Incomplete Information*

Robert J. Aumann[†] and Aviad Heifetz[‡]

July 3, 2001

Abstract

In interactive contexts such as games and economies, it is important to take account not only of what the players believe about substantive matters (such as payoffs), but also of what they believe about the beliefs of other players. Two different but equivalent ways of dealing with this matter, the semantic and the syntactic, are set forth. Canonical and universal semantic systems are then defined and constructed, and the concepts of common knowledge and common priors formulated and characterized. The last two sections discuss relations with Bayesian games of incomplete information and their applications, and with interactive epistemology — the theory of multi-agent knowledge and belief as formulated in mathematical logic.

Journal of Economic Literature Classification Codes: D82, C70.

Mathematics Subject Classification 2000: Primary: 03B42. Secondary: 91A10, 91A40, 91A80, 91B44.

Key Words: Incomplete or Differential Information, Interactive Epistemology, Semantic Belief Systems, Syntactic Belief Systems, Common Priors.

1. Introduction

In interactive contexts such as games and economies, it is important to take account not only of what the players believe about substantive matters (such as

*To appear in the Handbook of Game Theory, vol. 3, Elsevier/North Holland. Important input from Sergiu Hart, Martin Meier, and Dubi Samet is gratefully acknowledged.

[†]Center for Rationality and Institute of Mathematics, The Hebrew University of Jerusalem, Center for Game Theory, State University of New York at Stony Brook, and Stanford University. Research partially supported under NSF grant SES-9730205.

[‡]Tel Aviv University and California Institute of Technology

payoffs), but also of what they believe about the beliefs of other players. This chapter sets forth several ways of dealing with this matter, all essentially equivalent.

There are two basic approaches, the syntactic and the semantic; while the former is conceptually more straightforward, the latter is more prevalent, especially in game and economic contexts. Each appears in the literature in several variations.

In the **syntactic** approach, the beliefs are set forth explicitly: One specifies what each player believes about the substantive matters in question, about the beliefs of the others about these substantive matters, about the beliefs of the others about the beliefs of the others about the substantive matters, and so on ad infinitum. This sounds — and is — cumbersome and unwieldy, and from the beginning of research into the area, a more compact, manageable way was sought to represent interactive beliefs.

Such a way was found in the **semantic** approach, which is "leaner" and less elaborate, but also less transparent. It consists of a set of **states of the world** (or simply **states**), and for each player and each state, a probability distribution on the set of all states. We will see that this provides the same information as the syntactic approach.

The subject of **common priors** is discussed in Section 9; applications to Game Theory, in Section 10; and finally, Section 11 contains a brief discussion of the theory from the viewpoint of mathematical logic.

2. The Semantic Approach

Given a set N of players, a **(finite) semantic belief system** consists of

(i) a finite set Ω, the **state space**, whose elements are called **states of the world**, or simply **states**, and whose subsets are called **events**; and

(ii) for each player i and state ω, a probability distribution $\pi_i(\cdot;\omega)$ on Ω; if E is an event, $\pi_i(E;\omega)$ is called **i's probability for E in ω**.

Conceptually, a "state of the world" is meant to encompass all aspects of reality that are relevant to the matter under consideration, including the beliefs of all players in that state. As in probability theory, an event is a set of states; thus the event "it will snow tomorrow" is represented by a set of states — those in which it snows tomorrow.

We assume that

(2.1) if $\pi_i(\{\nu\};\omega) > 0$, then $\pi_i(E;\nu) = \pi_i(E;\omega)$ for all E.

In words: If in state ω, player i considers state ν possible (in the sense of assigning to it positive probability), then his probabilities for all events are the same in state ν as in state ω. That is, he does not seriously entertain the possibility that his own probability for some event E is different from what it actually is; he is sure[1] of his own probabilities.

The restriction to finite systems in this section is for simplicity only; for a general treatment, see Section 6.

3. An Example

Here $N = \{Ann, Bob\}$, $\Omega = \{\alpha, \beta, \gamma\}$, and the probabilities $\pi_i(\{\nu\}; \omega)$ are as follows:

π_{Ann}:

ω \ $\nu =$	α	β	γ
α	1/2	1/2	0
β	1/2	1/2	0
γ	0	0	1

π_{Bob}:

ω \ $\nu =$	α	β	γ
α	1	0	0
β	0	1/2	1/2
γ	0	1/2	1/2

In each of the states α and β, Ann attributes probability 1/2 to each of α and β, whereas in γ, she knows that γ is the state; and Bob, in each of the states β and γ, attributes probability 1/2 to each of β and γ, whereas in α, he knows that α is the state.

Now let $E = \{\alpha, \gamma\}$, and let us consider the probabilities of the players in state β. To start with, each player assigns probability 1/2 to E. At the next level, each player assigns probability 1/2 to the other assigning probability 1/2 to E, and probability 1/2 to the other being sure[2] of E. At the next level, Ann assigns probability 1/2 to Bob being sure that she assigns probability 1/2 to E; and probability 1/2 to Bob's assigning probability 1/2 to her being sure of E, and probability 1/2 to her assigning probability 1/2 to E. The same holds if Ann and

[1] By "sure" we mean "assigns probability 1 to" (probabilists use "almost sure" for this, but here it would make the text unnecessarily cumbersome and opaque.) Actually, in these models players "know" their own probabilities; i.e., they do not admit any possibility, even with probability 0, of having a different probability (see Section 7). Indeed, a person's own probability judgments would appear to be among the few things of which he can justifiably be absolutely certain.

[2] Indeed, Ann assigns $1/2 - 1/2$ probabilities to the states α and β. In β, Bob assigns probability 1/2 to E, and in α, Bob is sure of E. This demonstrates our assertion about Ann's probabilities for Bob's probabilities. The calculation of Bob's probabilities for Ann's probabilities is similar.

Bob are interchanged. That's already quite a mouthful; we refrain from describing the subsequent levels explicitly.

The previous paragraph, with its explicit description of each player's probabilities for the other player's probabilities, is typical of the syntactic approach. The complexity of the description increases exponentially with the "depth" of the level, and there are infinitely many levels. Because this example is particularly symmetric, the description is relatively simple; in general, things are even more complex, by far. Note, moreover, that the syntactic treatment corresponds to just *one* of the states in the semantic model.

By contrast, the semantic model encapsulates the whole mess, simultaneously for *all* the states, in just two compact (3×3) tables.

4. The Syntactic Approach

The syntactic approach may be formalized by "belief hierarchies." One starts with an exhaustive set of mutually exclusive "states of nature" (like "snow", "rain", "cloudy", or "clear" at noon tomorrow). The first level of the hierarchy specifies, for each player, a probability distribution on the states of nature. The second level specifies, for each player, a joint probability distribution on the states of nature and the others' probability distributions. The third level specifies, for each player, a joint probability distribution on the states of nature and the probability distributions of the others at the first and second levels. And so on. Certain consistency conditions are required (Mertens and Zamir 1985; see Section 8 below).

Since there is a continuum of possibilities for the probability distributions at the first level, the probability distributions at the second and higher levels may well[3] be continuous (i.e., have nonatomic components). So one needs some kind of additional structure — a topology or a measurable structure (σ-field) — on the space of probability distributions at each level. With finitely many states of nature, this offers no difficulty at the first and second levels. But already the third level consists of a probability distribution on the space of probability distributions on a continuum, which requires a topology or measurable structure on the space of probability distributions on a continuum — a nontrivial matter. Of course, this applies also to higher levels.

[3] A discrete distribution would mean, e.g., that Ann *knows* that Bob assigns probability precisely 0, 1/2 or 1 to snow. While possible, this seems unlikely. When modelling beliefs about the beliefs of others, some fuzziness seems natural.

An alternative syntactic formalism, which avoids these complications, works with sentences[4] or assertions rather than with probability distributions.[5] Start with a set $\{x, y, ...\}$ of "natural sentences" (like "rain", "warm", and "humid" at noon tomorrow), which need be neither exhaustive nor mutually exclusive. One may operate on these in two ways: by logical operators and connectives, like "not" ($\neg$), "or" ($\vee$), and "and" ($\wedge$), and by belief operators p_i^α, whose interpretation is "player i attributes probability at least α to ... ." The operations may be concatenated in any way one wishes; for example, $p_{Ann}^{1/2}(x \vee\, p_{Bob}^{3/4}(\neg p_{Ann}^{1/4} y))$ means that Ann ascribes probability at least 1/2 to the contingency that either x or that Bob ascribes probability at least 3/4 to her ascribing probability less than 1/4 to y. A **syntactic belief system** is a collection of such sentences that satisfies certain natural completeness, coherence, and consistency conditions (Section 8).

It may be seen that a syntactic belief system contains precisely the same substantive information as a belief hierarchy, without the topological or measure-theoretic complications.

5. From Semantics to Syntax

We now indicate how the syntactic formalism is derived from the semantic one in a general framework.

Suppose given a semantic belief system with state space Ω, and a family of "distinguished" events $X, Y, ...$ (corresponding[6] to the above "natural sentences" $x, y, ...$); such a pair is called an **augmented** semantic belief system. Let $\pi_i(E; \omega)$ be the probability of player i for event E in state ω. For numbers α between 0 and 1, let $P_i^\alpha E$ be the event that i's probability for E is at least α; i.e., the set of all states ω at which $\pi_i(E; \omega) \geq \alpha$. Thus the operator P_i^α in the semantic model corresponds to the operator p_i^α in the syntactic model. As usual, the set operations $\cup$ (union) and $\Omega\backslash$ (complementation) correspond to the logical operations $\vee$ (disjunction) and $\neg$ (negation). Thus each sentence in the syntactic formalism corresponds to an

[4] Usually called **formulas** in the literature.

[5] Some workers reserve the term "syntactic" for this kind of system only — i.e., one using sentences and syntactic rules. We prefer a more substantive terminology, in which **syntactic** refers to any system that directly describes the actual beliefs of the players, whereas **semantic** refers to a (states-of-the-world) model from which these beliefs can be derived.

[6] I.e., with the same conceptual content. Thus if in the syntactic treatment, x is the sentence "it will snow tomorrow," then in the semantic treatment, X is the event "it will snow tomorrow" (the set of all states in which it snows tomorrow).

event in the semantic formalism. For example, the sentence $p^{1/2}_{Ann}(x \vee p^{3/4}_{Bob}(\neg p^{1/4}_{Ann} y))$ discussed above corresponds to $P^{1/2}_{Ann}(X \cup P^{3/4}_{Bob}(\Omega \backslash P^{1/4}_{Ann} Y))$, namely the event that Ann ascribes probability at least 1/2 to the event that either X or that Bob ascribes probability at least 3/4 to her ascribing probability less than 1/4 to Y.

Now fix a state ω. If e is a sentence and E the corresponding event, say that e *holds in* ω if ω is in E. If we think of ω as the "true" state of the world, then the "true" sentences are precisely those that hold in ω. These sentences constitute a syntactic belief system in the sense of the previous section.

To summarize: Starting from a pair consisting of an augmented semantic belief system and a state ω, we have constructed a syntactic belief system L. We call this pair (or just the state ω) a *model* for L.

6. Removing the Finiteness Restriction

Given a set N of players, a *(general) semantic belief system* consists of

(ia) a set Ω, the *state space*, whose elements are the *states;*

(ib) a σ-field $\mathcal{F}$ of subsets of Ω — the *events;* and

(ii) for each player i and state ω, a probability distribution $\pi_i(\cdot;\omega)$ on $\mathcal{F}$.

We assume that

(6.1) $\pi_i(E;\omega)$ is $\mathcal{F}$-measurable in ω for each fixed event E, and

(6.2) $\pi_i(\{\nu : \pi_i(E;\nu) \neq \pi_i(E;\omega)\};\omega) = 0$ for each event E and state ω.

The interpretation is as in Section 2.

7. Knowledge and Common Knowledge

In the formalism of Sections 2 and 6, the concept of knowledge — in the sense of absolute certainty[7] rather than probability 1 — plays no explicit role. However, this concept can be *derived* from that formalism. Indeed, for each state ω and player i, let $\mathsf{I}_i(\omega)$ be the set of all states in which i's probabilities (for all events) are the same as in ω; call it i*'s information set in* ω. The information sets of i form a partition $\mathcal{P}_i$ of Ω, called i*'s information partition.* Say that i *knows* an event E in a state ω if E includes $\mathsf{I}_i(\omega)$, and denote by K_iE the set of all states ω in which i knows E. Conceptually, this formulation of knowledge presupposes that players know their own probabilities with absolute certainty, which is not

[7]I.e., with error impossible: If i "knows" an event E at a state ω, then ω *must* be in E.

unreasonable. It may be seen that

(7.1) $K_i E \subset P_i^1 E$,
i.e., that players ascribe probability 1 to whatever they know, and that

(7.2) $P_i^\alpha E \subset K_i P_i^\alpha E$,
which formally expresses the principle, enunciated above, that players know their own probabilities.

Let KE be the set $\cap_i K_i E$ of all states in which all players know E, and set

$$K^\infty E := KE \cap KKE \cap KKKE \cap ...;$$

thus $K^\infty E$ is the set of all states in which all players know E, all players know that, all players know that, and so on ad infinitum. We say that E is commonly known in ω if $\omega \in K^\infty E$ (Lewis 1969); for a comprehensive survey of common knowledge, see Geanakoplos, this Handbook, Volume 2, Chapter 40. If $\mathcal{P}^\infty$ is the meet (finest common coarsening) of the information partitions $\mathcal{P}_i$ of all the players i, then it may be seen (Aumann 1976, 1999a) that $K^\infty E$ is the union of all the atoms of $\mathcal{P}^\infty$ that are included in E. The atoms of $\mathcal{P}^\infty$ are called common knowledge components (or simply components) of Ω. In a sense, one can always restrict the discussion to such a component: The "true" state is always in some component, and then it is commonly known that it is, and all considerations relating to other components become irrelevant.

Some interactive probability formalisms (like Aumann 1999b) use a separate, exogenous, concept of knowledge, and assume analogues of 7.1 and 7.2 (see 7.3 and 7.4 below). In the semantic framework, such a concept is redundant: Surprisingly, knowledge is implicit in probability.

To see this, let $\mathcal{P}_i$, I_i, and K_i, as above, be the knowledge concepts derived from the probabilities $\pi_i(\cdot;\omega)$, and let $\mathcal{Q}_i$, J_i, and L_i be the corresponding exogenous concepts. Assume

(7.3) $L_i E \subset P_i^1 E$ (players ascribe probability 1 to what they know[8]), and
(7.4) $P_i^\alpha E \subset L_i P_i^\alpha E$ (players know their own probabilities).

It suffices to show that $\mathsf{J}_i(\omega) = \mathsf{I}_i(\omega)$ for all ω. This follows from 7.3 and 7.4; we argue from the verbal formulations. If $\nu \in \mathsf{J}_i(\omega)$, then by 7.4, i's probabilities for all events must be the same in ν and ω, so $\nu \in \mathsf{I}_i(\omega)$. Conversely, i knows in ν that he is in $\mathsf{J}_i(\nu)$, so by 7.3, his probability for $\mathsf{J}_i(\nu)$ is 1. So if $\nu \notin \mathsf{J}_i(\omega)$, then $\mathsf{J}_i(\omega)$ and $\mathsf{J}_i(\nu)$ are disjoint, so in ν his probability for $\mathsf{J}_i(\omega)$ is 0, while in ω it

[8] In the remainder of this section we use "know" in the exogenous sense.

is 1. So his probabilities are different in ω and ν, so $\nu \notin \mathsf{I}_i(\omega)$. Thus $\nu \in \mathsf{J}_i(\omega)$ if and only if $\nu \in \mathsf{I}_i(\omega)$, as claimed.

8. Canonical Semantic Systems

One may think of a semantic belief system purely technically, simply as providing the setting for a convenient, compact representation of a syntactic belief system, as in the example of Section 3. But one may also think of it substantively, as setting forth a model of interactive epistemology that reflects reality, including those states of the world that are "actually" possible,[9] and what the players know about them. This substantive viewpoint raises several questions, foremost among them being what the players know about the model itself. Does each know the space of states? Does he know the others' probabilities (in each state)? If so, from where does this knowledge derive? If not, how can the formalism indicate what each player believes about the others' beliefs? For example, why would the event $P^{1/2}_{Ann}P^{3/4}_{Bob}E$ then signify that Ann ascribes probability at least 1/2 to Bob ascribing probability at least 3/4 to E?

More generally, the whole idea of "state of the world," and of probabilities that accurately reflect the players' beliefs about other players' beliefs, is not transparent. What **are** the states? Can they be explicitly described? Where do they come from? Where do the probabilities come from? What justifies positing this kind of model, and what justifies a particular array of probabilities?

One way of overcoming these problems is by means of "canonical" or "universal" semantic belief systems. Such a system comprises a standard state space with standard probabilities for each player in each state. The system does not depend on reality; it is a framework, it fits any reality, so to speak, like the frames that one buys in photo shops, which do not depend on who is in the photo — they fit any photo with any subject, as long as the size is right. In brief, there is no substantive information in the system.

There are basically two ways of doing this. Both depend on two parameters: the set (or simply number n) of players and the set (or simply number) of "natural" eventualities. The first way (Mertens and Zamir 1985) is hierarchical. The zero'th level of the hierarchy is a (finite) set $\mathsf{X} := \{x, y, ...\}$, whose members represent mutually exclusive and exhaustive "states of nature," but formally are just abstract symbols. The first level H^1 is the cartesian product of X with n copies

[9] Philosophers like the term "possible worlds."

of the X-simplex (the set of all probability distributions on X); a point in this cartesian product describes the "true" state of nature, and also the probabilities of each player for each state of nature.

A point h^2 at the second level H^2 consists of a first-level point h^1 together with an n-tuple of probability distributions on H^1, one for each player; this describes the "true" state of nature, the probabilities of each player for the state of nature, and what each player believes about the other players' beliefs about the state of nature. There are two "consistency" conditions: First, the distribution on H^1 that h^2 assigns to i must attribute[10] probability 1 to the distribution on X that h^1 assigns to i; that is, i must know what he himself believes. Second, the distribution[11] on X that h^2 assigns to i must coincide with the one that h^1 assigns to i.

A third-level point consists of a second-level point h^2 together with an n-tuple of probability distributions[12] on H^2, one for each player; again, i's distribution on H^2 must assign probability 1 to the distribution on H^1 that h^2 assigns to i, and its marginal on H^1 must coincide with the distribution on H^1 that h^1 assigns to i.

And so on. Thus an $(m+1)$'th-level point h^{m+1} is a pair consisting of an m'th-level point h^m and a distribution on H^m; we say that h^{m+1} *elaborates* h^m. Define a state h in the canonical semantic belief system H as a sequence $h^1, h^2, ...$ in which each term elaborates the previous one. To define the probabilities, proceed as follows: For each subset S^m of H^m, set $\hat{S}^m = \{h \in H : h^m \in S^m\}$; let $\mathcal{F}$ be the σ-field generated by all the $\hat{S}^m$, where S^m is measurable; and let $\pi_i(\hat{S}^m; h)$ be the probability that i's component of h^{m+1} assigns to S^m. In words, a state consists of a sequence of probability distributions of ever-increasing depth; to get the probability of a set of such sequences in a given state, one simply reads off the probabilities specified by that state. That this indeed yields a probability on $\mathcal{F}_\sigma$ follows from the Kolmogorov extension theorem, which requires some topological assumptions.[13]

The second construction (Heifetz and Samet 1998, Aumann 1999b) of a canon-

[10] Technically, h^2 assigns to i a distribution on $H^1 = \mathsf{X} \times \Delta(\mathsf{X})$, where Δ denotes "the set of all distributions on ...". The first consistency condition says that the marginal of this distribution on the second factor (i.e., on $\Delta(\mathsf{X})$) is a unit mass concentrated on the distribution on X that h^1 assigns to i.

[11] Technically, the marginal on X of the distribution on $H^1 = \mathsf{X} \times \Delta(\mathsf{X})$ that h^2 assigns to i. See the previous footnote.

[12] As indicated in Section 5, this involves topological issues, which we will not discuss.

[13] Without which things might not work (Heifetz and Samet 1998).

ical belief system is based on sentences. It uses an "alphabet" $x, y, \ldots$, as well as more elaborate sentences constructed from the alphabet by means of logical operations and probability operators p_i^α (as in Section 4). The letters of the alphabet represent "natural sentences," not necessarily mutually exclusive or exhaustive; but formally, as above, they are just abstract symbols. In this construction, the states γ in the canonical space Γ are simply lists of sentences; specifically, lists that are complete, coherent, and closed in a sense presently to be specified. The point is that a state is determined — or better, *defined* — by the sentences that hold there; conceptually, the state *is* simply what happens.

More precisely: A list is *complete* if for each sentence f, it contains either f itself or its negation $\neg f$; *coherent,* if for each f, it does not contain both f and $\neg f$; and *closed,* if it contains every logical consequence[14] of the sentences in it.

This defines the states γ in the canonical semantic belief system; to complete the definition of the system, we must define the σ-field $\mathcal{F}$ of events and the probabilities $\pi_i(\cdot;\gamma)$. For any sentence f, let $E_f := \{\delta \in \Gamma : f \in \delta\}$; that is, E_f is the set of all states[15] in the canonical system that contain f. We define $\mathcal{F}$ as the σ-field generated by all the events E_f for all sentences f. Then to define i's probabilities on $\mathcal{F}$ in the state γ, it suffices to define his probabilities for each E_f. Since the state is simply the list of all sentences that "hold in that state," it follows that E_f is the set of all states in which f holds — in brief, the event that f holds. The probability that i assigns to E_f in the state γ is then implicit in the specification of γ; namely, it is the supremum of all the rational α for which $p_i^\alpha f$ is in γ — the supremum α such that in γ, player i assigns probability at least α to f. This defines $\pi_i(E_f;\gamma)$ for each i, γ, and f; and since the E_f generate $\mathcal{F}$, it follows that it defines[16] $\pi_i(\cdot;\gamma)$. This completes the second construction, which has the advantage of requiring no topological machinery.

The canonical semantic belief system Γ is *universal* in a sense that we now

[14] As used here, the term "logical consequence" is *defined* purely syntactically, by axioms and inference rules (Meier 2001). But it may be *characterized* semantically: A sentence g is a logical consequence of sentences $f_1, f_2, \ldots$ if and only if g holds at any state of any augmented semantic belief system at which $f_1, f_2, \ldots$ hold; i.e., if any model for $f_1, f_2, \ldots$ is also a model for g (Meier 2001). The axiomatic system is infinitary, as indeed it must be: For example, $p_i^1 f$ is a logical consequence of $p_i^{1/2} f, p_i^{3/4} f, p_i^{7/8} f, \ldots$, but of no finite subset thereof. Meier's beautiful, path-breaking paper is not yet published.

[15] Recall that a state is a list of sentences.

[16] That is, since the E_f generate $\mathcal{F}$, there cannot be two different probability measures on $\mathcal{F}$ with the given values on the E_f. That there is *one* such measure follows from a theorem of Caratheodory on extending measures from a field to the σ-field it generates.

describe. Let $x, y, ...$ be letters in an alphabet, and Ω, Ω' semantic belief systems with the same players, augmented by "distinguished" events $X, Y, ...$ and $X', Y', ...$, corresponding to the letters $x, y, ...$ (simply "systems" for short). A mapping M from Ω into Ω' is called a **belief morphism** if $M^{-1}(X') = X$, $M^{-1}(Y') = Y, ...$, and $\pi_i(M^{-1}(E'); \omega) = \pi'_i(E'; M(\omega))$ for all players i, events E' in Ω', and states ω in Ω; i.e., if it preserves the relevant aspects of the system (as usual for morphisms). A system Υ is called **universal** if for each system Ω, there is a belief morphism from Ω into Υ.

To see that Γ is universal, let Ω be a system. Then for each state ω in Ω, the family $D(\omega)$ of sentences that hold in ω (see Section 5) is a state in the canonical space Γ, and the mapping $D : \Omega \to \Gamma$ is a belief morphism — indeed the only one (Heifetz and Samet 1998). The construction of Mertens and Zamir (1985) also leads to a universal system, and the two canonical systems are isomorphic.[17]

At the end of the previous section, we noted that a separate, exogenous notion of knowledge is semantically redundant. Syntactically, it is not. To understand why, consider two two-player semantic belief systems, Ω and Ω'. The first has just one state ω, to which, of course, both players assign probability 1. The second has two states, ω and ω'. In ω, both players assign probability 1 to ω; in ω', Ann assigns probability 1 to ω, whereas Bob assigns probability 1 to ω'. Now augment both systems by assigning x to ω. Then the syntactic belief system derived from ω in Ω — the family $D(\omega)$ of all sentences that hold there — is identical to that derived from ω in Ω'. So syntactically, the two situations are indistinguishable. But semantically, they are different: In Ω, Ann knows **for sure** in ω that Bob assigns probability 1 to x; in Ω', she does not. There is no way to capture this syntactically without explicitly introducing knowledge operators k_i for the various players i (as in Aumann 1999b; see also 1998). In particular, the canonical semantic system Γ has just one component[18] — the whole space.

9. Common Priors

Call a semantic belief system **regular**[19] if it is finite, has a single common knowledge component, and in each state, each player assigns positive probability to that state.

[17] I.e., there is a one-one belief morphism from the one onto the other.

[18] Except when there is only one player. This is because the probability syntax can only express beliefs of the players, not that a player knows anything about another one **for sure.**

[19] Regularity enables uncluttered statements of the definitions and results. No real loss of generality is involved.

A **common prior** on a regular semantic belief system Ω is defined as a probability distribution π on Ω such that

(9.1) $\pi(E \cap \mathsf{I}_i(\omega)) = \pi_i(E;\omega)\pi(\mathsf{I}_i(\omega))$

for each event E, player i, and state ω; that is, the personal probabilities π_i in a state ω are obtained from the common prior π by conditioning on each player's private information in that state. In Section 3, for example, the distribution that assigns probability 1/3 to each of the three states is the unique common prior. Common priors can easily fail to exist; for example, if Ω has just two states, in each of which Ann assigns 1/2 - 1/2 probabilities to the two states, and Bob, 1/3 - 2/3 probabilities. Conceptually, existence of a common prior signifies that differences in probability assessments are due to differences in information only (Aumann 1987, 1998); that people who have always been fed precisely the same information will not differ in their probability assessments.

With a common prior, players cannot "agree to disagree" about probabilities. That is, if it is commonly known in some state that for some specific α, β, and E, Ann and Bob assign probabilities α and β respectively to the event E, then $\alpha = \beta$. This is the "Agreement Theorem" (Aumann 1976).

The question arises (Gul 1998) whether the existence of a common prior can be characterized syntactically. That is, can one characterize those syntactic belief systems L that are associated with common priors?[20] By a **syntactic** characterization, we mean one couched directly in terms of the sentences in L.

The answer is "yes." Indeed, there are two totally different characterizations. The first is based on the following, which is both a generalization of and a converse to the agreement theorem.

Proposition 9.2 (Morris 1994, Samet[21] 1998b, Feinberg 2000).[22] Let Ω be a regular semantic belief system, ω a state in Ω. Suppose first that there are just two players, Ann and Bob. Then there is **no** common prior on Ω if and only if there is a random variable[23] x for which it is commonly known in ω that Ann's

[20] More precisely, those L for which there is a state ω in some finite augmented semantic belief system with a common prior such that L is the collection of sentences holding at ω.

[21] Samet's proof, which uses the elementary theory of convex polyhedra, is the briefest and most elegant.

[22] For an early related result, see Nau and McCardle (1990).

[23] A real function on Ω. Although the concept of random variable may seem essentially semantic, it is not; in an augmented semantic system, random variables can in general be expressed in terms of sentences (Feinberg 2000, Heifetz 2001).

expectation of x is positive and Bob's is negative.[24]

With n players, there is no common prior on Ω if and only if there are n random variables x_i identically[25] summing to 0 for whom it is commonly known in ω that each player i's expectation of x_i is positive.

In the two-person case, one may think of x as the amount of a bet between Ann and Bob, possibly with odds that depend on the outcome. Thus the proposition says that with a common prior, it is impossible for both players to expect to gain from such a bet. The interpretation in the n-person case is similar; one may think of pari-mutuel betting in horse-racing, the track being one of the players.

In brief, common priors exist if and only if it is commonly known that you can't make something out of nothing: I.e., in a situation that is objectively zero-sum, it cannot be commonly known that all sides expect to gain.

The importance of this proposition lies in that it is formulated in terms of one single state ω only. We are talking about what Ann expects *in that state,* what she thinks *in that state* that Bob may expect, what she thinks *in that state* that Bob thinks that she may expect, and so on; and similarly for Bob. That is precisely what is needed for a syntactic characterization.[26]

Actually to derive a fully syntactic characterization from this proposition is, however, a different matter. To do so, one must provide syntactic formulations of (i) common knowledge, (ii) random variables and their expectations, and (iii) finiteness of the state space. The first is not difficult; setting $k := \wedge_i k_i$, call a sentence e *syntactically commonly known* if $e, ke, kke, kkke, ...$ all obtain. But (ii) and (iii), though doable — indeed elegantly (Feinberg 2000, Heifetz 2001) — are more involved; we will not discuss the matter further here.

Proposition 9.2 holds also for infinite state spaces that are "compact" in a natural sense. As before, this enables a syntactic characterization in the compact case; to do this properly one must characterize compactness syntactically. Without compactness, the proposition fails. Feinberg (2000) discusses these matters fully.

The second syntactic characterization of common priors is in terms of iterated expectations. Again, consider first the two-person case. If ν is a state and y a random variable, then Ann's and Bob's expectations of y in ν are

[24] Explicitly, the commonly known event in question is
$\{\nu : \sum_{\xi\in\Omega} \pi_{Ann}(\{\xi\};\nu)\mathsf{x}(\xi) > 0 > \sum_{\xi\in\Omega} \pi_{Bob}(\{\xi\};\nu)\mathsf{x}(\xi)\}$.

[25] At each state in Ω.

[26] Recall that we use the term *syntactic* rather broadly, as referring to anything that "actually obtains," such as the beliefs — and hence expectations — of the players. See Footnote 5.

$(\mathsf{E}_{Ann}\mathsf{y})(\nu) := \sum_{\xi\in\Omega} \pi_{Ann}(\{\xi\};\nu)\mathsf{y}(\xi)$ and
$(\mathsf{E}_{Bob}\mathsf{y})(\nu) := \sum_{\xi\in\Omega} \pi_{Bob}(\{\xi\};\nu)\mathsf{y}(\xi)$.

Both $\mathsf{E}_{Ann}\mathsf{y}$ and $\mathsf{E}_{Bob}\mathsf{y}$ are themselves random variables, as they are functions of the state ν. So one may form the iterated expectations

(9.3) $\mathsf{E}_{Ann}\mathsf{y}, \mathsf{E}_{Bob}\mathsf{E}_{Ann}\mathsf{y}, \mathsf{E}_{Ann}\mathsf{E}_{Bob}\mathsf{E}_{Ann}\mathsf{y}, \mathsf{E}_{Bob}\mathsf{E}_{Ann}\mathsf{E}_{Bob}\mathsf{E}_{Ann}\mathsf{y}$,... and
(9.4) $\mathsf{E}_{Bob}\mathsf{y}, \mathsf{E}_{Ann}\mathsf{E}_{Bob}\mathsf{y}, \mathsf{E}_{Bob}\mathsf{E}_{Ann}\mathsf{E}_{Bob}\mathsf{y}, \mathsf{E}_{Ann}\mathsf{E}_{Bob}\mathsf{E}_{Ann}\mathsf{E}_{Bob}\mathsf{y}$,... .

Proposition 9.5 (Samet 1998a). Each of these two sequences (9.3 and 9.4) of random variables converges to a constant (independent of the state); the system has a common prior if and only if for each random variable y, these two constants coincide; and in that case, the common value of the two constants is the expectation of y over Ω w.r.t. the common prior π.

The proof uses finite state Markov chains.

Samet illustrates the proposition with a story about two stock analysts. Ann has an expectation for the price y of IBM in one month from today. Bob does not know Ann's expectation, but has some idea of what it might be; his expectation of her expectation is $\mathsf{E}_{Bob}\mathsf{E}_{Ann}\mathsf{y}$. And so on.

Like the previous characterization of common priors (Proposition 9.2), this one appears semantic, but in fact is syntactic: All the iterated expectations can be read off from a single syntactic belief system. As before, actually deriving a fully syntactic characterization requires formulating finiteness of the state space syntactically, which we do not do here. Unlike the previous characterization, this one has not been established for compact state spaces.

With n players, one considers arbitrary infinite sequences $\mathsf{E}_{i_1}\mathsf{y}, \mathsf{E}_{i_2}\mathsf{E}_{i_1}\mathsf{y}, \mathsf{E}_{i_3}\mathsf{E}_{i_2}\mathsf{E}_{i_1}\mathsf{y}, \mathsf{E}_{i_4}\mathsf{E}_{i_3}\mathsf{E}_{i_2}\mathsf{E}_{i_1}\mathsf{y}$,..., where $i_1, i_2, ...$ is any sequence of players (with repetitions, of course), the only requirement being that each player appears in the sequence infinitely often. The result says that each such sequence converges to a constant; the system has a common prior if and only if all the constants coincide; and in that case, the common value of the constants is the expectation of y over Ω w.r.t. the common prior π.

10. Incomplete Information Games

The theory of incomplete information, whose general, abstract form is outlined in the foregoing sections of this chapter, has historical roots in a concrete application: games. Until the mid-sixties, game theorists did not think carefully about the informational underpinnings of their analyses. Luce and Raiffa (1957) did express some malaise on this score, but left the matter at that. The primary problem was each player's uncertainty about the payoff (or utility) functions of the others; to a lesser extent, there was also concern about the players' uncertainty about the strategies available to others, and about their own payoffs.

In path-breaking research in the mid-sixties, for which he got the Nobel Prize some thirty years later, John Harsanyi (1967-8) succeeded both in formulating the problem precisely, and in solving it. In brief, the formulation is the syntactic approach, whereas the solution is the semantic approach. Let us elaborate.

Harsanyi started by noting that though usually the players do not know the others' payoff functions, nevertheless, as good Bayesians, each has a (subjective) probability distribution over the possible payoff functions of the others. But that is not enough to analyze the situation. Each player must also take into account what the others think that he thinks about them. Even there it does not end; he must also take into account what the others think that he thinks that they think about him. And so on, ad infinitum. Harsanyi saw this infinite regress as a Gordian knot, not given to coherent, useful analysis.

To cut the knot, Harsanyi invented the notion of "type." Consider the case of two players, Ann and Bob. Each player, he said, may be one of several types. The type of a player determines his payoff function, and also a probability distribution on the other player's possible types. Since Bob's type determines his payoff function, Ann's probability distribution on his types induces a probability distribution on his payoff functions. But it also induces a probability distribution on his probability distributions on Ann's types, and so on Ann's payoff functions. And so on. Thus a "type structure" yields the whole infinite regress of payoff functions, distributions on payoff functions, distributions on distributions on payoff functions, and so on.

The reader will realize that the "infinite regress" is a syntactic belief hierarchy in the sense of Section 6, the "states of nature" being n-tuples of payoff functions; whereas a "type structure" is a semantic belief system, the "states of the world" being n-tuples of types. In modern terms, Harsanyi's insight was that a semantic system yields a syntactic system. Though this may seem obvious today (see Section 3), it was far from obvious at the time; indeed it was a major conceptual breakthrough, which enabled extending many of the fundamental concepts of game theory to the incomplete information case, and led to the opening of entirely new areas of research (see below).

The converse — that every syntactic system can be encapsulated in a semantic system — was not proved by Harsanyi. Harsanyi argued for the type structure on intuitive grounds. Roughly speaking, he reasoned that every player's brain must be configured in some way, and that this configuration should determine both the player's utility or payoff function and his probability distribution on the configuration of other players' brains.

The *proof* that indeed every syntactic system can be encapsulated in a semantic system is outlined in Section 8 above; the semantic system in question is simply the canonical one. This proof was developed over a number of years by several workers (Armbruster and Böge 1979, Böge and Eisele 1979), culminating in the work of Mertens and Zamir (1985) cited in Section 8 above.

Also the assumption of common priors (Section 9) was introduced by Harsanyi (1967-8), who called this the *consistent* case. He pointed out that in this case — and *only* in this case — an n-person game of incomplete information in strategic (i.e., normal) form can be represented by an n-person game of complete information in extensive form, as follows: First, "chance" chooses an n-tuple of types (i.e., a state of the world), using the common prior π for probabilities. Then, each player is informed of his type, but not of the others' types. Then the n players simultaneously choose strategies. Finally, payoffs are made in accordance with the types chosen by nature and the strategies chosen by the players.

The concept of strategic ("Nash") equilibrium generalizes to incomplete information games in a natural way. Such an equilibrium — often called a *Bayesian Nash* equilibrium — assigns to each type t_i of each player i a (mixed or pure) strategy of i that is optimal for i given t_i's assessment of the probabilities of the others' types and the strategies that the equilibrium assigns to those types. In the consistent (common prior) case, the Bayesian Nash equilibria of the incomplete information strategic game are precisely the same as the ordinary Nash equilibria of the associated complete information extensive game (described in the previous paragraph).

Since Harsanyi's seminal work, the theory of incomplete information games has been widely developed and applied. Several areas of application are of particular interest. *Repeated* games of incomplete information deal with situations where the same game is played again and again, but the players have only partial information as to what it is. This is delicate because by taking advantage of his private information, a player may implicitly reveal it, possibly to his detriment. For surveys up to 1991, see this *Handbook*, Volume I, Chapter 5 (Zamir) and Chapter 6 (Forges). Since 1992, the literature on this subject has continued to grow; see Aumann and Maschler (1995), whose bibliography is fairly complete up to 1994.

A more complete and modern treatment,[27] unfortunately as yet unpublished, is Mertens, Sorin, and Zamir (1994).

Other important areas of application include auctions (see Wilson, this *Handbook*, Volume I, Chapter 8), bargaining with incomplete information (Binmore, Osborne, and Rubinstein I,7 and Ausubel, Cramton, and Deneckere III,50), principal-agent problems (Dutta and Radner II,26), inspection (Avenhaus, von Stengel, and Zamir III,51), communication and signalling (Myerson II,24 and Kreps and Sobel II,25), and entry deterrence (Wilson I,10).

Not surveyed in this *Handbook* are coalitional (cooperative) games of incomplete information. Initiated by Wilson (1978) and Myerson (1984), this area is to this day fraught with unresolved conceptual difficulties. Allen (1997) and Forges, Minelli, and Vohra (2001) are surveys.

Finally, games of incomplete information are useful in understanding games of *complete* information — ordinary, garden-variety games. Here the applications are of two kinds. In one, the given complete information game is "perturbed" by adding some small element of incomplete information. For example, Harsanyi (1973) uses this technique to address the question of the significance of mixed strategies: Why would a player wish to randomize, in view of the fact that whenever a mixed strategy μ is optimal, it is also optimal to use any pure strategy in the support of μ? His answer is that indeed players never actually use mixed strategies. Rather, even in a complete information game, the payoffs should be thought of as commonly known only approximately. In fact, there are small variations in the payoff to each player that are known only to that player himself; these small variations determine which pure strategy s in the support of μ he actually plays. It turns out that the probability with which a given pure strategy s is actually played in this scenario approximates the coefficient of s in μ. Thus, a mixed strategy of a player i appears not as a deliberate randomization on i's part, but as representing the estimate of *other* players as to what i will do.

Another application of this kind comes under the heading of reputational effects. Seminal in this genre was the work of the "gang of four" (Kreps, Milgrom, Roberts, and Wilson 1982), in which a repeated prisoner's dilemma is perturbed by assuming that with some arbitrarily small exogenous probability, the players are "irrational" automata who always play "tit-for-tat." It turns out that in equilibrium, the irrationality "takes over" in some sense: Almost until the end of the game, the rational players themselves play tit-for-tat. Fudenberg and Maskin (1986) generalize this to "irrational" strategies other than tit-for-tat; as before,

[27] Including also the general theory of incomplete information covered in this chapter.

equilibrium play mimics the irrational perturbation, even when (unlike tit-for-tat) it is inefficient. Aumann and Sorin (1989) allow any perturbation with "bounded recall;" in this case equilibrium play "automatically" selects an efficient equilibrium.

The second kind of application of incomplete information technology to complete information games is where the object of incomplete information is not the payoffs of the players but the actual strategies they use. For example, rather than perturbing payoffs a la Harsanyi (1973 — see above), one can say that even without perturbed payoffs, players other than i simply do not know what pure strategy i will play; i's mixed strategy represents the probabilities of the other players as to what i will do. Works of this kind include Aumann (1987), in which correlated equilibrium in a complete information game is characterized in terms of common priors and common knowledge of rationality; and Aumann and Brandenburger (1995), in which Nash equilibrium in a complete information game is characterized in terms of mutual knowledge of rationality and of the strategies being played, and when there are more than two players, also of common priors and common knowledge of the strategies being played (for a more detailed account, see Hillas and Kohlberg, this *Handbook*, Volume III, Chapter 42).

The key to *all* these applications is Harsanyi's "type" definition — the semantic representation — without which building a workable model for applications would be hopeless.

11. Interactive Epistemology

Historically, the theory of incomplete information outlined in this chapter has two parents: (i) incomplete information games, discussed in the previous section, and (ii) that part of mathematical logic, sometimes called *modal* logic, that treats knowledge and other epistemological issues, which we now discuss. In turn, (ii) itself has various ancestors. One is probability theory, in which the "sample space" (or "probability space") and its measurable subsets (there, like here, "events") play a role much like that of the space Ω of states of the world; whereas subfields of measurable sets, as in stochastic processes, play a role much like that of our information partitions $\mathcal{P}_i$ (Section 7). Another ancestor is the theory of extensive games, originated by von Neumann and Morgenstern (1944), whose "information sets" are closely related to our information sets $\mathrm{I}_i(\omega)$ (Section 7).

Formal epistemology, in the tradition of mathematical logic, began some forty years ago, with the work of Kripke (1959) and Hintikka (1962); these works were set in a single-person context. Lewis (1969) was the first to define common knowledge, which of course is a multi-person, interactive concept; though verbal, his treatment was entirely rigorous. Computer scientists became interested in the area in the mid-eighties; see Fagin et al. (1995).

Most work in formal epistemology concerns knowledge (Section 7) rather than probability. Though the two have much in common, knowledge is more elementary, in a sense that will be explained presently. Both interactive knowledge and interactive probability can be formalized in two ways: semantically, by a states-of-the-world model, or syntactically, either by a hierarchic model or by a formal language with sentences and logical inference governed by axioms and inference rules. The difference lies in the nature of logical inference. In the case of knowledge, sentences, axioms, and inference rules are all finitary (Fagin et al. 1995; Aumann 1999a). But probability is essentially infinitary (Footnote 14); there is no finitary syntactic model for probability.

Nevertheless, probability *does* have a finitary aspect. Heifetz and Mongin (2001) show that there is a finitary system $\mathcal{A}$ of axioms and inference rules such that if f and g are *finitary* probability sentences then g is a logical consequence[28] of f if and only if it is derivable from f via the *finitary* system $\mathcal{A}$. That is, though in general the notion of "logical consequence" is infinitary, for finitary sentences it *can* be embodied in a finitary framework.

Finally, we mention the matter of backward induction in perfect information games, which has been the subject of intense epistemic study, some of it based on probability 1 belief. This is a separate area, which should have been covered elsewhere in the *Handbook*, but is not.

References

Allen, B. (1997), "Cooperative Theory with Incomplete Information," in *Cooperation: Game-Theoretic Approaches*, edited by S. Hart and A. Mas-Colell, Berlin: Springer, 51-65.

Armbruster, W., and W. Böge (1979), "Bayesian Game Theory," in *Game Theory and Related Topics*, edited by O. Moeschlin and D. Pallaschke, Amsterdam: North Holland, 17-28.

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

—— (1987), "Correlated Equilibrium as an Expression of Bayesian Rationality," *Econometrica* 55, 1-18.

[28] In the infinitary sense of Meier (2001) discussed in Section 8 (Footnote 14).

—— (1998), "Common Priors: A Reply to Gul," Econometrica 66, 929-938.

—— (1999a), "Interactive Epistemology I: Knowledge," International Journal of Game Theory 28, 263-300.

—— (1999b), "Interactive Epistemology II: Probability," International Journal of Game Theory 28, 301-314.

—— and A. Brandenburger (1995), "Epistemic Conditions for Nash Equilibrium," Econometrica 63, 1161-1180.

—— and M. Maschler (1995), Repeated Games of Incomplete Information, Cambridge: MIT Press.

—— and S. Sorin (1989), "Cooperation and Bounded Recall," Games and Economic Behavior 1, 5-39.

Böge, W., and T. Eisele (1979), "On Solutions of Bayesian Games," International Journal of Game Theory 8, 193-215.

Fagin, R., J.Y. Halpern, M. Moses and M.Y. Vardi (1995), Reasoning about Knowledge, Cambridge: MIT Press.

Feinberg, Y. (2000), "Characterizing Common Priors in the Form of Posteriors," Journal of Economic Theory 91, 127-179.

Forges, F., E. Minelli and R.Vohra (2001), "Incentives and the Core of an Exchange Economy: A Survey," Journal of Mathematical Economics, forthcoming.

Fudenberg, D., and E. Maskin (1986), "The Folk Theorem in Repeated Games with Discounting and Incomplete Information," Econometrica 54, 533-554.

Gul, F. (1998), "A Comment on Aumann's Bayesian View," Econometrica 66, 923-927.

Harsanyi, J. (1967-8), "Games with Incomplete Information Played by 'Bayesian' Players," Parts I-III, Management Science 8, 159-182, 320-334, 486-502.

—— (1973), "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," International Journal of Game Theory 2, 1-23.

Heifetz, A. (2001), "The Positive Foundation of the Common Prior Assumption," California Institute of Technology, mimeo.

—— and P. Mongin (2001), "Probability Logic for Type Spaces," Games and Economic Behavior 35, 31-53.

—— and D. Samet (1998), "Topology-Free Typology of Beliefs," *Journal of Economic Theory* 82, 324-341.

Hintikka, J. (1962), *Knowledge and Belief*, Ithaca: Cornell University Press.

Kreps, D., P. Milgrom, J. Roberts and R. Wilson (1982), "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma," *Journal of Economic Theory* 27, 245-252.

Kripke, S. (1959), "A Completeness Theorem in Modal Logic," *Journal of Symbolic Logic* 24, 1-14.

Lewis, D. (1969), *Convention*, Cambridge: Harvard University Press.

Luce, R.D., and H. Raiffa (1957), *Games and Decisions*, New York: Wiley.

Meier, M. (2001), "An Infinitary Probability Logic for Type Spaces," Bielefeld and Caen, mimeo.

Mertens, J.-F., S. Sorin and S. Zamir (1994), *Repeated Games*, Center for Operations Research and Econometrics, Université Catholique de Louvain, Discussion Papers 9420, 9421, 9422.

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

Morris, S. (1994), "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica* 62, 1327-1347.

Myerson, R. B. (1984), "Cooperative Games with Incomplete Information," *International Journal of Game Theory* 13, 69-96.

Nau, R.F., and K. F. McCardle (1990),"Coherent Behavior in Noncooperative Games," *Journal of Economic Theory* 50, 424-444.

Samet, D. (1990), "Ignoring Ignorance and Agreeing to Disagree," *Journal of Economic Theory* 52, 190-207.

—— (1998a), "Iterative Expectations and Common Priors," *Games and Economic Behavior* 24, 131-141.

—— (1998b), "Common Priors and Separation of Convex Sets," *Games and Economic Behavior* 24, 172-174.

Von Neumann, J., and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.

Wilson, R. (1978), "Information, Efficiency, and the Core of an Economy," *Econometrica* 46, 807-816.

Appendix: Limitations of the Syntactic Approach

Aviad Heifetz

The interpretation of the universal model of Section 8 as "canonical," with the association of beliefs to states "self-evident" by virtue of the states' inner structure, relies nevertheless on the non-trivial assumption that the players' beliefs cannot be but σ-additive. This is highlighted by the following example.

In the Mertens-Zamir (1985) construction mentioned in Section 8, we shall focus our attention on a subspace Ω whose states ω can be represented in the form

$$\omega = (s,\, i_1 i_2 i_3 \ldots,\, j_1 j_2 j_3 \ldots)$$

where each digit s, i_k, j_k may assume the value 0 or 1. The state of nature is s. The sequences $i_1 i_2 i_3 \ldots$, $j_1 j_2 j_3 \ldots$ encode the beliefs of the players i and j, respectively, in the following way. If the sequence starts with 1, the player assigns probability 1 to the true state of nature s in ω. Otherwise, if her sequence starts with 0, she assigns probabilities half-half to the two possible states of nature. Inductively, suppose we have already described the beliefs of the players up to level n. If $i_{n+1} = 1$, i believes with probability 1 that the n-th digit of j equals the actual value of j_n in ω, and otherwise — if $i_{n+1} = 0$ — she assigns equal probabilities to each of the possible values of this digit, 0 or 1. This belief of i is independent of i's lower-level beliefs. In addition, i assigns probability 1 to her own lower-level beliefs. The $n+1$-th level of belief of individual j is defined symmetrically.

Notice that up to every finite level n there are finitely many events to consider, so the beliefs are trivially σ-additive. Therefore, by the Kolmogorov extension theorem, for the hierarchy of beliefs of every sequence, there is a unique σ-additive coherent extension to a probability measure[29] over Ω. When $i_n = 0$ for all n, the strong law of large numbers asserts that this limit extension assigns probability 1 to the sequences of j where

$$\lim_{n\to\infty} \frac{j_1 + j_2 + \ldots j_n}{n} = \frac{1}{2}.$$

[29] With the σ-field generated by the events that depend on finitely many digits.

However, there are finitely additive coherent extensions to i's finite level beliefs that are concentrated on the disjoint set of j's sequences where

$$\liminf_{n\to\infty} \frac{j_1 + j_2 + \ldots j_n}{n} > \frac{1}{2} \tag{A.1}$$

and in fact there are finitely additive coherent extensions concentrated on any tail event of sequences $j_1 j_2 j_3 \ldots$, i.e., an event where every possible initial segment $j_1 \ldots j_n$ appears in some sequence.[30]

Thus, though the finite-level beliefs single out unique σ-additive limit beliefs over Ω, nothing in them can specify that these limit beliefs must be σ-additive. If finitely additive beliefs are not ruled out in the players' minds, we cannot assume that the inner structure of the states $\omega \in \Omega$ specifies uniquely the beliefs of the players.

A similar problem presents itself if we restrict our attention to knowledge. In the syntactic formalism of Section 4, replace the belief operators p_i^α with knowledge

[30] To see this, observe that to any event defined by k out of j's digits, the sequence

$$i_1 i_2 i_3 \ldots = 000 \ldots \tag{A.2}$$

assigns the probability 2^{-k}. Therefore, the integral I w.r.t. this belief is well defined over the vector space $\mathcal{V}$ of real-valued functions which are each measurable w.r.t. finitely many of j's digits. Now, for every real-valued function g on j's sequences, define the functional

$$\overline{I}(g) = \inf\{I(f) : f \in \mathcal{V},\, g \leq f\}.$$

Then $\overline{I}$ is clearly sub-additive, $\overline{I}(\alpha g) = \alpha \overline{I}(g)$ for $\alpha \geq 0$, and $\overline{I} = I$ on $\mathcal{V}$. Therefore, by the Hahn-Banach theorem, there is an extension (in fact, many extensions) of I as a positive linear functional to the limits of sequences of functions in $\mathcal{V}$, and further to all the real-valued functions on Ω, satisfying

$$I(g) \leq \overline{I}(g).$$

Restricting our attention to characteristic functions g, we thus get a finitely additive coherent extension of (A.2) over Ω.

The proof of the Hahn-Banach theorem proceeds by consecutively considering functions g to which I is not yet extended, and defining

$$I(g) = \overline{I}(g).$$

If the first function g_1 to which I is extended is a characteristic function of a tail event (like A.1), the smallest $f \in \mathcal{F}$ majorizing g_1 is the constant function $f \equiv \mathsf{1}$, so

$$I(g) = I(\mathsf{1}) = 1,$$

and the resulting coherent extension of (A.2) assigns probability 1 to the chosen tail event.

operators k_i. When associating sentences to states of an augmented semantic belief system in Section 5, say that the sentence $k_i e$ holds in state ω if $\omega \in K_i E$,[31] where E is the event that corresponds to the sentence e. The canonical knowledge system Γ will now consist of those lists of sentences that hold in some state of some augmented semantic belief system. The information set $\mathsf{I}_i(\gamma)$ of player i at the list $\gamma \in \Gamma$ will consist of all the lists $\gamma' \in \Gamma$ that contain exactly the same sentences of the form $k_i f$ as in γ. It now follows that

$$k_i f \in \gamma \Leftrightarrow \gamma \in K_i E_f \tag{A.3}$$

where $E_f \subseteq \Gamma$ is the event that f holds,[32] and the knowledge operator K_i is as in Section 7: $K_i E = \{\gamma \in \Gamma : \mathsf{I}_i(\gamma) \subseteq E\}$.

However, there are many alternative definitions for the information sets $\mathsf{I}_i(\gamma)$ for which (A.3) would still obtain.[33] Thus, by no means can we say that the information sets $\mathsf{I}_i(\gamma)$ are "self-evident" from the inner structure of the lists γ which constitute the canonical knowledge system.

To see this, consider the following example (similar to that in Fagin et al. 1991). Ana, Bjorn and Christina participate in a computer forum over the web. At some point Ana invites Bjorn to meet the next evening. At that stage they leave the forum to continue the chat in private. If they eventually exchange n messages back and forth regarding the meeting, there is mutual knowledge of level n between them that they will meet, but not common knowledge, which they could attain, say, by eventually talking over the phone.

Christina doesn't know how many messages were eventually exchanged between Ana and Bjorn, so she does not exclude any finite level of mutual knowledge between them about the meeting. Nevertheless, Christina could still rule out the possibility of common knowledge (say, if she could peep into Bjorn's room and see that he was glued to his computer the whole day and spoke with nobody). But if this situation is formalized by a list of sentences γ, Christina does not exclude the possibility of common knowledge between Ana and Bjorn with the above definition of $I_{Christina}(\gamma)$. This is because there exists an augmented belief system Ω with a state ω' in which there is common knowledge between Ana and Bjorn, while Christina does not exclude any finite level of mutual knowledge between them, exactly as in the situation above. We then have $\gamma' \in \mathsf{I}_{Christina}(\gamma)$, where γ' is the list of sentences that hold in ω'. Thus, if we redefine $\mathsf{I}_{Christina}(\gamma)$ by omitting γ'

[31] The semantic knowledge operator $K_i E$ is defined in Section 7.

[32] As in Section 8.

[33] In fact, as many as there are subsets of Γ! (Heifetz 1999.)

from it, (A.3) would still obtain, because (A.3) refers only to sentences or events that describe finite levels of mutual knowledge.

At first it may seem that this problem can be mitigated by enriching the syntax with a common knowledge operator (for every subgroup of two or more players). Such an operator would be shorthand for the infinite conjunction "everybody (in the subgroup) knows, and everybody knows that everybody knows, and...". This would settle things in the above example, but create numerous new, analogous problems. *The discontinuity of knowledge is essential:* The same phenomena persist (i.e., (A.3) does not pin down the information sets $I_i(\gamma)$) even if the syntax explicitly allows for infinite conjunctions and disjunctions of sentences of whatever chosen cardinality (Heifetz 1994).

References

Fagin, R., Y. J. Halpern and M.Y. Vardi (1991), "A Model-Theoretic Analysis of Knowledge," *Journal of the Association for Computing Machinery (ACM) 91*, 382-428.

Heifetz, A. (1994), "Infinitary Epistemic Logic," in *Proceedings of the Fifth Conference on Theoretical Aspects of Reasoning about Knowledge (TARK 5)*, Los Altos, California: Morgan Kaufmann, 95-107. Extended version in *Mathematical Logic Quarterly 43* (1997), 333-342.

—— (1999), "How Canonical is the Canonical Model? A Comment on Aumann's Interactive Epistemology," *International Journal of Game Theory 28*, 435-442.