Project Report

KNOWLEDGE IN DIFFERENT SYSTEMS OF MODAL LOGIC

15-816 Modal Logic

Patricia Rich

May 7, 2010

Abstract

Modal logic is used by game theorists in the form of epistemic modal logic, where the $\Box$ operator is interpreted as a knowledge operator. This is useful because it allows the knowledge of players in games to be formalized and included in models of games. Then outcomes can be characterized by this knowledge, leading to theorems of the form "in games of type x, if all players have knowledge y, then solution z will be realized." The system of modal logic used in this context is classical S5, a very strong system which translates into an epistemic logic with a very strong knowledge operator which may not be appropriate for all situations. The purpose of this project, therefore, is to examine the results of using other systems of modal logic, S4 and T, as a basis for the epistemic logic used by game theorists. A natural starting place for inquiry is Aumann's framework and his well-known result that common knowledge of rationality implies the backward inductive outcome in perfect information games. It is shown in this paper that Aumann's result does not depend on an S5-based framework; his result holds for S4- and T-based frameworks as well.

1 Introduction

1.1 Motivation

Jaakko Hintikka's 1962 book *Knowledge and Belief* [11] is regarded as the foundation of epistemic logic, using modal logic as a means to formalize and study the logical properties of knowledge. Not long after, Robert Aumann

[2]-[6] and others [8] realized that being able to represent knowledge formally is important in fields such as economics and game theory, where the beliefs and knowledge of different agents about each other and about their situation play a significant role in how those agents make and should make decisions.

As an intuitive example of the importance of taking account of players' knowledge when studying a game situation, consider poker games. Studying such games and calculating optimal strategies clearly requires not just a formal representation of the game itself, but also of the knowledge of the players. Which cards are dealt to which players and whose hand is most likely to be the best are important questions to consider when deciding whether to fold, call or raise in response to an opponent's bet; at least as important, however, are questions about what each player knows about the distribution of cards and about their opponents, including their opponents' knowledge. Player A's optimal action depends just as much on whether A knows her opponent B is bluffing, whether A knows that B knows that A knows that B is bluffing, and whether A knows that B knows that A is also bluffing, as on the probability of A's actual hand beating B's actual hand in the end.

Reflection on this and other common game situations makes clear that studying interactive situations as game theorists and economists do requires taking account of the knowledge and beliefs that the agents in the situation have. To study such situations formally therefore necessitates formalizing knowledge and belief as done by Hintikka and continued by Aumann, Fagin et al [8] and others [17]. The problem is that in such formalisms the definition of knowledge is based on the S5 system of classical modal logic.

1.2 The task

This universal adherence to S5 as a basis for defining knowledge is a problem for several reasons. One is that the S5 definition of knowledge may not be appropriate for all situations, as it assumes knowledge to have some very strong properties. The purpose of this project, however, is not to determine once and for all which system of modal logic is the best for defining knowledge; the purpose is to explore how our choice of a logical system on which to base our definition of knowledge affects our descriptions of games and the theorems that can be proved about them. Specifically, the task in this project is to determine whether important theorems in game theory depend on an S5 definition of knowledge, in particular by testing Aumann's well known theorem that common knowledge of rationality entails the backward inductive outcome in perfect information games [3], where the backward inductive outcome is the result of all players making the choice at every decision point that is optimal given that all subsequent choices will be made according to the same principle. The theorem will be tested in an S4-based and a T-based system. It will be shown that Aumann's Theorem holds for such systems, and thus that it does not depend on an S5 definition of knowledge.

1.3 Previous work, and how it suggests focusing on Aumann

The use of Aumann's work as a base is appropriate partly because Aumann's papers are foundational. Aumann's most sophisticated framework is largely the product of lectures given in 1989 and it is essentially the culmination of work done by Aumann and others on the formal study of knowledge during a period of about 20 years following Hintikka's [11], as Aumann explains in [4]. See, for example, [8]. The standard model of knowledge in games is that of Aumann. Other knowledge-based models build on that of Aumann and use extended information structures which include a *hypothetical* knowledge operator in addition to the usual knowledge operator; see [1], [10], and [14]. For the purposes of this project, Aumann's model is the clear choice for an object of study since it uses only the knowledge operator which is the intended focus; it is best to see how the choice of a logical system for knowledge influences frameworks in the simple case before investigating what influence it has when other operators are added.

Besides being simple, standard and foundational, Aumann's framework has some other properties that make it a desirable subject of study: Aumann approaches knowledge both semantically through a possible worlds semantics and syntactically through epistemic modal logic in [4]. He discusses the advantages of building a semantic knowledge system from a syntax, allowing the semantic system to be constructed canonically and bringing transparency to the logical properties of knowledge. This transparency will prove useful to the current project, since it will be very clear in what way the S4- and T-based canonical semantic knowledge systems that I construct incorporate the axioms of S4 and T. The proof of Aumann's Theorem in [3] is not carried out in a canonical semantic knowledge system constructed from a syntax; its use of S5 properties of knowledge is visible only in that the knowledge operator of the semantic system partitions the state space, implying that the knowledge operator is an equivalence relation and is therefore an S5 operator. Because of the above described virtues of beginning with a syntax, I re-prove Aumann's Theorem in semantic systems that are canonical and syntax-based.

Much of the related prior work not due directly to Aumann is focused on another important theorem that Aumann proved, again assuming an S5 knowledge operator. This theorem, the Agreeing to Disagree theorem [2], says that if two people have the same prior probability distribution over possible states of the world, and if their posterior probabilities for some event (calculated based on private information) are common knowledge between them, then they have equal posterior probabilities for the event. The details of this theorem are not important for this project; the two important points regarding the theorem are firstly that it is proved using information partitions of the possible states of the world which implies that the knowledge operator defined is that of the modal logic S5, and secondly that there is a body of literature investigating Aumann's results, especially ways of weakening Aumann's assumptions (such as the partitional structure of information) while still enabling the proof of the agreeing to disagree impossibility result. See [7], [9], [13], and [15] for some important papers in that literature. These papers are similar to the current project in that they test one of Aumann's important theorems and demonstrate that the theorem is provable with a definition of knowledge weaker than the S5 definition. In this sense, there is a precedent for my project. My project is unlike the previous work in that it is concerned with game situations rather than economics, and in that I will be starting from syntax and focusing specifically on the underlying modal logical structure of the definition of knowledge, whereas other previous work has focused on whether the state space (set of possible worlds) must be assumed to be partitioned by the knowledge operator, often not mentioning the underlying logical properties of knowledge.

1.4 Obtaining the results

The project will begin by stating Aumann's Theorem more clearly and giving some more detail on what he accomplishes in his most relevant papers. Then I will build a framework for formalizing games based on these papers, but with modifications to incorporate game situations into the formalism and to base the knowledge operator on S4 rather than S5. Then I will prove an analog of Aumann's Theorem in the new S4-based system. I will show that with little difficulty, the same can be done with a T-based knowledge operator.

2 Aumann's work

2.1 Backward Induction and Common Knowledge of Rationality

In *Backward Induction and Common Knowledge of Rationality* [3], Aumann proves the theorem which will be a focal point of this project, referred to here as Aumann's Theorem:

Aumann's Theorem (Theorem 1): Let $\mathfrak{G}$ be a perfect information game and $\mathfrak{K} = (\Omega, \mathbf{s}, (\mathcal{K}_i)_{i \in N})$ be a knowledge system for $\mathfrak{G}$, where Ω is a set of states taken as primitive, $\mathbf{s}$ is a function from states to strategies for each player, and the $\mathcal{K}_i$'s are partitions of the state space for each player. It is assumed that $\mathbf{s}_i$ is measurable with respect to $\mathcal{K}_i$ (i.e. the players have the same strategy for all worlds in a given element of their information partition, so they all know their own strategies at all worlds). Then $CR \subseteq BI$ (those worlds at which there is common knowledge of rationality are a subset of those at which the backward inductive outcome is realized).

The proof of Aumann's Theorem in [3] is carried out in a semantic knowledge system which takes states of the world as primitive and assumes that the set of states is partitioned by the knowledge operator. This means that the knowledge operator divides the state space into sets of worlds which are subjectively indistinguishable from one another; the agent's knowledge at any world in the set is the same as at any other in the set. This partitional structure implies that the knowledge operator obeys the axioms of S5: (T) everything known by an agent is true, (4) an agent's knowing something implies their knowing that it is known, (5) an agent's not knowing something implies their knowing that it is not known, and (K) if it is known that A implies B then knowing A implies knowing B. See [8], for example, for an explanation of the axioms of S5 interpreted as knowledge axioms.

2.2 Interactive Epistemology

Aumann's *Interactive Epistemology I: Knowledge* [4] is a nice paper using Dov Samet's idea of characterizing possible worlds by the propositions that hold there [15]. In it Aumann gives syntactic and semantic formalisms for knowledge and shows them to be equivalent in a sense, and observes that the syntactic approach is more straightforward than the semantic and clears up conceptual difficulties. I will use this approach, but I will need to augment it quite a bit since Aumann does not incorporate game situations or the concept of rationality into this paper and focuses on single player knowledge. So, my framework will be a greatly-augmented version of Aumann's in [4], proving the new versions of Aumann's Theorem in the constructed canonical semantic knowledge systems rather than in the kind of semantic knowledge system used in [3].

3 An S4-based Aumann-style framework

3.1 The game

This project's definition of a game is based on Aumann's in [3]. Let $\mathfrak{G}$ be a perfect information game with non-degenerative payoffs. For $\mathfrak{G}$ to be a perfect information game means that at every point in the game all players know what has happened previously. For the payoffs to be non-degenerative means that the payoffs to each player differ depending on which outcome of the game is realized (where the outcome depends on the strategy choices of the players). Define

$\mathfrak{G} := (N, V, (U_i, S_i)_{i \in N})$.

N is a set of players, assumed here to be finite. V is a set of nodes, points at which a player makes a decision, selecting an action; V_i will refer to those nodes at which player $i \in N$ makes a decision. For $w, v \in V$, $w > v$ indicates that the node w follows the node v, or in other words a playing of the game that includes a decision at w also includes a decision at v. Each player i has a set of strategies, S_i, each of which specifies a choice of action for every $v \in V_i$. Crucially, each strategy s_i selects an action for all of the player's decision nodes, whether or not the player expects that node to be reached during the actual playing of the game. The players select full strategies before the start of play, so they must decide what they would do in every situation that could possibly arise, if they found themselves in that situation. A player's choice of action *at* a particular node v according to strategy s_i will be written s_i^v, and at nodes *after* v as $s_i^{>v}$. U_i is a utility function

$U_i : \underset{j \in N}{\times} S_j \to \mathbb{N}$;

the payoffs to a player are determined by the strategy choices of all the players.

3.2 The syntax

3.2.1 The language

The language presented here is a modification of Aumann's [4], altered to reflect the shift from S5 to S4 and the application of the formalism to a game situation, as Aumann's formalism deals very generally with formalizing knowledge alone. Assume that a game $\mathfrak{G}$ is given. Then construct a language $\mathcal{L}$ starting with an alphabet

$\mathfrak{X} := \{x, y, z, ..., \mathcal{G}, (\mathrm{s}_i^x)_{1 \leq x \leq |S_i|, i \in N}\}$.

The alphabet is assumed to be finite or denumerable, as in [4]. The language will have symbols $\vee$, $\neg$, (,), k_i for all $i \in N$, and e^n for $n \in \mathbb{N}$. The members of $\mathfrak{X}$ are primitive propositions representing relevant simple facts relevant to the game situation.

Then a formula of $\mathcal{L}$ is defined as follows:

For all letters of the alphabet $x \in \mathfrak{X}$, x is a formula.

If f and g are formulae, then so is $(f) \vee (g)$.

If f is a formula, then so is $\neg(f)$.

If f is a formula, then so is $k_i(f)$ for every i.

If f is a formula, then so is $e^n(f)$ for every $n \in \mathbb{N}$.

Parentheses will often be omitted in the usual way, and the familiar connectives $\wedge$,$\rightarrow$, $\leftrightarrow$, and $\bigwedge$ will be used as abbreviations.

3.2.2 Tautologies

Following Aumann again (with the appropriate modifications), define a tautology as a formula with one of the following forms, for any f, g, $h \in \mathcal{L}$, $i \in N$, $n \in \mathbb{N}$, $1 \leq x \leq |S_i|$, and $1 \leq y \leq |S_i|$ with x and y distinct in any formula.

(a) $(f \vee f) \rightarrow f$

(b) $f \rightarrow (f \vee g)$

(c) $(f \vee g) \rightarrow (g \vee f)$

(d) $(f \rightarrow g) \rightarrow ((h \vee f) \rightarrow (h \vee g))$

(e) $k_i f \rightarrow f$

(f) $k_i(f \to g) \to ((k_i f) \to (k_i g))$

(g) $k_i f \to k_i k_i f$

(h) $\mathrm{s}_i^x \to k_i \mathrm{s}_i^x$

(i) $\mathrm{s}_i^x \to (\bigwedge_y \neg \mathrm{s}_i^y)$

(j) $\mathcal{G}$

(k) $(\bigwedge_{i \in N} k_i f) \leftrightarrow e^1 f$

(l) $(\bigwedge_{i \in N} k_i e^n f) \leftrightarrow e^{n+1} f$

An intuitive explanation for the list of tautologies and their intended semantic interpretation may be useful: (a) through (d) are propositional tautologies, which should hold in any model of any game. Note that Aumann's tautology $\neg k_i f \to k_i \neg k_i f$ [4] corresponding to the axiom 5 of S5 has been replaced with (g) $k_i f \to k_i k_i f$ corresponding to axiom 4 of S4. Also note that (e) is the axiom T and (f) is the axiom K, or the distribution axiom. (h) specifies intuitively that the players know their own strategies, while (i) ensures that this strategy choice is unique. (j) represents the fact that propositions about the game itself are universally true within the framework. (k) and (l) define the e^n operator, which represents n'th level mutual knowledge: $e^1 f$ is true if all players know f , or in other words if there is first level mutual knowledge of f. There is $n+1$'th level mutual knowledge if all players know that there is n'th level mutual knowledge.

3.2.3 Lists and the syntax

In a manner similar to that of [4], call the set of all formulae in our language $\mathcal{L}$, given the game $\mathfrak{G}$ and alphabet $\mathfrak{X}$, a syntax. Then let $\mathfrak{S}(\mathfrak{G}, \mathfrak{X})$ be the syntax for $\mathcal{L}$. The syntax will be abbreviated as $\mathfrak{S}$.

A list is a set of formulae. A list $\mathfrak{L}$ is logically closed if $f \in \mathfrak{L}$ and $f \to g \in \mathfrak{L}$ implies $g \in \mathfrak{L}$; in other words, a list is logically closed if it is closed under modus ponens. $\mathfrak{L}$ is epistemically closed if $f \in \mathfrak{L}$ implies $k_i f \in \mathfrak{L}$. A list is strongly closed if it is logically and epistemically closed. The strong closure of a list $\mathfrak{L}$ is the smallest strongly closed list including $\mathfrak{L}$. A list $\mathfrak{L}$ is coherent if $\neg f \in \mathfrak{L}$ implies $f \notin \mathfrak{L}$. It is complete if the reverse is true.

3.3 The semantic knowledge system

3.3.1 Possible worlds

The connection between the syntax developed above and the semantic knowledge system to be developed below and used in proofs is via *possible worlds*. In Aumann [3], for instance, possible worlds are taken as primitives, and therefore their nature is somewhat mysterious. By constructing possible worlds out of a syntax, as in [4] and in this paper, the mystery is dissipated as it is made explicit exactly what possible worlds are: possible worlds are defined as lists of formulae that could all possibly be true at once. In other words, a possible world is defined by the list of the formulae that are true at that world. Another advantage of constructing possible worlds from the syntax, besides increased clarity, is that there is a canonical construction. The canonical construction delivers a unique set of possible worlds, namely all those possible worlds which are both internally consistent and give a definite answer as to whether any formula in the language is true. This circumvents the problem with other potential construction methods of deciding how much information each world should contain (how many formulae should be included) and how many possible worlds there should be. There is no particular good reason for a construction to construct only some possible worlds and not others, or to make the worlds less informative than the language makes possible. The canonical construction is therefore less ad hoc and less prone to the resulting problems than other constructions. Aumann [4] credits Samet [15] with the idea of characterizing worlds in this way rather than taking them as primitives.

Construct the set of possible worlds from the syntax $\mathfrak{S}$ above as follows:

Define a possible world ω as a list $\mathfrak{L}$ with the following properties: $\mathfrak{L}$ contains the list $\mathfrak{L}^\star$ of all tautologies; the list $\mathfrak{L}^\star$ of tautologies is epistemically closed. $\mathfrak{L}$ is logically closed. $\mathfrak{L}$ is coherent and complete. Note that the epistemic closure of $\mathfrak{L}^\star$ corresponds to the necessitation rule of modal logic, which is present in all normal systems including S4, S5, and T; here it corresponds to all players knowing all tautologies.

The set of all possible worlds (all such lists of formulae) is Ω.

3.3.2 The canonical system

Given any world $\omega \in \Omega$, there is a set of formulae known by any player i to be true at that world. Following Samet [15], call this set the ken of the agent at ω, written here as

$K_i^\bullet(\omega) := \{f \in \omega | f = k_i g \text{ for some } g\}$.

Then those other possible worlds $\omega' \in \Omega$ considered possible by an agent at world ω are those worlds at which everything the agent knows at ω – everything in the agent's ken at ω – is true. Therefore I define a possibility relation $\omega p_i \omega'$, saying that at ω agent i considers ω' possible:

$\omega p_i \omega'$ iff $K_i^\bullet(\omega) \subseteq \omega'$.

Then the set of worlds considered possible by the agent at ω,

$P_i(\omega)$, is just $\{\omega' | \omega p_i \omega'\}$.

Note that given the choice of tautologies, the relation p_i is reflexive (due to (e)) and transitive (due to (g)). It is not, however, symmetric (and therefore not an equivalence relation, nor does it partition Ω). See section 2.4 of [8] for proofs of these familiar correspondences between possibility relations between possible worlds and modal logical axioms.

Then

$\mathfrak{C} = (\Omega, \mathfrak{G}, (K_i^\bullet)_{i \in N})$

is the canonical semantic knowledge system for the game $\mathfrak{G}$.

3.3.3 The formula-event correspondence

With the semantic knowledge system it is possible to reason about events, which are subsets of Ω. An event is a set of worlds, corresponding to the formulae that are true at every one of those worlds. This is now defined precisely. Upper-case letters will be used to denote events, while the corresponding lower-case letters will continue to be used to denote formulae, for convenience. Sometimes an event will also be denoted by square brackets, following Aumann [3]. This will be the case in particular when an event corresponds to a formula in the syntax which doesn't have a specific name.

For any formula f, define the semantic event F:

$F := \{\omega \in \Omega | f \in \omega\}$.

Note first of all that since f can be any formula, it could in particular be a formula of the form $k_i g$, and so F could be an event $K_i G$ representing an agent's knowledge of some other event G. Also note that since the tautologies are all true at all ω, if a formula t is a tautology, then the event $T = \Omega$. Since the formula $\mathcal{G}$ is one such tautology, interpreted in the semantic knowledge system as all the information about the game itself, $\mathcal{G} = \Omega$.

The formulae s_i^x remain to be interpreted; it was suggested earlier that they correspond to strategies of player i; now it must be determined which formulae correspond to which strategies. First, define the semantic event that some generic strategy S_i^x is played as usual:

$\mathrm{S}_i^x := \{\omega \in \Omega | \mathrm{s}_i^x \in \omega\}$.

Then let the function

$\mathbf{s}_i : \{(\mathrm{S}_i^x)_{1 \leq x \leq |S_i|}\} \rightarrow S_i$

be a bijection; each event that some strategy is played by i is mapped to exactly one of i's strategies in the game, and all strategies are chosen in some set of worlds. Denote by $\mathbf{s}(\omega)$ the n-tuple of strategies chosen by the players at ω. This will be abbreviated as $\mathbf{s}$ when it is not necessary to explicitly name ω.

4 A proof in S4

4.1 The theorem

Aumann proved in his partitional framework in [3] that in perfect information games if there is common knowledge of rationality among players then the backward inductive outcome is realized. Here is Aumann's Theorem, in the notation of this paper. The formal definitions of rationality, common knowledge of rationality and backward induction will be set forth in the subsequent subsection.

Aumann's Theorem (Theorem 1): Let $\mathfrak{G}$ be a perfect information game and $\mathfrak{K} = (\Omega, \mathbf{s}, (\mathcal{K}_i)_{i \in N})$ be a knowledge system for $\mathfrak{G}$, where Ω is a set of states taken as primitive, $\mathbf{s}$ is a function from states to strategies for each player, and the $\mathcal{K}_i$'s are partitions of the state space for each player. It is assumed that $\mathbf{s}_i$ is measurable with respect to $\mathcal{K}_i$ (i.e. the players have the same strategy for all worlds in a given element of their information partition, so they all know their own strategies at all worlds). Then $CR \subseteq BI$ (those worlds at which there is common knowledge of rationality are a subset of those at which the backward induction outcome is realized).

Aumann's Theorem is proved not in a canonical semantic knowledge system like that constructed above, but rather in a semantic knowledge system which takes possible worlds as primitive. It also takes an S5 information partition for each player as primitive. In this section, I will prove an analogue of Aumann's Theorem in the canonical semantic knowledge system developed above, which is S4-based and therefore non-partitional. This new theorem will be called S4-A's Theorem.

S4-A's Theorem (Theorem 2): Let $\mathfrak{G}$ be a perfect information game, $\mathfrak{S}$ a syntax, and $\mathfrak{C}$ the canonical semantic knowledge system as defined above. Then where CR is the event that there is common knowledge of rationality and BI is the even that the backward induction strategy is chosen by every player, $CR \subseteq BI$.

4.2 Prerequisite definitions and lemmas

4.2.1 Common knowledge

For a proposition to be common knowledge means that all players know it, all know that all know it, all know this, and so on ad infinitum. In other words, a proposition is common knowledge exactly when it is mutual knowledge at every level $n \in \mathbb{N}$. A common knowledge operator is not included in the language $\mathcal{L}$ for a few reasons. One is that it can be straightforwardly defined from mutual knowledge in the semantic knowledge system; the event that a proposition is common knowledge is the same event as that proposition being every level of mutual knowledge. The other reason is that including a common knowledge operator in the language would require stating the axioms of common knowledge as tautologies, which is problematic since common knowledge is an infinitary concept, while formulae of $\mathcal{L}$ must be finite.

Recall that for any event F and level of mutual knowledge m, the event that F is m'th level mutual knowledge is $E^m F$, where

$E^m F := \{\omega \in \Omega | e^m f \in \omega\}$.

Then the event that F is common knowledge, CF, is defined by

$CF := \bigcap_{n \in \mathbb{N}} E^n F$.

It is worth noting in passing that since every closed, coherent and complete set of formulae of $\mathcal{L}$ (containing the tautologies, etc.) constitutes a possible world ω, there will exist worlds where some event is common knowledge. Aumann proves in [3] that the set of worlds in his state space Ω contains at least one in which rationality is common knowledge, but as the purpose of this paper is not to discuss or prove the possibility of common knowledge, I will not do the same here.

4.2.2 Rationality

Rationality, like common knowledge, is not included in the language. To build rationality into $\mathcal{L}$ would require including tautologies governing when it obtains, which would unduly complicate the language since rationality can only be defined from other complex concepts which would themselves need to be explicitly added to the language. The event of a player i being rational is the same as another event, the event that as far as i knows, at all i's decision nodes, i's chosen action based on i's strategy yields at least as high of a conditional payoff as any other action that could be taken. Every player's full strategy is explicitly given for every possible world, and all relevant propositions about the game itself (including strategy options and utilities) are listed in each possible world; therefore, in the semantic knowledge system in which syntactic strategy choices have been interpreted as actual strategies, the conditional payoffs to each player at a given world for their actual strategy and for deviations from it can be calculated. The set of worlds where the conditional payoffs for each node for a player's actual strategy are higher than the conditional payoffs for any deviation from that strategy is an event. The event that the player knows this event is the event that the player is rational.

The conditional payoff to i at v given an n-tuple of strategies $s \in \times_{j \in N} S_j$ is written $h_i^v(s)$. The conditional payoff if i were to select strategy t_i instead is $h_i^v(s, t_i)$. The conditional payoff at a node v given s is the payoff a player would receive if, starting at node v, the players selected the given strategies s_i. The event that i's conditional payoff at v would be greater if i selected some strategy t_i rather than the strategy $\mathbf{s}_i$ specified by $\mathbf{s}$ (those worlds at which the actual strategy choice is not the best, conditional on v) is written $[h_i^v(\mathbf{s}; t_i) > h_i^v(\mathbf{s})]$. The event R_i^v that i is rational at v, then, is the event that i does not know that this is the case:

$$R_i^v := \bigcap_{t_i \in S_i} (\sim K_i[h_i^v(\mathbf{s}; t_i) > h_i^v(\mathbf{s})]).$$

i is rational, R_i, if i is rational at every node:

$$R_i := \bigcap_{v \in V_i} \bigcap_{t_i \in S_i} (\sim K_i[h_i^v(\mathbf{s}; t_i) > h_i^v(\mathbf{s})]).$$

The event that all players are rational at all nodes is

$R := \bigcap_{i \in N} R_i$.

4.2.3 Backward induction

The backward inductive outcome is defined as the outcome of the game when at every decision node, the player to act at that node makes the (backward) inductive choice. The inductive choice is the choice that would result in the highest payoff to the chooser given that all players at all subsequent nodes would also make the inductive choice. This bottoms out in the last possible decision node of the game, since the inductive choice at the last node is trivially just the choice that yields the highest payoff for the acting player. Since our game is a perfect information game, the player to move at the last node knows exactly what payoffs will result from each possible action, and furthermore these payoffs must be different for each action as part of our specification of the game. Therefore there will be a single choice at the last node of the game which maximizes the payoff to the acting player, and this is the unique inductive choice at that node. Given this, the inductive choice at all previous nodes is also determined. Following Aumann, I call the inductive choice (at a given decision node v) b^v. The event that the inductive choice is made at v is $[\mathbf{s}^v = b^v]$. The outcome of the game that results from all players choosing strategies that make the inductive choice at all nodes is referred to as BI:

$BI := \bigcap_{v \in V} [\mathbf{s}^v = b^v]$.

4.2.4 Lemmas

The following lemmas 4-10 will be used in the proof; they are numbered so as to match the numbering in [3]. 4-7 and 9-10 are taken directly from [3], as they are not affected by the change from S5 to S4. Lemma 8* is a modification of Aumann's original Lemma 8: $K_i \sim K_i E = \sim K_i E$, which no longer holds in our system, as it corresponds to the axiom 5 of S5. The new Lemma 8* is trivially an instance of Lemma 9, but it is included because it is used in the proof and the change makes clear an important difference between Aumann's system and the current one.

Lemma 4. $CF = K_i CF$

Lemma 5. If $G \subseteq F$ then $K_i G \subseteq K_i F$

Lemma 6. $K_i G \cap K_i F = K_i(G \cap F)$

Lemma 7. $CF \subseteq F$

Lemma 8*. $K_i \sim K_i F \subseteq \sim K_i F$

Lemma 9. $K_i G \subseteq G$

Lemma 10. $BI^v \subseteq K_i BI^v$ for all nodes v of player i.

The proofs of the lemmas are included in the Appendix, since they are of less independent interest than the proof of S4-A's Theorem. It should be noted, however, that they are essential to the proof and, unlike the proof of the theorem, the proofs of the lemmas in the current S4 system are substantially different from (and less automatic than) the proofs Aumann would have given for his system; Aumann himself shows only the proof of Lemma 10 in [3].

4.3 The proof

The proof of S4-A's Theorem is nearly identical to the proof of Aumann's Theorem, given the lemmas. The difference is near the end of the proof, and it noted there.

S4-A's Theorem (Theorem 2): Let $\mathfrak{G}$ be a perfect information game, $\mathfrak{S}$ a syntax, and $\mathfrak{C}$ the canonical semantic knowledge system as defined above. Then $CR \subseteq BI$.

Proof. First I show that for all nodes v, $CR \subseteq BI^v$. Assume therefore that for all $w > v$ for some v, $CR \subseteq BI^w$. Let i be the player to act at the node v. By assumption $CR \subseteq BI^w$, and so by Lemma 5, $K_i CR \subseteq K_i BI^w$. Since by Lemma 4, $CR = K_i CR$, it follows that $CR \subseteq K_i BI^w$, still for all $w > v$. Then by Lemma 6, $CR \subseteq \bigcap_{w>v} K_i BI^w$. By the definition of BI^w, and Lemma 6 again, $\bigcap_{w>v} K_i BI^w = \bigcap_{w>v} K_i[\mathbf{s}^w = b^w] = K_i \bigcap_{w>v} [\mathbf{s}^w = b^w] = K_i[\mathbf{s}^{>v} = b^{>v}]$. Therefore, $CR \subseteq K_i[\mathbf{s}^{>v} = b^{>v}]$.

By Lemma 7 and the definitions of R and R_i,
$CR \subseteq R \subseteq R_i \subseteq \sim K_i[h_i^v(\mathbf{s}; b_i) > h_i^v(\mathbf{s})]$.

Using Lemma 6 again, we prove that $K_i[\mathbf{s}^{>v} = b^{>v}] \cap K_i[h_i^v(\mathbf{s}; b_i) > h_i^v(\mathbf{s})] = K_i[\mathbf{s}^{>v} = b^{>v} \wedge h_i^v(\mathbf{s}; b_i) > h_i^v(\mathbf{s})] = K_i[\mathbf{s}^{>v} = b^{>v} \wedge h_i^v(b) > h_i^v(b; \mathbf{s}^v)$ $= K_i[\mathbf{s}^{>v} = b^{>v}] \cap [h_i^v(b) > h_i^v(b; \mathbf{s}^v)]$. The move from the second to the third equality here is justified by the fact that the conditional payoffs $h_i^v(x)$

depend only on node v and those that follow it, and the backward inductive strategy choices are known to be made at all $w > v$. Therefore if there is a higher conditional payoff for making the backward inductive choice at v for player i, then the backward inductive strategy choices at and after v have a higher conditional payoff than would switching to some other strategy choice at v.

Given the above equality, $K_i[\mathbf{s}^{>v} = b^{>v}] \cap \sim K_i[h_i^v(\mathbf{s}; b_i) > h_i^v(\mathbf{s})]$
$= K_i[\mathbf{s}^{>v} = b^{>v}] \cap \sim K_i[h_i^v(b) > h_i^v(b; \mathbf{s}^v)]$. Therefore since
$CR \subseteq K_i[\mathbf{s}^{>v} = b^{>v}] \cap \sim K_i[h_i^v(\mathbf{s}; b_i) > h_i^v(\mathbf{s})]$, also
$CR \subseteq K_i[\mathbf{s}^{>v} = b^{>v}] \cap \sim K_i[h_i^v(b) > h_i^v(b; \mathbf{s}^v)]$, and so
$CR \subseteq \sim K_i[h_i^v(b) > h_i^v(b; \mathbf{s}^v)]$. Since b is defined so that b^v is optimal given $b^{>v}$, $\sim K_i[h_i^v(b) > h_i^v(b; \mathbf{s}^v) = \sim K_i[\mathbf{s}^v \neq b^v]$. By definition, $\sim K_i[\mathbf{s}^v \neq b^v] = \sim K_i \sim BI^v$, and $\sim K_i \sim BI^v = \sim K_i \sim K_i BI^v$ by Lemma 10.

Now the proof diverges from Aumann's, but only in that his next step was an equality whereas mine is a subset relation:

By Lemma 8*, $\sim K_i \sim K_i BI^v \subseteq \sim\sim K_i BI^v = K_i BI^v$. By Lemmas 9 and 10, $K_i BI^v = BI^v$, and so $CR \subseteq BI^v$. Since this is true for all nodes v, $CR \subseteq BI$. □

5 The T-based Aumann-style framework, and proof

5.1 The difference between S4 and T

The above shows that S4-A's Theorem, an analog of Aumann's Theorem, can be proved in the canonical S4 semantic knowledge system. Thus the substance of Aumann's Theorem, that if there is common knowledge of rationality in a perfect information game then the backward inductive outcome is realized, does not depend either on a partitional structure of knowledge or on the axiom 5 of S5. The next question is whether the axioms of S4 might also include more than is necessary to prove such a theorem, and therefore whether the axioms of T might be sufficient. The observant reader will already suspect an affirmative answer to this question. The difference between the modal logical systems S4 and T is that T lacks the axiom $\Box\phi \to \Box\Box\phi$; this axiom is present in the S4 system above in the list of tautologies of the system, in the form of (g) $k_i f \to k_i k_i f$. It is included because it replaces the tautology $\neg k_i f \to k_i \neg k_i f$, corresponding to the axiom 5 of S5, which was present in Aumann's development of the syntax for knowledge in [4]. Yet

(g) is used neither in proving any of the lemmas nor in proving S4-A's Theorem. Therefore the T-based system resulting from simply removing (g) from the list of tautologies and leaving the rest of the syntax and the canonical construction of the semantic knowledge system intact results in a semantic knowledge system in which all of the above lemmas and an analog to S4-A's Theorem can still be proved.

5.2 A T-based knowledge system

Let $\mathfrak{S}(\mathfrak{G}, \mathfrak{X})$ be the syntax for the language $\mathcal{L}$ as above. Remove (g) $k_i f \to k_i k_i f$ from the list of tautologies. Then define a possible world ω^T as a list $\mathfrak{L}^T$ of formulae that is logically closed, complete, coherent, contains all tautologies, and in which the set of tautologies is epistemically closed. The set of all such ω^T is called Ω^T. Define the ken of an agent i at ω^T as $K_i^{T\bullet}(\omega^T) := \{f \in \omega^T | f = k_i g \text{ for some } g\}$. Then $\mathfrak{C}^T = (\Omega^T, \mathfrak{G}, (K_i^{T\bullet})_{i \in N})$ is the canonical (T-based) semantic knowledge system for the game $\mathfrak{G}$. Other definitions can be stated as for the system $\mathfrak{C}$; there is no substantial difference, but only a difference in notation.

5.3 A T-based theorem and proof

T-A's Theorem (Theorem 3): Let $\mathfrak{G}$ be a perfect information game, $\mathfrak{S}$ a syntax, and $\mathfrak{C}^T$ the canonical semantic knowledge system as defined above. Then $CR \subseteq BI$.

Proof. Just as the proof of S4-A's Theorem above. □

6 Conclusion

The above proves that Aumann's Theorem does not require a strong S5-definition of knowledge; common knowledge of rationality implies backward induction in perfect information games in systems with weaker definitions of knowledge. The systems I have constructed are based on S4 and T, which have the advantage compared to S5 of incorporating fewer assumptions about players' knowledge. There are still weaker systems of modal logic which would make fewer assumptions about knowledge, but these systems would in fact assume too little about knowledge and are not suitable for defining it.

This is primarily because systems weaker than T do not include the axiom T, which in the context of knowledge says that if an agent knows something, then it is true. Unlike the axioms 4 and 5, the axiom T is essential to any definition of knowledge, since truth is the main feature distinguishing belief in a proposition from knowledge of it. Therefore by proving that analogs of Aumann's Theorem hold when knowledge is based on S4 or T I have proved essentially that Aumann's Theorem holds under any reasonable definition of knowledge. This discovery strengthens the theorem.

A natural extension of this project would be to investigate the influence of S4- or T-based definitions of knowledge on systems with both a knowledge operator and a hypothetical knowledge operator, as discussed by Horacio Arló-Costa and Cristina Bicchieri [1], Joseph Halpern [10], and Samet [14]. Such systems are extensions of the standard Aumann-type system which have the advantage of explicitly representing the hypothetical reasoning that game players engage in; such reasoning takes the form of determining what an agent would know in the event that some sequence of game actions took place, and using this hypothetical knowledge to select optimal strategies. It would be worthwhile to extend the current project to such sophisticated systems.

The reader interested in the connection between axioms of modal logic and the semantic knowledge systems employed by Aumann and others may find useful a recent paper by Samet, *S5 Knowledge Without Partitions* [16]. In this paper, Samet makes the surprising observation that having an S5-based knowledge operator does not entail a partitional structure of the state space unless the knowledge operator is defined for all subsets of possible worlds or the state space is finite. If the knowledge operator partitions the state space, however, then it must obey the axioms of S5. Also of interest is Michael Bacharach's *Some extensions of a claim of Aumann in an axiomatic model of knowledge* [7], which contains some philosophical discussion about the universality in economics of the assumption that agents have information partitions and the justification (or lack thereof) for employing strong epistemic models incorporating such assumptions.

7 Appendix: Proofs of the lemmas

Lemma 9. $K_iG \subseteq G$

Proof. $K_iG = \{\omega | k_ig \in \omega\} = \{\omega | (k_ig) \wedge (k_ig \to g) \in \omega\}$, since $k_ig \to g$ is a tautology and is therefore in all ω. Then since each ω is a list closed under modus ponens, $\{\omega|(k_i g) \wedge (k_i g \rightarrow g) \in \omega\} =$

$\{\omega|(k_i g) \wedge (k_i g \rightarrow g) \wedge (g) \in \omega\} \subseteq \{\omega|g \in \omega\} = G$. □

Lemma 8*. $K_i \sim K_i F \subseteq \sim K_i F$

Proof. By Lemma 9, with $G = \sim K_i E$. □

Lemma 7. $CF \subseteq F$

Proof. $CF = \bigcap_{n\in\mathbb{N}} E^n F = \bigcap_{n\in\mathbb{N}} \{\omega|e^n f \in \omega\}$. From tautologies (k), (l), and (e) another tautology, $e^n f \rightarrow f$, is derivable, and this fact in conjunction with the closure of the ω's under modus ponens yields that $\bigcap_{n\in\mathbb{N}} \{\omega|e^n f \in \omega\} = \bigcap_{n\in\mathbb{N}} \{\omega|(e^n f) \wedge (e^n f \rightarrow f) \wedge (f) \in \omega\} \subseteq \bigcap_{n\in\mathbb{N}} \{\omega|f \in \omega\} = \{\omega|f \in \omega\} = F$. □

Lemma 6. $K_i G \cap K_i F = K_i(G \cap F)$

Proof. $K_i G \cap K_i F = \{\omega|k_i g \in \omega\} \cap \{\omega|k_i f \in \omega\}$. $g \rightarrow (f \rightarrow (g \wedge f)) \in \omega$ for all ω, since it is a propositional tautology, and since the set of tautologies is epistemically closed, also $k_i(g \rightarrow (f \rightarrow (g \wedge f))) \in \omega$ for all ω. By tautology (f), then, $k_i f \rightarrow k_i(g \rightarrow g \wedge f) \in \omega$ for all ω, and by propositional logic and (f) again, also for all ω, $k_i f \rightarrow (k_i g \rightarrow k_i g \wedge f) \in \omega$. Then $K_i G \cap K_i F = \{\omega|k_i g \wedge k_i f \in \omega\} = \{\omega|(k_i g \wedge k_i f) \wedge (k_i f \rightarrow (k_i g \rightarrow k_i g \wedge f)) \in \omega\}$, and since all ω are logically closed, this is equal to $\{\omega|k_i f \wedge g \in \omega\} = K_i(G \cap F)$. □

The following corollary to Lemma 6 is needed for the proof of Lemma 4:

Corollary 1. *For any $|X| \in \mathbb{N}$, where X is any index set,* $\bigcap_{x\in X} K_i F_x = K_i \bigcap_{x\in X} F_x$.

Proof. By induction on $|X|$. For the base case, Lemma 6 shows that when $|X| = 2$, $\bigcap_{x\in X} K_i F_x = K_i \bigcap_{x\in X} F_x$ since $K_i G_1 \cap K_i G_2 = K_i(G_1 \cap G_2)$. Now suppose that for some $|X| = k$, $\bigcap_{x\in X} K_i F_x = K_i \bigcap_{x\in X} F_x$. It remains to be show that the equality $\bigcap_{y\in X^\star} K_i F_y = K_i \bigcap_{y\in X^\star} F_y$ holds for $X^\star = X \cup \{z\}$, where $|X^\star| = k + 1$.

$\bigcap_{y\in X^\star} K_iF_y = (\bigcap_{x\in X} K_iF_x)\cap(K_iF_z)$. By the induction hypothesis, this is equal to $(K_i(\bigcap_{x\in X} F_x))\cap K_iF_z$. Since this reduces the problem to the base case, this is equal to $K_i((\bigcap_{x\in X} F_x)\cap(F_z)) = K_i\bigcap_{y\in X^\star} F_y$. Then for any $|X|$, $\bigcap_{x\in X} K_iF_x = K_i\bigcap_{x\in X} F_x$. □

Lemma 4. $CF = K_iCF$

Proof. By Lemma 9, $K_iCF \subseteq CF$.

For the other inclusion: $CF = \bigcap_{n\in\mathbb{N}} E^nF = \bigcap_{n\in\mathbb{N}} \{\omega|e^nf \in \omega\}$. By the tautology (n), this is equal to $\bigcap_{n\in\mathbb{N}} \{\omega|\bigwedge_{i\in N} k_ie^{n-1}f \in \omega\} = \bigcap_{i\in N}\bigcap_{n\in\mathbb{N}} K_iE^{n-1}F \subseteq \bigcap_{n\in\mathbb{N}} K_iE^{n-1}F$. By Corollary 1, this is equal to $K_i\bigcap_{n\in\mathbb{N}} E^{n-1}F$. Since $|\mathbb{N}|-1 = |\mathbb{N}|$, $\bigcap_{n\in\mathbb{N}} E^{n-1}F = \bigcap_{n\in\mathbb{N}} E^nF$. Therefore $K_i\bigcap_{n\in\mathbb{N}} E^{n-1}F = K_i\bigcap_{n\in\mathbb{N}} E^nF = K_iCF$. Then $CF \subseteq K_iCF$.

It follows that $CF = K_iCF$. □

Lemma 5. If $G \subseteq F$ then $K_iG \subseteq K_iF$

Proof. If $G \subseteq F$ then $\{\omega|g \in \omega\} \subseteq \{\omega|f \in \omega\}$. Then $g \in \omega$ implies $f \in \omega$, and so $g \to f \in \omega$ for all ω since the ω are complete and coherent. Then $\{\omega|g \to f \in \omega\} = \Omega$, so $g \to f$ is a tautology. Since the tautologies are epistemically closed, then $\{\omega|k_i(g \to f) \in \omega\} = \Omega$. Then by tautology (f), the distribution axiom, $\{\omega|k_ig \to k_if \in \omega\} = \Omega$, so $K_iG \subseteq K_iF$. □

Lemma 10. $BI^v \subseteq K_iBI^v$ for all nodes v of player i.

Proof. Suppose that for any node v, player i is to act at v. Then the event that i makes the backward inductive choice at v, $BI^v = [\mathbf{s}_i^v = b^v]$. Call $B_i \subseteq S_i$ the subset of i's strategies at which $\mathbf{s}_i^v = b^v$. Then the event that $\mathbf{s}_i^v = b^v$ is the set of worlds at which a strategy $t_i \in B_i$ is chosen. So where $\mathbf{s}_i : \{(\mathrm{S}_i^x)_{1\leq x\leq|S_i|}\} \to S_i$ is the function from strategy choice-events to strategies, $[\mathbf{s}_i^v = b^v] = \bigcup_{t_i\in B_i} \mathbf{s}_i^{-1}(t_i)$. Each $\mathbf{s}_i^{-1}(t_i) = \mathrm{S}_i^z$ for some $1 \leq z \leq |S_i|$, where $\mathrm{S}_i^z = \{\omega|\mathrm{s}_i^z \in \omega\}$. Then by tautologies (e) and (h) and the fact that each ω is logically closed, $\mathrm{S}_i^z = \{\omega|\mathrm{s}_i^z \wedge \mathrm{s}_i^z \leftrightarrow k_i\mathrm{s}_i^z \in \omega\} = \{\omega|k_i\mathrm{s}_i^z \in \omega\}$.

Therefore $\mathrm{S}_i^z = K_i\mathrm{S}_i^z$. Since each S_i^z is an event that i's strategy makes the backward inductive choice at the node v, $\mathrm{S}_i^z \subseteq BI^v$. Then by Lemma 5, since $\mathrm{S}_i^z = K_i\mathrm{S}_i^z$, $K_i\mathrm{S}_i^z \subseteq K_iBI^v$ for all z. Therefore $BI^v = \bigcup_z \mathrm{S}_i^z = \bigcup_z K_i\mathrm{S}_i^z \subseteq K_iBI^v$ (where $z = \{r | \mathbf{s}_i^{-1}(t_i) = \mathrm{S}_i^r$ for some $t_i \in B_i\}$). Then $BI^v \subseteq K_iBI^v$ for all nodes v of each player i. □

References

[1] Horacio Arló-Costa and Cristina Bicchieri. *Knowing and supposing in games of perfect information.* In Studia Logica, pages 353-373, 2007.

[2] Robert Aumann. *Agreeing to Disagree.* In The Annals of Statistics, Vol. 4, No. 6, pages 1236-1239, 1976.

[3] Robert Aumann. *Backward Induction and Common Knowledge of Rationality.* In Games and Economic Behavior 8, pages 6-19, 1995.

[4] Robert Aumann. *Interactive Epistemology I: Knowledge.* In International Journal of Game Theory, pages 263-300, 1999.

[5] Robert Aumann. *Interactive Epistemology II: Probability.* In International Journal of Game Theory, pages 301-314, 1999.

[6] Robert Aumann. *On the Centipede game.* In Games and Economic Behavior, pages 97-105, 1998.

[7] M. Bacharach. *Some extensions of a claim of Aumann in an axiomatic model of knowledge.* In Journal of Economic Theory, pages 167-190, 1985.

[8] R. Fagin, JY Halpern, M. Moses, and MY Vardi. *Reasoning about knowledge.* MIT Press, Cambridge, 1995.

[9] John Geanakoplos. *Game theory without partitions, and applications to speculation and consensus.* Cowles Foundation Discussion Paper NO. 914, 1989.

[10] Joseph Y. Halpern. *Hypothetical knowledge and counterfactual reasoning.* In International Journal of Game Theory, pages 315-330, 1999.

[11] Jaakko Hintikka. *Knowledge and belief.* Cornell University Press, Ithaca, 1962.

[12] P. Milgrom and N. Stokey. *Information, trade and common knowledge.* In Journal of Economic Theory, pages 17-27, 1982.

[13] Ariel Rubinstein and Asher Wolinsky. *A comment on the logic of "Agreeing to disagree" type results*, mimeo, 1989.

[14] Dov Samet. *Hypothetical knowledge and games with perfect information.* In Games and Economic Behavior, pages 230-251, 1996.

[15] Dov Samet. *Ignoring ignorance and agreeing to disagree.* In Journal of Economic Theory, pages 190-207, 1990.

[16] Dov Samet. *S5 Knowledge without partitions.* In Synthese, pages 145-155, 2010.

[17] Hans van Ditmarsch, Wiebe van der Hoek, and Barteld Kooi. *Dynamic Epistemic Logic.* Springer, Dordrecht, 2008.