LUC LISMONT AND PHILIPPE MONGIN

ON THE LOGIC OF COMMON BELIEF AND COMMON KNOWLEDGE

ABSTRACT. The paper surveys the currently available axiomatizations of common belief (CB) and common knowledge (CK) by means of modal propositional logics. (Throughout, knowledge – whether individual or common – is defined as true belief.) Section 1 introduces the formal method of axiomatization followed by epistemic logicians, especially the syntax-semantics distinction, and the notion of a soundness and completeness theorem. Section 2 explains the syntactical concepts, while briefly discussing their motivations. Two standard semantic constructions, Kripke structures and neighbourhood structures, are introduced in Sections 3 and 4, respectively. It is recalled that Aumann's partitional model of CK is a particular case of a definition in terms of Kripke structures. The paper also restates the well-known fact that Kripke structures can be regarded as particular cases of neighbourhood structures. Section 3 reviews the soundness and completeness theorems proved w.r.t. the former structures by Fagin, Halpern, Moses and Vardi, as well as related results by Lismont. Section 4 reviews the corresponding theorems derived w.r.t. the latter structures by Lismont and Mongin. A general conclusion of the paper is that the axiomatization of CB does not require as strong systems of individual belief as was originally thought – only *monotonicity* has thusfar proved indispensable. Section 5 explains another consequence of general relevance: despite the "infinitary" nature of CB, the axiom systems of this paper admit of effective decision procedures, i.e., they are *decidable* in the logician's sense.

Keywords: common belief, common knowledge, Kripke structures, neighbourhood structures, partitional model, modal propositional logic, epistemic logic.

1. INTRODUCTION. WHY A FORMAL LANGUAGE?

An event is said to be *common belief* (CB) if every individual in the group believes it, believes that every individual in the group believes it, and so on *ad infinitum*. Following an equally well-received (albeit questionable) view, a known event must be true. Hence the standard definition of *common knowledge* (CK), as perhaps first introduced by Lewis (1969) and as formalized in Aumann's (1976) classic paper: an event is said to be CK if it is true, every individual in the group knows it, etc.

The notions of CK and CB are pervasive in today's game theory. There is an ongoing discussion on what it means (and whether it is at

Theory and Decision **37**: 75–106, 1994.

all necessary) to assume that players have CK of the *rules of the game*; see Binmore and Brandenburger (1990) for a survey and references. An especially important application of this problem relates to the notion of a *type* in games of incomplete information. The standard answer here is suggested by the very construction of types from infinite sequences of mutual beliefs as in Mertens and Zamir (1985): some subset of the type set (a "belief space") must be CB among the players. Game theorists have also long tried to understand the meaning and role of the assumption that players have CB or CK of *rationality*. Efforts have been made to analyze equilibrium concepts in terms of such assumptions: a whole range of solutions should be expected to result from varying the meaning of "rationality" and of either "belief" or "knowledge". Here as elsewhere, Aumann's (1987) work has proved influential. Stalnaker's (1994) recent results clarify and further extend this continuing line of research.[1]

To mention but an economic example, CK assumptions underlie the *no trade theorems* that are derived in Milgrom and Stokey (1982) as well as numerous papers in the same vein. These negative results appear to relate to Aumann's demonstration in 1976 that if the prior probabilities of two agents are the same, and they have CK of their posterior probabilities, these probabilities must also be the same – however different the conditioning information. Aumann's finding uncovered a curious and important problem for the theory of markets under asymmetric information.

The pervasiveness of CB and CK concepts in the current research led some game theorists and economists to analyze the abstract or "logical" properties of these concepts in much detail. As is well-known, Aumann (1976, p. 1237) gave two equivalent definitions of common knowledge: an event E is CK at a state ω if

(1) E contains all ω' "reachable" from ω, given the individuals' information partitions, or

(2) E includes that member of the meet of the individuals' partitions that contains ω.

Definition (1) prompted Geanakoplos and Polemarchakis's (1982) dynamic version of Aumann's impossibility-of-probabilistic-disagreement theorem. Parikh and Krasucki (1990) follow up this line of research in a framework adapted from distributed computing analysis.

Definition (2) is static in character, which makes it perhaps less intuitive than the former. It has suggested an equivalent, "axiomatic" restatement of CK by Milgrom (1981), which illuminates the connection between "public" knowledge and common knowledge. Bacharach (1985) provided a further (again static) restatement using the notion of an "epistemic model".

The papers by Milgrom and Bacharach introduced economists and game theorists to the use of *belief operators*, defined as functions from the algebra of events to itself. The problem of "axiomatizing" CK then became the technical problem of defining suitable constraints on the CK as well as the individual belief operators. This method of analysis has been taken up in a number of later papers. For instance, Monderer and Samet (1989) offer another restatement of Aumann's definition in terms of "evident knowledge". A further variant, which does not quite lead to an equivalence, is Brandenburger and Dekel's (1987): they express CK in the language of measurable sets having conditional probability 1.

Aumann as well as several of his followers had taken for granted that individual belief is *partitional*. This assumption is of course part and parcel of the theoretical economist's modelling of information. Upon reflection, it proved disputable. Hence there arose a research programme of criticizing it and reformulating Aumann's concept of CK, as well as (if possible at all) his impossibility-of-disagreement theorem. A good example is Samet's (1990) demonstration that the theorem does not require the strong *negative introspection* assumption which underlies the partitional model. Geanakoplos (1992) surveys further results in this vein. In economics and game theory, Bacharach (1985, p. 189) was among the first to emphasize that the information partition assumption was both disputable and perhaps dispensable; see also his (forthcoming) discussion.

For all its merits, the foundational literature on CK and CB which has just been discussed has a technical shortcoming. It claims to clarify the "logical" or "axiomatic" basis of these concepts but bears little relation to the logician's method of axiomatization. Logicians distinguish between a *syntax* and a *semantics*. A syntax consists of an artificial, highly constrained, not very expressive but well-understood language; of sentences in that artificial language (axioms); and of rules of inference which generate theorems from axioms and previously generated theorems. A semantics is made out of structures, which are descriptions of the objects of interest. In the context of an epistemic logic structures could involve individual partitions as in Aumann, operators in the Milgrom–Bacharach sense, as well as lesser-known constructions. Structures are defined using the resources of ordinary mathematics, in effect set theory. Hence, the semantic language, as opposed to the syntactical one, is natural, powerfully expressive, but – unsurprisingly – neither well-regimented nor well-understood. The logician's task is to relate the syntax and semantics to each other. The obvious requirements are (i) that the axioms (along with the inference rules) do refer to the given class of structures, and (ii) that they provide an exhaustive account of the properties of that class. These twin requirements correspond to a *soundness* and *completeness* theorems, respectively. Only with the proof of the two theorems does the axiomatization process, in the logician's sense, come to an end. Plainly, what has been labelled "axiomatization of CK" by economists and game-theorists is a more modest undertaking. The above literature uses ordinary mathematics exclusively; it has no syntax. From the logical point of view, the characterization results derived in this literature (such as the equivalences between variously stated definitions of CK) count only as semantic clarifications. The characterization results should by no means be underrated, as one must understand epistemic structures before embarking on the more formalistic steps of axiomatization. However, they are heuristic and preliminary in character.

The axiomatization of belief and knowledge in the sense just sketched is the subject matter of a recently developed subbranch of logic – epistemic logic – as well as a lively area of research in artificial intelligence and parts of computer science. From the syntactical point of view, it has become standard since Hintikka (1962) to use propositional languages enriched with unary operators acting on sentences. These *modal* operators are designed to capture various epistemic qualities; note carefully that they differ from the Milgrom–Bacharach operators since they are syntactical, therefore subjected to more stringent constraints. Various axiom systems have been investigated; many of them were suggested by earlier research on modal operators with non-epistemic interpretations, such as "it is necessary that", "it is possible that", etc. Moving now to semantics, the standard concept of a structure involves a basic set of states (called *possible worlds*, *p.w.*), an assignment of truth values to elementary sentences, and additional mathematical entities defined on the set of p.w. The most popular among those entities are Kripke binary relations. They serve as set-theoretic counterparts to the syntactical belief operators. Properties of Kripke relations – reflexivity, transitivity, symmetry, as well as lesser-known properties – can be shown to correspond exactly to axioms, a feature which probably accounts for the widespread use of this semantics. Economists and game theorists should find it congenial: the partitional model of individual knowledge turns out to be that particular case of Kripke structures in which the individuals' binary relations are equivalence relations. There are, however, alternative semantics to Kripke's that are more expressive and no less elegant, as will be explained at some length in this paper.

Initially, epistemic modal logic was concerned with a single individual's beliefs and knowledge. The move to the multi-agent setting is largely the contribution of Fagin, Halpern, Moses and Vardi (FHMV) in a number of papers, starting with Halpern and Moses (1984, 1985), Fagin, Halpern and Vardi (1984), and Fagin and Vardi (1985). For an updated survey of their work, see Halpern and Moses (1992). The introduction of n belief operators B_a, $a \in A$ (where A is the finite set of agents) into the syntax, along with that of n Kripke relations into the semantics, proved to make little difference to already known completeness and soundness theorems. (It does make a difference, however, to the *complexity* properties of the logics.) Much more delicate was the axiomatization of CB or CK by means of a suitable syntactical operator C. The technical and conceptual problem was the *finiteness* constraint that standard logic imposes on both the length of permissible sentences and the number of axiom schemata. Hence the commonsense definition of C through an infinite conjunction of higher-order belief sentences could not be expressed directly in the formal language. This problem was eventually circumvented by the introduction of a *fixed-point axiom* and an *induction rule*. FHMV have managed to prove soundness and completeness *vis-à-vis* Kripke structures of an axiom system that includes these two components. The remaining problem of how to axiomatize CB and CK in a more general semantics was then tackled by Lismont (1993a) and Lismont and Mongin (1993). The present paper is largely concerned with this issue.

It may be asked, why should economists and game-theorists become interested in the logician's demanding concept of axiomatization instead of pursuing their intuitive methods? There are two *prima facie* answers to this question, having to do with *explicitness* and *effectivity*. A discussion of axioms and inference rules currently used in epistemic modal logic would illustrate the former point. Axioms and rules are a very precise tool to spot hidden assumptions behind ordinary economic or game-theoretic reasoning and to analyze them. In particular, the programme initiated in the 80's of weakening the partitional model of information could be rejuvenated by the injection of a formal language and the accompanying search for sound and complete axiomatizations. In connection with the latter point, recall the logician's usual concern with finiteness. Sentences in the formal language are finite and proofs are defined as finite sequences of some sort. Depending on the system at hand, these constraints might or might not be reflected in the availability of an *effective decision procedure*; that is, a procedure to decide in a finite number of steps whether or not a sentence is a theorem of the system. The search for effective decision procedures connects epistemic model logic with *computability theory*, a subbranch of logic that has recently attracted attention from game theorists and mathematical economists.[2] FHMV and the present writers independently showed that effective decision procedures exist in the case of CB and CK systems. Thus, the infinitary nature of these concepts, as suggested by their commonsense interpretation, does not prevent them from being *effective* in some well-defined sense.

Rather than pursuing a methodological discussion in the abstract, the present paper aims at presenting a sample of the epistemic logician's method of analysis. Hopefully, some of its technical and conceptual advantages will transpire. The strategy of this paper is as follows. Section 2 deals with the syntactical concepts. It briefly reviews standard axioms of individual belief and knowledge as well as axioms for CB and CK. Economists and game theorists will notice both the analogies and disanalogies between our syntactically-based presentation and the current epistemic discussions in their fields. Section 3 reviews the axiomatization of CB and CK in the Kripke semantics, in particular the above-mentioned work by FHMV. It will be seen that sound and complete axiomatizations *à la* Kripke require no more on the semantic side than the intuitive definition of CB in terms of a countable sequence of higher-order shared beliefs. Section 4 introduces neighbourhood structures as a more powerful alternative to Kripke's and reviews some of the authors' work in this framework. Sound and complete axiomatizations of CB exist, using a weak (though not quite minimal) system of individual belief. Importantly, the naive semantic definition of CB would fail to deliver soundness. To circumvent the problem one has either to extend sequences of higher-order shared beliefs beyond denumerable infinity or to translate the fixed-point property of the syntax directly into the semantics. Section 5 states the result on effective decision procedures and adds a few conceptual comments.

2. SYNTACTICAL CONCEPTS

The systems of epistemic logic covered in this paper are specially devised variants of well-known systems of propositional modal logic [see the textbooks by Chellas (1980) or Hughes and Cresswell (1984); Stigum (1990) includes an overview]. Their specific features stem from the fact that there is one belief operator B_a for each "agent" a, and more importantly, there is a C operator, to be interpreted as "it is common belief that". There will be no separate *knowledge* operator. Throughout this paper, as in most of the work of epistemic modal logicians, the difference between belief and knowledge will hinge on whether or not the given operator satisfies the *truth axiom* ("what is believed is true"). This (philosophically objectionable[3]) simplification also applies to CB versus CK. For convenience, we shall also introduce the shared belief operator E, to be understood as "everybody believes that". FHMV have a richer set of syntactical operators than ours. They introduce E_G and C_G operators to capture shared and common belief among a *subgroup* G, as well as a *distributed belief* operator D, to be thought of as a dual to C.[4]

The set Φ of *sentences* or *well-formed formulae* (w.f.f.) of our systems is standardly obtained from the following building blocks: a set PV of propositional variables (of any cardinality); the logical connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$ ("not", "and", "or", "implies", "is equivalent to"); the above-discussed operators $(B_a)_{a \in A}$, C and E.[5] Recall that w.f.f. are finite sequences of symbols. There is no comparable restriction in the work by Aumann and his followers (who are free to use infinite intersections and unions). On the other hand, both their work and the logical one crucially assume that the set A of agents is finite.

Throughout this paper, we shall take for granted the logician's distinction between *axioms* and *rules*. Any logical system below is understood to include an axiomatization of *propositional calculus* by means of the *modus ponens* rule and suitable tautologies. We leave this part of the systems implicit and just explain the *modal* part of them. As usual, both rules and axioms are to be understood as *schemata* in which φ and χ stand for any w.f.f. The logician's requirement of finite axiomatization has of course to do with the number of schemata.

2.1. *Axioms and Rules of Individual Belief*

For any $a \in A$ and any $\varphi, \chi \in \Phi$:

(RM$_a$) $\dfrac{\varphi \rightarrow \chi}{B_a\varphi \rightarrow B_a\chi}$ ("from $\varphi \rightarrow \chi$, infer $B_a\varphi \rightarrow B_a\chi$")

(RN$_a$) $\dfrac{\varphi}{B_a\varphi}$

(C$_a$) $B_a\varphi \wedge B_a\chi \rightarrow B_a(\varphi \wedge \chi)$

In words, the *monotonicity rule* (RM$_a$) says that if it is a theorem of the system that φ implies χ, it follows as another theorem that belief in φ implies belief in χ. This is tantamount to saying that the agent can reproduce all and every inference in the system. (RM$_a$) is compatible with the limiting case in which a does not believe anything. The so-called *necessitation rule* (RN$_a$) stipulates that the agent believes any theorem in the system and thus takes care of this limiting case. The *conjunctiveness axiom* (C$_a$) requires that if a believes φ and believes χ, he believes their conjunction. Note that the converse implication, $B_a(\varphi \wedge \chi) \rightarrow B_a\varphi \wedge B_a\chi$, is implied by (RM$_a$).

These three components together define the *K-system for agent a*, to be denoted by K_a. This system is the minimal one among those warranting a Kripke semantics. Hence, it *has* to be included in any axiomatization of CB of the FHMV type. The epistemic objections against using K_a (and *a fortiori* any stronger system) have recently been discussed in a number of papers under the heading of *logical omniscience* [e.g. Stalnaker (1991), Dubucs (1992), and Gillet and Gochet (1993): see also Bacharach (1994), for an application of this problem to game theory]. It would seem desirable to endow agents with a much less powerful logic than K_a. An attractive proposal is to replace (RM_a) with the weaker *equivalence rule*:

$$(RE_a) \quad \frac{\varphi \leftrightarrow \chi}{B_a\varphi \leftrightarrow B_a\chi}.$$

This says that *a* must reproduce a strict subset of the system's inferences (those involving equivalences). One also looses the unpleasant consequence of (RM_a) that *a* believes all of the system's theorems as soon as he believes anything at all. Instead, he should believe all theorems whenever he believes one of them – a still unpleasant, but weaker consequence. Unfortunately, there are a number of technical problems with the axiomatization of CB under (RE_a). We shall argue below for a weak axiomatization of CB which dispenses with (RN_a) and (C_a) but retains (RM_a).

The careful reader might have noticed that K_a is compatible with *a*'s believing a logical contradiction. This limiting case can be excluded by adding:

$$(D_a) \quad B_a\varphi \rightarrow \neg B_a \neg \varphi .$$

K_aD_a-systems have been widely used as definitions of belief *stricto sensu*, i.e., belief considered irrespective of its truth or falsity content. To move from belief *stricto sensu* to knowledge, one would have to replace (D_a) with the much stronger *truth axiom*:

$$(T_a) \quad B_a\varphi \rightarrow \varphi .$$

This axiom schema simply says that *a* cannot believe falsehoods (in particular: logical contradictions). (RN_a) is a limited converse to (T_a); for it says that all *logical* truths are known. (Of course it is pointless to insist on a full converse to (T_a); for it would make the B_a operator altogether superfluous.) K_aT_a-systems are probably the most common definitions of knowledge in epistemic logic. The following axiom schemata have been included in a number of applications:

(4_a) $B_a\varphi \rightarrow B_aB_a\varphi$

(T'_a) $B_aB_a\varphi \rightarrow B_a\varphi$

(5_a) $\neg B_a\varphi \rightarrow B_a\neg B_a\varphi$

(4_a) and (5_a) are the well-known *positive* and *negative introspection axioms*. (4_a) says in effect that a knows what he believes. A weakening of (T_a) which may be suited for K_aD_a-systems, (T'_a) expresses the converse principle: a's believing that he believes something implies a's actually believing it. (On this principle, higher-order belief should count as a special kind of knowledge.) The intuitive content of both (4_a) and (T'_a) seems to be taken for granted in the game-theoretic analysis of infinite hierarchies of mutual beliefs.[6] Axiom schema (5_a) has a notorious history of criticism in epistemic logic and A.I. It says that if a does not believe something, at least a believes the very fact of not believing it. There no doubt are cases in which the antecedent and consequent clauses of this implication are simultaneously satisfied; but one feels reluctant to elevate it to a universal principle. (5_a) carries with it the unpleasant suggestion that a might be aware of things or propositions towards which he has no epistemic attitudes whatsoever, such as the capital of Vanuatu or arithmetic geometry conjectures. The root of the problem here seems to lie with the dual informal interpretation of $\neg B_a\varphi$ as "a disbelieves φ" *and* "a does not believe φ while not disbelieving it either".[7] There is, however, a positive argument for (5_a) which will be mentioned below.

Systems of modal propositional logic that are (weakly) stronger than K are called *normal*. Prominent among the normal systems for agent a are $K_aD_a4_a$, $K_aT_a4_a$ (to be denoted by $S4_a$), and (for all their demerits) $K_aD_a4_a5_a$ as well as $K_aT_a4_a5_a$ (equivalently $K_aT_a5_a$; to be denoted by $S5_a$). In the sequel any axiom or rule of individual belief will be assumed to hold for either every $a \in A$ or for none. Hence the following notations for the multi-agent setting: RM_A, K_A, $\mathrm{K}_A\mathrm{D}_A4_A$, etc.

2.2. *Axioms and Rules of Common Belief*

For every $\varphi, \chi \in \Phi$:

(Def.E) $E\varphi \leftrightarrow \bigwedge_{a\in A} B_a\varphi$

(FP) $C\varphi \rightarrow E(\varphi \wedge C\varphi)$

(RI$_1$) $\dfrac{\varphi \rightarrow E\varphi}{E\varphi \rightarrow C\varphi}$

(RM$_C$) $\dfrac{\varphi \rightarrow \chi}{C\varphi \rightarrow C\chi}$

(RN$_C$) $\dfrac{\varphi}{C\varphi}$

(C$_C$) $C\varphi \wedge C\chi \rightarrow C(\varphi \wedge \chi)$.

We shall also use:

(IN) $C(\varphi \rightarrow E\varphi) \rightarrow (E\varphi \rightarrow C\varphi)$

(RI$_2$) $\dfrac{\varphi \rightarrow E(\varphi \wedge \chi)}{\varphi \rightarrow C\chi}$

as alternatives to (RI$_1$).

(Def.E) defines *shared belief* by the obvious conjunction and is just introduced into the formal language for convenience reasons. (FP) means that common belief implies shared belief of the statement of interest as well as shared belief of the statement that there is common belief; this is why the axiom is usually referred to as a *fixed-point* one. Note carefully that the informal explanation just given assumes that the shared belief operator distributes over conjunctions in the relevant way, i.e., that (RM_A) holds. The fixed-point part of the implication in (FP) can be viewed in two ways. On the one hand, it indirectly captures the basic, iterative intuition of CB within the axioms. For (FP) in conjunction with suitable monotonicity requirements on E and C will deliver the desired sequence of inferences: $C\varphi \rightarrow E^k\varphi$, $\forall k > 1$, where E^k is $\overbrace{E \cdots E}^{k}$. On the other hand, (FP) directly captures the conceptually important feature that CB involves well-behaved circularities. One might expect CB to imply (and perhaps to be equivalent to) everybody's belief of CB, CB of everybody's belief, as well as higher-order properties where C and E are mixed.

There is indeed a sense in which CB completes the infinite regress of belief (i.e., there is no further infinite regress to be feared on the part of CB itself). This side of the picture has been usefully emphasized by Barwise (1989, ch. 9) in his comparison of the "iterate" and "fixed-point" (or "circular") views of CK.

Formally, (RI_1) says that if $\varphi \rightarrow E\varphi$ is a theorem of the system, then $E\varphi \rightarrow C\varphi$ also is. In words, if a statement is *inherently* shared belief, then it is *inherently* common belief – a by no means implausible view of the way in which CB proceeds from natural evidence. Milgrom (1981, p. 221) was among the first to argue for this view, using the example of the Walrasian auctioneer. Our rule (RI_1) inherits its label *rule of induction* from the related rule (RI_2) in the current FHMV system [earlier axiomatizations by Halpern and Moses (1984, 1985) had "induction axioms" such as (IN); see also Lehmann (1984)]. To see why the label is justified in the case of (RI_1) assume that (RM_a) holds for any $a \in A$. Then, applying (Def.E) one checks that E is monotonic:

$$(\mathrm{RM}_E) \quad \frac{\varphi \rightarrow \chi}{E\varphi \rightarrow E\chi}$$

A simple inductive argument leads to:

$$\frac{\varphi \rightarrow E\varphi}{E\varphi \rightarrow E^k\varphi}, \quad k > 1 .$$

Comparison with (RI_1) suggest that the latter is tantamount to carrying the inference process to the limit.

K_C [$= (\mathrm{RM}_C) + (\mathrm{RN}_C) + (\mathrm{C}_C)$] is a K-system for the C operator. It is an interesting conceptual issue whether or not common and individual belief operators should be subjected to parallel constraints. In Section 3 we shall assume *both* K_A and K_C. As the further results of Section 4 show, the crucial epistemic assumptions are in fact (RM_A) and (RM_C). The latter rule plays a special role in the derivation of the desirable implications $C\varphi \rightarrow E^k\varphi$.

Axiom schema (IN) can be defended on similar intuitions as (RI_1) or (RI_2). The FHMV rule (RI_2) is clearly more powerful than our rule (RI_1). It can be seen that (RI_2) implies the whole of K_C in the presence of K_A and (FP).[8]

2.3. *Formal Inference*

Given an axiom system the formal inference relation $\vdash$ is defined in the following, usual way. If $\varphi \in \Phi$, then $\vdash \varphi$ holds if there is a *finite* sequence of w.f.f. that terminates at φ and is such that every element in it is either the instantiation of an axiom schema or the result of applying a rule to an earlier element. The formula φ is then said to be a *theorem* of the system.[9]

The systems of individual and common belief to be analyzed in Section 3 are:

$$\mathrm{K}_A\mathrm{C}_1 = \mathrm{K}_A + (\mathrm{Def.}E) + (\mathrm{FP}) + (\mathrm{RI}_1) + (\mathrm{RM}_C)$$
$$\mathrm{K}_A\mathrm{C}_2 = \mathrm{K}_A + (\mathrm{Def.}E) + (\mathrm{FP}) + (\mathrm{IN}) + (\mathrm{RM}_C) + (\mathrm{RN}_C)$$
$$\mathrm{K}_A\mathrm{C}_3 = \mathrm{K}_A + (\mathrm{Def.}E) + (\mathrm{FP}) + (\mathrm{RI}_2)$$

as well as occasional strengthenings of these. In Section 4 we shall explore weak variants in which (RM_A) is substituted for K_A. To each system is attached an inference relation $\vdash$; subscripts will be used when necessary.

3. COMMON BELIEF IN THE KRIPKE SEMANTICS

In a multi-agent framework a *Kripke structure* is any $(|A| + 2)$-tuple:

$$m = \langle W, (R_a)_{a \in A}, v \rangle ,$$

where W is a nonempty set (the members of which are referred to as *possible worlds*, *p.w.*); for any $a \in A$, R_a is a binary relation on W (agent a's *Kripke* or *accessibility relation*); and v is a mapping $W \times PV \rightarrow \{0, 1\}$ (the *valuation function*). For convenience we introduce the following derived entities:

$$R_E = \bigcup_{a \in A} R_a \text{ and } R_C = \text{the transitive closure of } R_E.^{10}$$

The class of all Kripke structures will be denoted by $\mathcal{M}^K$.

Given a system of individual and common belief and its associated set Φ of w.f.f. we define the *relation of semantic validation* $\langle m, w \rangle \models$

φ, for any $m \in \mathcal{M}^K$, any w in the p.w. set W of m, and any $\varphi \in \Phi$. This is achieved by means of the following clauses:

- if $\varphi \in PV$, $\langle m, w\rangle \models \varphi \Leftrightarrow v(w, \varphi) = 1$;
- if $\varphi = \neg\chi$, $\langle m, w\rangle \models \varphi \Leftrightarrow$ not $\langle m, w\rangle \models \chi$;
- if $\varphi = \chi \wedge \psi[\chi \vee \psi]$, $\langle m, w\rangle \models \varphi \Leftrightarrow \langle m, w\rangle \models \chi$ and [resp. or] $\langle m, w\rangle \models \psi$;
- if $\varphi = \chi \rightarrow \psi[\chi \leftrightarrow \psi]$, $\langle m, w\rangle \models \varphi \Leftrightarrow \langle m, w\rangle \models \psi$ whenever [resp. iff] $\langle m, w\rangle \models \chi$;
- if $\varphi = B_a\chi$, $\langle m, w\rangle \models \varphi \Leftrightarrow \forall w' \in W$, $wR_aw' \Rightarrow \langle m, w'\rangle \models \chi$;
- if $\varphi = E\chi$, $\langle m, w\rangle \models \varphi \Leftrightarrow \forall w' \in W$, $wR_Ew' \Rightarrow \langle m, w'\rangle \models \chi$;
- if $\varphi = C\chi$, $\langle m, w\rangle \models \varphi \Leftrightarrow \forall w' \in W$, $wR_Cw' \Rightarrow \langle m, w'\rangle \models \chi$.

This list should be seen as a condensed inductive definition, where the induction variable is the *syntactical complexity* of w.f.f.'s, as defined in the obvious way. The first clause settles the case of 0-complexity sentences while the remaining ones enter the inductive step. The clause relative to $B_a\chi$ is usually commented upon as follows: a (semantically) believes χ at a world w iff χ holds true of all those worlds w' which a regards as possible when the actual world is w. Thus, the standard construal is that R_a describes *subjective possibility*, i.e., possibility viewed by a. The present writers think that this informal explication of Kripke relations is obscure and that the neighbourhood semantics, to be introduced in Section 4, is in general better suited than the Kripke semantics when it comes to specifically epistemic applications of modal logic. However, an advantage of the present approach is that it delivers simple and elegant results, some of which happen to connect with Aumann's and his followers' analyses of CB and CK.

The clause relative to $E\chi$ reflects the intended meaning of E as everybody's belief. More technically, it validates the (Def.E) schema:

$$\langle m, w\rangle \models E\chi \Leftrightarrow \langle m, w\rangle \models \bigwedge_{a \in A} B_a\chi \, .$$

The $C\chi$ clause can usefully be reformulated in two equivalent ways. For any $w, w' \in W$ define w' to be *reachable from* w if there is a finite sequence of p.w. in W starting with w and ending with w' such that, for every two consecutive w_i, w_{i+1} in the sequence, $w_iR_aw_{i+1}$ holds for some $a \in A$. Using this definition the following observation is easily made:

OBSERVATION 1. *For any* $m \in \mathcal{M}^K$ *and any* $w \in W$, $\langle m, w \rangle \models C\chi$ *iff* $\langle m, w' \rangle \models \chi$ *for all* w' *reachable from* w *iff* $\langle m, w \rangle \models E^k\chi$ *for all* $k \geq 1$.

Interestingly, the above definition of reachability is very nearly Aumann's (1976, p. 1237), assuming the special case in which there are two agents and the R_a are equivalence relations. The minute difference is that Aumann only considers sequences of alternating relations, as in $w_1 R_1 w_2 R_2 w_3 R_1 w_4$.

As usual, $m \models \varphi$ denotes $\langle m, w \rangle \models \varphi$ for all $w \in W$, and $\models \varphi$ denotes $m \models \varphi$ for all m in the relevant class of structures. In the latter case, φ is called to be *valid*. To avoid referential ambiguities we might have to state the class of structures explicitly and write for instance $\mathcal{M}^K \models \varphi$. The logician's task of axiomatization was informally explained in Section 1. It can now be made precise. One aim is to prove that the axiom system is *sound* relative to the given class of structures:

$$\vdash \varphi \Rightarrow \models \varphi \text{ ("every theorem is valid".)}$$

This often turns out to be a routine verification. The other aim is to show that the system is *complete* with respect to the class of structures:

$$\models \varphi \Rightarrow \vdash \varphi \text{ ("every valid w.f.f. is a theorem".)}$$

Here lies the more technical part of axiomatization. No proof of soundness-and-completeness theorems (or *determination* theorems) will be provided in this paper. We invite the reader to check soundness by himself and refer to the original papers for the completeness part.

We now proceed to determination results:

THEOREM 1 (FHMV). *The system* $K_A C_3$ *is a sound and complete axiomatization of* $\mathcal{M}^K$.

Proof. See Halpern and Moses (1992, pp. 328–329 and 343–345).[11] ■

Lismont provided an independent proof of determination for $K_A C_1$ and (derivately) for $K_A C_2$:

THEOREM 2(Lismont). *The systems* $K_A C_1$ *and* $K_A C_2$ *are sound and complete axiomatizations of* $\mathcal{M}^K$.

Proof. See Lismont (1993a, pp. 120–123) for the completeness of $K_A C_1$.[12] The completeness proof for $K_A C_2$ results from adapting the previous one. ■

Combining Theorems 1 and 2 leads to a quick verification of the suspected syntactical equivalences:

COROLLARY 1. $\vdash_{K_A C_1} \varphi \Leftrightarrow \vdash_{K_A C_2} \varphi \Leftrightarrow \vdash_{K_A C_3} \varphi$.

The determination results above extend unproblematically to a variety of normal systems for A. Exactly as in elementary (= one-operator) modal logic, axioms added on top of K_A correspond – in the technical sense of determination theorems – to simple relational properties. For instance, D_a, T_a, 4_a, 5_a correspond to R_a being *serial*, *reflexive*, *transitive* and *Euclidean*.[13] The reader interested in these facts could pursue them in Chellas (1980, ch. 5) for the one-operator setting, and in Halpern and Moses (1992) for the relevant applications to CB and CK systems. Here we single out the partitional case in honour of Aumann.

Define an *Aumann structure* to be any member of $\mathcal{M}^K$ in which the R_a are equivalence relations, for every $a \in A$. To take care of partitions in the usual sense we also define a *strict Aumann structure* to be an Aumann structure in which each R_a has a finite number of equivalence classes. Then:

PROPOSITION 3. $K_A T_A 5_A = S5_A$ *is a sound and complete axiomatization of the class of Aumann structures. It is also a sound and complete axiomatization of the class of strict Aumann structures.*

Proof. By suitably adapting one of the completeness proofs above and using the fact that the reflexive, transitive and Euclidean properties together characterize equivalence relations.[14] ■

We may now follow Aumann and make the simplifying (but inessential) assumption that there are two agents. Denote by $\overline{R_1}, \overline{R_2}$ the partitions canonically associated with their equivalence relations R_1, R_2. Observation 1 can be refined as follows:

OBSERVATION 2. *If* $m = \langle W, R_1, R_2, v\rangle$ *is a strict Aumann structure and* $w \in W$, $\langle m, w\rangle \models C\chi$ *iff* $\langle m, w'\rangle \models \chi$ *for all* w' *reachable from* w (*in* either *Aumann's sense* or *ours*) *iff* $\langle m, w\rangle \models E^k\chi$, $\forall k \geqslant 1$ *iff the set* $\{w' \in W \mid \langle m, w'\rangle \models \chi\}$ *includes that member of the meet of* $\overline{R_1}$ *and* $\overline{R_2}$ *which contains* w.

The fact that the two definitions of reachability have become indistinguishable from each other results from the strong properties of equivalence relations (syntactically: $S5_A$).[15]

Proposition 3 delivers an axiomatization of the partitional model of individual belief. It therefore clarifies the role of the notorious principle of negative introspection in that model – a role that economists and game theorists had come to realize by purely semantic means. In fairness, it should be mentioned that the relation between $S5_A$ and the partitional model heuristically goes in two directions. On the one hand, uncovering the formidable (5_A) behind the seemingly innocuous assumption that "agents have partitions over the state set" is damaging for that assumption. But there *are* situations in which the partitional assumption is truly innocuous; this means that (5_A) is not always as formidable as it looks. FHMV argue that equivalence classes occur most naturally in distributed computing [e.g., Halpern and Moses (1990, pp. 559–561)]. If a is a process in a system and $w \in W$ is a *global* state of the system, i.e., a vector of local states, one for each process, then $w' \sim_a w''$ means that a's local states are the same in w' and in w''. FHMV add the point that the use of the $\sim_a$ relation exemplifies an "external" view of knowledge, i.e., knowledge as ascribed by the scientist rather than computed by the agent. The remaining conceptual issue is whether or not such a concept of knowledge is at all applicable to game-theoretic reasoning.

We end up this section by listing properties of CB that can (or cannot) be proved in normal systems of the type discussed here.

PROPOSITION 4.

- $\vdash_{K_AC_1} C\varphi \leftrightarrow E(\varphi \wedge C\varphi)$;
- $\vdash_{K_AC_1} EC\varphi \leftrightarrow CE\varphi$;

- $\vdash_{K_A C_1} C\varphi \leftrightarrow E\varphi \wedge CE\varphi$;
- $\vdash_{K_A C_1} C\varphi \leftrightarrow C(\varphi \wedge E\varphi)$;
- $\vdash_{K_A C_1} C\varphi \rightarrow E^k\varphi,\ \forall k \geqslant 1$;
- $\vdash_{K_A C_1} C\varphi \rightarrow C^k\varphi,\ \forall k \geqslant 1$;
- $\vdash_{K_A T_A C_1} C\varphi \rightarrow \varphi$;
- $\vdash_{K_A T_A C_1} C\varphi \leftrightarrow C^k\varphi,\ \forall k \geqslant 1$;
- $\nvdash_{K_A 5_A C_1} \neg C\varphi \rightarrow C \neg C\varphi$.

Proof. Syntactical proofs of most of these results are available (Lismont and Mongin, 1993) but it is much easier to check them by semantic means, relying on the above determination theorem.[16] ■

The above properties along with the axiom system itself (or any equivalent restatement such as K_AC_2 and K_AC_3) appear to capture the essential features of CB in both the "iterative" and the "circular" interpretations of this concept (to repeat Barwise's useful distinction). This would suggest that except for arguments to the contrary (as in the case of distributed computing), one should select a $K_AD_AC_1$ (perhaps $K_AD_A4_AC_1$) axiomatization of individual and common belief, and a $K_AT_AC_1$ (perhaps $K_AT_A4_AC_1$) axiomatization of individual and common knowledge.

4. COMMON BELIEF IN THE NEIGHBOURHOOD SEMANTICS

We start with introducing and discussing neighbourhood semantics in the case of a pure multi-agent logic, i.e., when the only modal operators are B_a, $a \in A$, and E.

A *neighbourhood structure* is defined to be any $(|A| + 2)$-*tuple*:

$$m = \langle W, (N_a)_{a\in A}, v \rangle ,$$

where W and v are as in section 3 and for any $a \in A$, N_a is a mapping $W \rightarrow \mathcal{P}(\mathcal{P}(W))$ – where $\mathcal{P}(\cdot)$ denotes the power set. The class $N_a(w)$ of subsets of W is referred to as a *neighbourhood system* for a at w. For convenience, we introduce the mapping N_E, as defined by:

$$N_E(w) = \bigcap_{a\in A} N_a(w) , \quad \forall w \in W .$$

As in Section 3 the relation of semantic validation $\langle m, w\rangle \models \varphi$ will be defined inductively. The inductive definition includes the previously stated clauses when $\varphi \in PV$, $\varphi = \neg \chi$, and $\varphi = \chi_1 * \chi_2$, where $*$ is any two-place connective. It has the following specific clauses:

– if $\varphi = B_a\chi$, $\langle m, w \rangle \models \varphi$ iff $[\![\chi]\!]^m \overset{\text{def}}{=} \{w' \in W \mid \langle m, w' \rangle \models \chi\} \in N_a(w)$;

– if $\varphi = E\chi$, $\langle m, w \rangle \models \varphi$ iff $[\![\chi]\!]^m \in N_E(w)$.

In words, $[\![\chi]\!]^m$ is the set of all worlds at which χ holds. It is usually called the *truth set* of χ but one could as aptly refer to it as the *proposition* corresponding to χ in m. Recall the philosopher's basic distinction between a sentence, viewed as a symbol or string of symbols in a more or less formal language, and the proposition corresponding to it. The latter is an extensional (set-theoretic) entity – the possibly empty set of states of affairs at which the given sentence holds true. Using this terminology, the neighbourhood system $N_a(w)$ may be seen as listing the propositions that a accepts at w. This system $N_a(w)$ provides an *informal* description of beliefs, to be compared with the *formal* use of belief operators. The above clauses then just say that a (everybody) believes χ iff a (resp. everybody) accepts the proposition associated with χ. On the face of it, this is a relatively non-committal semantics. The neighbourhood account of belief strikes one as less philosophically exacting than the Kripke one in terms of subjective possibility.

There is another reason for preferring the former to the latter in many applications. The neighbourhood semantics is well-known to be *more general* than the Kripke one. Technically, the class $\mathcal{M}^N$ of all neighbourhood structures as above defined is axiomatized by the weak equivalence rules (RE_a), $a \in A$, and the insubstantial (Def.E).

PROPOSITION 5. *$RE_A Def.E$ is a sound and complete axiomatization of $\mathcal{M}^N$.*

Proof. By minor adjustments in the ordinary proof of the one-operator case, for example in Chellas (1980, pp. 252–257). ∎

Further determination theorems show how axioms and rules over and above (RE_A) are to be interpreted in terms of specific *membership or closure conditions* on classes of subsets. Importantly:

PROPOSITION 6. *$K_A Def.E$ is a sound and complete axiomatization* *of the subclass* $\mathcal{M}^{NK}$ *of models* $m = \langle W, (N_a)_{a \in A}, v \rangle$ *in which, for any* $a \in A$ *and any* $w \in W$:

(i) $N_a(w)$ *is closed under supersets* [*cf.* RM_A];

(ii) $N_a(w)$ *includes* W[*cf.* RN_A];

(iii) $N_a(w)$ *is closed under any* (*finite or infinite*) *intersections* [*cf.* C_A].

Proof. By adapting Chellas (1980, p. 260). ■

Since K_ADef.E is also a sound and complete axiomatization of multi-agent Kripke structures, Proposition 6 suggests a close connection between $\mathcal{M}^K$ and $\mathcal{M}^{NK}$. There is indeed a *logical isomorphism* between $\mathcal{M}^K$ and $\mathcal{M}^{NK}$, i.e., a one-to-one mapping $\mathcal{M}^K \to \mathcal{M}^{NK}: m = \langle W, (R_a)_{a \in A}, v \rangle \to m' = \langle W, (N_a)_{a \in A}, v \rangle$ such that, for any $w \in W$ and any $\varphi \in \Phi$: $\langle m, w \rangle \models \varphi$ (in the Kripke sense) iff $\langle m', w \rangle \models \varphi$ (in the neighbourhood semantics sense). On this crucial result, see Chellas (1980, p. 222). It clarifies the sense in which Kripke structures should be regarded as *particular cases* of neighbourhood structures.[17]

We now return to the rich language of Section 2 and embark upon the task of providing $\varphi = C\chi$ with a neighbourhood semantics. This raised a number of problems that were solved only recently. The rest of this section is devoted to surveying solutions offered by Lismont (1993a) and Lismont and Mongin (1993). All of these solutions involve one's restricting attention to the subclass $\mathcal{M}^{NM}$ of *monotonic neighbourhood structures*, i.e., of those $\langle W, (N_a)_{a \in A}, v \rangle$ in which for every a and w, $N_a(w)$ is closed under supersets. Accordingly, it will not be possible to use the full force of Proposition 5. (RM_A) rather than (RE_A) should be taken as the minimal system of individual belief [see Proposition 6, (i)]. The corresponding minimal system is RM_AC_1.

We want our semantic construction to embody the commonsense notion of CB. Technically, the construction must satisfy the *Minimum Semantic Requirement* (MSR) that whenever $\langle m, w \rangle \models C\chi$, $\langle m, w \rangle \models E^k\chi$, $\forall k \geq 1$. It is tempting to try the MSR as a *definition* (i.e. necessary *and sufficient* condition) for CB:

$$* \qquad \text{if } \varphi = C\chi\,, \quad \langle m, w \rangle \models \varphi \text{ iff } \langle m, w \rangle \models E^k\chi,\ \forall k \geq 1\,.$$

Lismont (1983b) has shown that this intuitive approach fails to deliver a soundness proof for RM_AC_1. Such a negative result makes a striking difference between the neighbourhood semantics approach to CB and the Kripkean one. For the Kripkean systems covered in this paper were complete systems, granting the $*$ definition of common belief (see again Observation 1). When it comes to neighbourhood structures, even granting the monotonicity requirement, the MSR can be met only as a *necessary* condition. The basic reason for the failure of the intuitive approach can only be alluded here. It is because the infinite *denumerable* sequence of higher-order shared belief sentences $\langle E^k\chi\rangle_{k\geq 1}$ does not have as much semantic force as does the fixed-point axiom (FP). Such a discrepancy did not occur in the Kripkean case.

There are two solutions to this problem. Solution 1 consists roughly speaking in constructing sequences of semantic analogues of relevant higher-order belief sentences that extend sufficiently far into the ordinals. In essence Lismont's (1993a) construction adopts solution 1. As a preliminary step, for any given $m \in \mathcal{M}^{NM}$, he defines a sequence of functions N_η inductively on ordinals as follows: $N_0 = N_E$ and $N_\eta = N_E \circ (\bigcap_{\zeta<\eta} N_\zeta \cap \mathscr{E})$ for any $\eta > 0$, where $\mathscr{E}$ is the *neutral* neighbourhood, and $\circ$ is the *composition* operation $\circ$ on the neighbourhood functions. They are defined by: $P \in \mathscr{E}(w) \Leftrightarrow w \in P$ and $P \in N_1 \circ N_2(w) \Leftrightarrow \{w' \in W \mid P \in N_2(w')\} \in N_1(w)$, for any $P \subseteq W$ and any $w \in W$.[18] The epistemic undertone of the composition operation must be clear. Assume that N_1 and N_2 are the neighbourhoods associated with agents 1 and 2 respectively, and $P = [\![\chi]\!]^m$ for some χ. Then, $P \in N_1 \circ N_2(w)$ iff $\langle m, w\rangle \models B_1B_2\chi$. This example shows that $\circ$ is indeed introduced to analyze higher-order belief sentences.[19]

Given the monotonicity property of N_E, the sequence N_η is seen to be decreasing, i.e., for any ordinals η, ζ, if $\eta < \zeta$ then $N_\zeta \subseteq N_\eta$. Elementary set-theoretic facts ensure that there is a smallest ordinal *min* such that $N_\eta(w) = N_{min}(w)$ for every $\eta \geq min$ and every $w \in W$. The neighbourhood N_C for CB can now be defined to be N_{min}. It is seen to satisfy the important *fixed-point property* that:

$$N_C = N_E \circ (N_C \cap \mathscr{E}),$$

a semantic analogue of (FP). We can now define the long-awaited validation clause of $C\chi$ as:

** $\quad$ if $\varphi = C\chi$, $\langle m, w\rangle \models \varphi \Leftrightarrow [\![\chi]\!]^m \in N_C(w)$

and state another determination theorem:

THEOREM 7 (Lismont). *RM_AC_1 is a sound and complete axiomatization of $\mathcal{M}^{NM}$, given the ** validation clause for $C\chi$.*

Proof. See Lismont (1993a, pp. 128–130). ■

As an alternative to the above construction, solution 2 attempts to mimic the properties of the syntax by a straightforward semantic clause. This is the method adopted in Lismont and Mongin (1993). The following concept has proved surprisingly powerful. Given $m \in \mathcal{M}^{NM}$ define $P \subseteq W$ to be *belief closed* (b.c.) if $\forall w \in P$, $P \in N_E(w)$. Using the semantic clause for operator E, the following holds: whenever $P = [\![\chi]\!]^m$ for some $\chi \in \Phi$, belief closure of P is equivalent to the property that $m \models \chi \rightarrow E\chi$. In words, a *proposition* is b.c. iff it is believed by everybody at every world where it is true. We proceed to define CB using the belief closure concept:

*** $\quad$ if $\varphi = C\chi$, $\langle m, w\rangle \models \varphi \Leftrightarrow \exists P \in N_E(w)$ s.t. $P \subseteq [\![\chi]\!]^m$ and P is b.c.

It is important to check that this definition agrees with intuitive modelling purposes. Because the N_a are closed under supersets, N_E also is, and the Minimum Semantic Requirement holds:

OBSERVATION 3. *In any $m \in \mathcal{M}^{NM}$ clause *** implies the MSR.*

The following property further illuminates the reasonableness of defining CB in terms of belief closure:

OBSERVATION 4. *For any $m \in \mathcal{M}^{NM}$ and $\chi \in \Phi$, using the *** definition of $C\chi$, if $[\![\chi]\!]^m$ is b.c. then $[\![E\chi]\!]^m \subseteq [\![C\chi]\!]^m$.*

This observation provides a semantic rendering of the syntactical rule (RI_1). By and large it can be said that solution 1 involves translating the *fixed-point axiom* into the semantics, whereas solution 2 amounts to mimicking the *induction rule*. Notice how simpler the *** stipulation looks compared with ** and its implied set-theoretic construction. There is an apparent disadvantage, however. Solution 1 had endowed the C operator with a semantics that exactly parallels that of the B_a and E operators, i.e., a neighbourhood function had been defined for C. In its explicit wording at least, solution 2 sacrifices this elegant feature.

The upshot of the present construction is

THEOREM 8 (Lismont and Mongin). *$RM_A C_1$ is a sound and complete axiomatization of $\mathcal{M}^{NM}$, given the *** clause for $C\chi$.*

Proof. See Lismont and Mongin (1993, Section 3). ∎

Something can be learned from combining Theorem 7 with Theorem 8: the two semantics ** and *** define the same *valid sentences* exactly. More than that is true:

PROPOSITION 9. *For any $m \in \mathcal{M}^{NM}$ and any $w \in W$, $\langle m, w \rangle \models C\chi$ in the ** sense iff $\langle m, w \rangle \models C\chi$ in the *** sense.*

Proof. See Lismont and Mongin (1993, Section 5). ∎

Hence, the difficulty raised at the end of the last paragraph reduces to a presentation problem. There is no loss of information after all when solution 2 is adhered to. It delivers a *completely* equivalent semantics to solution 1 (although this might not be apparent).

Interestingly, belief closure and related notions have already been used in game-theoretic contexts. The first occurence is probably Mertens and Zamir (1985). In essence, they construct a set W of worlds endowed with much internal structure (any $w \in W$ is an infinite sequence stating a value for an objective parameter θ, a subjective probability on the space of parameters for each player, a subjective probability on the spaces of those subjective probabilities for each player, and so on). Mertens and Zamir's isomorphism theorem implies that each world w can be paired with a vector $\langle \theta, (\theta_a(w))_{a \in A} \rangle$, where $\theta_a(w)$ is a *subjective probability on* W for each a. They investigate subsets P of W – "*belief subspaces*" – *that have the property that for any* $w \in P$ and any $a \in A$, $\theta_a(w)(P) = 1$. In words: a belief subspace is an event that is believed (in the sense of having probability 1) by everybody at every world at which it occurs. This notion is the authors' main tool to analyze CB and CK in their construction. There are visible analogies between their approach and the semantics of the present section.

Closure properties have also been discussed in a more elementary context by Monderer and Samet (1989), and Binmore and Brandenburger (1990). These papers introduce the notions of "evident events" and "truisms", respectively, which are analogous to our belief closed propositions. The latter paper retains Aumann's partitional assumption, whereas the former modifies it just as is required to allow for probabilistic belief (instead of knowledge) and probabilistic CB (instead of CK). Both papers note that Aumann's definitions of CK admit of a further restatement in terms of "evident events" or "truisms", respectively. Shin (1993) elaborates on a related point.

We end up this section by listing syntactical properties of CB. As the following proposition shows, the rather weak system RM_AC_1 is enough to capture the intuitively desirable properties of the C operator. Perhaps surprisingly, most of Proposition 4 survives the weakening of K_AC_1 into RM_AC_1.

PROPOSITION 10.

- $\vdash_{RM_AC_1} C\varphi \leftrightarrow E(\varphi \wedge C\varphi)$;
- $\vdash_{RM_AC_1} C\varphi \rightarrow E\varphi \wedge CE\varphi$;
- $\vdash_{RM_AC_1} C\varphi \rightarrow E^k\varphi,\ \forall k \geq 1$;
- $\vdash_{RM_AT_AC_1} C\varphi \rightarrow \varphi$;
- $\vdash_{RM_AC_1} EC\varphi \rightarrow CE\varphi$;
- $\vdash_{RM_AC_1} C\varphi \rightarrow EC\varphi$;
- $\vdash_{RM_AC_1} C\varphi \rightarrow C^k\varphi,\ \forall k \geq 1$;
- $\vdash_{RM_AT_AC_1} C\varphi \leftrightarrow C^k\varphi,\ \forall k \geq 1$.

5. A DECIDABILITY THEOREM AND FURTHER COMMENTS

An axiom system in a formal language is said to be *decidable* if it admits of an effective decision procedure, i.e., if there is a finitary procedure for deciding of any w.f.f. in the language whether or not it is a theorem of that system. It is well-known that the (nonmodal) propositional calculus is decidable, whereas the predicate calculus is not (e.g. Boolos and Jeffrey, 1974, ch. 10). What about systems of modal propositional logic such as those analyzed in this paper? They are typically decidable in spite of the complications created by modal operators, especially C. Chellas (1980) states the classic decidability results relative to the one-agent case. These results readily extend to multi-agent systems such as RM_A, K_A, $\mathrm{K}_A\mathrm{D}_A4_A$, $\mathrm{S4}_A$, $\mathrm{S5}_A$. By contrast, the fact that the relevant systems of individual *and common* belief are decidable is a novel result that deserves emphasis. In the work of FHMV as well as in the present writers', the decidability conclusion is a by-product of the technique used to prove completeness.

THEOREM 11. *The systems* $K_AC_1 = K_AC_2 = K_AC_3$ *and* RM_AC_1 *are decidable.*

Proof. For the former see Halpern and Moses (1992, Section 3), and Lismont (1993a). For the latter see Lismont (1993a), and Lismont and Mongin (1993). ∎

As an illustration of the finitary procedure that Theorem 11 claims to exist, take for instance $\mathrm{RM}_A\mathrm{C}_1$. The proof that it is complete with respect to the belief closure semantics consists in showing that for any w.f.f. φ there exists a special ("canonical") model m_φ such that

$$m_\varphi \models \varphi \Rightarrow \vdash_{RM_AC_1} \varphi .$$

The proof even implies that there is an upper bound k_φ on the cardinality of W, i.e., the p.w. set of m_φ. Knowing k_φ the following procedure is applied. Construct the class S of all neighbourhood structures having set of p.w. of cardinality at most k_φ. If φ is a theorem of $\mathrm{RM}_A\mathrm{C}_1$, the soundness theorem implies that φ is true of all of these structures. If φ is *not* a theorem, from the above implication φ fails in some element of S (since $m_\varphi \in S$). That is to say, φ must be false of a p.w. in a structure in S. Whatever is the case, the procedure delivers a conclusion after a finite number of steps.

The decidability theorem might be claimed to provide the notion of common belief, as properly axiomatized, with *operational meaning*. It contradicts the following *prima facie* intuition of the case: given the semantic force of the CB operator as at least equal to that of an infinite conjunction, one would have expected that properties of this operator could *not* be falsified by referring to finite models only.

If decidability means some kind of methodological warrant for the use of CB assumptions by social scientists, it does not say much on the currently debated issue of whether or not CB assumptions are epistemically plausible. According to a standard argument in cognitive philosophy they are not.[20] For – it is alleged – the epistemic state of CB and CK can only be reached after the agents have performed an infinite number of steps, and this is impossible. The present paper is not concerned with clarifying the demerits of this view – which the authors regard as naive – but the logical results above might help to provide some perspective. Take for instance system K_AC_1. The soundness of K_AC_1 with respect to the iterative semantics informally means that this system does refer to the "natural" concept of CB, and its completeness that it is an exhaustive description of that "natural" concept. Thus, one is entitled to reason on CB from the axiom system as legitimately as from the "natural" concept. The axiom system explicates CB as a potential rather than actual infinity of shared belief sentences. It is suggested that agents in a state of CB of φ *may* derive $E^{1000}\varphi$; it is not suggested that they *must* have derived $E^{1000}\varphi$ in order to qualify as common believers. Some of the writers attacking the use of CB assumptions in social sciences just overlook this very simple distinction. Others insist that any CB assumption involves the implausible claim that agents be able to perform an *actually* infinite number of steps. Both groups of writers appear to ignore the logical elaboration of common belief by current epistemic logic.

To conclude on a more positive note, the present writers feel that CB assumptions can be justified for two different modelling purposes. On the one hand, there are contexts in which the agent is required to perform actual inference steps as in the "muddy children puzzle" (Barwise, 1981) and various game-theoretic problems (typically those involving backward reasoning on decision trees).[21] In such cases a CB assumption can be defended as a convenient artefact. Any particular model should involve a finite sequence of shared belief operators, $E^1, \ldots, E^k$, but it is easier and more elegant to encompass all particular models at once by introducing C. The role of the infinitary operator here is to serve as an idealization and summary for finitely defined operators of any order. On the other hand, CB assumptions can be, and have been, claimed to have an application to *public* events – see Milgrom's (1981) auction example and Bacharach's (1992) analysis of how CK is acquired. The essential point here seems to be that *a proposition which is inherently shared belief* (such as perhaps "an ambulance is roaring in the street") *is ipso facto common belief*. The induction rule and belief closure definition of CB faithfully capture this interpretation of the C operator at the syntactical and semantic levels, respectively. The use of C here is not intended as a modelling device; rather it should directly account for a phenomenon in social psychology.

ACKNOWLEDGEMENTS

The authors have benefited from useful interchanges with M. Bacharach, J. Dubucs, L.A. Gérard-Varet, A. Heifetz, T. Lucas, A. Rustichini, B. Walliser, as well as many other colleagues interested in epistemic logic and its applications. Special thanks are due to E. Minelli, S. Morris, H. Polemarchakis and T. Van Zandt, who commented on an earlier version.

The two authors acknowledge financial support from the Union Européenne programmes "Capital humain et mobilité" and SPES, respectively.

NOTES

[1] Earlier results in the same research programme are surveyed in Walliser (1991) or Brandenburger (1992).

[2] See Anderlini (1990), Binmore and Shin (1992), Canning (1992), Shin and Williamson (1994). There is also a more remote (but by no means nonexistent) connection with *complexity theory*, a tool of analysis that has over the years become familiar to a number of game theorists.

[3] A common view among philosophers is that knowledge is *justified* true belief but some have objected that this is not yet the correct definition [see Gettier's (1963) classic paper and the recent survey by Usberti (1992)]. At least there is a consensus view that it is inadequate to define knowledge as true belief. Epistemic modal logicians are aware of this objection but find it difficult to take it into account within their formalism.

[4] See Halpern and Moses (1990, Section 6) for the E_G and C_G operators, and Halpern and Moses (1992, Section 5) for the D operator. Informally, distributed belief corresponds to what a fictitious individual would believe if he shared in every belief of every individual.

[5] The set Φ of w.f.f. is the smallest set that contains PV and is closed under the action of logical connectives and modal operators.

[6] The major paper in this field is Mertens and Zamir (1985); it was anticipated by Böge and Eisele (1979); published variants include Brandenburger and Dekel (1993), and

Heifetz (1993); Mongin (forthcoming) provides some comparison with logical hierarchies.
[7] Modica and Rustichini (1994) make the related point that (5_a) should be weakened into a "symmetry of awareness" axiom.
[8] The two rules (RM_C) and (RN_C) follow from (RI_2), (FP), (RM_A), and (RI_2), (RE_A), (RN_A), respectively. Axiom schema (C_C) follows from (RI_2), (RE_A), (C_A).
[9] If $\Sigma \subseteq \Phi$, then $\Sigma \vdash \varphi$ holds if there is a finite number of w.f.f. $\varphi_1, \ldots, \varphi_n \in \Sigma$ such that $\varphi_1 \wedge \cdots \wedge \varphi_n \rightarrow \varphi$ is a theorem of the system.
[10] Formally, for any $w, w' \in W$, wR_Cw' holds iff there is $n > 1$ and a sequence of n p.w., $w_1, \ldots, w_n$, such that $w_1 = w$, $w_n = w'$ and $w_iR_Ew_{i+1}$, $1 \leqslant i \leqslant n-1$.
[11] The proof of the completeness part relies on a selective filtration technique, i.e., the proof is carried out for a given formula φ at a time, using maximal consistent sets of w.f.f. relative to a finite sublanguage $\mathscr{L}$ generated by φ.
[12] The completeness proof is again by selective filtration through a finite sublanguage $\mathscr{L}'$. (Here $\mathscr{L}'$ is taken to be the sublanguage of w.f.f. constructed from the propositional variables of φ and having depth less than or equal to that of φ.)
[13] A serial binary relation is one in which every element has a successor. A Euclidean relation is such that every two elements with a common predecessor are mutually related. Recall that w may happen to be its own predecessor or successor.
[14] The conclusion relative to *strict* Aumann structures is secured by the fact that the above completeness proofs use a filtration technique. See also Lismont and Mongin (1993, section 4).
[15] Observation 2 can be extended to the case of Aumann structures in general, i.e., to the case in which the R_a have a *possibly infinite* number of equivalence classes. This extension is easy enough but not quite trivial. By contrast, when the partitional model is weakened, some results reached in the finite case do not carry through to the infinite case; see Samet (1992),
[16] To prove the theorems, use completeness. To prove that the last formula is *not* a theorem use soundness, as in the following argument. Take a Kripke structure m in which $W = \{w_1, w_2, w_3\}$ and the two individuals' Kripke relations are defined by: $w_1R_aw_2$, $w_2R_aw_2$, $w_1R_bw_3$ and $w_3R_bw_2$. Take any propositional variable p and a valuation function v such that $v(w_2, p) = 0$. Individual relations are Euclidean. Hence, for any $a \in A$, (5_a) is valid in this structure. If $\neg C\varphi \rightarrow C\neg C\varphi$ were a theorem of the system $K_A5_AC_1$, from the soundness theorem it would be valid in this structure. But this is not the case: $\langle m, w_1\rangle \models \neg Cp$ and $\langle m, w_1\rangle \not\models C\neg Cp$.
[17] More should be said on the properties of neighbourhood structures. The reader is again referred to Chellas (1980, ch. 7–9). As a simple example of their flexibility in epistemic applications, Mongin (1994) offers a neighbourhood semantics interpretation of sets having probability 1 or belief function (in Shafer's sense) 1.
[18] The composition operation on neighbourhood functions was first introduced by R. Lavendhomme and T. Lucas.
[19] More generally, the following correspondence holds: given any modality $\mu = \mu_1 \cdots \mu_k$, where the $\mu_i \in \{B_a\}_{a \in A} \cup \{E\}$, and the sequence of neighbourhood functions $N_1, \ldots, N_k$ that are pointwise associated with the μ_i,

$$\langle m, w\rangle \models \mu_1 \cdots \mu_k \chi \text{ iff } [\![\chi]\!]^m \in N_1 \circ \cdots \circ N_k(w).$$

Accordingly, what the construction does is to reproduce, and extend into the transfinite, sequences of sentences that are related to, but more "complicated" than, $E^k\chi$, $\forall k > 1$. The "complication" stems from the fact that we cannot rely on a semantic analogue of the conjunctiveness axiom (C_A).

[20] An example is Sperber and Wilson (1986), echoed in Dupuy (1989).

[21] See Reny (1992) and Bacharach (1994) for references to the work currently done on CB in extensive-form games.

REFERENCES

Anderlini, L.: 1990, "Some Notes on Church's Thesis and the Theory of the Games', *Theory and Decision* **29**, 19–52.

Aumann, R.J.: 1976, 'Agreeing to Disagree', *The Annals of Statistics* **4**, 1236–1239.

Aumann, R.J.: 1987, 'Correlated Equilibrium as an Expression of Bayesian Rationality', *Econometrica* **55**, 1–18.

Bacharach, M.: 1985, 'Some Extensions of a Claim of Aumann in an Axiomatic Model of Knowledge', *Journal of Economic Theory* **37**, 167–190.

Bacharach, M.: 1992, 'The Acquisition of Common Knowledge', in C. Bicchieri and M.L. Dalla Chiara (editors), *Knowledge, Belief and Strategic Interaction*, 285–316, Cambridge University Press, Cambridge.

Bacharach, M.: 1994, 'The Epistemic Structure of a Theory of a Game', *Theory and Decision* **37**, 7–48 (this issue).

Bacharach, M.: forthcoming, 'When Do We Have Information Partitions?', in M. Dempster *et al.* (editors), *Mathematical Models in Economics*, Oxford University Press, Oxford.

Barwise, J.: 1981, 'Scenes and Other Situations', *Journal of Philosophy* **78**, 369–397.

Barwise, J.: 1989, *The Situation in Logic*, Center for the Study of Language and Information, Stanford.

Binmore, K. and A. Brandenburger: 1990, 'Common Knowledge and Game Theory', in K. Binmore, *Essays on the Foundations of Game Theory*, Blackwell, Oxford.

Binmore, K. and H.S. Shin: 1992, 'Algorithmic Knowledge and Game Theory', in C. Bicchieri and M.L. Dalla Chiara (editors), *Knowledge, Belief and Strategic Interaction*, 141–154, Cambridge University Press, Cambridge.

Böge, W. and T. Eisele: 1979, 'On Solutions of Bayesian Games', *International Journal of Game Theory* **8**, 193–215.

Boolos, G. and R.C. Jeffrey: 1974, *Computability and Logic*, Cambridge University Press, Cambridge.

Brandenburger, A.: 1992, 'Knowledge and Equilibria in Games', *Journal of Economic Perspectives* **6**, 83–101.

Brandenburger, A. and E. Dekel: 1987, 'Common Knowledge with Probability 1', *Journal of Mathematical Economics* **16**, 237–245.

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

Canning, D.: 1992, 'Rationality, Computability, and Nash Equilibrium', *Econometrica* **60**, 877–888.

Chellas, B.F.: 1980, *Modal Logic, an Introduction*, Cambridge University Press, Cambridge.

Dubucs, J.: 1992, 'Omniscience logique et frictions cognitives', in D. Andler *et al.* (editors), *Epistémologie et cognition*, 115–131, Mardaga, Liège.

Dupuy, J.P.: 1989, 'Convention et Common Knowledge', *Revue économique* **40**, 361–400.

Fagin, R., J.Y. Halpern, and M.Y. Vardi: 1984, 'A Model-Theoretic Analysis of Knowledge', in *Proceedings of the 25th IEEE Symposium on Foundations of Computer Science*, 268–278. A revised and expanded version appeared in *Journal of the Association for Computing Machinery* **38**, 382–428, 1991.

Fagin, R. and M.Y. Vardi: 1985, 'An Internal Semantics for Modal Logic', in *Proceedings of the 17th ACM Symposium on Theory of Computing*, 305–315, Providence.

Geanakoplos, J.D.: 1992, 'Common Knowledge', *Journal of Economic Perspectives* **6**, 53–82.

Geanakoplos, J.D. and H.M. Polemarchakis: 1982, 'We Can't Disagree Forever', *Journal of Economic Theory* **26**, 363–390.

Gettier, E.: 1963, 'Is Justified True Belief Knowledge?', *Analysis* **23**, 121–123.

Gillet, E. and P. Gochet: 1993, 'La logique de la connaissance: le problème de l'omniscience logique', *Dialectica* **47**, 143–171.

Halpern, J.Y. and Y.O. Moses: 1984, 'Knowledge and Common Knowledge in a Distributed Environment', in *Proceedings of the 3rd ACM Conference on Principles of Distributed Computing*, 50–61, New York. A revised and expanded version appeared in *Journal of the Association for Computing Machinery* **37**, 549–587, 1990.

Halpern, J.Y. and Y.O. Moses: 1985, 'A Guide to the Modal Logic of Knowledge', in *Proceedings of the 9th International Joint Conference on Artificial Intelligence* (*IJCAI-85*), 480–490, New York.

Halpern, J.Y. and Y.O. Moses: 1992, 'A Guide to Completeness and Complexity for Modal Logics of Knowledge and Belief', *Artificial Intelligence* **54**, 319–379.

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

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

Hughes, G.E. and M.J. Cresswell: 1984, *A Companion to Modal Logic*, Methuen, London.

Lehmann, D.: 1984, 'Knowledge, Common Knowledge and Related Puzzles', in *Proceedings of the 3rd ACM Conference on Principles of Distributed Computing*, 62–67, New York.

Lewis, D.K.: 1969, *Convention. A Philosophical Study*, Harvard University Press, Harvard.

Lismont, L.: 1993a, 'La connaissance commune en logique modale', *Mathematical Logic Quarterly* **39**, 115–130.

Lismont, L.: 1993b, 'Common Knowledge. Relating Anti-Founded Situation Semantics to Neighbourhood Semantics', Prépublications du Département de Mathématiques, Université Catholique de Louvain.

Lismont, L. and P. Mongin: 1993, 'Belief Closure: A Semantics of Common Knowledge for Modal Propositional Logic', CORE Discussion Paper 9339, Université Catholique de Louvain.

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

Milgrom, P.: 1981, 'An Axiomatic Characterization of Common Knowledge', *Econometrica* **49**, 219–222.
Milgrom, P. and N. Stokey: 1982, 'Information, Trade and Common Knowledge', *Journal of Economic Theory* **26**, 17–27.
Modica, S. and A. Rustichini: 1994, 'Awareness and Partitional Information Structures', *Theory and Decision* **37**, 107–124 (this issue).
Monderer, D. and D. Samet: 1989, 'Approximating Common Knowledge with Common Beliefs', *Games and Economic Behavior* **1**, 170–190.
Mongin, P.: 1994, 'Some Connections between Epistemic Logic and the Theory of Nonadditive Probability', in P. Humphreys (editor), *Patrick Suppes, Scientific Philosopher*, Kluwer, Dordrecht.
Mongin, P.: forthcoming, 'Une interprétation logique du théorème de Mertens et Zamir', in J. Dubucs and F. Lepage (editors), *Intelligence artificielle et logiques non-classiques: le point de vue philosophique*, Hermès, Paris.
Parikh, R. and P. Krasucki: 1990, 'Communication, Consensus and Knowledge', *Journal of Economic Theory* **52**, 178–189.
Reny, P.J.: 1992, 'Rationality in Extensive-Form Games', *Journal of Economic Perspectives* **6**, 103–118.
Samet, D.: 1990, 'Ignoring Ignorance and Agreeing to Disagree', *Journal of Economic Theory* **52**, 190–207.
Samet, D.: 1992, 'Agreeing to Disagree in Infinite Information Structures', *International Journal of Game Theory* **21**, 213–218.
Shin, H.S.: 1993, 'Logical Structure of Common Knowledge', *Journal of Economic Theory* **60**, 1–13.
Shin, H.S. and T. Williamson: 1994, 'Representing the Knowledge of Turing Machines', *Theory and Decision* **37**, 125–146 (this issue).
Sperber, D. and D. Wilson: 1986, *Relevance*, Blackwell, Oxford.
Stalnaker, R.: 1991, 'The Problem of Logical Omniscience, I', *Synthese* **89**, 425–440.
Stalnaker, R.: 1994, 'On the Evaluation of Solution Concepts', *Theory and Decision* **37**, 49–73 (this issue).
Stigum, B.: 1990, *Towards a Formal Science of Economics*, The MIT Press, Cambridge, Mass.
Usberti, G.: 1992, 'Connaissance et croyance: pour une épistémologie dualiste', in D. Andler *et al.* (editors), *Epistémologie et cognition*, Mardaga, Liège.
Walliser, B.: 1991, 'Logique épistémique et théorie des jeux', *Revue économique* **42**, 801–831.

Luc Lismont
G.R.E.Q.E.,
Ecoles des Hautes Études en Sciences Sociales,
Marseille, France

and

Philippe Mongin
Centre National de la Recherche Scientifique (France) and C.O.R.E.,
Université Catholique de Louvain,
34 voie du Roman Pays,
1348 Louvain-la-Neuve,
Belgium

[*Fax*: (32)10 474301; *E-mail*: corsec@core.ucl.ac.be]