is precisely what it should be, namely μ_i^{k+1}. The proof of these results builds upon Kolmogorov's Extension Theorem, as may be suggested by the similarity of the coherency condition in Eq. (5) with the notion of Kolmogorov consistency: cf. e.g. [4, Theorem 14.26].

This result does not quite imply that all coherent hierarchies can be generated in a suitable type space; however, it suggests a way to obtain this result. Notice that the belief on $S_{-i} \times H_{-i}$ associated by the homeomorphism g_i^c to a coherent hierarchy μ_i may include *in*coherent hierarchies $\nu_{-i} \in H_{-i} \setminus H_{-i}^c$ in its support. This can be interpreted in the following terms: if Player i's hierarchical beliefs are given by μ_i, then she is coherent, but she is not certain that her opponent is. On the other hand, consider a type space $(T_i, g_i)_{i=1,2}$; as noted above, for every player i, each type $t_i \in T_i$ generates a coherent hierarchy $h_i(t_i) \in H_i^c$. So, for instance, if (s_1, t_1) is in the support of $g_2(t_2)$, then t_1 also generates a coherent hierarchy. Thus, not only is type t_2 of Player 2 coherent: he is also certain (believes with probability one) that Player

1 is coherent. Iterating this argument suggests that *hierarchies of beliefs generated by type spaces display common certainty of coherency.*

Motivated by these considerations, let

$$H_i^0 = H_i^c \quad \text{and} \quad \forall k > 0,\ H_i^k = \{\mu_i \in H_i^{k-1} : g_i^c(\mu_i)(S_{-i} \times H_{-i}^{k-1}) = 1\}. \tag{6}$$

Thus, H_i^0 is the set of coherent hierarchies for Player i; H_i^1 is the set of hierarchies that are coherent and correspond to beliefs that display certainty of the opponent's coherency; and so on. Finally, let $H_i^* = \bigcap_{k\geq 0} H_i^k$. Each element of H_i^* is intuitively consistent with coherency and common certainty of coherency.

Brandenburger and Dekel [10, Proposition 2] show that the restriction g_i^* of g_i^c to H_i^* is a homeomorphism between H_i^* and $\Delta(S_{-i} \times H_{-i}^*)$; furthermore, it is canonical in the sense described above. This implies that the tuple $(H_i^*, g_i^*)_{i=1,2}$ is a type space in its own right—the (S_1, S_2)-based ***universal type space***.

The existence of a universal type space fully addresses the issue of richness. Since the homeomorphism g_i^* is canonical, it is easy to see that the hierarchy generated as per Eqs. (1) and (2) by any "type" $t_i = (\mu^1, \mu^2, \ldots) \in H_i^*$ in the universal type space $(H_i^*, g_i^*)_{i=1,2}$ is t_i itself; thus, since H_i^* *consists* of all hierarchies that are coherent and display common certainty of consistency, the universal type space also *generates* all such hierarchies.

The type space $(H_i^*, g_i^*)_{i=1,2}$ is rich in two additional, related senses. First, as may be expected, every belief hierarchy for Player i generated by an arbitrary type space is an element of H_i^*; this implies that every type space $(T_i, g_i)_{i=1,2}$ can be uniquely embedded in $(H_i^*, g_i^*)_{i=1,2}$ as a "belief-closed" subset: see Battigalli and Siniscalchi [5, Proposition 3.8]. Call a type space ***terminal*** if, like $(H_i^*, g_i^*)_{i=1,2}$, it embeds all other type spaces as belief-closed subsets.

Second, since each function g_i^* is a homeomorphism, in particular it is a surjection (i.e. onto). Call a type space $(T_i, g_i)_{i=1,2}$ ***complete*** if every map g_i is onto. (This should not be confused with the topological notion of completeness). Thus, the universal type space $(H_i^*, g_i^*)_{i=1,2}$ is complete. It is often the case that, when a universal type space is employed in the epistemic analysis of solution concepts, the objective is precisely to exploit its completeness. Furthermore, for certain representations of beliefs, it is not known whether universal type spaces can be constructed; however, the existence of complete type spaces can be established, and is sufficient for the purposes of epistemic analysis. The next Section provides examples.

4 Alternative Constructions and Extensions

The preceding discussion adopts the approach proposed by Brandenburger and Dekel [10], which has the virtue of relying on familiar ideas from the theory of stochastic processes. However, the first constructions of universal type spaces consisting of hierarchies of beliefs are due to Armbruster and Böge [2], Böge and Eisele [9] and Mertens and Zamir [24].

From a technical point of view, Mertens and Zamir [24] assume that the state space S is compact Hausdorff and beliefs are regular probability measures. Heifetz and Samet [21] instead drop topological assumptions altogether: in their approach, both the underlying set of states and the sets of types of each player are modeled as measurable spaces. They show that a terminal type space can be explicitly constructed in this environment.

In all the contributions mentioned so far, beliefs are modeled as countably additive probabilities. The literature has also examined other representations of beliefs, broadly defined.

A *partitional structure* (Aumann [3]) is a tuple $(\Omega,(\sigma_i,P_i)_{i=1,2})$, where Ω is a (typically finite) space of "possible worlds," every $\sigma_i:\Omega\to S_i$ indicates the realization of the basic uncertainty corresponding to each element of Ω, and every P_i is a partition of Ω. The interpretation is that, at any world $\omega\in\Omega$, Player i is only informed that the true world lies in the cell of the partition P_i containing ω, denoted $P_i(\omega)$. The *knowledge operator* for Player i can then be defined as

$$\forall E\subset\Omega,\quad K_i(E)=\{\omega\in\Omega:P_i(\omega)\subseteq E\}.$$

Notice that no probabilistic information is provided in this environment (although it can be easily added).

Heifetz and Samet [20] show that a terminal partitional structure does not exist. This result was extended to more general "possibility" structures by Meier [23]. Brandenburger and Keisler [12] establish related non-existence results for complete structures. However, recent contributions show that topological assumptions, which play a key role in the constructions of Mertens and Zamir [24] and Brandenburger and Dekel [10], can also deliver existence results in non-probabilistic settings. For instance, Mariotti, Meier and Piccione [22] construct a structure that is universal, complete and terminal for possibility structures.

Other authors investigate richer probabilistic representations of beliefs. Battigalli and Siniscalchi [5] construct a universal, terminal, and complete type space for *conditional probability system*, or collections of probability measures indexed by relevant conditioning events (such as histories in an extensive game) and related by a version of Bayes' Rule. This type space is used in [6] to provide an epistemic analysis of forward induction. Brandenburger, Friedenberg and Keisler [11] construct a complete type space for *lexicographic sequences*, which may be thought of as an extension of lexicographic probability systems (Blume, Brandenburger and Dekel [8]) for infinite domains. They then use it to provide an epistemic characterization of iterated admissibility.

Non-probabilistic representations of beliefs that reflect a concern for ambiguity (Ellsberg [14]) have also been considered. Heifetz and Samet [21] observe that their measure-theoretic construction extends to beliefs represented by continuous *capacities*, i.e. non-additive set functions that preserve monotonicity with respect to set inclusion. Motivated by the multiple-priors model of Gilboa and Schmeidler [17], Ahn [1] constructs a universal type space for sets of probabilities.

Epstein and Wang [15] approach the richness issue taking *preferences*, rather than beliefs, as primitive objects. In their setting, an S-based type space is a tuple $(T_i, g_i)_{i=1,2}$, where, for every type t_i, $g_i(t_i)$ is a suitably regular preference over *acts* defined on the set $S \times T_{-i}$. The analysis in the preceding section can be viewed as a special case of [15], where preferences conform to expected-utility theory. Epstein and Wang construct a universal type space in this framework; see also Di Tillio [13].

Finally, constructions analogous to that of a universal type space appear in other, unrelated contexts. For instance, Epstein and Zin [16] develop a class of recursive preferences over infinite-horizon temporal lotteries; to construct the domain of such preferences, they employ arguments related to Mertens and Zamir's. Gul and Pesendorfer [18] employ analogous techniques to analyze self-control preferences over infinite-horizon consumption problems.

References

[1] Ahn. D. (2006): "Hierarchies of Ambiguous Beliefs," *Journal of Economic Theory*, forthcoming.

[2] Armbruster, W., and W. Boge (1979): "Bayesian Game Theory," in O. Moeschlin and D. Pallaschke (eds.), *Game Theory and Related Topics*, North-Holland, Amsterdam.

[3] Aumann, R. (1976): "Agreeing to Disagree," *Annals of Statistics* 4, 1236-1239.

[4] Aliprantis, C and K. Border (1999): *Infinite Dimensional Analysis*, 2nd. ed., Springer Verlag, Berlin.

[5] Battigalli, P. and M. Siniscalchi (1999): "Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games," *Journal of Economic Theory*, 88, 188-230.

[6] Battigalli, P. and M. Siniscalchi (2002):"Strong belief and forward induction reasoning," *Journal of Economic Theory* 106, 356-391.

[7] Bergemann, D. and S Morris (2005): "Robust Mechanism Design," *Econometrica* 73, 1521-1534.

[8] Blume, L., A. Brandenburger, and E. Dekel (1991): "Lexicographic probabilities and choice under uncertainty," *Econometrica* 59, 61-79.

[9] Boge, W., and T. Eisele (1979): "On Solutions of Bayesian Games," *International Journal of Game Theory*, 8, 193-215.

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

[11] Brandenburger, A., A. Friedenberg, and H.J. Keisler, "Admissibility in Games," mimeo, `http://www.stern.nyu.edu/~abranden`.

[12] Brandenburger, A. and J. Keisler (2006): "An Impossibility Theorem on Beliefs in Games," *Studia Logica* 84, 211-240.

[13] Di Tillio, A. (2006): "Subjective Expected Utility in Games," IGIER Working Paper # 311, Università Bocconi.

[14] Ellsberg, D. (1961): "Risk, ambiguity, and the Savage axioms, " *Quarterly Journal of Economics* 75, 643-669.

[15] Epstein, L., and T. Wang (1996): "Beliefs about Beliefs Without Probabilities," *Econometrica* 64, 1343-1373.

[16] Epstein, L. and S. Zin (1989): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework," *Econometrica* 57, pp. 937-969.

[17] Gilboa, I. and D. Schmeidler (1989): "Maxmin-Expected Utility with a non-unique prior," *Journal of Mathematical Economics* 18, 141-153.

[18] Gul, F. and W. Pesendorfer (2004): "Self-Control and the Theory of Consumption," *Econometrica* 72, 119-158.

[19] Harsanyi, J. (1967-68): "Games of Incomplete Information Played by Bayesian Players. Parts I, II, III," *Management Science* 14, 159-182, 320-334, 486-502.

[20] Heifetz, A., and D. Samet (1998): "Knowledge Spaces with Arbitrarily High Rank," *Games and Economic Behavior* 22, 260-273.

[21] Heifetz, A., and D. Samet (1998): "Topology-free Typology of Beliefs," *Journal of Economic Theory* 82, 324-381.

[22] Mariotti, T., M. Meier and M. Piccione (2005): "Hierarchies of Beliefs for Compact Possibility Models," *Journal of Mathematical Economics* 41, 303-324.

[23] Meier, M. (2005): "On the Nonexistence of Universal Information Structures," *Journal of Economic Theory* 122, 132-139.

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

[25] Morris, S. and Hyun Shin (2003): "Global Games: Theory and Applications," in *Advances in Economics and Econometrics (Proceedings of the Eight World Congress of the Econometric Society)*, edited by M. Dewatripont, L. Hansen and S. Turnovsky. Pages 56-114.