Int J Game Theory (1999) 28:263–300

International Journal of Game Theory

Interactive epistemology I: Knowledge*

Robert J. Aumann

Center for Rationality and Interactive Decision Theory, The Hebrew University, Feldman Building, Givat Ram, 91904 Jerusalem, Israel (e-mail: raumann@math.huji.ac.il)

Abstract. Formal Interactive Epistemology deals with the logic of knowledge and belief when there is more than one agent or "player." One is interested not only in each person's knowledge about substantive matters, but also in his knowledge about the others' knowledge. This paper examines two parallel approaches to the subject. The first is the *semantic* approach, in which knowledge is represented by a space Ω of *states of the world*, together with partitions $\mathscr{I}_i$ of Ω for each player i; the atom of $\mathscr{I}_i$ containing a given state ω of the world represents i's knowledge at that state – the set of those other states that i cannot distinguish from ω. The second is the *syntactic* approach, in which knowledge is embodied in sentences constructed according to certain syntactic rules. This paper examines the relation between the two approaches, and shows that they are in a sense equivalent.

In game theory and economics, the semantic approach has heretofore been most prevalent. A question that often arises in this connection is whether, in what sense, and why the space Ω and the partitions $\mathscr{I}_i$ can be taken as given and commonly known by the players. An answer to this question is provided by the syntactic approach.

Key words: Epistemology, interactive epistemology, knowledge, common knowledge, semantic, syntactic, model

---

* This paper is based on notes originally distributed in connection with a sequence of six lectures on Interactive Epistemology given at the Cowles Foundation for Research in Economics, Yale University, in the spring of 1989, and expanded and corrected several times since then. Previous versions were distributed in connection with the tutorial and workshop on "Knowledge and Game Theory" held at the Summer Institute in Game Theory, State University of New York at Stony Brook, July 20–24, 1992, and in connection with the Fifth Jerusalem Summer School in Economic Theory, on "Rationality of Action and Belief in the Economy," June 13–23, 1994. The author is grateful to Prof. Dov Samet for important help in preparing the current version. Research support from the USNSF under grant SES-9209277 and previous grants is gratefully acknowledged.

0. Introduction

In interactive contexts like game theory and economics, it is important to consider what each player knows and believes about what the other players know and believe. Here we discuss and relate two formalisms – the *semantic* and the *syntactic* – for analyzing these matters. The current paper treats knowledge only; probability is left to a companion paper (Aumann 1999).

The semantic formalism consists of a "partition structure:" a space Ω of *states of the world*, together with a partition of Ω for each player, whose atoms represent *information sets*[1] of that player; Ω is called the *universe*. Like in probability theory, *events* are subsets of Ω; intuitively, an event is identified with the set of all those states of the world at which the event obtains. Thus an event E obtains at a state ω if and only if $\omega \in E$, and a player i "knows" E at ω if and only if E includes his information set at ω. For i to know E is itself an event, denoted K_iE: it obtains at some states ω, and at others does not. The knowledge operators K_i can therefore be concatenated; for example, K_jK_iE denotes the event "j knows that i knows E."

The syntactic formalism, on the other hand, is built on propositions, expressed in a formal language. The language has logical operators and connectives, and also operators k_i expressing knowledge: If e is a sentence, then k_ie is also a sentence, whose meaning is "i knows e." The operators k_i can be iterated: The sentence k_jk_ie means "j knows that i knows e." Logical relations between the various propositions are expressed by formal rules.

There is a rough correspondence between the two formalisms: Events correspond to sentences, unions to disjunctions, intersections to conjunctions, inclusions to implications, complementation to negation, and semantic knowledge operators K_i to syntactic knowledge operators k_i. But the correspondence really *is* quite rough; for example, only some – not all – events correspond to syntactically admissible sentences.

While the semantic formalism is the more convenient and widely used of the two, it is conceptually not quite straightforward. One question that often arises is, what do the players know about the formalism itself? Does each know the others' partitions? If so, from where does this knowledge derive? If not, how can the formalism indicate what each player knows about the others' knowledge? For example, why would the event K_jK_iE then signify that j knows that i knows E?

Another point concerns the interpretation of the concept "state of the world." In game theoretic applications, for example, the state of the world often specifies the players' actions. Some feel that this restricts the freedom of action of a player, by "forcing" him, at the state ω, to choose the action that ω specifies for him. Why, they ask, should he not be allowed, if he wishes, to choose a different action at that state?

More generally, the whole idea of "state of the world," and of a partition structure that accurately reflects the players' knowledge about other players' knowledge, is not transparent. What *are* the states? Can they be explicitly described? Where do they come from? Where do the information partitions

[1] I.e., he can distinguish between states ω and ω' if and only if they are in different atoms of his partition.

come from? What justifies positing this kind of model, and what justifies a particular partition structure?

One may also ask about the global properties of the partition structure. We have seen that there is a rough correspondence between the semantic and syntactic models. Do global properties of the partition structure correspond to anything syntactic? For example, if the partition structure is finite, does this correspond to anything in the syntactic formalism? Can finiteness of the partition structure be formulated in syntactic terms – that is, in plain English?

A related point is that the semantic formalism is not *canonical*: There are many essentially different partition structures. Which is the "right" one?

These questions are not unanswerable. For example, an answer to the last question (which is the "right" partition structure?) is that it usually doesn't matter; proofs using the semantic formalism generally work with *any* partition structure. We also hold that the issue of the players' "freedom of action," while superficially puzzling, poses no real problem (Aumann and Brandenburger 1995, Section 7a, p. 1174). Yet all in all, there is little doubt that conceptually and philosophically, the semantic formalism is comparatively subtle.

In contrast, the syntactic formalism is entirely straightforward. One reasons from explicit hypotheses to explicit conclusions using explicit, transparent rules of deduction. There are no philosophical subtleties, no problems with the meaning of the sentences that enter the syntactic formalism. And *this* formalism *is* canonical.

Thus the semantic formalism appears to involve nontrivial conceptual issues. There are two approaches to these issues. One is to deal with them on the verbal, conceptual level, while retaining the semantic formalism as a formal primitive. In the other, the semantic formalism – the partition structure – is no longer taken as primitive, but is explicitly constructed from another formalism.

In the past, such explicit constructions have employed a hierarchical procedure (see Section 10c). While perfectly coherent, these hierarchical procedures are cumbersome and far from transparent. Here, we construct a canonical semantic partition structure in a more straightforward way, directly in terms of the syntactic formalism. In this construction, the states are simply lists of propositions. Each state ω is *defined* by specifying those propositions that hold at ω; any list of propositions that is complete, coherent and consistent in the appropriate sense constitutes a state.

Before presenting this construction, we set forth both the semantic (Sections 1, 2, 3) and the syntactic (Sections 4, 5) formalisms in detail. Section 1 presents the foundations of the semantic formalism for the fundamental case of one player, including an axiomatic development of the knowledge operator. Section 2 develops the many-player case, concentrating on a succinct presentation of the fundamental properties of common knowledge. Section 3 is a conceptual discussion of the semantic formalism, presenting in more detail some of the issues discussed above, and setting the stage for the subsequent canonical construction in terms of the syntactic formalism. Section 4 formally presents the syntactic formalism, and Section 5 discusses it. Section 6 constructs the canonical semantic partition structure in terms of the syntactic formalism, and Section 7 is a conceptual discussion of this construction. Section 8 establishes in precise terms the above-discussed "rough" correspondence between the syntactic and the semantic formalisms. In Section 9, this is used to show that a list of sentences in the syntactic formalism is logically consistent if and only if it has a semantic "model." Section 10 is devoted to additional discussion, including discussion of the relationship between the "direct" construction of the canonical semantic partition structure that is presented here, and the hierarchical construction to which we alluded above. An appendix discusses the cardinality of the canonical universe (which, when there is more than one player, is at least that of the continuum).

This paper originated in notes for a series of lectures given at Yale University in 1989, and informally circulated, in various forms, for many years thereafter. Much of the material is basically due to others, and has been known for some time[2], though perhaps not in the form presented here. Nevertheless, this is not a review; there is no attempt to cover any significant part of the literature, nor even to cite it. Rather, it is largely expository; the purpose is to make the basic facts of interactive epistemology accessible in a succinct form that should be useful for game theorists and economists.

Doing justice to all the relevant previous work would take more effort, time and space than we can afford here. However, we do wish to cite one particular work to which we owe the greatest immediate debt: Dov Samet's "Ignoring Ignorance and Agreeing to Disagree" (1990), from which we drew the simple but ingenious and fundamental idea of formally characterizing a state of the world by the sentences that hold there.

1. The semantics of knowledge for a single individual

In this section we present and examine the relationships between five different but equivalent semantic formalizations of knowledge: knowledge functions, information functions, information partitions, knowledge operators, and knowledge ufields (universal fields).

We start with a set Ω whose members are called *states of the world*, or simply *states*. An *event* is defined as a subset of Ω; the family of all events is denoted $\mathscr{E}$. This formalism may be familiar to some readers from probability theory; an event in the ordinary sense of the word, like "it will be sunny tomorrow," is identified in the formalism with the set of all states at which it is sunny tomorrow. Union of events corresponds to disjunction, intersection to conjunction, and complementation ($\sim$) to negation. For example, the event "it will snow or rain tomorrow" is the same as the union of the event "it will snow tomorrow" with the event "it will rain tomorrow." When E and F are events with $E \subset F$, we will say that *E entails F*.

We will assume given a function κ on Ω, which we call the *knowledge function*. The range of κ is an abstract set, whose members represent different possible "states of knowledge" of some fixed individual i. Thus, $\kappa(\omega)$ represents the knowledge that i possesses when the true state of the world is ω. For example, this knowledge might include today's weather, but not tomorrow's. A related way of thinking about $\kappa(\omega)$ is as the signal that i receives from the outside world when the true state of the world is ω.

Define a function $\mathbf{I}$ on Ω by

$$\mathbf{I}(\omega) := \{\omega' \in \Omega : \kappa(\omega') = \kappa(\omega)\}. \tag{1.1}$$

[2] E.g., Fagin et al. (1995); Hintikka (1992); Kripke (1959); Lewis (1969).

In words, $\mathbf{I}(\omega)$ is the set of all those states that i cannot distinguish from ω. Thus if ω is the true state of the world, i will usually not know this; he will know only that *some* state in $\mathbf{I}(\omega)$ is the true one. $\mathbf{I}(\omega)$ consists of those states ω' that i considers possible.

The reader may verify that for any states ω and ω',

$$\omega \in \mathbf{I}(\omega), \text{ and} \tag{1.11}$$

$$\mathbf{I}(\omega) \text{ and } \mathbf{I}(\omega') \text{ are either disjoint or identical.} \tag{1.12}$$

Any function $\mathbf{I}$ on Ω satisfying 1.11 and 1.12 is called an *information function*. From 1.11 and 1.12 it follows that the distinct $\mathbf{I}(\omega)$ form a *partition* of Ω; that is, they are pairwise disjoint and their union is Ω. We call this partition i's *information partition*, and denote it $\mathscr{I}$; atoms of $\mathscr{I}$ are called *information sets*. The family of all unions of events in $\mathscr{I}$, denoted $\mathscr{K}$, is a *universal field*, or *ufield* for short; that is, it is closed under complementation and under arbitrary (not necessarily finite or denumerable) unions, and therefore also under arbitrary intersections. It is called i's *knowledge ufield*.

Now define an operator $K : \mathscr{E} \to \mathscr{E}$ as follows: for any event E,

$$KE \text{ is the union of all the events in } \mathscr{I} \text{ that are included in } E. \tag{1.2}$$

Equivalent forms of this definition are

$$\omega \in KE \text{ if and only if } \mathrm{I}(\omega) \subset E, \quad \text{and} \tag{1.21}$$

$$KE \text{ is the largest element of } \mathscr{K} \text{ that is included in } E. \tag{1.22}$$

In words, KE is the event that i knows that the event E obtains[3]; more explicitly, the set of all states ω at which i knows that E contains ω (he usually will not know the true ω).

From 1.22 it follows that $KE \in \mathscr{K}$ for all E. Conversely, if $E \in \mathscr{K}$, then 1.22 yields $E = KE$, so each member of $\mathscr{K}$ has the form KE. Thus we conclude that

$$\mathscr{K} = K\mathscr{E} := \{KE : E \in \mathscr{E}\}. \tag{1.23}$$

In words, $\mathscr{K}$ is the family of events that express i's knowing some particular event.

The reader may verify, directly from its definition, that the operator K has the following properties (for all events E and F):

$$KE \subset E \tag{1.31}$$

$$E \subset F \quad \text{implies} \quad KE \subset KF \tag{1.32}$$

$$\sim KE \subset K \sim KE. \tag{1.33}$$

[3] By "obtains" we mean "happened, is happening, or will happen." Time is irrelevant; we want to avoid tenses. If the reader wishes, he can think of a state ω of the world as a complete possible history, from the big bang until the end of the world. Thus "it will snow tomorrow" and "it snowed yesterday" are both events, which may or may not "obtain" at a given ω.

In words, 1.31 says that a person can only know something if it is true; and 1.33, that if he does *not* know something, then he knows that he does not know it. 1.32 expresses the intuition that if E entails F, and i knows that E, then he may conclude that F, so he knows F. Any operator $K : \mathscr{E} \to \mathscr{E}$ satisfying 1.31 through 1.33 is called a *knowledge operator*.

Note also that

$$K\left(\bigcap_{\alpha} E_{\alpha}\right) = \bigcap_{\alpha} K(E_{\alpha}), \tag{1.4}$$

where the intersection may be over an arbitrary index set; this, too, follows directly from 1.2. In words, 1.4 says that i knows each of several events if and only if he knows that they all obtain.

Rather than starting with the knowledge function κ as a primitive, and deriving from it the information function $\mathbf{I}$, the information partition $\mathscr{I}$, the knowledge operator K, and the knowledge ufield $\mathscr{K}$, one can start from any other one of these five and derive the other four from it. The only one of these derivations that is not entirely straightforward[4] is that starting with K; we show below how to do this. To have a term for the relationship between the objects $\kappa, \mathbf{I}, \mathscr{I}, K, \mathscr{K}$ defined in this section, call any two of them *associated*.

We now show how $\mathbf{I}, \kappa, \mathscr{I}$ and $\mathscr{K}$ can be derived if we start with a knowledge operator K as a primitive. So suppose we are given such a K, i.e., an operator $K : \mathscr{E} \to \mathscr{E}$ satisfying 1.31 through 1.33. It is enough to define $\mathbf{I}$, as we have already noted that $\mathscr{I}$, κ, and $\mathscr{K}$ can be derived from $\mathbf{I}$. We define

$$\mathbf{I}(\omega) := \sim K \sim \{\omega\}. \tag{1.5}$$

In words, 1.5 defines $\mathbf{I}(\omega)$ as the set of all those states at which i does not know that ω did not occur; i.e., those he cannot distinguish from ω, the ones he considers possible. This is precisely the intuitive meaning of $\mathbf{I}$ that is embodied in the original definition 1.1. To justify the definition (1.5), we must show

Proposition 1.6. *As defined from K by 1.5,* $\mathbf{I}$ *is an information function and satisfies 1.21.*

Indeed, by proving that $\mathbf{I}$ satisfies 1.21, we show that it is associated with K in the sense explained at the end of Section 1. Thus if we start with $\mathbf{I}$, define K from it by means of 1.2, and then apply the definition (1.5), we get back an information function identical to that with which we started.

Proof of Proposition 1.6: In this proof, we may use only 1.5 and 1.31 through 1.33, not formulas derived from the definitions in the body of the text.

To show that $\mathbf{I}$ is an information function, we must establish 1.11 and 1.12. To prove 1.11, note that $\sim\{\omega\} \supset K \sim \{\omega\}$ by 1.31, so $\{\omega\} \subset \sim K \sim \{\omega\} = \mathbf{I}(\omega)$. 1.12 will be established below.

[4] Starting from $\mathscr{K}$, we call two states *equivalent* iff they are not separated by $\mathscr{K}$, and define $\mathscr{I}$ to consist of the equivalence classes.

Now *define* $\mathscr{K}$ by 1.23 (remember that K is now the primitive); we then assert that

$$\mathbf{I}(\omega) \in \mathscr{K}, \text{ and} \tag{1.61}$$

$$\mathbf{I}(\omega) \text{ is the intersection of all members of } \mathscr{K} \text{ containing } \omega. \tag{1.62}$$

To prove 1.61, note that by 1.33, $\mathbf{I}(\omega) = \sim K \sim \{\omega\} = K \sim K \sim \{\omega\} \in K\mathscr{E} = \mathscr{K}$. To prove 1.62, suppose $\omega \in F \in \mathscr{K}$. Then $F = KE$ for some E in $\mathscr{E}$. Thus $\sim KE = \sim F \subset \sim \{\omega\}$. So by 1.33 and 1.32, $\sim KE \subset K \sim KE \subset K \sim \{\omega\}$. Hence $\mathbf{I}(\omega) = \sim K \sim \{\omega\} \subset KE = F$. Thus $\mathbf{I}(\omega)$ is included in every member F of $\mathscr{K}$ that contains ω, and so in their intersection. For the reverse inclusion, note that by 1.11 and 1.61 (which we have already proved), $\mathbf{I}(\omega)$ itself is an element of $\mathscr{K}$ containing ω, and therefore it includes the intersection of all such elements.

To prove 1.12, note that either ω is contained in precisely the same members of $\mathscr{K}$ as ω', or not. If yes, then 1.62 yields $\mathbf{I}(\omega) = \mathbf{I}(\omega')$. If not, then w.l.o.g. there is an F in $\mathscr{K}$ with $\omega \in F$ and $\omega' \in \sim F$; then by 1.62, $\mathbf{I}(\omega) \cap \mathbf{I}(\omega') = \varnothing$.

To prove 1.2, we first show

$$KE \text{ is the disjoint union of those events } \mathbf{I}(\omega) \text{ that are included in E.} \tag{1.63}$$

Indeed, that the union is disjoint follows from 1.12. Next, if $\mathbf{I}(\omega) \subset E$, then by 1.5, 1.33, and 1.32, $\mathbf{I}(\omega) \subset K\mathbf{I}(\omega) \subset KE$. Hence the union in question is included in KE. On the other hand, if $\omega \in KE$, then by 1.11 and 1.62, $\omega \in \mathbf{I}(\omega) \subset KE$, so every element of KE is in the union in question, so the union includes KE.

1.2 now follows from 1.11, 1.12, and 1.63. ■

Notes to Section 1

(i) An alternative for 1.32 is the intuitively more transparent

$$K(E \cap F) = K(E) \cap K(F), \tag{1.64}$$

which says that i knows two things if and only if he knows each. Note that 1.64 implies[5] 1.32; indeed, if $E \subset F$ then 1.64 yields $KE = K(E \cap F) = KE \cap KF \subset KF$. But from Note (i) it follows that in the presence of the other two conditions, 1.32 implies 1.4 and a fortiori 1.64, so that the two systems (with 1.32 and with 1.64) are equivalent.

(ii) Note that

$$K\Omega = \Omega. \tag{1.7}$$

[5] The converse is false (as a referee pointed out): take $\Omega := \{1, 2, 3\}$, $KE := E \backslash \{\min E\}$.

Indeed, $K\emptyset = \emptyset$, by 1.31; so $\Omega \supset K\Omega = K \sim \emptyset = K \sim K\emptyset \supset \sim K\emptyset = \sim \emptyset = \Omega$, by 1.33, and 1.7 follows. In words, 1.7 says that if an event obtains at all states of the world, then i knows it.

(iii) Condition 1.33 is sometimes called *negative introspection*. A related condition is

$$KE = KKE, \tag{1.8}$$

sometimes called *positive introspection*, which says that if a person knows something, then he knows that he knows it. 1.8 follows from 1.31 and 1.33, as follows: $\sim KE = K \sim KE$, so $KE = \sim K \sim KE$, so $KKE = K \sim K \sim KE = \sim K \sim KE = KE$. From 1.8 and 1.23 it follows that

$$\mathscr{K} = \{E : E = KE\}. \tag{1.9}$$

2. Common knowledge

Let N be a collection of several individuals i, each with his own knowledge function κ_i, knowledge operator K_i, knowledge ufield $\mathscr{K}_i$, information partition $\mathscr{I}_i$, and information function $\mathbf{I}_i$; the space Ω of states of the world and the ufield $\mathscr{E}$ of events remain the same for all individuals. We call N the *population*.

Call an event E *common knowledge* (in the population) if all know E, all know that all know it, all know that all know that all know it, and so on ad infinitum. Formally, define operators K^m from $\mathscr{E}$ to $\mathscr{E}$ by

$$K^1E := \bigcap_{i \in N} K_iE, \quad K^{m+1}E := K^1K^mE, \tag{2.1}$$

then define K^∞ by

$$K^\infty E := K^1E \cap K^2E \cap \dots. \tag{2.2}$$

If $\omega \in K_iE$, we say that "i knows E at ω." Thus K_iE is the set of all states of the world at which i knows E; in other words, K_iE is the event that i knows E. If $\omega \in K^\infty E$, we say that "E is commonly known at ω;" and $K^\infty E$ is the event that E is commonly known. Similarly, K^1E is the event that E is *mutually* known (between the individuals in N), and K^mE that E is *m'th level mutual* knowledge. Thus m'th level mutual knowledge of E means that all relevant individuals know E, all know that all know it, and so on, but only m times; and an event is common knowledge iff it is mutual knowledge at all levels.

Lemma 2.3. *$K_iK^\infty E = K^\infty E$ for all i.*

Proof: The inclusion $\subset$ follows from 1.31. For the opposite inclusion, 2.1 and 1.4 yield $K_iK^\infty E = K_i\bigcap_{m=1}^\infty K^mE = \bigcap_{m=1}^\infty K_iK^mE \supset \bigcap_{m=1}^\infty K^1K^mE = \bigcap_{m=2}^\infty K^mE \supset \bigcap_{m=1}^\infty K^mE = K^\infty E$. ■

Lemma 2.4.

$$K^\infty E \subset E, \text{ and} \tag{2.41}$$

$$E \subset F \quad \text{implies} \quad K^\infty E \subset K^\infty F. \tag{2.42}$$

Proof: Follows from 2.1, 2.2, 1.31, and 1.32.

Call an event F *self-evident* if $K_i F = F$ for all i; that is, if the event itself entails everyone knowing it (in other words, whenever it obtains, everyone knows that it obtains). For example, if two people enter into a contract[6], then it is self-evident to them that they do. From 1.9 we get

$$\text{An event is self-evident if and only if it is in } \bigcap_{i \in N} \mathscr{K}_i. \tag{2.5}$$

Theorem 2.6. *$K^\infty E$ is the largest event in $\bigcap_{i \in N} \mathscr{K}_i$ that is included in E.*
In words: E is commonly known if and only if some self-evident event that entails *E obtains.*

Proof: By 2.3, 1.8, and 2.41, $K^\infty E$ is in $\bigcap_{i \in N} \mathscr{K}_i$ and is included in E. If $E \supset F \in \bigcap_{i \in N} \mathscr{K}_i$, then $F \in \mathscr{K}_i$, so $K_i F = F$ by 1.8, so $K^\infty F = F$ by 2.1, so $F = K^\infty F \subset K^\infty E$ by 2.42. ■

Corollary 2.7. *K^∞ is a knowledge operator, and the associated knowledge ufield is $\bigcap_{i \in N} \mathscr{K}_i$.*

Remark. The events in $\bigcap_{i \in N} \mathscr{K}_i$ are those that are in all the $\mathscr{K}_i$; those that are *common* to them all. This provides an additional rationale for the term "common knowledge;" it is knowledge that is, in a sense, common[7] to the protagonists. Of course, the main rationale for the term "common knowledge" is that its everyday meaning[8] corresponds nicely to the formal definition at 2.1.

---

[6] In common law, a contract is sometimes characterized as a "meeting of the minds." Interpreted more broadly – i.e., epistemically only, without necessarily involving the element of mutual commitment – the idea of "a meeting of the minds" nicely encapsulates the notion of a self-evident event.

[7] To be sharply distinguished from the knowledge that results if the protagonists share (tell each other) what they know. The latter corresponds to the ufield, sometimes denoted $\bigvee_{i \in N} \mathscr{K}_i$, comprising unions of events of the form $\bigcap_{i \in N} E_i$, where $E_i \in \mathscr{K}_i$ for each i. This is quite different from $\bigcap_{i \in N} \mathscr{K}_i$; in particular, $\bigcap_{i \in N} \mathscr{K}_i$ is included in (represents less information than) each of the $\mathscr{K}_i$, whereas $\bigvee_{i \in N} \mathscr{K}_i$ includes each of them (represents more information).

[8] Suppose you are told "Ann and Bob are going together," and respond "sure, that's common knowledge." What you mean is not only that everyone knows this, but also that the announcement is pointless, occasions no surprise, reveals nothing new; in effect, that the situation after the announcement does not differ from that before. This is precisely what is described in 2.6; the event "Ann and Bob are going together" – call it E – is common knowledge if and only if some event – call it F – happened that entails E and also entails all players' knowing F (like if all players met Ann and Bob at an intimate party). Contrast the familiar missionary story, in which announcing the facts changes the situation dramatically, in spite of everybody's knowing the facts, knowing that everybody knows them, and so on up to the 36'th order. If the facts had been common knowledge, in the everyday meaning of the phrase, the announcement would have changed nothing.

Proof: Each ufield $\mathscr{K}$ is associated with a unique information partition $\mathscr{I}$, and so by 1.2, with a unique knowledge operator K (given by 1.22). By 2.6, K^∞ is the knowledge operator associated with the ufield $\bigcap_{i \in N} \mathscr{K}_i$. ∎

Corollary 2.8. $K^\infty E \subset K^\infty(K^\infty E)$, *and* $\sim K^\infty E \subset K^\infty(\sim K^\infty E)$.
In words: If an event is commonly known, then it is commonly known that it is commonly known; if it is not commonly known, then it is commonly known that it is not commonly known.

Proof: A consequence of 2.7, using 1.8 and 1.33. ∎

Remark 2.9: $K^\infty \Omega = \Omega$.
In words: An event obtaining at all states is commonly known at each state.

Proof: 1.7 and 2.7. ∎

The information partition and information ufield corresponding to the knowledge operator K^∞ are denoted $\mathscr{I}^\infty$ and $\mathscr{K}^\infty$ respectively. Ω is called the *universe*. The member of $\mathscr{K}^\infty$ are called *common knowledge subuniverses*, or simply *subuniverses*; they are the self-evident events. The atoms of $\mathscr{I}^\infty$ are called *common knowledge components* (of Ω); they are the minimal non-empty subuniverses. The structure consisting of the universe Ω, the population N, and the knowledge functions κ_i of the individuals i is called a *semantic knowledge system*.

In the sequel, we sometimes consider common and mutual knowledge among the individuals in a specified proper subset N' of N. Thus in Section 10A, we treat a world with three individuals, and consider common and mutual knowledge among two out of the three. In that case, the same definitions and notation as above apply, with N replaced by N'.

3. Discussion

When we come to interpret the model introduced in Section 2, an inevitable question is, "what do the participants know about the model itself?" Does each "know" the information partitions $\mathscr{I}_i$ of the others? Are the $\mathscr{I}_i$ themselves in some sense "common knowledge"? If so, how does this common knowledge come about – how does each individual get to know what the others' partitions are? If not, how does the model reflect each individual's information – or lack of information – about the others' partitions? To do this right, doesn't one need to superpose another such model over the current one, to deal with knowledge of (or uncertainty about) the $\mathscr{I}_i$? But then, wouldn't one need another and yet another such model, without end even in the transfinite domain?

Addressing this question in 1976, we wrote as follows: "... the implicit assumption that the information partitions ... are themselves common knowledge ... constitutes no loss of generality. Included in the full description of a state ω of the world is the manner in which information is imparted to the two persons. This implies that the information sets $\mathbf{I}_1(\omega)$ and $\mathbf{I}_2(\omega)$ are indeed defined unambiguously as functions of ω, and that these functions are known to both players."

In 1987, we were more expansive: "While Player 1 may well be ignorant of what Player 2 knows – i.e., of the element of $\mathscr{I}_2$ that contains the 'true' state ω of the world – 1 cannot be ignorant of the *partition* $\mathscr{I}_2$ itself. . . . Indeed, since the specification of each ω includes a complete description of the state of the world, it includes also a list of those other states ω' of the world that are, for Player 2, indistinguishable from ω. If there were uncertainty about this list on the part of Player 1 (or any other player), then the description of ω would not be complete; one should then split ω into several states, depending on which states are, for 2, indistinguishable from ω. Therefore the very description of the ω's implies the structure of $\mathscr{I}_2$, and similarly for all the $\mathscr{I}_i$. The description of the ω's involves no 'real' knowledge; it is only a kind of code book or dictionary. The structure of the $\mathscr{I}_i$ also involves no real knowledge; it simply represents different methods of classification in the dictionary."

That is all well and good as far as it goes; but it leaves some important questions unanswered.

Can such a "dictionary" actually be constructed? Isn't there some kind of self-reference implicit in the very idea of such a dictionary? If it can nevertheless be constructed, is the construction in some sense unique, "canonical"? If not, which of the possible "dictionaries" would the participants use?

The most convincing way to remove all these questions and doubts is to construct Ω and the $\mathscr{I}_i$ – or equivalently, the κ_i – in an explicit, canonical, manner, so that it is clear from the construction itself that the knowledge operators are "common knowledge" in the appropriate sense. This will be done in Section 6, making use of the syntactic formalism. But first, we introduce the syntactic formalism for its own sake.

4. The syntactic knowledge formalism

As in Section 2, we assume given a set N of *individuals*, called the *population*. We start by constructing a certain "language", in purely formal terms; afterwards we interpret it. The building blocks of the language are the following symbols, constituting the *keyboard*:

Letters from an *alphabet* $\mathfrak{X} := \{x, y, z, \ldots\}$, taken as fixed throughout; and the symbols $\vee$, $\neg$, (,), and k_i (for all i in N).

A *formula* is a finite string of symbols obtained by applying the following three rules in some order finitely often:

Every letter in the alphabet is a formula. (4.11)

If f and g are formulas, so is $(f) \vee (g)$. (4.12)

If f is a formula, so are $\neg(f)$ and $k_i(f)$ for each i. (4.13)

In the sequel, we often omit parentheses when the intended meaning is clear, and we use $f \Rightarrow g$ as an abbreviation for $(\neg f) \vee g$.

A *list* is a set of formulas. A list $\mathfrak{L}$ is called *logically closed*, or simply *closed*, if

$$(f \in \mathfrak{L} \text{ and } f \Rightarrow g \in \mathfrak{L}) \quad \text{implies} \quad g \in \mathfrak{L}. \tag{4.2}$$

It is called *epistemically closed* if

$$f \in \mathfrak{L} \quad \text{implies} \quad k_i f \in \mathfrak{L}, \tag{4.3}$$

and *strongly closed* if it is both logically and epistemically closed. The *strong closure* of a list $\mathfrak{L}$ is the smallest strongly closed list[9] that includes $\mathfrak{L}$. A *tautology*[10] is a formula in the strong closure of the list of all formulas having one of the following seven forms (for some f, g, h and i):

$$(f \vee f) \Rightarrow f \tag{4.41}$$

$$f \Rightarrow (f \vee g) \tag{4.42}$$

$$(f \vee g) \Rightarrow (g \vee f) \tag{4.43}$$

$$(f \Rightarrow g) \Rightarrow ((h \vee f) \Rightarrow (h \vee g)) \tag{4.44}$$

$$k_i f \Rightarrow f \tag{4.51}$$

$$k_i(f \Rightarrow g) \Rightarrow ((k_i f) \Rightarrow (k_i g)) \tag{4.52}$$

$$\neg k_i f \Rightarrow k_i \neg k_i f. \tag{4.53}$$

A formula g is a *consequence* of (or *follows* from) a formula f if $f \Rightarrow g$ is a tautology.

The set of all formulas with a given population N and alphabet $\mathfrak{X}$ is called a *syntax*, and is denoted $\mathfrak{S}(N, \mathfrak{X})$, or just $\mathfrak{S}$. For convenience[11], assume that N and $\mathfrak{X}$ are finite or denumerable; it follows that $\mathfrak{S}$ is denumerable.

5. Interpretation of the syntactic formalism

The syntactic formalism is subject to different interpretations; for now, we present only one. The letters of the alphabet represent what may be called "natural occurrences": substantive happenings that are not themselves described either in terms of people knowing something, or as combinations of other natural occurrences using the connectives of the propositional calculus[12]. Such a "natural" occurrence might be, "it will snow tomorrow." Not all possible natural occurrences need be represented in the alphabet; normally, one restricts oneself to natural occurrences that are "relevant" to the matter under discussion. In many cases of interest these constitute a finite set, perhaps even with just one element.

---

[9] The intersection of all strongly closed lists including $\mathfrak{L}$; it is itself strongly closed, and is included in all strongly closed lists that include $\mathfrak{L}$.

[10] See the next section for an informal explanation of this concept.

[11] This is used only twice – in 8.7 and in the appendix – and even there can be circumvented.

[12] The distinction between "natural occurrences" and arbitrary "occurrences" is analogous to that between "states of nature" and "states of the world" that one sometimes sees in the literature. We ourselves don't like this distinction; in the context of the current general treatment, we consider it artificial. But it is often used by economists, game theorists, and others who treat knowledge formally, and for the present we go along (but see Section 10b).

The symbol "k_i" means "i knows that ...". So if x stands for "it will snow tomorrow," then $k_i x$ stands for "i knows that it will snow tomorrow." The symbol "$\neg$" means "it is not true that", and the symbol "$\vee$" means "or". Parentheses have their usual meaning. Thus a formula is a **finite** concatenation of natural occurrences, using the operators and connectives[13] of the propositional calculus, and the knowledge operators k_i.

In ordinary discourse, a "tautology" is a statement that is logically necessary, one that embodies no empiric knowledge, nothing substantive about the real world, one whose truth is inherent in the meaning of the terms involved (like "a circle is round"). That is the meaning intended here. For example, formulas of the form 4.4·, which embody the axioms of the propositional calculus, are tautologies; so are formulas of the form 4.5·, which embody fundamental properties of knowledge.[14]

A logical deduction from a tautology is also a tautology. This is embodied in the condition that the set of tautologies be logically closed, i.e., obeys 4.2 (called the rule of *modus ponens*): if f and $f \Rightarrow g$ are tautologies, so is g.

In addition, the set of tautologies is epistemically closed, i.e., obeys 4.3 (this is called the rule of *necessitation*[15]): if f is a tautology, so is $k_i f$. In effect, this says that each individual knows each tautology, and that this is a logical necessity. Thus it is part of the logical infrastructure that the individuals are logically consistent. It follows that everybody knows this itself as a logical necessity, so that, for example, if f is a tautology, then so are $k_j k_i f$, $k_i k_j k_i f$, and so on. In this sense, one might say that it is "commonly known" that all individuals reason logically, though in the syntactic formalism, common knowledge has not been formally defined.

The formal definition of tautology in Section 4 embodies precisely these principles: A tautology is *defined* as a formula that follows from the axioms of logic (including the logic of knowledge) by repeated use of modus ponens and the rule of necessitation.

In applications, the semantic and the syntactic formalisms have similar functions: Either may be used formally to analyze interactive situations – such as games or economies – in which the knowledge of the protagonists plays a role. In principle, the syntactic formalism is the more straightforward; it simply represents the usual methods of logical deduction. Thus given a hypothesis

---

[13] As is well known, all the connectives of the propositional calculus can be defined in terms of $\vee$ and $\neg$.

[14] Substantively, the knowledge "axioms" 4.5· correspond roughly to the defining properties 1.3· of the semantic knowledge operator K; but there is a noteworthy difference between 1.32 and 4.52. Both have the same conclusion – that if i knows the hypothesis of an implication, then he knows the conclusion – but their hypotheses are different. The hypothesis of 1.32 is the implication itself; that of 4.52, that i knows the implication, which seems stronger. To reconcile the two, note that the set inclusion used in the semantic formalism denotes *logical* – or better, *tauto*logical – implication; to say that $E \subset F$ means that by the very definitions of E and F, it cannot be that F happens unless E does. Indeed, $E \subset F$ is not an event, and so cannot hold at only some states of the world; it is either true – a logical necessity – or false. On the other hand, the formula $f \Rightarrow g$ – which simply stands for $\neg f \vee g$ – is not a logical necessity; it may hold *sometimes* – under certain circumstances – and sometimes not. Unlike with $E \subset F$, therefore, it makes sense to talk of i's knowing $f \Rightarrow g$; and then one cannot conclude that i knows g from his knowing f unless $f \Rightarrow g$ is not merely true, but in fact *i knows* $f \Rightarrow g$.

[15] This term is inherited from modal logic, where "necessity" plays a role like that of "knowledge" in epistemic logic. In that context, the rule of necessitation says that tautologies are not just true, but "necessarily" true.

f, we may conclude that g if and only if $f \Rightarrow g$ is a tautology. The semantic formalism is more roundabout. In it, sentences are replaced by "events" – sets of states of the world. Specifically, a sentence f is replaced by the set of those states of the world at which f obtains. To conclude g from f, one shows that the corresponding events G and F satisfy $F \subset G$; i.e., that F entails G, that g obtains at each state of the world at which f does.

What we here call a "tautology" is in formal logic sometimes called a "theorem," the term "tautology" being reserved for sentences that "hold at" each state of each semantic knowledge system. Actually, the two meanings are equivalent, as will be shown below (9.4). Using the term "theorem" in this connection would cause confusion with the more usual kind of theorem – the kind that appears in papers like this, and in particular in this paper itself.

Though in principle the syntactic formalism is more straightforward, in practice the semantic formalism is often more useful.[16] On the other hand, the semantic formalism is beset by the conceptual difficulties discussed in Section 3 and in the introduction. To overcome these difficulties, we now construct an explicit canonical semantic knowledge system, using the syntactic formalism introduced in Section 4.

6. The canonical semantic knowledge system

As in Section 4, assume given a finite population N and an alphabet $\mathfrak{X}$. Call a list $\mathfrak{L}$ of formulas *coherent* if

$$\neg f \in \mathfrak{L} \quad \text{implies} \quad f \notin \mathfrak{L}, \tag{6.1}$$

complete if

$$f \notin \mathfrak{L} \quad \text{implies} \quad \neg f \in \mathfrak{L}. \tag{6.2}$$

Define a *state ω of the world*, or simply a *state*, as a closed, coherent, and complete list of formulas that contains all tautologies. Denote the set of all states $\Omega(N, \mathfrak{X})$, or simply Ω. For all individuals i, define a knowledge function κ_i on Ω by specifying that for all states ω,

$$\kappa_i(\omega) \text{ is the set of all formulas in } \omega \text{ that start with } k_i. \tag{6.3}$$

The system comprising Ω, N, and the κ_i is called the *canonical semantic knowledge system* for the population N and the alphabet $\mathfrak{X}$ (or simply the *canonical system*).

7. Interpretation of the canonical system

Intuitively, once a state ω of the world has been specified, there can be no remaining uncertainty, or at least no "relevant" uncertainty. This means that at ω, each of the denumerably many possible formulas is either definitely true

[16] This is an empirical observation; we are not sure of the reason. Part of it may be that the semantic formalism is set theoretic, and set theory has more immediacy and transparency than formal logic. For example, a syntactic tautology, no matter how complex, corresponds in the semantic formalism simply to the set Ω of all states.

or definitely false. Our device is formally to *define* a state ω as the list of those formulas that are true at ω.

Thus for each formula f, either f itself or its negation $\neg f$ must be in the list that defines ω (completeness), and only one of these two alternatives can obtain (coherence). Also, the ordinary processes of logical deduction are assumed valid at each state; anything that follows logically from formulas true at a certain state is also true at that state. This is embodied in two conditions. The first is that each state ω is logically closed, which means that *modus ponens* (4.2) is satisfied at ω ($\Rightarrow$ stands for "implies"). The second is that all tautologies are true at each state.

At a given state ω, which formulas does individual i know? Well, it is all written in the list that defines ω. He knows exactly those formulas that it says in the list that he knows: namely, those formulas f such that $k_i f$ is in the list. The list of all such $k_i f$ is $\kappa_i(\omega)$. So an individual can distinguish between two states ω and ω' if and only if $\kappa_i(\omega) \neq \kappa_i(\omega')$, i.e., if and only if he knows something at one state that he does not know at the other. The other representations of knowledge – $\mathbf{I}_i, \mathscr{I}_i, K_i$ and $\mathscr{K}_i$ – are then defined from κ_i as in Section 1.

We can now address the question raised in Section 3, whether each individual "knows" the information partitions $\mathscr{I}_i$ of the others. The $\mathscr{I}_i$ are defined in terms of the information functions κ_i, and so in terms of the operators k_i (which, in the formulation of Section 6, replace the κ_i as primitives of the system). Thus the question becomes, does each individual "know" the operators k_i of the others (in addition, of course, to his own)?

The answer is "yes." The operator k_i operates on formulas; it takes each formula f to another formula. Which other formula? What is the result of operating on f with the operator k_i? Well, it is simply the formula $k_i f$. "Knowing" the operator k_i just means knowing this definition. Intuitively, for an individual j to "know" k_i means that j knows what it means for i to know something. It does not imply that j knows any specific *formula* $k_i f$.

Suppose, for example, that f stands for "it will snow tomorrow." For j to know the operator k_i implies that j knows that $k_i f$ stands for "i knows that it will snow tomorrow;" it does *not* imply that j knows that i knows that it will snow tomorrow (indeed, i may not know this, and perhaps it really will not snow).

In brief, j's knowing the operator k_i means simply that j knows what it means for i to know something, not that j knows anything specific about what i knows.

Thus the assertion that each individual "knows" the knowledge operators k_i of all individuals has no real substance; it is part of the framework. If j did not "know" the operators k_i, he would be unable even to consider formulas in the language of Section 4, to say nothing of knowing or not knowing them.

From this we conclude that all individuals indeed "know"[17] the functions κ_i, as well as the partitions $\mathscr{I}_j$.

---

[17] It should be recognized that "knowledge" in this connection has a meaning that is somewhat different from that embodied in the operators k_i and K_i; that is why we have been using quotation marks when talking about "knowing" an operator or a partition. That an individual i knows an event E or a formula f can be embodied in formal statements ($\omega \in K_i E$ or $k_i f \in \omega$) that are well defined within the formal systems we have constructed, and whose truth or provability can be explicitly discussed in the context of these formal systems. This is not the case for "knowing" an operator or a partition. For discussion of a related point, see Section 10a.

We conclude this section by recalling the end of Section 3, where we said that "the description of the ω's involves no 'real' knowledge; it is only a kind of code book or dictionary." We have now explicitly constructed this "dictionary;" the construction is canonical (given the set N of individuals and the alphabet $\mathfrak{X}$); no self-reference is involved; and all individuals "know" the description of the model, including the partitions $\mathscr{I}_i$.

8. The relationship between the syntax and semantics of knowledge

In the foregoing we developed two related but distinct approaches to interactive epistemology. The states-of-the-world construction presented in Sections 1 and 2 embodies the *semantic* approach. The approach of Section 4, involving a formal "language," is *syntactic*. In our development, the syntactic approach is the more fundamental: In Section 6, we define semantics (states ω and knowledge functions κ_i) in terms of syntax (the formal language).

There are some obvious correspondences between the two approaches. Formulas (syntactic) correspond to events (semantic); disjunction, conjunction, and negation correspond to union, intersection, and complementation; the knowledge operators k_i to the knowledge operators K_i. By "correspond", we mean "have similar substantive content". Thus "i knows that it will snow or rain tomorrow" is expressed by $k_i(x \vee y)$ in the syntactic formalism and by $K_i(E_x \cup E_y)$ in the semantic formalism, where x and E_x stand for "it will snow tomorrow", and y and E_y for "it will rain tomorrow". Formally, $k_i(x \vee y)$ and $K_i(E_x \cup E_y)$ are distinct objects; the first is a string of letters, connectives, operators and parentheses, the second a set of states. But substantively, they express the same thing.

The purpose of this section is to formalize the correspondence between syntactic and semantic entities, to embody it in precise theorems. We will see that the correspondence is not complete; whereas every formula f corresponds to a unique event E_f, it is not the case that every event corresponds to some formula. Roughly, the reason is that formulas are finite concatenations, whereas events can also express infinite concatenations (e.g., infinite unions). But subject to this caveat (i.e., if one considers events of the form E_f only), we will see that there is a perfect "isomorphism" between the syntactic and semantic approaches.

Unless otherwise stated, the population N and the alphabet $\mathfrak{X}$ – and so also the syntax $\mathfrak{S} = \mathfrak{S}(N, \mathfrak{X})$ – will be taken as fixed, as in Section 4.

In the sequel, we use the following standard notations (for arbitrary formulas f and g):

$$f \wedge g := \neg(\neg f \vee \neg g), \tag{8.01}$$

$$f \Leftrightarrow g := (f \Rightarrow g) \wedge (g \Rightarrow f). \tag{8.02}$$

The following syntactic formulation of "positive introspection" will be needed in the sequel:

Proposition 8.1. *Any formula of the form $k_i f \Leftrightarrow k_i k_i f$ is a tautology.*

Proof: Analogous to the proof of 1.8.

We proceed now to a precise formulation of the "isomorphism" between the syntactic and semantic approaches. If f is a formula, define

$$E_f := \{\omega \in \Omega : f \in \omega\}. \tag{8.2}$$

In words, E_f is the set of all states of the world containing the formula f, i.e., the set of all states at which f is true; in other words, it is the event that f obtains.

The "isomorphism" is embodied in the following three propositions (f and g denote arbitrary formulas, i an arbitrary individual):

Proposition 8.3.

$$\sim E_f = E_{\neg f}; \tag{8.31}$$

$$E_f \cup E_g = E_{f \vee g}; \tag{8.32}$$

$$E_f \cap E_g = E_{f \wedge g}. \tag{8.33}$$

Proposition 8.4. $K_i E_f = E_{k_i f}$.

Proposition 8.5. $E_f \subset E_g$ *if and only if* $f \Rightarrow g$ *is a tautology.*

Though 8.4 is analogous to the components of 8.3, we have stated it separately because its proof lies considerably deeper. From 8.4 we will deduce that for *any* event E (not necessarily of the form E_f),

$$\omega \in K_i E \text{ iff } E \supset \bigcap_{k_i f \in \omega} E_f. \tag{8.51}$$

In words, 8.51 says that $K_i E$ is the event that E follows logically[18] from the formulas f that i knows. For a discussion of this result, see Section 10a.

From 8.5 it follows[19] that

$$E_f = E_g \text{ if and only if } f \Leftrightarrow g \text{ is a tautology.} \tag{8.52}$$

In words, 8.52 says that logically equivalent formulas correspond to the same event[20].

The proofs require some preliminary work. Call a formula f *elementary* if it contains no k_i. A *theorem of the propositional calculus* is defined as an elementary formula that yields the truth value T for each assignment of truth values to the letters appearing in it (using the usual rules for calculating truth values).

---

[18] $E \supset F$ means that E is a *logical* consequence of F. Compare Proposition 8.5 and Footnote 15.

[19] A formal derivation is given below.

[20] The term "isomorphism" usually denotes a one-one correspondence that preserves the relevant operations. In our case, the correspondence between formulas and events is not one-one, even when one considers events of the form E_f only. But from 8.52, it follows that there is a one-one correspondence between events of the form E_f and *equivalence classes* of formulas (under logical equivalence); and by 8.3 and 8.4, this correspondence indeed preserves the relevant operations.

Lemma 8.61. *Let e be a theorem of the propositional calculus, possibly in a syntax different from the given one*[21]*. Let f be a formula (in the given syntax) that is obtained by substitution*[22] *from e. Then f is a tautology, and so is in each state* ω.

Proof: Let $\mathfrak{S}$ be the given syntax, $\mathfrak{S}'$ a syntax containing e. Since e is a theorem of the propositional calculus, it can be derived[23] by repeatedly applying *modus ponens* to formulas of the form 4.4. (recall that the rule of *modus ponens* allows one to derive h from g and $g \Rightarrow h$). The formal expression of modus ponens is 4.2, so e is in the closure of the list of formulas of the form 4.4· in $\mathfrak{S}'$. Now f is obtained from e by substituting appropriate formulas in $\mathfrak{S}$ for the letters in e. If we make these substitutions already in the original formulas of the form 4.4· in $\mathfrak{S}'$ from which e was derived, then we obtain formulas of the form 4.4· in $\mathfrak{S}$. If we then follow exactly the same derivation step by step, we obtain f. Thus f is in the closure of the list of formulas of the form 4.4· in $\mathfrak{S}$. A fortiori it is in the strong closure of the list of formulas of the form 4.4· and 4.5·, i.e., the set of tautologies. ■

Remark 8.62: If f and $f \Rightarrow g$ are both in a state ω, then $g \in \omega$.

Proof: Follows from 4.2, since each state is by definition closed.

The following proposition is important in its own right, and also as a lemma for 8.3.

Proposition 8.63. *For all states* ω *and formulas* f *and* g,

$$\neg f \in \omega \text{ if and only if } f \notin \omega, \tag{8.631}$$

$$f \vee g \in \omega \text{ if and only if } (f \in \omega \text{ or } g \in \omega), \tag{8.632}$$

$$f \wedge g \in \omega \text{ if and only if } (f \in \omega \text{ and } g \in \omega), \tag{8.633}$$

$$f \Rightarrow g \in \omega \text{ if and only if } (f \in \omega \text{ implies } g \in \omega), \text{ and} \tag{8.634}$$

$$f \Leftrightarrow g \in \omega \text{ if and only if } (f \in \omega \text{ if and only if } g \in \omega). \tag{8.635}$$

Proof: 8.631 restates the coherence (6.1) and completeness (6.2) of ω. Next, by 8.61, $f \Rightarrow (f \vee g)$ and $g \Rightarrow (f \vee g)$ are in each state, so "if" in 8.632 follows

[21] I.e., with a different alphabet. The reason for this caveat is that we may need more letters than are available in the given alphabet for which to substitute different formulas, though each of these formulas is in the given syntax. For example, let there be two individuals 1, 2, and let the given alphabet have only one letter x. Then 8.61 implies that $k_1x \wedge k_2x \Rightarrow k_1x$ is a tautology, but this would not be so if the e of the lemma had to be in the given syntax.
[22] Of course, if the substitutes (the formulas being substituted for the letters in e) are elementary, then f itself is a theorem of the propositional calculus. Our lemma applies to the more general case in which the substitutes may contain knowledge operators k_i.
[23] See, e.g., Hilbert and Ackermann (1928), p. 22ff. They use the rule of substitution as well as modus ponens; our approach obviates the need for the rule of substitution, by using axiom schemas (families of axioms obtained by substituting arbitrary formulas for f, g, and h) rather than axioms.

from 8.62. To prove "only if," suppose that on the contrary, $f \vee g \in \omega$ and $f \notin \omega$ and $g \notin \omega$. Since ω is complete, it follows that $\neg f \in \omega$ and $\neg g \in \omega$. Now by 8.61, $(\neg f \Rightarrow (\neg g \Rightarrow \neg (f \vee g))) \in \omega$. So $\neg f \in \omega$ and 8.62 yield $(\neg g \Rightarrow \neg (f \vee g)) \in \omega$. So $\neg g \in \omega$ and 8.62 yield $\neg (f \vee g) \in \omega$. So the coherence of ω yields $(f \vee g) \notin \omega$, contrary to what we assumed. This proves 8.632, and the other assertions follow from it, from 8.631, and from the definitions of $\wedge, \Rightarrow$, and $\Leftrightarrow$. ■

Proof of Proposition 8.3: Let $\omega \in \Omega$. By the meaning of the symbol $\sim$ (complementation), 8.2, 8.631, and again 8.2, we have

$$\omega \in \sim E_f \text{ iff } \omega \notin E_f \text{ iff } f \notin \omega \text{ iff } \neg f \in \omega \text{ iff } \omega \in E_{\neg f};$$

this proves 8.31. The proofs of 8.32 and 8.33 are similar; instead of 8.631, we use 8.632 and 8.633 respectively. ■

Proposition 8.4 requires more preparation. Denote the list of all tautologies by $\mathfrak{T}$. If $\mathfrak{L}$ is any list, denote by $\mathfrak{L}^*$ the smallest closed list that includes[24] $\mathfrak{L} \cup \mathfrak{T}$.

Lemma 8.64. *If $f \in \mathfrak{L}$ and $g \in \mathfrak{L}$, then $f \wedge g \in \mathfrak{L}^*$.*

Proof: $\mathfrak{L}^*$ contains $f \Rightarrow (g \Rightarrow (f \wedge g))$, which is a tautology by 8.61, and by hypothesis it contains f. So since it is closed (4.2), it contains $(g \Rightarrow (f \wedge g))$. But it also contains g. So since it is closed, it contains $f \wedge g$. ■

Corollary 8.641. *If f and g are tautologies, so is $f \wedge g$.*

Proof: Since $\mathfrak{T}$ is by definition strongly closed, it is closed. So $\mathfrak{T}^* = \mathfrak{T}$. The corollary now follows by setting $\mathfrak{L} := \mathfrak{T}$ in 8.64. ■

Lemma 8.65. *$\mathfrak{L}^*$ consists of all consequences of finite conjunctions of formulas in $\mathfrak{L}$.*

Proof: By 8.64, all finite conjunctions of formulas in $\mathfrak{L}$ are in $\mathfrak{L}^*$. Also all tautologies are in $\mathfrak{L}^*$, so since $\mathfrak{L}^*$ is closed, all consequences of finite conjunctions of formulas in $\mathfrak{L}$ are in $\mathfrak{L}^*$.

It remains to show that the set of all such consequences is closed. So suppose that f is a finite conjunction of formulas in $\mathfrak{L}$, that g is a consequence of f, and that $g \Rightarrow h$ is a tautology; it is sufficient to show that h is also a consequence of f. But g being a consequence of f means that $f \Rightarrow g$ is a tautology. So by 8.641, $(f \Rightarrow g) \wedge (g \Rightarrow h)$ is a tautology. By 8.61, $((f \Rightarrow g) \wedge (g \Rightarrow h)) \Rightarrow (f \Rightarrow h)$ is a tautology. So since $\mathfrak{T}$ is closed, $f \Rightarrow h$ is a tautology. So h is a consequence of f. ■

Call $\mathfrak{L}$ *consistent* if for all formulas f,

$$f \in \mathfrak{L}^* \quad \text{implies} \quad \neg f \notin \mathfrak{L}^*. \tag{8.66}$$

[24] The intersection of all closed lists including $\mathfrak{L} \cup \mathfrak{T}$; it is itself closed, and is included in all closed lists that include $\mathfrak{L} \cup \mathfrak{T}$.

Lemma 8.67. *If* $\mathfrak{L}_1 \subset \mathfrak{L}_2 \subset \cdots$ *is a nested sequence (possibly finite) of consistent lists, then its union* $\mathfrak{L}_\infty := \bigcup_m \mathfrak{L}_m$ *is consistent.*

Proof: By 8.65, $\mathfrak{L}^*_\infty = \bigcup_m \mathfrak{L}^*_m$. So if 8.66 is false, then there are m, n, and f with $f \in \mathfrak{L}^*_m$ and $\neg f \in \mathfrak{L}^*_n$. W.l.o.g. $n \leq m$, so $\mathfrak{L}_n \subset \mathfrak{L}_m$, so $\mathfrak{L}^*_n \subset \mathfrak{L}^*_m$, so $f \in \mathfrak{L}^*_m$ and $\neg f \in \mathfrak{L}^*_m$, contrary to the hypothesis that $\mathfrak{L}_m$ is consistent. ■

Lemma 8.68. *Let* $\mathfrak{L}$ *be a consistent list,* f *a formula with* $\neg f \notin \mathfrak{L}^*$. *Then* $\mathfrak{L} \cup \{f\}$ *is consistent.*

Proof: If not, then by 8.65, there are formulas $g_1, g_2, \ldots, g_m$ and $h_1, h_2, \ldots, h_n$ in $\mathfrak{L}$, and a formula e, such that $(g \wedge f) \Rightarrow e$ and $(h \wedge f) \Rightarrow \neg e$ are tautologies, where $g := g_1 \wedge g_2 \wedge \cdots \wedge g_m$ and $h := h_1 \wedge h_2 \wedge \cdots \wedge h_n$. Hence by 8.641, $((g \wedge f) \Rightarrow e) \wedge ((h \wedge f) \Rightarrow \neg e)$ is a tautology. By 8.61, $(((g \wedge f) \Rightarrow e) \wedge ((h \wedge f) \Rightarrow \neg e)) \Rightarrow ((g \wedge h) \Rightarrow \neg f))$ is a tautology. Since the set of tautologies is closed, we deduce that $(g \wedge h) \Rightarrow \neg f$ is a tautology. Now $g \wedge h$ is a finite conjunction of formulas in $\mathfrak{L}$, so we deduce that $\neg f \in \mathfrak{L}^*$, contrary to our hypothesis. ■

Lemma 8.69. *Every state* ω *is consistent.*

Proof: By definition, ω is closed and includes $\mathfrak{T}$, so $\omega = \omega^*$. But then the consistency of ω follows from its coherence. ■

Proposition 8.7. *A list* $\mathfrak{L}$ *is consistent if and only if there is a state* ω *with* $\mathfrak{L} \subset \omega$.

Proof: If: Follows from 8.69.

Only if: Suppose that $\mathfrak{L}$ is consistent. We must find a state that includes $\mathfrak{L}$. Let $(f_1, f_2, \ldots)$ be an enumeration of all formulas. We will define a nested sequence $\mathfrak{L}_1 \subset \mathfrak{L}_2 \subset \cdots$ (possibly finite) such that $\mathfrak{L}^*_\infty$ is a state, where $\mathfrak{L}_\infty := \bigcup_m \mathfrak{L}_m$. Set $\mathfrak{L}_1 := \mathfrak{L}$. Suppose $\mathfrak{L}_1, \mathfrak{L}_2, \ldots, \mathfrak{L}_m$ have been defined. If for each formula f, either $f \in \mathfrak{L}^*_m$ or $\neg f \in \mathfrak{L}^*_m$, then the sequence ends with $\mathfrak{L}_m$. Otherwise, let f^m be the first f_n such that $f \notin \mathfrak{L}^*_m$ and $\neg f \notin \mathfrak{L}^*_m$, and define $\mathfrak{L}_{m+1} := \mathfrak{L}_m \cup \{f^m\}$. An induction using 8.68 then shows that all the $\mathfrak{L}_m$ are consistent. So by 8.67, their union $\mathfrak{L}_\infty$ is consistent. Now $\mathfrak{L}^*_\infty$ contains all the $\mathfrak{L}^*_m$, so by construction, it contains, for each formula f_n, either f_n or $\neg f_n$; i.e., it is complete. Moreover, since $\mathfrak{L}_\infty$ is consistent, $\mathfrak{L}^*_\infty$ cannot contain both. Thus $\mathfrak{L}^*_\infty$ is complete and coherent. Moreover it is closed and contains all tautologies, so it is a state; denoting it ω completes the proof. ■

Lemma 8.71. *Let* e *be a theorem of the propositional calculus, possibly in a syntax different from the given one*[25]. *Let* f *be a formula (in the given syntax) that is obtained by substitution from* e. *Then* $k_i f$ *is a tautology for all* i.

Proof: By 8.61, f is a tautology; since the set $\mathfrak{T}$ of tautologies is by definition strongly closed, 4.3 yields $k_i f \in \mathfrak{T}$. ■

Lemma 8.72. $k_i(f \wedge g) \Leftrightarrow (k_i f) \wedge (k_i g)$ *is a tautology.*

---

[25] See Footnote 21.

Proof: By 8.71, $k_i((f \wedge g) \Rightarrow f)$ is a tautology, and so a member of each ω. So by 4.52 and 4.2, $k_i(f \wedge g) \Rightarrow k_i f$ is a tautology. Similarly, $k_i(f \wedge g) \Rightarrow k_i g$ is a tautology, and it follows that

$$k_i(f \wedge g) \Rightarrow (k_i f) \wedge (k_i g) \text{ is a tautology.} \tag{8.721}$$

For the other direction, note first that by 8.61, $f \Rightarrow (g \Rightarrow (f \wedge g))$ is a tautology. So by 8.71, $k_i(f \Rightarrow (g \Rightarrow (f \wedge g)))$ is a tautology. So by 4.52, $(k_i f) \Rightarrow k_i(g \Rightarrow (f \wedge g))$ is a tautology. Again by 4.52, $(k_i(g \Rightarrow (f \wedge g))) \Rightarrow ((k_i g) \Rightarrow k_i(f \wedge g))$. Combining the last two observations, we deduce that $(k_i f) \Rightarrow ((k_i g) \Rightarrow k_i(f \wedge g))$ is a tautology, and from this it follows that $((k_i f) \wedge (k_i g)) \Rightarrow k_i(f \wedge g)$ is a tautology. The proof may be completed by applying 8.641 to this and to 8.721. ■

Lemma 8.73. *Let ω and ω' be states, and let $\kappa_i(\omega) \subset \omega'$. Then $\kappa_i(\omega') = \kappa_i(\omega)$.*

Proof: Any formula in either $\kappa_i(\omega)$ or $\kappa_i(\omega')$ has the form $k_i f$. Suppose, therefore, that $k_i f \in \kappa_i(\omega)$. Then since $\kappa_i(\omega) \subset \omega'$, it follows that $k_i f \in \omega'$, so by the definition of $\kappa_i(\omega')$, we get $k_i f \in \kappa_i(\omega')$. Conversely, suppose that $k_i f \in \kappa_i(\omega')$. If $k_i f \in \omega$, then also $k_i f \in \kappa_i(\omega)$, and we are done. Otherwise, $\neg k_i f \in \omega$, by 6.2. So by 4.53 and 4.2, $k_i \neg k_i f \in \omega$. So $k_i \neg k_i f \in k_i(\omega)$. Since $\kappa_i(\omega) \subset \omega'$, we deduce $k_i \neg k_i f \in \omega'$. So by 4.51 and 4.2, $\neg k_i f \in \omega'$. Since ω is coherent, it follows that $k_i f \notin \omega'$, contradicting $k_i f \in \kappa_i(\omega') \subset \omega'$. ■

Proof of Proposition 8.4: We start with the inclusion $\supset$. Let $\omega \in E_{k_i f}$, which means that $k_i f \in \omega$ (by 8.2). We must show that $\omega \in K_i E_f$, i.e. (by 1.21) that $\mathbf{I}_i(\omega) \subset E_f$. So let ω' be a state with $\kappa_i(\omega') = \kappa_i(\omega)$. Since $k_i f \in \omega$, it follows from 6.3 that $k_i f \in \omega'$. So $f \in \omega'$ by 4.51 and 4.2; so again by 8.2, $\omega' \in E_f$. This holds for all ω' with $\kappa_i(\omega') = \kappa_i(\omega)$, i.e., for all ω' in $\mathbf{I}_i(\omega)$. Thus $\mathbf{I}_i(\omega) \subset E_f$, as asserted. This demonstrates $\supset$.

To demonstrate $\subset$, let $\omega \in K_i E_f$. Then by 1.21,

$$\mathbf{I}_i(\omega) \subset E_f. \tag{8.74}$$

We wish to show that $\omega \in E_{k_i f}$, which by 8.2 is the same as $k_i f \in \omega$. Setting $\mathfrak{L} := \kappa_i(\omega)$, we obtain from 8.74, 1.1, 8.2, and 8.73 that for all states ω',

$$\mathfrak{L} \subset \omega' \text{ implies } f \in \omega'. \tag{8.75}$$

From this it follows that $f \in \mathfrak{L}^*$. For if not, then by 8.68, $\mathfrak{L} \cup \{\neg f\}$ is consistent. So by 8.7, there is a state ω' including $\mathfrak{L} \cup \{\neg f\}$, which contradicts 8.75 (because of the coherence of ω').

Since $f \in \mathfrak{L}^*$, it follows from 8.65 that f is a consequence of a finite conjunction of formulas in $\mathfrak{L}$. Since $\mathfrak{L}$ is defined as $\kappa_i(\omega)$, all formulas in $\mathfrak{L}$ start with k_i. Thus there are formulas $k_i f_1, k_i f_2, \ldots, k_i f_m$ in $\mathfrak{L}$ such that $(k_i f_1 \wedge k_i f_2 \wedge \cdots \wedge k_i f_m) \Rightarrow f$ is a tautology. So by 4.3 and $\mathfrak{T}$ being strongly closed, $k_i((k_i f_1 \wedge k_i f_2 \wedge \cdots \wedge k_i f_m) \Rightarrow f)$ is a tautology. So by 4.52 and 4.2, $k_i(k_i f_1 \wedge k_i f_2 \wedge \cdots \wedge k_i f_m) \Rightarrow k_i f$ is a tautology. Using 8.72 and that $k_i k_i g \Leftrightarrow k_i g$ is a tautology (8.1), we deduce that $(k_i f_1 \wedge k_i f_2 \wedge \cdots \wedge k_i f_m) \Rightarrow k_i f$ is a tautology. So $k_i f \in \mathfrak{L}^*$, by 8.65. Now $\mathfrak{L} = \kappa_i(\omega) \subset \omega$, and ω is closed, so ω includes the smallest closed list $\mathfrak{L}^*$ that includes $\mathfrak{L}$. So $k_i f \in \omega$, as was to be proved. ■

We proceed next to prove 8.5. First, an auxiliary proposition that is of interest on its own.

Proposition 8.8. *A formula is a tautology if and only if it is in every state.*

Proof: "Only if" is part of the definition of state. To prove "if", suppose that $g \in \omega$ for all states ω. Then $\neg g \notin \omega$ for all ω, so by 8.7, the list $\{\neg g\}$ is not consistent. So by 8.66, there is an f in $\{\neg g\}^*$ with $\neg f \in \{\neg g\}^*$. By 8.65, it follows that both f and $\neg f$ are consequences of $\neg g$. Thus both $\neg g \Rightarrow f$ and $\neg g \Rightarrow \neg f$ are tautologies, so $(\neg g \Rightarrow f) \wedge (\neg g \Rightarrow \neg f)$ is a tautology. Now $(\neg y \Rightarrow x) \wedge (\neg y \Rightarrow \neg x) \Rightarrow y$ is a theorem of the propositional calculus, so by 8.61, $(\neg g \Rightarrow f) \wedge (\neg g \Rightarrow \neg f) \Rightarrow g$ is a tautology. But the list of all tautologies is closed, since it is strongly closed; so g is a tautology. ■

Proof of Proposition 8.5: Suppose first that $f \Rightarrow g$ is a tautology. Then it is in every state ω. Hence if $f \in \omega$, then since ω is closed, it follows that $g \in \omega$. Hence $E_f = \{\omega \in \Omega : f \in \omega\} \subset \{\omega \in \Omega : g \in \omega\} = E_g$. This proves "if".

Conversely, suppose that $f \Rightarrow g$ is not a tautology. Then by 8.8, there is a state ω that does not contain $f \Rightarrow g$, so contains $\neg (f \Rightarrow g)$. Now $\neg (f \Rightarrow g) \Rightarrow (f \wedge \neg g)$ is a tautology by 8.61, so it is in ω. Since ω is closed, it follows that $f \wedge \neg g \in \omega$; so by 8.633, $f \in \omega$ and $\neg g \in \omega$, so $g \notin \omega$. So E_f is not included in E_g. This proves "only if". ■

We come now to the proofs of 8.51 and 8.52; it is convenient to start with 8.52.

Proof of 8.52: $f \Leftrightarrow g$ is a tautology iff $f \Rightarrow g$ and $g \Rightarrow f$ are tautologies, which is iff $E_f \subset E_g$ and $E_g \subset E_f$, i.e., $E_f = E_g$. ■

For the proof of 8.51, two more lemmas are needed (again of interest in their own right).

Lemma 8.81. $\mathbf{I}_i(\omega) = \bigcap_{g \in \kappa_i(\omega)} E_g$.

Proof: To show $\subset$, suppose $\omega' \in \mathbf{I}_i(\omega)$ and $g \in \kappa_i(\omega)$; we must prove $\omega' \in E_g$. By 1.1, $\kappa_i(\omega') = \kappa_i(\omega)$, so $g \in \kappa_i(\omega') \subset \omega'$, so $\omega' \in E_g$ by 8.2.

To show $\supset$, suppose $\omega' \in E_g$ for all $g \in \kappa_i(\omega)$. Thus by 8.2, $g \in \omega'$ for all $g \in \kappa_i(\omega)$; i.e., $\kappa_i(\omega) \subset \omega'$. Hence by 8.73, $\kappa_i(\omega') = \kappa_i(\omega)$, so by 1.1, $\omega' \in \mathbf{I}_i(\omega)$. ■

Lemma 8.82. $\bigcap_{g \in \kappa_i(\omega)} E_g = \bigcap_{k_i f \in \omega} E_f$.

Proof: To show $\subset$, note that by 1.4 and 8.4, $K_i \bigcap_{k_i f \in \omega} E_f = \bigcap_{k_i f \in \omega} K_i E_f = \bigcap_{k_i f \in \omega} E_{k_i f} = \bigcap_{g \in \kappa_i(\omega)} E_g$, and then use 1.31. To show $\supset$, note that $\bigcap_{k_i f \in \omega} E_f \subset \bigcap_{k_i k_i h \in \omega} E_{k_i h} = \bigcap_{h : k_i h \in \omega} E_{k_i h} = \bigcap_{g \in \kappa_i(\omega)} E_g$, where $\subset$ follows from the intersection of a smaller family being larger, and the first $=$ sign from 8.1 and 8.52 (which has already been proved). ■

Proof of 8.51: If: Assume $E \supset \bigcap_{f:k_i f \in \omega} E_f$; from 1.32, 1.4, 8.4, and 8.2 we get $K_i E \supset K_i \bigcap_{f:k_i f \in \omega} E_f = \bigcap_{f:k_i f \in \omega} K_i E_f = \bigcap_{f:k_i f \in \omega} E_{k_i f} \supset \bigcap_{f:k_i f \in \omega} \{\omega\} = \{\omega\}$. Only if: Assume $\omega \in K_i E$. Then by 1.21, $\mathbf{I}_i(\omega) \subset E$. So by 8.82 and 8.81, $\bigcap_{f:k_i f \in \omega} E_f = \bigcap_{g \in \kappa_i(\omega)} E_g \subset E$. ■

We conclude this section with

Proposition 8.9. *A list is a state if and only if it is complete and consistent.*

Proof: By 8.69, every state is complete and consistent. Conversely, let $\mathfrak{L}$ be a complete consistent list. By 8.7, there is a state ω with $\mathfrak{L} \subset \omega$. If the inclusion is strict, there is a formula f with $f \in \omega$ and $f \notin \mathfrak{L}$. Since $\mathfrak{L}$ is complete, $\neg f \in \mathfrak{L}$, so $\neg f \in \omega$, so $f \notin \omega$ by the coherence of ω, contradicting $f \in \omega$. ■

9. Representations and models

This section treats the notion of "model." The idea is to "interpret" syntactic formulas as events in a semantic knowledge system $\mathscr{S}$ (not necessarily the canonical system). The letters $x, y, z, \ldots$ of the alphabet are interpreted as certain distinguished events $X, Y, Z, \ldots$ in $\mathscr{S}$; the syntactic knowledge operators k_i as semantic knowledge operators K_i; and disjunctions, conjunctions, and negations as unions, intersections, and complements respectively. Thus each formula f corresponds to a unique event F in $\mathscr{S}$, constructed from the distinguished events $X, Y, Z, \ldots$ like f is constructed from the alphabet letters $x, y, z, \ldots$.

Substantively, f and F have similar content; to say them in English, one uses the same words. There is, however, a subtle difference between them. A syntactic formula f is, so to speak, "context-free;" the semantic system $\mathscr{S}$ provides the context, the environment. Thus, f entails *nothing* except the formula itself and its logical consequences – strictly in the sense of Section 4. The corresponding event F may entail much more, depending on the system $\mathscr{S}$ in which F is imbedded. Specifically, F entails all those events G that include F, although G may correspond to a formula g that does not follow from f "logically" – in the sense of Section 4 – but only, so to speak, in the "context" of $\mathscr{S}$.

A formula f is said to *hold at* a state ω in $\mathscr{S}$ if ω is a member of the event F corresponding to f. A *model* of a list $\mathfrak{L}$ of formulas is a state ω in a semantic knowledge system $\mathscr{S}$ at which each formula in $\mathfrak{L}$ holds. The main result of this section, Proposition 9.2, is that a list is consistent if and only if it has a model. In practice, pointing to a model is the major practical tool for proving consistency of lists – or indeed, proving that an individual formula is not tautologically false; proving consistency directly, by syntactic methods only, is notoriously cumbersome. The other side of the same coin is Corollary 9.3, which says that a formula g "follows" from a list $\mathfrak{L}$ if and only if g holds in *every* model of $\mathfrak{L}$. This, too, is a major tool in practical applications. As promised in Section 5, we also show that the tautologies are precisely those formulas that hold at each state of each semantic knowledge system (Corollary 9.4).

One simple application of these ideas is Corollary 9.5 – that the canonical knowledge system Ω defined in Section 6 is non-empty, i.e., that there exists at least one complete coherent list for every population and alphabet. Obvious as this may sound, it is not easily established by purely syntactic methods. But using the idea of model, it is immediate. One need only point to a state (*any* state) in a semantic knowledge system (*any* such system – for example, the system with one state only). The set of all formulas holding at that state is closed, complete and coherent, and contains all tautologies (Lemma 9.21).

A more substantial application is to issues of cardinality. Using 9.2, one may show that when there are at least two players, the canonical system has a continuum of states, even when the alphabet has just one letter. Two proofs are known. One is by Hart, Heifetz, and Samet (1996) – henceforth HHS; the other is in the appendix to this paper. Both are based on Proposition 9.2. HHS prove that there is a continuum of states by exhibiting a continuum of lists, each of which is consistent in itself, but any two of which are mutually inconsistent; by 8.7, each of these lists can be expanded to a state. HHS's proof is shorter and more elegant than ours; we adduce the latter here "for the record," and because its method, though clumsier than HHS's, may sometimes be applicable where theirs is not (needless to say, the opposite is also true).

To complete the picture, we will show in the appendix that when there is only one individual (and the alphabet is finite), then Ω is finite. This underscores the relative complexity of many-person epistemology when compared with one-person epistemology.

We proceed now to the formal treatment. Fix a population N and an alphabet $\mathfrak{X} := \{x, y, \ldots\}$; let $\mathfrak{S} := \mathfrak{S}(N, X)$ be the corresponding syntax. Recall (from Section 2) that a *semantic knowledge system* (or simply *knowledge system* for short) consists of a population, a universe $\hat{\Omega}$, and knowledge functions[26] $\hat{\kappa}_i$ for each individual i; denote the knowledge operators in such a system by $\hat{K}_i$, and the ufield of events by $\hat{\mathscr{E}}$. For brevity, we will sometimes denote the whole knowledge system simply by $\hat{\Omega}$, rather than $(\hat{\Omega}, N, (\hat{\kappa}_i)_i)$. This is standard mathematical usage: A group is defined as a pair consisting of a set G and a multiplication, but the group (i.e., the pair) is itself often denoted G; a metric space is a set together with a metric, but the space and the set are often denoted by the same letter; and so on.

Define a *representation* of $\mathfrak{S}$ as a knowledge system $(\hat{\Omega}, N, (\hat{\kappa}_i)_i)$, together with a function $\varphi : \mathfrak{S} \to \hat{\mathscr{E}}$ satisfying

$$\begin{aligned} &\varphi(\neg f) = \sim\varphi(f), \quad \varphi(f \vee g) = \varphi(f) \cup \varphi(g), \quad \text{and} \\ &\varphi(k_i f) = \hat{K}_i \varphi(f). \end{aligned} \tag{9.1}$$

We will slightly abuse our terminology by using "representation" to refer also to just $\hat{\Omega}$ itself, or just φ itself (not just to the pair consisting of the knowledge system and φ). Given φ and a particular formula f, we will also call $\varphi(f)$ the "representation of f."

Remark 9.11: For each knowledge system $\hat{\Omega}$, each function from the alphabet $\mathfrak{X}$ into $\hat{\mathscr{E}}$ has a unique extension to a representation of $\mathfrak{S}$.

---

[26] The carets (hats) are to distinguish these objects from the specific universe Ω and knowledge functions κ_i defined in Section 6.

Proof: Follows from 9.1, as each formula f results from a *unique* sequence of operations $\neg$, $\vee$, and k_i. ■

Remark 9.11 implies that one can think of a representation of $\mathfrak{S}$ simply as an (arbitrary) knowledge system $\hat{\Omega}$, together with an (arbitrary) specification of events $\varphi(x), \varphi(y), \ldots$ corresponding to the letters of the alphabet. Intuitively, the events $\varphi(x), \varphi(y), \ldots$ may be thought of as the "natural" events.

Given a representation φ of $\mathfrak{S}$ and a state $\hat{\omega}$ in $\hat{\Omega}$, define

$$\phi(\hat{\omega}) := \{f \in \mathfrak{S} : \hat{\omega} \in \varphi(f)\}. \tag{9.12}$$

In words, $\phi(\hat{\omega})$ is the list of all formulas that *hold* at $\hat{\omega}$, i.e., whose representation $\varphi(f)$ obtains at $\hat{\omega}$.

Let $\mathfrak{L} \subset \mathfrak{S}$ be a list. Define a *model* for $\mathfrak{L}$ as a representation $(\hat{\Omega}, \varphi)$ of $\mathfrak{S}$ together with a state $\hat{\omega}$ in $\hat{\Omega}$ at which all formulas in $\mathfrak{L}$ hold (i.e., $\hat{\omega} \in \varphi(f)$ for all f in $\mathfrak{L}$).

Proposition 9.2. *A list $\mathfrak{L}$ of formulas is consistent if and only if it has a model.*

We first prove

Lemma 9.21. *Let $(\hat{\Omega}, \varphi)$ be a representation of the syntax $\mathfrak{S}$, and let $\hat{\omega} \in \hat{\Omega}$. Then $\phi(\hat{\omega}) \in \Omega$.*

In words: For a fixed representation, the list of all formulas that hold at a given state $\hat{\omega}$ is closed, coherent, complete, and contains all tautologies; i.e., it is a state in the canonical knowledge system.

Proof: Set $\omega := \phi(\hat{\omega})$. We must prove that the list ω is closed, coherent, complete, and contains all tautologies.

To say that ω is closed means that it satisfies 4.2; i.e., that ($f \in \omega$ and $f \Rightarrow g \in \omega$) implies $g \in \omega$; i.e., that

$$(\hat{\omega} \in \varphi(f) \text{ and } \hat{\omega} \in \varphi((\neg f) \vee g))) \quad \text{implies} \quad \hat{\omega} \in \varphi(g). \tag{9.22}$$

Now

$$\hat{\omega} \in \varphi(f) \text{ and } \hat{\omega} \in \varphi((\neg f) \vee g)) \text{ iff } \hat{\omega} \in \varphi(f) \cap \varphi((\neg f) \vee g)), \tag{9.23}$$

and since $\varphi(f) \cap \varphi(h) = \varphi(f \wedge h)$ by 9.1, we have

$$\varphi(f) \cap (\varphi(\neg f) \cup \varphi(g)) = \varphi(f) \cap (\sim\varphi(f) \cup \varphi(g)) \subset \varphi(g); \tag{9.24}$$

putting together 9.23 and 9.24 yields 9.22.

We prove next that ω contains all tautologies. First we show

$$\text{if } e \text{ is a tautology, then } \varphi(e) = \hat{\Omega}. \tag{9.25}$$

Indeed, let $\hat{\mathfrak{T}}$ be the list of all formulas e such that $\varphi(e) = \hat{\Omega}$. Using 9.1 and 1.31 through 1.33, we verify directly that every formula with one of the seven forms 4.41 through 4.53 is in $\hat{\mathfrak{T}}$. Moreover, if $f \in \hat{\mathfrak{T}}$ and $f \Rightarrow g \in \hat{\mathfrak{T}}$ then $\hat{\Omega} = \varphi(f)$ and $\hat{\Omega} = \varphi(f \Rightarrow g)$, so $\hat{\Omega} = \varphi((\neg f) \vee g) = (\sim\varphi(f)) \cup \varphi(g) = (\sim\hat{\Omega}) \cup \varphi(g) = \emptyset \cup \varphi(g) = \varphi(g)$, so $g \in \hat{\mathfrak{T}}$; thus $\hat{\mathfrak{T}}$ is closed. Finally, if $f \in \hat{\mathfrak{T}}$, then $\varphi(k_i f) = \hat{K}_i \varphi(f) = \hat{K}_i \hat{\Omega} = \hat{\Omega}$ (by 1.7), so $\hat{\mathfrak{T}}$ is strongly closed. Thus $\hat{\mathfrak{T}}$ is strongly closed and contains all formulas of the form 4.41 through 4.53, so it includes the strong closure of the list of all formulas of this form, i.e., the list of all tautologies; this proves 9.25. It follows that $\hat{\omega} \in \hat{\Omega} = \varphi(e)$ for all tautologies e, so ω contains all tautologies, as asserted.

Finally, to verify completeness (6.2) and coherence (6.1), note that

$$\neg f \in \omega \text{ iff } \hat{\omega} \in \varphi(\neg f) \text{ iff } \hat{\omega} \in \sim\varphi(f) \text{ iff } \hat{\omega} \notin \varphi(f) \text{ iff } f \notin \omega.$$

Thus ω is a closed, coherent, complete list that contains all tautologies; i.e., $\omega = \phi(\hat{\omega})$ is a state. ■

Proof of Proposition 9.2: Only if: Let the knowledge system consist of the state space Ω together with N and the knowledge functions κ_i constructed in Section 6. Define $\varphi(f) := \{\omega \in \Omega : f \in \omega\}$ $(= E_f$, by 8.2). By 8.3 and 8.4, φ satisfies 9.1. Since $\mathfrak{L}$ is consistent, there is a state ω that includes $\mathfrak{L}$ (8.7). So each formula f in $\mathfrak{L}$ is in ω, so ω is in all the corresponding $\varphi(f)$.

If: Let the model be $(\hat{\Omega}, \varphi, \hat{\omega})$. Then $\mathfrak{L} \subset \phi(\hat{\omega})$. By 9.21, $\phi(\hat{\omega})$ is a state, so by 8.7, $\mathfrak{L}$ is consistent.

Corollary 9.3. *Let $\mathfrak{L}$ be a list, g a formula. Then g is a consequence of a finite conjunction of formulas in $\mathfrak{L}$ if and only if every model for $\mathfrak{L}$ is also a model for $\{g\}$. (In words, a formula "follows" from a list if and only if it holds in* **every** *model of that list.)*

Proof: Recall that $\mathfrak{L}^*$ denotes the list of all consequences of finite conjunctions of formulas in a list $\mathfrak{L}$ (8.65). If g is *not* in $\mathfrak{L}^*$, then by 8.68, $\mathfrak{L} \cup \{\neg g\}$ is consistent, so by 9.2 has a model, which by 9.13 is not a model for $\{g\}$. In the opposite direction, if it is not the case that every model for $\mathfrak{L}$ is also a model for $\{g\}$, then $\mathfrak{L}$ has a model that is *not* a model for $\{g\}$, and so by 9.21 *is* a model for $\{\neg g\}$, hence also for $\mathfrak{L} \cup \{\neg g\}$. So by 9.2, $\mathfrak{L} \cup \{\neg g\}$ is consistent, so by 8.68, $g \notin \mathfrak{L}^*$. ■

Corollary 9.4. *A formula f is a tautology if and only if for any representation, $\varphi(f) = \hat{\Omega}$ (in words, if and only if f holds at each state in each representation).*

Proof: "Only if" follows from 9.21. To show "if," note first that by 8.3 and 8.4, the canonical semantic knowledge system Ω, together with the function φ on $\mathfrak{S}$ defined by $\varphi(f) := E_f$, constitute a representation of $\mathfrak{S}$. So by hypothesis, $E_f = \varphi(f) = \Omega$, which means that f is in every state of Ω. So by 8.8, f is a tautology. ■

Corollary 9.5. *The canonical knowledge system Ω is non-empty.*

Proof: By Lemma 9.21, $\phi(\hat{\omega})$ is always in Ω; so it is enough to show that there exists *some* semantic knowledge system, and that is obvious (for example, the one with one state only).

Note to Section 9

In this note we show that every representation of $\mathfrak{S}$ can be thought of as a common knowledge subuniverse of Ω. More precisely, for every representation there is a common knowledge subuniverse of Ω such that the given representation is isomorphic, in a natural sense, to this common knowledge subuniverse.

Given a representation $(\hat{\Omega}, \varphi)$ of the syntax $\mathfrak{S}$, denote $\hat{\mathscr{S}} := \varphi(\mathfrak{S}) \subset \hat{\mathscr{E}}$, and call the events in $\hat{\mathscr{S}}$ *syntactic*. In words, a syntactic event is one of the form $\varphi(f)$; one that corresponds, under the representation φ, to a formula in the syntax $\mathfrak{S}$.

Now let $(\hat{\Omega}_1, \varphi_1)$ and $(\hat{\Omega}_2, \varphi_2)$ be two representations of $\mathfrak{S}$; set $\hat{\mathscr{S}}_j := \varphi_j(\mathfrak{S})$, $j = 1, 2$. Call φ_1 and φ_2 *isomorphic* if there is a one-one mapping F (an *isomorphism*) from $\hat{\mathscr{S}}_1$ onto $\hat{\mathscr{S}}_2$ such that $\varphi_2 = F \circ \varphi_1$ (see the diagram at 9.6).

$$\begin{array}{ccc} & \mathfrak{S} & \\ {}^{\varphi_1}\swarrow & & \searrow^{\varphi_2} \\ \hat{\mathscr{S}}_1 & \xrightarrow[F]{} & \hat{\mathscr{S}}_2 \end{array} \tag{9.6}$$

Let $(\hat{\Omega}, \varphi)$ be a representation of $\mathfrak{S}$. By 9.21, for each state $\hat{\omega}$ in $\hat{\Omega}$, the list $\phi(\hat{\omega})$ is a state ω in Ω. The set of all ω obtained in this way is a subset of Ω, which we denote Ω^{φ}. Formally, $\Omega^{\varphi} := \phi(\hat{\Omega})$. In words, a state in Ω is in Ω^{φ} iff it corresponds to some $\hat{\omega}$ in $\hat{\Omega}$ under the representation φ; i.e., iff it consists of precisely those formulas f for which the representation $\varphi(f)$ holds at $\hat{\omega}$.

Denote by Γ^{φ} the smallest common knowledge subuniverse[27] of Ω that includes Ω^{φ}. By associating with Γ^{φ} the population N and the knowledge functions $\kappa_i | \Gamma^{\varphi}$, we obtain a knowledge system; denote the set of events in this system by $\mathscr{E}^{\varphi}$. Define $\varphi^* : \mathfrak{S} \to \mathscr{E}^{\varphi}$ by

$$\varphi^*(f) := E_f \cap \Gamma^{\varphi}. \tag{9.61}$$

Proposition 9.7. *$(\Gamma^{\varphi}, \varphi^*)$ is a representation of $\mathfrak{S}$, which is isomorphic to $(\hat{\Omega}, \varphi)$.*

For the proof of 9.7, we need several lemmas. In the statements of these lemmas, f, g, and h denote arbitrary formulas.

Lemma 9.71. $\phi(\varphi(f)) = E_f \cap \Omega^{\varphi}$.

Proof: We have $\phi(\varphi(f)) = \{\phi(\hat{\omega}) : \hat{\omega} \in \varphi(f)\}$, so $\omega \in \phi(\varphi(f))$ iff there is an $\hat{\omega} \in \varphi(f)$ with $\omega = \phi(\hat{\omega}) = \{g : \hat{\omega} \in \varphi(g)\}$, and this is so iff $f \in \omega \in \Omega^{\varphi}$; thus $\phi(\varphi(f)) = \{\omega \in \Omega^{\varphi} : \omega \in f\} = E_f \cap \Omega^{\varphi}$. ■

Lemma 9.72. *$\Omega^{\varphi} \subset E_f$ iff $\hat{\Omega} = \varphi(f)$.*

[27] See the end of Section 2.

Proof: $\Omega^{\varphi} \subset E_f$ iff $f \in \omega$ for all ω in Ω^{φ}; that is, iff for each $\hat{\omega}$ in $\hat{\Omega}$, we have $f \in \phi(\hat{\omega}) = \{g : \hat{\omega} \in \varphi(g)\}$; that is, iff $\hat{\omega} \in \varphi(f)$ for all $\hat{\omega}$ in $\hat{\Omega}$; i.e., $\hat{\Omega} = \varphi(f)$. ■

Corollary 9.73. *If $\Omega^{\varphi} \subset E_f$, then $\Omega^{\varphi} \subset K_i E_f$.*

Proof: By 9.72, $\Omega^{\varphi} \subset E_f$ yields $\hat{\Omega} = \varphi(f)$. Hence by 9.1, $\hat{\Omega} = \hat{K}_i \hat{\Omega} = \hat{K}_i \varphi(f) = \varphi(k_i f)$. So again by 9.72, $\Omega^{\varphi} \subset K_i E_f$. ■

Lemma 9.74. *If $\Omega^{\varphi} \subset E_f$, then $\Gamma^{\varphi} \subset E_f$.*

Proof: Set $F := \bigcap E_f$, where the intersection is over all f with $\Omega^{\varphi} \subset E_f$. From 1.4, 8.4, and 9.73, it follows that $K_i F = F$ for all i. Thus F is self-evident, and so a subuniverse of Ω. By definition, $F \supset \Omega^{\varphi}$. Hence F includes the smallest subuniverse that includes Ω^{φ}, namely Γ^{φ}. But by its definition, F is included in each E_f for which $\Omega^{\varphi} \subset E_f$, so the lemma is proved. ■

Lemma 9.75. *$\varphi^*(g) = \varphi^*(h)$ iff $\varphi(g) = \varphi(h)$.*

Proof: If: Assume $\varphi(g) = \varphi(h)$. Then by 9.71, $E_g \cap \Omega^{\varphi} = \phi(\varphi(g)) = \phi(\varphi(h)) = E_h \cap \Omega^{\varphi}$. Thus for each ω in Ω^{φ}, we have $g \in \omega$ iff $h \in \omega$, which, by 8.635, is equivalent to $(g \Leftrightarrow h) \in \omega$. Now set $f := (g \Leftrightarrow h)$. Then $f \in \omega$ for each ω in Ω^{φ}; that is, $\Omega^{\varphi} \subset E_f$. So by 9.74, $\Gamma^{\varphi} \subset E_f = E_{g \Leftrightarrow h}$. Thus $(g \Leftrightarrow h) \in \omega$ for each ω in Γ^{φ}; that is, $g \in \omega$ iff $h \in \omega$ for each such ω; that is, $\varphi^*(g) = E_g \cap \Gamma^{\varphi} = E_h \cap \Gamma^{\varphi} = \varphi^*(h)$.

Only if: Suppose $\varphi(g) \neq \varphi(h)$; w.l.o.g. let $\hat{\omega} \in \varphi(g) \backslash \varphi(h)$. Set $\omega := \phi(\hat{\omega}) = \{f : \hat{\omega} \in \varphi(f)\}$. Then $\omega \in \phi(\hat{\Omega}) = \Omega^{\varphi} \subset \Gamma^{\varphi}$, and $g \in \omega$ but $h \notin \omega$. Hence $\omega \in E_g \cap \Gamma^{\varphi} = \varphi^*(g)$ but $\omega \notin E_h \cap \Gamma^{\varphi} = \varphi^*(h)$, so $\varphi^*(g) \neq \varphi^*(h)$. ■

Proof of 9.7: To prove that $(\Omega^{\varphi}, \varphi^*)$ is a representation of $\mathfrak{S}$, we establish 9.1 for φ^*. From 9.61 and 8.31 we get

$$\begin{aligned}\varphi^*(\neg f) &= E_{\neg f} \cap \Gamma^{\varphi} = (\sim E_f) \cap \Gamma^{\varphi} = \Gamma^{\varphi} \backslash (\Gamma^{\varphi} \cap E_f) \\ &= \Gamma^{\varphi} \backslash \varphi^*(f).\end{aligned} \tag{9.76}$$

From 9.61 and 8.32 we get

$$\varphi^*(f \vee g) = E_{f \vee g} \cap \Gamma^{\varphi} = (E_f \cup E_g) \cap \Gamma^{\varphi} = \varphi^*(f) \cup \varphi^*(g). \tag{9.77}$$

From 9.61, 8.4, 1.4, and Γ^{φ} being a subworld, we get

$$\begin{aligned}\varphi^*(k_i f) &= E_{k_i f} \cap \Gamma^{\varphi} = K_i E_f \cap \Gamma^{\varphi} = K_i E_f \cap K_i \Gamma^{\varphi} \\ &= K_i (E_f \cap \Gamma^{\varphi}) = K_i \varphi^*(f).\end{aligned} \tag{9.78}$$

Combining 9.76 through 9.78, we conclude[28] that φ^* is indeed a representation of $\mathfrak{S}$.

[28] The conclusion from 9.38 uses the fact that the knowledge operator for the universe Γ^{φ} is $K_i | \mathscr{E}^{\varphi}$; this follows from Γ^{φ} being a subuniverse.

That φ^* is isomorphic with φ follows from 9.75. ■

10. Discussion

(a) The meaning of the knowledge operators K_i

Proposition 8.4 says that when K_i is applied to a syntactic event E_f, its meaning corresponds precisely to that of the knowledge operator k_i on the formula f. But the domain of K_i includes *all* events – all subsets of Ω – not just syntactic events. Thus one may ask how to interpret K_iE for events E that are not syntactic (e.g., infinite unions of syntactic events). Specifically, is it still correct that K_iE means "i knows E"?

This question is in the interpretive realm; the answer depends on how we think about it. On the straightforward intuitive level, the answer is "no." With the definition in Section 6, the correct interpretation of K_iE is not that "i knows E", but that "E follows logically from the syntactic events that i knows" (see 8.51). An individual i may know something without its following logically from the syntactic events (formulas) that he knows; in that case, K_iE does not obtain, though it is the case that "i knows E".

For example, i might know an infinite disjunction without knowing any of the partial finite disjunctions. Thus suppose that the alphabet contains infinitely many letters $x_1, x_2, \ldots$. The event $E_{x_1} \cup E_{x_2} \cup \cdots$ signifies that "at least one of the x_m obtains". But $K_i(E_{x_1} \cup E_{x_2} \cup \cdots)$ does *not* signify that "i knows that at least one of the x_m obtains"; rather, it signifies the stronger statement that "it follows from the *formulas* known to i that at least one of the x_m obtains." For this, i must know some *finite* disjunction $x_1 \vee x_2 \vee \cdots \vee x_m$.

More generally, we have

Proposition 10.1. *If $f_1, f_2, \ldots,$ are formulas, then*

$$K_i(E_{f_1} \cup E_{f_2} \cup E_{f_3} \cup \cdots) = E_{k_i f_1} \cup E_{k_i(f_1 \vee f_2)} \cup E_{k_i(f_1 \vee f_2 \vee f_3)} \cdots.$$

Proof: That the right side is included in the left follows from 1.32, 8.3 and 8.4. To prove the opposite inclusion, let $\omega \in K_i(E_{f_1} \cup E_{f_2} \cup E_{f_3} \cup \cdots)$. Then by 1.21,

$$\mathbf{I}_i(\omega) \subset E_{f_1} \cup E_{f_2} \cup E_{f_3} \cup \cdots. \tag{10.11}$$

Setting $\mathfrak{L} := \kappa_i(\omega)$, we obtain from 8.74, 1.1, 8.2, and 8.73 that for all states ω',

$$\mathfrak{L} \subset \omega' \quad \text{implies} \quad f_1 \in \omega' \quad \text{or } f_2 \in \omega' \quad \text{or } \cdots. \tag{10.12}$$

Therefore the list $\mathfrak{L} \cup \{\neg f_1, \neg f_2, \ldots\}$ is not consistent; for if it were, then by 8.7, it would be included in a state, contrary to 10.12. So by 8.67, already one of the lists $\mathfrak{L} \cup \{\neg f_1, \ldots \neg f_m\}$ is not consistent. Therefore $\mathfrak{L} \cup \{\neg f_1 \wedge \cdots \wedge \neg f_m\}$ – and so also $\mathfrak{L} \cup \{\neg(f_1 \vee \cdots \vee f_m)\}$ – are not consistent. So $\neg(f_1 \vee \cdots \vee f_m)$ is not in any state ω' that includes $\mathfrak{L}$. So $f_1 \vee \cdots \vee f_m$ is in every state ω' that includes $\mathfrak{L}$. In particular, $f_1 \vee \cdots \vee f_m$ is in every state ω'

with $\kappa_i(\omega') = \kappa_i(\omega)$. Thus $\mathbf{I}_i(\omega) \subset E_{f_1 \vee \cdots \vee f_m}$. So by 1.21 and 8.4,

$$\omega \in K_i(E_{f_1 \vee \cdots \vee f_m}) = E_{k_i(f_1 \vee \cdots \vee f_m)}$$
$$\subset E_{k_i f_1} \cup E_{k_i(f_1 \vee f_2)} \cup E_{k_i(f_1 \vee f_2 \vee f_3)} \cdots. \qquad \blacksquare$$

For another illustration[29], consider an alphabet with just one letter x. Let $N = \{1, 2, 3\}$ (three individuals). The statement "x is common knowledge between 1 and 2" describes the event[30] $K_{12}^{\infty} E_x$. But "3 knows that x is not common knowledge between 1 and 2" does *not* describe the event $K_3 \sim K_{12}^{\infty} E_x$. Indeed, "3 knows that x is not common knowledge between 1 and 2" means that 3 knows that mutual knowledge of x between 1 and 2 (Section 2) fails at some level, but he (3) need not know at *which* level it fails[31]. On the other hand, $K_3 \sim K_{12}^{\infty} E_x$ signifies that mutual knowledge of x between 1 and 2 fails at some specified level m that 3 knows.

Formally, $K_{12}^{\infty} E_x := K_{12}^1 \cap K_{12}^2 E_x \cap \cdots$, where K_{12}^m denotes m'th level mutual knowledge between 1 and 2. Thus $\sim K_{12}^{\infty} E_x = \sim K_{12}^1 E_x \cup \sim K_{12}^2 E_x \cup \cdots$. So as before, $K_3 \sim K_{12}^{\infty} E_x$ signifies that 3 knows some finite disjunction $\sim K_{12}^1 E_x \cup \sim K_{12}^2 E_x \cup \cdots \cup \sim K_{12}^m E_x$. Since this disjunction equals[32] $\sim K_{12}^m E_x$, it follows that $K_3 \sim K_{12}^{\infty} E_x$ signifies that mutual knowledge of x between 1 and 2 fails at a level known to 3, as asserted.

Note that there is *no* event $E \subset \Omega$ that signifies "3 knows that x is not common knowledge between 1 and 2". Events can only represent statements that are constructed from (finite) formulas by means of negation, conjunction and disjunction (possibly infinite or even non-denumerable). The statement "3 knows that x is not common knowledge between 1 and 2" is not of this kind. In contrast, "x is common knowledge between 1 and 2" does describe an event (i.e., a subset of Ω), though not a syntactic event; and of course, the same holds for "x is not common knowledge between 1 and 2". Also, "3 knows that x is common knowledge between 1 and 2" describes an event, namely the event $K_3 \sim K_{12}^{\infty} E_x$; this is because knowledge operators commute with intersections (1.4), but not with unions.

Up to now in this section, we have taken the "intuitive" view: that there is some informal notion of knowledge, not embodied in the syntactic knowledge operators k_i, that enables us to speak of, say, "knowing" an infinite disjunction even when we don't know any of the partial finite disjunctions. But it is possible also to take a different, more formal view: that in an essay of this kind, it is not appropriate to discuss concepts that are not given formal, mathematical expressions. With this view, all relevant knowledge is embodied in formulas; the term "knowledge" is *meaningful* only when it is derived from knowledge of formulas.

If we take this view, we must adjust our answer to the question that begins this section, "does $K_i E$ mean 'i knows E'?" The answer now becomes "yes." Substantively, nothing has changed; as before, $K_i E$ still means "E follows

---

[29] Communicated by Moshe Vardi.

[30] See Section 2; we do not distinguish between K_{12}^{∞} and $K_{\{1,2\}}^{\infty}$.

[31] For example, this could happen if 2 tells 3, "x is not common knowledge between 1 and me," and 3 knows that 2 tells the truth.

[32] If x is mutual knowledge at the m'th level, it certainly is at any lower level.

logically from the formulas that i knows." But with the new view, there *is* no other knowledge; *all* knowledge is embodied in formulas; you *can't* know something *unless* it follows logically from the finite formulas that you know.

Under this view, 10.1 may be stated verbally as follows: "i knows an infinite disjunction of formulas if and only if he knows the disjunction of some finite subset of these formulas."

One may take this view as representing a procedural position only, according to which one should not discuss a concept (such as knowledge of infinite disjunctions) that is not part of the formal framework. But one can also take it as representing a deeper philosophical position, somewhat reminiscent of the intuitionistic school of mathematics that was popular at the beginning of this century. According to this position, the infinite is at best a useful abstraction – a shorthand for avoiding complex, cumbersome formulations. Substantively, all real knowledge must, in the end, be finitely describable.

While we find this position attractive, we do not take an unequivocal stand on these matters. We are pragmatists; our purpose is to develop epistemic machinery that will be useful in the applications. When all is said and done, there is little real difference between the views and positions we have described; they only represent different "angles," different ways of looking at the same thing. The reader may choose the approach he prefers – or supply his own.

(b) An alternative interpretation of the letters of the alphabet

In Section 5, we interpreted the letters in the alphabet as "natural occurrences": substantive happenings that are not themselves described either in terms of people knowing something, or as combinations of other natural happenings using the connectives of the propositional calculus. An alternative interpretation is that a letter of the alphabet may represent any occurrence whatsoever. Usually, the letters will repre sent occurrences in which one is particularly interested, the objects of the analysis. If they happen to refer to an individual knowing something, or to involve logical connectives, that is alright.

A basic rationale for the first interpretation is to avoid logically inequivalent representations of the same occurrence, which would lead to "states of the world" that are in practice impossible. Thus if x denotes "it will snow tomorrow", and y denotes "2 knows it will snow tomorrow," then formally there is a state ω at which both k_2x and $\neg y$ hold, and this is nonsensical. By restricting the alphabet to "natural" occurrences, we seek to avoid this awkwardness.

The trouble with this is that though a natural occurrence is most directly described without knowledge operators or logical connectives, it may have practical implications that do involve them; and then we are back with the original problem. For example, "Alice and Bob played a game of chess" implies that "Bob knows that Alice won or Bob knows that Alice did not win", and this implication is commonly known. A formal system that always avoided this awkwardness would be quite cumbersome, and it does not seem worthwhile to try to construct one.

We have seen that in the alternative interpretation, where a letter may represent any occurrence whatsoever, there may well be "impossible" states:

states that conform to the formal consistency requirements of Section 6, but that cannot occur when the meaning of the letters is taken into account (like the state ω discussed above, containing both $\neg y$ and k_2x, even though y and k_2x have the same meaning). Differently put, it may happen that in a specific application, there are logical connections between the formulas constituting a syntax; connections that are inherent in the interpretations of the letters of the alphabet within the application – and so are commonly known – but that do not follow from the axioms of the propositional calculus (4.4) and of knowledge (4.5). In such an application, we will wish to reserve the term "state" for those states that are actually "possible" – states at which the logical connections that are inherent in the interpretations of the letters are satisfied and are commonly known. The states that are in this sense "possible" constitute a common knowledge subuniverse of Ω.

(c) Knowledge hierarchies

The first explicit constructions[33] of canonical semantic formalisms were hierarchical. For simplicity, suppose there are just two players. Start with a set $\mathfrak{H}$ of mutually exclusive and exhaustive "states of nature," which describe some aspect of reality (like tomorrow's temperature in Jerusalem) in terms not involving knowledge. Let $\mathfrak{H}^1$ be the set of non-empty subsets of $\mathfrak{H}, \mathfrak{H}^2$ the set of non-empty subsets of $\mathfrak{H} \times \mathfrak{H}^1, \mathfrak{H}^3$ the set of non-empty subsets of $\mathfrak{H} \times \mathfrak{H}^1 \times \mathfrak{H}^2$, and so on. A *knowledge hierarchy* h_i of a player i has infinitely many levels. Level 1 describes i's information about the true state s of nature; formally, it is a member h_i^1 of $\mathfrak{H}^1$ (the set of members of $\mathfrak{H}$ that i thinks could be the true s). Level 2 describes i's information about the state of nature *and* the information h_j^1 of the *other* player j about the state of nature;[34] formally, it is a member h_i^2 of $\mathfrak{H}^2$ (the set of members of $\mathfrak{H} \times \mathfrak{H}^1$ that i thinks could be the true (s, h_j^1)). Level 3 describes i's information about the triple consisting of the state of nature, j's information h_j^1 about the state of nature, and j's information h_j^2 about i's information h_i^1 about the state of nature; formally, it is a member h_i^3 of $\mathfrak{H}^3$ (the set of members of $\mathfrak{H} \times \mathfrak{H}^1 \times \mathfrak{H}^2$ that i thinks could be the true (s, h_j^1, h_j^2)). Proceeding in this way, we generate the entire hierarchy $h_i := (h_i^1, h_i^2, \ldots)$; it embodies i's information about the state of nature, about the players' knowledge about the state of nature, about the players' knowledge about *that*, and so on.

Not all sequences h_i in $\mathfrak{H}^1 \times \mathfrak{H}^2 \times \cdots$ are feasible hierarchies of i; some consistency conditions must be met. For h_i to be feasible, it must be that (i) i's information at each level is consistent with his information at all previous levels, (ii) i knows that j considers i's information possible, and (iii) i knows that j's hierarchy h_j is feasible. Formally, we first define feasibility for *finite* sequences $(h_i^1, \ldots, h_i^m)$, using induction on the level m. At level 1, all non-empty subsets of $\mathfrak{H}$ are feasible. For $m > 1$, call $(h_i^1, \ldots, h_i^m)$ *feasible* if

[33] Which were in the probabilistic context; see the companion paper to this one (Aumann 1999).
[34] Though i's information about the state of nature is already embodied in his Level 1, it is not enough at Level 2 to specify just his information about j's Level 1; we need his information about the *pair* (h^0, h_i^1).

(i) h_i^{m-1} is the projection of h_i^m on $\mathfrak{H}^{m-1}$,
(ii) if $(s, h_j^1, \ldots, h_j^{m-1})$ is in h_i^m, then $(s, h_i^1, \ldots, h_i^{m-2})$ is in h_j^{m-1}, and
(iii) if $(s, h_j^1, \ldots, h_j^{m-1})$ is in h_i^m, then $(h_j^1, \ldots, h_j^{m-1})$ is feasible.

Define a *hierarchy for i* as a sequence h_i in $\mathfrak{H}^1 \times \mathfrak{H}^2 \times \cdots$ with $(h_i^1, \ldots, h_i^m)$ feasible for each m.

Call a pair (h_i, h_j) of hierarchies of i and j *mutually consistent* if each player considers the other's hierarchy possible; formally, if $(h_j^1, \ldots, h_j^{m-1})$ is in h_i^m, and $(h_i^1, \ldots, h_i^{m-1})$ is in h_j^m, at each level m.

This yields an explicit construction of a canonical semantic formalism, where a state is identified with a mutually consistent pair of hierarchies, the universe is the set of all states, and i's partition of the universe separates between two such pairs g and h if and only if $g_i \neq h_i$.

While this construction looks quite different from that of Section 6, they are in fact equivalent. If, say, the alphabet has r letters $x_1, \ldots x_r$, then $\mathfrak{H}$ consists of 2^r "states of nature," corresponding to the 2^r possible specifications of x_j or $\neg x_j$ for each $j = 1, \ldots, r$. At each ω in Ω, a unique such state of nature obtains. Moreover, ω determines, for each player i, the states of nature that i considers possible at ω; for example, if $k_i x_1$ is in ω, then i can exclude any state of nature with $\neg x_1$. Thus h_i^1 – and so also h_j^1 – can be read off from ω; similarly, $h^2, h^3, \ldots$ can be read off from ω. One can reason similarly in the opposite direction, and conclude that the two constructions are equivalent.

Though they are equivalent, the construction in Section 6 above is far simpler, more transparent and more straightforward.[35]

Note that two states of the world ω and ω' are in the same element of i's partition of Ω if and only if they correspond to the same knowledge hierarchy of i. Thus i's knowledge hierarchies correspond precisely to the atoms of his information partition; each can be read off from the other. In particular, it follows that the knowledge hierarchy of any *one* player determines precisely the common knowledge component of the true state of the world.

Appendix: Cardinality of the canonical state space

Here we show that when there are at least two players, the canonical system has at least a continuum of states, even when the alphabet has just one letter. By contrast, with only one player (and finitely many letters), the number of states is finite. For background, see the beginning of Section 9.

Define a *context* as a pair consisting of a population N and an alphabet $\mathfrak{X}$. Call a context $(M, \mathfrak{Y})$ *larger* than a context $(N, \mathfrak{X})$ if they are not the same and each component of $(M, \mathfrak{Y})$ includes the corresponding component of $(N, \mathfrak{X})$. Proposition 9.2 yields:

Remark A.1: A list $\mathfrak{L}$ that is consistent in a context $(N, \mathfrak{X})$ is also consistent in any larger context.

[35] In fact, the hierarchy construction is so convoluted that we present it here with some diffidence. Specifically, we have not checked carefully that the three conditions (i), (ii), (iii) really are the "right" conditions for feasibility.

Proof: Since $\mathfrak{L}$ is consistent in $(N, \mathfrak{X})$, it has a model in that context. This can be converted to a model for $\mathfrak{L}$ in the larger context by defining $\hat{\kappa}_j$ and $\varphi(z)$ in an arbitrary fashion when j is not in N and z is a letter not in $\mathfrak{X}$ (e.g., one can define the $\hat{\kappa}_j$ as constants and the $\varphi(z)$ as $\hat{\Omega}$). ■

Proposition A.2. *If* $2 \leq |N| \leq \aleph_0$ *and* $1 \leq |\mathfrak{X}| \leq \aleph_0$, *then* $|\Omega| = 2^{\aleph_0}$.

Before proving A.2, we develop some machinery. Let $\hat{\Omega}$ be a universe, i an individual. We shall say that two states $\hat{\omega}$ and $\hat{\eta}$ in $\hat{\Omega}$ are *i-adjacent* if $\hat{\omega} \in \hat{\mathbf{I}}_i(\hat{\eta})$, *adjacent* if they are i-adjacent for some i. By 1.1, adjacency is symmetric and reflexive. Define the *distance* $d(\hat{\omega}, \hat{\eta})$ as the minimal length of a chain from $\hat{\omega}$ to $\hat{\eta}$ in which successive members are adjacent; formally, as the smallest integer $m \geq 0$ for which there exist states $\hat{\omega}_0, \hat{\omega}_1, \ldots, \hat{\omega}_r$ in $\hat{\Omega}$ with $\hat{\omega} = \hat{\omega}_0, \hat{\omega}_r = \hat{\eta}$, and $\hat{\omega}_{j-1}$ adjacent to $\hat{\omega}_j$ for $j = 1, \ldots, r$. Note that the distance between a state and itself is 0, between different adjacent states 1, and between states in different common knowledge components ∞; and that d is a metric on $\hat{\Omega}$. Denote by $B(\hat{\omega}, r)$ the ball with center $\hat{\omega}$ and radius r, i.e., the set $\{\hat{\eta} : d(\hat{\omega}, \hat{\eta}) \leq r\}$.

Recall (Section 2) that $\hat{K}^1\hat{E} : \bigcap_{i \in N} \hat{K}_i\hat{E}, \hat{K}^m E := \hat{K}^1(\hat{K}^{m-1}\hat{E})$.

Lemma A.31. $\hat{\omega} \in \hat{K}_i\hat{E}$ *iff all states i-adjacent to* $\hat{\omega}$ *are in* E.

Proof: By 1.23, $\hat{\omega} \in \hat{K}_i\hat{E}$ iff $\hat{\mathbf{I}}_i(\hat{\omega}) \subset \hat{E}$; but $\hat{\mathbf{I}}_i(\hat{\omega})$ consists of precisely those states that are i-adjacent to $\hat{\omega}$, so it follows that $\hat{\omega} \in \hat{K}_i\hat{E}$ iff all states i-adjacent to $\hat{\omega}$ are in $\hat{E}$. ■

Corollary A.32. $\hat{\omega} \in \hat{K}^1\hat{E}$ *iff* $\hat{E} \supset B(\hat{\omega}, 1)$ *(i.e., iff each state adjacent to* $\hat{\omega}$ *is in* E*).*

Proof: We have $\hat{\omega} \in \hat{K}^1\hat{E}$ iff $\hat{\omega} \in \hat{K}_i\hat{E}$ for all i; by A.31, this holds iff for all i, all states i-adjacent to $\hat{\omega}$ are in E; and this, in turn, holds iff each state adjacent to $\hat{\omega}$ is in E. ■

Lemma A.33. $\hat{\omega} \in \hat{K}^n\hat{E}$ *iff* $\hat{E} \supset B(\hat{\omega}, n)$.

Proof: By induction on n. For $n = 1$ this is A.32. Suppose the lemma true up to $n - 1$. Then $\hat{\omega} \in \hat{K}^n\hat{E}$ iff $\hat{\omega} \in \hat{K}^1(\hat{K}^{n-1}\hat{E})$; by A.32, this is so iff $\hat{K}^{n-1}\hat{E} \supset B(\hat{\omega}, 1)$, which means that

$$d(\hat{\eta}, \hat{\omega}) \leq 1 \Rightarrow \hat{\eta} \in \hat{K}^{n-1}\hat{E}. \tag{A.331}$$

But by the induction hypothesis, $\hat{\eta} \in \hat{K}^{n-1}\hat{E}$ iff $\hat{E} \supset B(\hat{\eta}, n-1)$, i.e., iff $d(\hat{\xi}, \hat{\eta}) \leq n - 1 \Rightarrow \hat{\xi} \in \hat{E}$. Thus A.331 holds iff

$$d(\hat{\eta}, \hat{\omega}) \leq 1 \Rightarrow (d(\hat{\xi}, \hat{\eta}) \leq n - 1 \Rightarrow \hat{\xi} \in \hat{E}),$$

which is the same as

$$((d(\hat{\eta}, \hat{\omega}) \leq 1) \wedge (d(\hat{\xi}, \hat{\eta}) \leq n - 1)) \Rightarrow \hat{\xi} \in \hat{E}. \tag{A.332}$$

We thus conclude that

$$\hat{\omega} \in \hat{K}^n \hat{E} \quad \text{iff A.332 holds for all } \hat{\eta} \text{ and } \hat{\xi}. \tag{A.333}$$

Now it is a theorem of the first order predicate calculus (and may be verified directly) that

$$(\forall \hat{\xi})(\forall \hat{\eta})(P(\hat{\eta}, \hat{\xi}) \Rightarrow Q(\hat{\xi})) \Leftrightarrow (\forall \hat{\xi})((\exists \hat{\eta})(P(\hat{\eta}, \hat{\xi})) \Rightarrow Q(\hat{\xi})) \tag{A.334}$$

for any predicates P and Q. Also, it may be verified that

$$(\exists \hat{\eta})((d(\hat{\eta}, \hat{\omega}) \leq 1) \wedge (d(\hat{\xi}, \hat{\eta}) \leq n-1)) \Leftrightarrow d(\hat{\xi}, \hat{\omega}) \leq n. \tag{A.335}$$

Together, A.333, A.334, and A.335 yield

$$\hat{\omega} \in \hat{K}^n \hat{E} \quad \text{iff } (\forall \hat{\xi})(d(\hat{\xi}, \hat{\omega}) \leq n \Rightarrow \hat{\xi} \in \hat{E}). \quad \blacksquare \tag{A.336}$$

Corollary A.34. *$\hat{\omega} \in \sim\hat{K}^n \sim \hat{E}$ iff $\hat{E} \cap B(\hat{\omega}, n) \neq \emptyset$ (i.e., iff there is a state in $\hat{E}$ whose distance from $\hat{\omega}$ is at most n).*

Proof: $\hat{\omega} \in \sim\hat{K}^n \sim \hat{E}$ iff $\hat{\omega} \notin \hat{K}^n \sim \hat{E}$; by A.33, this holds iff it is not the case that $\sim\hat{E} \supset B(\hat{\omega}, n)$, i.e., iff $\hat{E} \cap B(\hat{\omega}, n) \neq \emptyset$. ■

Proof of Proposition A.2: Since $|N| \leq \aleph_0$ and $|\mathfrak{X}| \leq \aleph_0$, the set of formulas is at most denumerable, so the set of sets of formulas has cardinality $\leq 2^{\aleph_0}$. Since a state ω is a set of formulas, it follows that $|\Omega| \leq 2^{\aleph_0}$.

To prove $|\Omega| \geq 2^{\aleph_0}$, we will construct a continuum of different lists, each of which is consistent, but any two of which contradict each other. Since any consistent list can be extended to a state (8.7), it will follow that there is a continuum of states. To show that the lists are indeed consistent, we construct models for them (9.2).

By A.1, we may assume w.l.o.g. that $|N| = 2$ and $|\mathfrak{X}| = 1$. The models we construct all use the same universe $\hat{\Omega}$. This universe has $\aleph_0$ states, denoted $\hat{\omega}_0, \hat{\omega}_1, \hat{\omega}_2, \ldots$. The information partition $\hat{\mathscr{I}}_1$ of individual 1 consists of the events $\{\hat{\omega}_0, \hat{\omega}_1\}, (\hat{\omega}_2, \hat{\omega}_3\}, \{\hat{\omega}_4, \hat{\omega}_5\}, \ldots$. The information partition $\hat{\mathscr{I}}_2$ of individual 2 consists of $\{\hat{\omega}_0\}, \{\hat{\omega}_1, \hat{\omega}_2\}, \{\hat{\omega}_3, \hat{\omega}_4\}, \ldots$. One can think of $\hat{\Omega}$ as a sequence of lattice points in the plane, $(0,0), (0,1), (1,1), (1,2), (2,2), (2,3), \ldots$; then $\hat{\kappa}_1$ is the first coordinate (the vertical projection), and $\hat{\kappa}_2$ the second coordinate (the horizontal projection). Note that $\hat{\omega}_n$ is adjacent only to itself, to $\hat{\omega}_{n-1}$ and to $\hat{\omega}_{n+1}$, and that therefore the distance $d(\hat{\omega}_n, \hat{\omega}_m)$ between states $\hat{\omega}_n$ and $\hat{\omega}_m$ is $|n-m|$.

Since there is only one letter x in the alphabet, in order to define a representation, we need only specify to which $\hat{\omega}_n$ we assign x and to which $\neg x$ (9.11). Start out by assigning $\neg x$ to $\hat{\omega}_0$ and x to $\hat{\omega}_1$. After that, define a sequence of blocks with successive lengths $3, 3^2, 3^3, \ldots$. The last state in the ℓ'th block D_ℓ is thus $\hat{\omega}_{1+3+3^2+\cdots+3^\ell} = \hat{\omega}_{(3^{\ell+1}-1)/2}$, the first state in D_ℓ is $\hat{\omega}_{1+3+3^2+\cdots+3^{\ell-1}+1} = \hat{\omega}_{(3^\ell+1)/2}$, and the middle state in D_ℓ is $\hat{\omega}_{((3^\ell+1)/2)+((3^\ell-1)/2)} = \hat{\omega}_{3^\ell}$. Hence

$$B(\hat{\omega}_{3^\ell}, (3^\ell - 1)/2) = D_\ell. \tag{A.35}$$

For each ℓ, define an operator $Q^\ell : \hat{\mathscr{C}} \to \hat{\mathscr{C}}$ (i.e., from events in $\hat{\Omega}$ to events in $\hat{\Omega}$) by $Q^\ell(\hat{E}) := \sim\hat{K}^{3^\ell} \sim\hat{K}^{(3^\ell-1)/2}(\hat{E})$. From A.33 and A.34 it follows that

$$\hat{\omega} \in Q^\ell(\hat{E}) \quad \text{iff there is a ball of radius } (3^\ell - 1)/2 \text{ included in } \hat{E} \text{ whose center is within a distance } 3^\ell \text{ of } \hat{\omega}. \tag{A.36}$$

Now suppose given an arbitrary infinite sequence $\delta := (\delta_1, \delta_2, \ldots)$ of 0's and 1's. Denote by φ_δ the representation that assigns x to each $\hat{\omega}$ in D_ℓ if $\delta_\ell = 1$, and assigns $\neg x$ to each $\hat{\omega}$ in D_ℓ if $\delta_\ell = 0$. From A.35 and A.36 it follows that

$$\hat{\omega}_0 \in Q^\ell \varphi_\delta(x) \quad \text{if } \delta_\ell = 1; \tag{A.37}$$

the ball required by A.36 is simply D_ℓ. Next, we assert

$$\hat{\omega}_0 \in \sim Q^\ell \varphi_\delta(x) \quad \text{if } \delta_\ell = 0. \tag{A.38}$$

Contrariwise, suppose that $\delta_\ell = 0$ and $\hat{\omega}_0 \in Q^\ell \varphi_\delta(x)$. Since $\hat{\omega}_0 \in Q^\ell \varphi_\delta(x)$, it follows from A.36 that within a distance at most 3^ℓ from $\hat{\omega}_0$, there lies the center of a ball – say C_ℓ – of radius $(3^\ell - 1)/2$, all states in which are assigned x. But since $\delta_\ell = 0$, the entire ℓ'th block D_ℓ is assigned $\neg x$; since ω_0 is also assigned $\neg x$, C_ℓ must be included in $\{\omega_1\} \cup D_1 \cup D_2 \cup \cdots \cup D_{\ell-1} = \{\omega_1, \ldots, \omega_{(3^\ell-1)/2}\}$. This set has diameter[36] $(3^\ell - 3)/2$, so the diameter of C_ℓ is $\leq (3^\ell - 3)/2$. But C_ℓ has radius $(3^\ell - 1)/2$, and so diameter $\geq (3^\ell - 1)/2$; this contradiction proves A.38.

Define an operator $k^1 : \mathfrak{S} \to \mathfrak{S}$ (i.e., from formulas to formulas) by $k^1 f := k_1 f \wedge k_2 f$. Define k^m inductively by $k^m f := k^1(k^{m-1} f)$, and define $q^\ell := \neg k^{3^\ell} \neg k^{(3^\ell-1)/2}$. From 9.1 we get

$$Q^\ell \varphi_\delta(x) = \varphi_\delta(q^\ell x) \quad \text{and} \quad \sim Q^\ell \varphi_\delta(x) = \varphi_\delta(\neg q^\ell x). \tag{A.39}$$

For each sequence δ, let $\mathfrak{L}_\delta$ be the list $(f_\delta^1, f_\delta^2, \ldots)$, where $f_\delta^\ell := q^\ell x$ if $\delta = 1$, and $f_\delta^\ell := \neg q^\ell x$ if $\delta = 0$. From A.37, A.38, and A.39, it follows that for each δ, the intersection of the events $\varphi_\delta(f_\delta^\ell)$ contains $\hat{\omega}_0$, and so is non-empty. Hence $(\hat{\Omega}, \varphi_\delta)$ is a model for $\mathfrak{L}_\delta$, and so $\mathfrak{L}_\delta$ is consistent (9.2). But any two lists $\mathscr{L}_\delta$ contradict each other. For, there must be some ℓ such that $\delta_\ell = 1$ for one of the lists whereas $\delta_\ell = 0$ for the other; and for that ℓ, the two corresponding formulas f_δ^ℓ are precisely the negatives of each other. Since there is a continuum of different δ, we have constructed a continuum of lists, each of which is consistent, but any two of which contradict each other. ■

In contrast to A.2, we have

Proposition A.4. *If there is only one individual* ($|N| = 1$) *and the alphabet* $\mathfrak{X}$ *is finite, then* Ω *is finite.*

[36] The *diameter* of a set in a metric space is defined as the maximum (or supremum) distance between two points.

To prove A.4, we need some lemmas. Since there is only one individual, we write k for k_i.

Lemma A.41. *$k(f \vee kg) \Leftrightarrow (kf \vee kg)$ is a tautology for all f and g.*

Proof: This is most quickly proved directly from the definitions in Section 4; but we prefer verbal reasoning[37], which makes the proof more transparent. We use 8.8, which says that a formula is a tautology iff it is in every state. Thus given an arbitrary state ω, we must show $(k(f \vee kg) \Leftrightarrow (kf \vee kg)) \in \omega$. To this end, we use 8.63, which enables us to replace symbols like $\vee, \neg$, and $\Rightarrow$ by the corresponding words. The process is facilitated by writing "f" for the more cumbersome $f \in \omega$; thus "f" means "f holds at ω".

We wish to prove that "$k(f \vee kg)$" iff "$kf \vee kg$". To show "if", assume "$kf \vee kg$". If "kf", then "$k(f \vee kg)$" (4.52), so we are done. If "kg", then "$k(kg)$" (8.1), so "$k(f \vee kg)$" (4.52), and again we are done.

To show "only if", assume "$k(f \vee kg)$". Then "$f \vee kg$" (4.51). If "kg", then "$kf \vee kg$", so we are done. Otherwise "$\neg kg$", so "$k\neg kg$" (4.53), so "$k(f \vee kg)$" and "$k\neg kg$", so "$k(f \vee kg) \wedge k(\neg kg)$", i.e., "$k((f \vee kg) \wedge \neg kg)$" (8.72), i.e., "$k(f \wedge \neg kg)$", so "$kf$" (4.52), so "$kf \vee kg$", and again we are done. ■

Corollary A.42. *$k(kf \vee kg) \Leftrightarrow (kf \vee kg)$ is a tautology.*

Proof: A.41 and 8.1. ■

Corollary A.43. *$k(f \vee \neg kg) \Leftrightarrow (kf \vee \neg kg)$ is a tautology.*

Proof: A.41, 4.51, and 4.53. ■

Call a formula *fundamental* if it is either a letter in the alphabet or has the form kg, where g is elementary (contains no knowledge operators). Say that a formula has *depth* 1 if it is constructed from atomic formulas by using the symbols $\vee, \neg$,), and (only. Thus a formula has depth 1 if and only if the knowledge operator is not concatenated.

Lemma A.44. *If $|N| = 1$, every formula is equivalent*[38] *to a formula of depth 1.*

Proof: Define a formula of *depth n* inductively by specifying that it be constructed from formulas of the form h and kh, where h is a formula of depth $n-1$, by using the symbols $\vee, \neg$,), and (only. Clearly, each formula is of depth n for some n. Thus it suffices to prove that a formula of depth 2 is equivalent to one of depth 1; and for this, it suffices to prove that if h is of depth 1, then kh is equivalent to a depth 1 formula.

So let h be a formula of depth 1. Thus h is constructed from letters and from formulas of the form kg – where g is an elementary formula – by using the symbols of the propositional calculus only. Replace each occurrence in h

---

[37] E.g., using words like "or" and "not".

[38] Recall that f is *tautologically equivalent* to g – or simply *equivalent* for short – if $f \Leftrightarrow g$ is a tautology.

of a formula of the form kg by a new letter, one that does not occur in the original alphabet. We thus get a new formula h', which does not contain any knowledge operators. Write h' in its conjunctive normal form, i.e., as a conjunction of disjunctions of letters and their negatives[39]. Now replace the new letters – those that replaced formulas of the form kg – by the original formulas that they replaced. We thus conclude that

$$h \sim (f_1 \vee \pm kg_{11} \vee \pm kg_{12} \vee \ldots) \wedge (f_2 \vee \pm kg_{21} \vee \pm kg_{22} \vee \ldots) \wedge \ldots,$$

where $\sim$ stands for equivalence, $\pm$ stands for either nothing or $\neg$, and in this instance, three does (...) indicate a *finite* sequence. From 8.72, A.41, A.42 and A.43 we then get

$$kh \sim (kf_1 \vee \pm kg_{11} \vee \pm kg_{12} \vee \ldots) \wedge (kf_2 \vee \pm kg_{21} \vee \pm kg_{22} \vee \ldots) \wedge \ldots,$$

and the right side is a formula of depth 1. ■

Proof of Proposition A.4: From the conjunctive normal form it follows that with a given finite alphabet, there are at most finitely many inequivalent elementary formulas (only finitely many inequivalent disjunctions, so only finitely many inequivalent conjunctions of these disjunctions). Since $f \sim g$ implies[40] $kf \sim kg$, it follows that there are only finitely many formulas of the form kf where f is elementary. Hence there are only finitely many inequivalent first order formulas. The proposition now follows from A.44. ■

References

Aumann RJ (1976) Agreeing to disagree. Annals of Statistics 4:1236–1239

Aumann RJ (1987) Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55:1–18

Aumann RJ (1999) Interactive epistemology II: Probability. International Journal of Game Theory 28:301–314

Aumann RJ, Brandenburger A (1995) Epistemic conditions for Nash equilibrium. Econometrica 63:1161–1180

Fagin R, Halpern JY, Moses M, Vardi MY (1995) Reasoning about knowledge. MIT Press, Cambridge

Hart S, Heifetz A, Samet D (1996) "Knowing whether", "knowing that", and the cardinality of state spaces. Journal of Economic Theory 70:249–256

Hilbert D, Ackermann W (1928) Grundzüge der theoretischen Logik. Springer, Berlin

Hintikka J (1962) Knowledge and belief. Cornell University Press, Ithaca

Kripke S (1959) A completeness theorem in modal logic. J Symbolic Logic 24:1–14

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

Samet D (1990) Ignoring ignorance and agreeing to disagree. Journal of Economic Theory 52:190–207

[39] Every elementary formula can be written in this way. See, for example, Hilbert and Ackermann (1928), p. 9ff.

[40] This follows from 4.52 and the set of tautologies being epistemically closed.