A LOGIC FOR CHANGING BELIEFS WITH APPLICATIONS TO REASONING ABOUT CHOICE AND GAMES

Arnis Vilks
Handelshochschule Leipzig
- Leipzig Graduate School of Management -
Jahnallee 59, 04109 Leipzig, Germany
vilks@microec.hhl.de

ABSTRACT. We suggest a belief set semantics for an epistemic logic with a sequence K_0, K_1, K_2,... of belief operators. The key idea consists in taking a model for the language to include a distinct "belief set" B_t for each point of time t, and to say that $K_t\varphi$ is satisfied in such a model iff $\varphi \in B_t$. By imposing certain conditions on the admissible models we arrive at a logic which resembles the normal modal system K45, but does not require that introspection and knowledge generalization are instantanuous. Nevertheless, a delayed version of knowledge generalization holds. As a first application of the logic we suggest a solution to the problem of self-knowledge of one's options and rationality, as discussed by Schick, 1979, and Levi, 1991. As a second application, we define a multi-agent version of the above logic, and give an epistemically explicit reformulation of the backward-induction argument in Vilks, 1997. We also use the multi-agent framework to explain that common knowledge of rationality and the game need not entail backward induction, if the players lack mutual knowledge of each other's reasoning processes.

This version: May 1999. An earlier version has been presented at the LOFT3 conference in Torino, and the conference on „Logic, Game Theory, and Social Choice" in Oisterwijk, and has appeared in the proceedings volume of the latter, ed. by H. de Swart, and published by Tilburg University Press 1999. I have benefitted from comments on earlier versions by Jelle Gerbrandy, Philippe Mongin, and Isaac Levi.

1. INTRODUCTION

In the literature, quite different approaches have been used to model beliefs formally. In the "belief revision" literature (e.g. Gärdenfors 1988), an "epistemic state" is typically identified with the set of formulas (of some formal language) which are believed by the agent under consideration. Attention is then focused on how this belief set will (or should) change when some new piece of information is received and the formula representing this new information must be somehow combined with the old belief set to arrive at a new one. In this literature, however, it is typically not made explicit that some beliefs may themselves be about beliefs.

By contrast, epistemic logic (e.g. Fagin, Halpern, Moses, and Vardi 1995) explicitly uses a formal language with belief operators and is thus well-suited for many applications where beliefs about beliefs are of importance. However, with very few exceptions (e.g., Battigalli and Bonanno 1996) the epistemic logic literature has just one belief (or knowledge) operator per agent, and is thus unable to express (in the object-language) that something is not known now, but is known later. Statements of this kind, however, are of central importance not only when beliefs change through "exogenous" information, but also when beliefs change "endogenously", i.e. by reasoning.

In this paper, we suggest a combination of the "belief set" approach with an epistemic logic that has a sequence K_0, K_1, K_2,... of belief operators. The key idea consists in taking a model for the language to include a distinct "belief set" B_t for each point of time t, and to say that $K_t\varphi$ is satisfied in such a model iff $\varphi \in B_t$. (A somewhat similar semantics underlies the "autoepistemic logic" of Moore, 1985, and Konolige, 1988.) By imposing certain rather perspicuous conditions on the admissible models we arrive at a logic which resembles K45, but does not require that introspection and knowledge generalization are instantanuous. (For K45 and other normal systems of modal logic cf. Chellas, 1980.)

Most of the paper is limited to the one-agent case, but we indicate the natural generalization to the multi-agent case, and applications to game theory in the final section.

2. THE FORMAL LANGUAGE AND BASIC SEMANTIC DEFINITION

We begin by defining the formal language: Its alphabet consists of: primitive propositions: p, q, r, ...; connectives: $\neg$, $\wedge$; belief operators: K_0, K_1, K_2, K_3,...

Belief operators K_t are thought of as referring to consecutive points (or periods) of time.

The well-formed formulas (wffs) are defined inductively in the usual manner: any primitive proposition is a wff; if φ and ψ are wffs, then $\neg\varphi$, $\wedge\varphi\psi$, and $K_t\varphi$ are wffs. We write $(\varphi\wedge\psi)$ instead of $\wedge\varphi\psi$, define $\vee$, $\rightarrow$, and $\leftrightarrow$ as usual, and adopt standard bracketing conventions. The set of all wffs is denoted by $\mathfrak{L}$, the set of primitive propositions by Φ. Given a set of wffs $S\subset\mathfrak{L}$, we write $K_t(S):=\{K_t\varphi|\ \varphi\in S\}$, and $\neg S:=\{\neg\varphi|\varphi\in S\}$.

A truth valuation for $\mathfrak{L}$ is a function $w:\mathfrak{L}\rightarrow\{TRUE, FALSE\}$ such that $w(\neg\varphi)\neq w(\varphi)$ and $w(\varphi\wedge\psi)=TRUE$ iff $w(\varphi)=w(\psi)=TRUE$. A wff φ is a tautology, if $w(\varphi)=TRUE$ for all truth valuations w, it is a tautological consequence of a set $S\subset\mathfrak{L}$, if $w(\varphi)=TRUE$ for all truth valuations w such that $w(\psi)=TRUE$ for all $\psi\in S$. For $S\subset\mathfrak{L}$, the set of all tautological consequences of S is denoted by Cn(S).

Definition 1. A model $M=(v,(B_t))$ for $\mathfrak{L}$ consists of a function $v:\Phi\rightarrow\{TRUE, FALSE\}$ and a sequence $B_0, B_1, B_2,...$ of sets $B_t\subset\mathfrak{L}$.

The set B_t is thought of as consisting of those formulas that are believed at time t.

Definition 2. The model $M=(v,(B_t))$ satisfies the wff φ, symbolically $M\models\varphi$, according to the following conditions:

For $\varphi\in\Phi$, $M\models\varphi$ iff $v(\varphi)=TRUE$,

$M\models\neg\varphi$ iff not: $M\models\varphi$,

$M\models\varphi\wedge\psi$ iff $M\models\varphi$ and $M\models\psi$,

$M\models K_t\varphi$ iff $\varphi\in B_t$ (for $t\in\mathbb{N}$).

Definition 3. A wff is valid in some class C of models, if it is satisfied by all $M \in C$.

It is easily seen that a wff is valid in the class of all models iff it is a tautology. However, by defining narrower classes of models, we arrive at more interesting notions of validity.

3. A SPECIAL CASE: DELAYED INTROSPECTION

In the remainder of we consider a particular class of models.

Definition 4. A model is said to be

(a) propositionally closed iff $B_t = Cn(B_t)$ for all t,

(b) with perfect memory iff $B_t \subset B_{t+1}$ for all t,

(c) with delayed positive introspection iff $K_t(B_t) \subset B_{t+1}$ for all t,

(d) with delayed negative introspection iff $\neg K_t(\mathfrak{L} \setminus B_t) \subset B_{t+1}$ for all t.

The class of all models which have all of the properties (a) through (d), will be denoted by C_0.

Next we specify an axiomatic system that turns out to be sound and complete w.r.t. C_0. This axiomatic system AX has *modus ponens* as its sole rule of inference, and the following six axiom schemes:

(A1) φ, whenever φ is a tautology,

(A2) $K_t\varphi$, whenever φ is a tautology,

(A3) $K_t\varphi \wedge K_t(\varphi \rightarrow \psi) \rightarrow K_t\psi$,

(A4) $K_t\varphi \rightarrow K_{t+1}\varphi$,

(A5) $K_t\varphi \rightarrow K_{t+1}K_t\varphi$,

(A6) $\neg K_t\varphi \rightarrow K_{t+1}\neg K_t\varphi$.

Theorem 1. (Soundness and completeness.)

A wff of $\mathfrak{L}$ is a theorem of AX, iff it is valid in C_0.

Proof. Soundness is straightforward. For completeness, it suffices to show that for any maximal AX-consistent set $F \subset \mathfrak{L}$ there is a model M such that $\psi \in F$ iff $M \models \psi$. (Fagin, Halpern, Moses, and Vardi, 1995, pp. 48-54.)

Let F be a maximal AX-consistent set. To define the appropriate model $M=(v,(B_t))$, let $v(\psi)$=TRUE iff $\psi \in F$ for $\psi \in \Phi$, and define $B_t := \{\varphi \in \mathfrak{L} | K_t\varphi \in F\}$. As F is maximal AX-consistent, it is easy to check that M is propositionally closed, and has perfect memory, and both kinds of delayed introspection. Moreover, $M \models \psi$ iff $\psi \in F$. This completes the proof. ☺

An important difference between AX and K45 is that the rule of epistemization (or "knowledge generalization") does not hold in AX. However, AX does have a "delayed" version of it: Every AX-theorem φ is believed from that time t onwards, at which all K_τ-operators which appear in φ either have $\tau \leq t-1$ or appear within the scope of such an operator. To state this formally, we recursively define sublanguages $\mathfrak{L}_t$ of $\mathfrak{L}$ as follows:

Definition 5. $\mathfrak{L}_0$ is the smallest subset of $\mathfrak{L}$ which contains $\Phi \cup K_0(\mathfrak{L})$, and is closed with respect to applications of $\neg$ and $\wedge$;

for $t \geq 1$: $\mathfrak{L}_t$ is the smallest subset of $\mathfrak{L}$ which contains $\mathfrak{L}_{t-1} \cup K_t(\mathfrak{L})$, and is closed with respect to applications of $\neg$ and $\wedge$.

With this definition we can express "delayed" epistemization as follows:

Theorem 2. If $\varphi \in \mathfrak{L}_t$ and $AX \vdash \varphi$, then $AX \vdash K_{t+1}\varphi$.

Proof. We exploit completeness of AX w.r.t. the class of all propositionally closed models with perfect memory, positive and negative introspection. In this proof, we write $\models \varphi$ for validity w.r.t this class. Assume that for some model $M=(v,(B_t))$ of this class $M \models K_{t+1}\varphi$ does not hold for $\varphi \in \mathfrak{L}_t$. This implies that $\varphi \notin Cn(B_t \cup K_t(B_t) \cup \neg K_t(\mathfrak{L} \setminus B_t))$. By definition of Cn this implies that there is some truth valuation w such that $w(\varphi)$=FALSE, but $w(\varphi)$=TRUE for all $\psi \in B_t \cup K_t(B_t) \cup \neg K_t(\mathfrak{L} \setminus B_t)$. Define $v': \Phi \to \{TRUE, FALSE\}$ to be the restriction of w to Φ, i.e. $v'(\psi)=w(\psi)$ for all $\psi \in \Phi$. Clearly, the model $M':=(v',(B_t))$ is still of the relevant class, and to

complete the proof we just have to show that M'$\models\neg\varphi$. To do so, we show that for any $\psi\in\mathfrak{L}_t$ we have M'$\models\psi$ iff w(ψ)=TRUE. This is clear for $\psi\in\Phi$. If $\psi=K_\tau\chi$ for some $\tau\leq t$, we have that M'$\models\psi$ iff $\chi\in B_\tau$. If $\chi\in B_\tau$, then $\psi\in K_\tau(B_\tau)$ and hence w(ψ)=TRUE. If $\chi\notin B_\tau$, i.e. $\chi\in\mathfrak{L}\setminus B_\tau$, then w($\neg K_\tau\chi$)=TRUE, hence w($\psi$)=FALSE. Assuming that $\psi=\neg\chi$ or $\psi=\chi\wedge\omega$, and that M'$\models\omega$ iff w(ω)=TRUE has already been proved for those $\omega\in\mathfrak{L}_t$ which are shorter than ψ, it is straightforward to conclude that, for any $\psi\in\mathfrak{L}_t$, we have M'$\models\psi$ iff w(ψ)=TRUE. As w(φ)=FALSE, it follows that M' does not satisfy φ. Thus φ cannot be valid in the relevant class of models, and by the soundness of AX, it cannot be an AX-theorem. ☺

One can apply delayed epistemization to the axioms of AX to get, e.g., AX $\vdash K_{t+1}(K_t\varphi \wedge K_t(\varphi\rightarrow\psi)\rightarrow K_t\psi)$; AX $\vdash K_{t+2}(K_t\varphi \rightarrow K_{t+1}\varphi)$. One can also prove AX $\vdash K_{t+1}(K_t\varphi \rightarrow \varphi)$.

4. An Open Problem

If M=(v,(B_t))$\in C_0$, we have:

$$Cn(B_t\cup K_t(B_t)\cup\neg K_t(\mathfrak{L}\setminus B_t)) \subset B_{t+1}.$$

An interesting subclass of C_0 models is defined by requiring (for all t) that

$$Cn(B_t\cup K_t(B_t)\cup\neg K_t(\mathfrak{L}\setminus B_t)) = B_{t+1}.$$

In models with this property, all belief changes are due to reasoning. An open question is the following: Is it possible to axiomatize validity with respect to this subclass of models? If so, what does the axiom system look like?

5. An Application to Reasoning About Choice: The Problem of Self-knowledge of Options and Rationality

There is a certain tension between three assumptions which are often made in Game Theory. When a player reasons about how to play a given game Γ, it seems natural to assume (1) that initially he considers all moves of Γ as possible, and (2) that the players have knowledge from which they can deduce, and thus know on reflection that some moves will not be taken. Moreover, it is standard practice in both epistemic logic and Game Theory to (3) identify "the

agent considers φ as possible" with "the agent does not know ¬φ" (e.g., Hintikka, 1962; Lenzen, 1980; Binmore, 1992; Fagin et al., 1995; Samet, 1996; Dekel and Gul, 1997).

The tension between these assumptions arises even in the one-agent case (see Schick, 1979, and Levi, 1991), and we want to argue in this section that the tension even turns into a formal inconsistency, as long as standard „static" epistemic logic is used for modelling belief.

Consider an agent who has to choose between action 1 and action 2, whereof he prefers the former. Let a_1 stand for "the agent takes action 1", a_2 for "the agent takes action 2", and $a_1 \succ a_2$ for „the agent prefers a_1 to a_2" (we assume this to be a primitive wff of our language). We can then describe the situation by the following formula S:

(S) $$(a_1 \vee a_2) \wedge \neg(a_1 \wedge a_2) \wedge (a_1 \succ a_2)$$

Should we expect a rational agent to choose action 1, whenever he is in situation S? If the agent might not know how to carry out action 1, or if might be convinced that he is physically unable to carry out action1 (none of which is precluded by S being satisfied), the answer depends on the notion of rationality we want to employ. A mild notion of rationality would require only that a most preferred action among those considered possible (by the agent) must be carried out. In accordance with (3) above, and using „K" as a belief operator, we could express the assumption that both actions are considered possible by the agent by the following formula P:

(P) $$\neg K(\neg a_1) \wedge \neg K(\neg a_2)$$

A mild condition of rationality could then be expressed by the formula R:

(R) $$S \wedge P \rightarrow a_1$$

Obviously, S∧P∧R tautologically implies a_1, and it is an easy exercise in modal logic to show that this formula S∧P∧R is consistent in, for instance, the normal system K45. However, it is just as easy to see the following:

Observation 1. $P \wedge K(S \wedge R)$ is inconsistent in K45 (and a fortiori in S5).

The formula $P \wedge K(S \wedge R)$ just seems to express that the agent considers both actions possible, that he knows them to be mutually exclusive and collectively exhaustive, and that he also knows about his own preferences and rationality. Thevertheless, because of the instantanuous negative introspection of K45, his beliefs allow him to deduce - and he therefore believes - that $\neg a_2$. But this plainly contradicts $\neg K(\neg a_2)$, which is assumed in P.

One might perhaps regard this problem as an argument against negative introspection. However, even in the weakest "normal" system K, which has no introspection axioms, a similar difficulty arises: The formula $S \wedge R \wedge P \wedge K(S \wedge R \wedge P)$ is inconsistent in K, while it just seems to express an innocuous implication of the common knowledge assumptions typically used in game theory. Thus, difficulties of this sort seem hard to avoid in a static epistemic logic which has the distribution axiom and instantanuous "epistemization" as a rule of inference.

The source of these difficulties seems to lie in the fact that (at least) two slightly different notions of belief are needed in theorizing about choice and rationality: One the one hand, one needs the *initial* beliefs of the agent which he holds before he has carried out all his reasoning, and on the other hand, one needs his beliefs *after* he has drawn all the relevant conclusions from his initial beliefs. An epistemic logic with just one belief operator (per agent) simply lacks the formal means required to represent these different stages of the reasoning process.

By contrast, our logic AX developed above allows one to resolve the problem in a very direct way. Define the wffs P_0 and R_0 to be the result of simply replacing the operator K by K_0 in P and R respectively. It is then easy to see the following.

Observation 2. $P_0 \wedge K_0(S \wedge R_0)$ is consistent in AX.

Moreover, $P_0 \wedge K_0(S \wedge R_0)$ implies (in AX) $K_1(\neg a_2) \wedge K_1(\neg K_0 \neg a_2)$: On reflection, the agent knows he will not take action 2, but remembers that he previously considered it possible that he would.

6. AN APPLICATION TO REASONING ABOUT GAMES: BACKWARDS INDUCTION IN THE CENTIPEDE

As a second application, we define a multi-agent version of the above logic, and indicate how to give an epistemically explicit reformulation of the backward-induction argument in Vilks (1997). We also use the multi-agent framework to explain that common knowledge of rationality and the game need not entail backward induction, if the players lack mutual knowledge of each other's reasoning processes. We limit ourselves to the four-legged version of the much-discussed Centipede Game (see, e.g., Aumann, 1998) as an example (and accordingly to two agents).

The formal language used now, $\mathfrak{L}$', differs from $\mathfrak{L}$ only by having a sequence K_0^i, K_1^i, K_2^i,... of belief operators for each agent i. Definitions 1 and 2 carry over directly to the two-agent case:

Definition 1'. A model $M=(v,(B_t^i))$ for $\mathfrak{L}$' consists of a function $v:\Phi\rightarrow\{TRUE, FALSE\}$ and, for each agent $i\in\{I, II\}$, a sequence B_0^i, B_1^i, B_2^i,... of sets $B_t^i\subset\mathfrak{L}$'.

Definition 2'. The model $M=(v,(B_t^i))$ satisfies the wff φ, symbolically $M\vDash\varphi$, according to the following conditions:

For $\varphi\in\Phi$, $M\vDash\varphi$ iff $v(\varphi)=TRUE$,

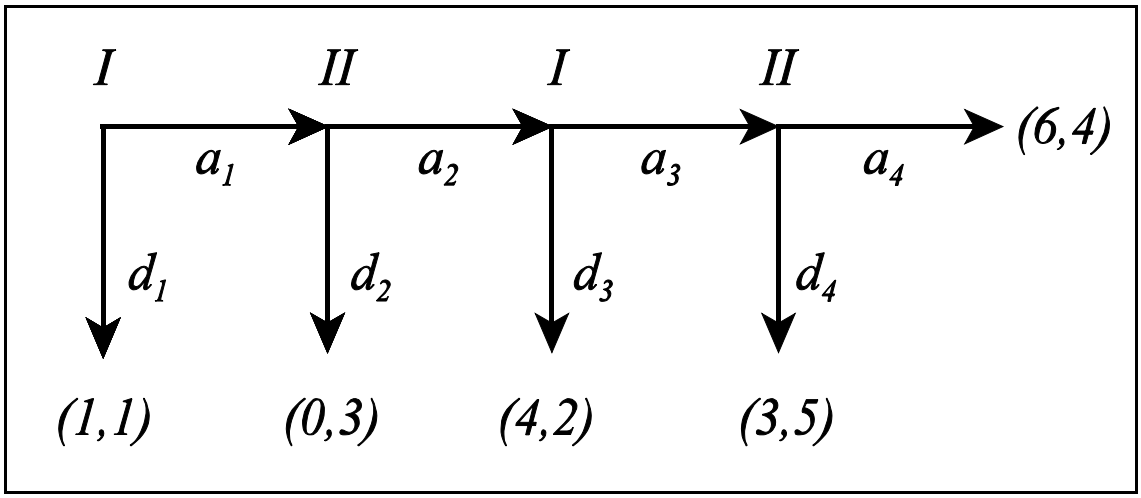
$M\vDash\neg\varphi$ iff not: $M\vDash\varphi$,

$M\vDash\varphi\wedge\psi$ iff $M\vDash\varphi$ and $M\vDash\psi$,

$M\vDash K_t^i\varphi$ iff $\varphi\in B_t^i$ (for $t\in\mathbb{N}$, $i\in\{I,II\}$).

The class of models in the sense of Definition 1', where both $(v,(B_t^I))$, and $(v,(B_t^{II}))$ belong to C_0, will be denoted by C_0'. An axiomatization of validity in C_0' is analogous to AX above, and we denote it by AX'.

To describe the (four-legged) centipede, let $\mathfrak{L}$' be such that $A:=\{a_1,d_1,...,a_4,d_4\}\subset\Phi$, and $B:=\{\varphi\succ_i\psi\mid \varphi\in A, \psi\in A, i\in\{I,II\}\}\subset\Phi$, with the interpretation that, e.g., a_2 stands for "at the second decision node, player 2 moves across", d_3 stands for "at the third decision node, player 1 moves down", $d_1\succ_I d_2$ stands for "player I prefers d_1 to d_2", etc.



The rules of the four-legged centipede can now be expressed by the following formula:

(G) $(a_1 \wedge \neg d_1 \wedge a_2 \wedge \neg d_2 \wedge a_3 \wedge \neg d_3 \wedge a_4 \wedge \neg d_4) \vee (a_1 \wedge \neg d_1 \wedge a_2 \wedge \neg d_2 \wedge a_3 \wedge \neg d_3 \wedge \neg a_4 \wedge d_4)$
$\vee (a_1 \wedge \neg d_1 \wedge a_2 \wedge \neg d_2 \wedge \neg a_3 \wedge d_3 \wedge \neg a_4 \wedge \neg d_4) \vee (a_1 \wedge \neg d_1 \wedge \neg a_2 \wedge d_2 \wedge \neg a_3 \wedge \neg d_3 \wedge \neg a_4 \wedge \neg d_4)$
$\vee (\neg a_1 \wedge d_1 \wedge \neg a_2 \wedge \neg d_2 \wedge \neg a_3 \wedge \neg d_3 \wedge \neg a_4 \wedge \neg d_4)$,

and the players' preferences satisfy (at least):

(U) $(d_4 \succ_{II} a_4) \wedge (d_3 \succ_I d_4) \wedge (d_2 \succ_{II} d_3) \wedge (d_1 \succ_I d_2)$

If players know the structure of the game as expressed by G∧U, it seems they should, if rational, also satisfy the following conditions (for convenience, we set $d_5 := a_4$):

($R_{i,j}^t$) $K_t^i(a_j \leftrightarrow d_{j+1}) \wedge \neg K_t^i(\neg d_j) \rightarrow \neg a_j$ (for j=1,3, and i=I; or j=2,4, and i=II)

We will make use of

$$R_T := \bigwedge_{t=0}^{T} (R_{I,1}^t \wedge R_{II,2}^t \wedge R_{I,3}^t \wedge R_{II,4}^t)$$

To ensure that the BI moves cannot be ruled out a priori - on the basis of additional knowledge the players might have - it is moreover natural to assume:

(CP^t) $\quad (\neg K_t^{I}\neg d_1) \wedge (a_1 \rightarrow \neg K_t^{II}\neg d_2) \wedge (a_2 \rightarrow \neg K_t^{I}\neg d_3) \wedge (a_3 \rightarrow \neg K_t^{II}\neg d_4)$

(The abbrevation „CP^t“ for „conditional possibility at stage t“ has been chosen to indicate that the BI moves are assumed to be doxastically possible conditional upon the relevant decision node being reached.)

We will make use of

$$CP_T := \bigwedge_{t=0}^{T} CP^t$$

We define initial mutual knowledge of φ by $E_0(\varphi) := K_0^{I}(\varphi) \wedge K_0^{II}(\varphi)$, and mutual knowledge at stage t of the reasoning by $E_t(\varphi) := K_t^{I}(\varphi) \wedge K_t^{II}(\varphi)$; moreover, m-th order iterated mutual knowledge at stage t is expressed by $E_t^{m}(\varphi) := E_t(E_t^{m-1}(\varphi))$, where, of course, $E_t^{1}(\varphi) := E_t(\varphi)$.

Now, we can express 3rd order iterated initial mutual knowledge of the game, initial rationality, and initial conditional possibility of BI moves by

$$G \wedge U \wedge R_0 \wedge CP_0 \wedge E_0^{3}(G \wedge U \wedge R_0 \wedge CP_0).$$

One can show (as in Vilks, 1998) that this condition would imply the BI play d_1, if the K_0^{i}-operators were Kripkean K45-operators. In AX', however, the formula

$$G \wedge U \wedge R_0 \wedge CP_0 \wedge E_0^{3}(G \wedge U \wedge R_0 \wedge CP_0) \rightarrow d_1$$

is *not* valid. (Replacing R_0 and CP_0 by, e.g., R_4 and CP_4 does not alter this conclusion.) The reason is that players' reasoning capabilities need not be mutually known. For instance, consider the rationality condition for the forth decison node:

$(R_{II,4}^{0})$ $\quad K_0^{II}(a_4 \leftrightarrow a_4) \wedge \neg K_0^{II}(\neg d_4) \rightarrow \neg a_4$

Although $E_0(R_0)$ implies that player I knows this condition to hold, and $E_0(CP_0)$ implies that I

knows $a_3 \rightarrow \neg K_0^{II}(\neg d_4)$, he cannot conclude from this that $a_3 \rightarrow \neg a_4$, as would be required for the BI argument. The missing link here is player I's knowledge of $K_0^{II}(a_4 \leftrightarrow a_4)$: While it is an axiom of AX', and known introspectively by II from t=1 onwards, $K_t^I(K_0^{II}(a_4 \leftrightarrow a_4))$ is not valid in C_0'.

Thus the logic AX' behaves quite differently from multi-agent K45: Although agents do not have much less reasoning powers, they may doubt the logical powers of others. For somewhat complicated perfect-information games this may well be a relevant reason why BI may fail.

The belief-set semantics developed above is flexible enough, however, to accomodate stronger epistemic systems. To illustrate this point, we consider a particularly simple subclass of C_0', which is defined as follows:

$$C^* := \{(v,(B_t^i)) \in C_0' \mid \forall t: B_t^I = B_t^{II}\}$$

It is straightforward to verify that an axiomatization of validity in C* is provided by adding to AX' the axiom scheme:

(*) $\quad K_t^I \varphi \leftrightarrow K_t^{II} \varphi$

This may be taken to express that the two agents start with the same initial beliefs, and reason in exactly the same manner. Clearly, the models of C* are essentially the single-agent models of C_0.

For the four-legged centipede, it is now easy to verify that the formula

$$G \wedge U \wedge R_3 \wedge CP_3 \wedge E_0(G \wedge U \wedge R_2 \wedge CP_2) \rightarrow d_1$$

is valid in C*, providing yet another sufficient epistemic condition for BI.

To be sure, (*) is a rather drastic simplification. However, our aim in this paper is not to "defend" BI, but to suggest a framework for epistemic logic which can help to analyse reasoning

about reasoning - a topic which is almost assumed away by relying on state-space models.

REFERENCES

Aumann, R. J. (1998), "On the Centipede Game," *Games and Economic Behavior* **23**, 97-105.

Battigalli, P., and Bonanno, G. (1997), "The Logic of Belief Persistence," *Economics and Philosophy* **13**: 39-59.

Binmore, K. (1992), *Fun and Games. A Text on Game Theory.* Lexington, MA: D.C.Heath.

Chellas, B. (1980), *Modal Logic: An Introduction.* Cambridge: Cambridge University Press.

Dekel, E., and F. Gul (1997), "Rationality and Knowledge in Game Theory," in D.M. Kreps, and K. F. Wallis (eds.), *Advances in Economics and Econometrics: theory and Applications, Volume I,* Cambridge: Cambridge University Press.

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

Gärdenfors, P. (1988), *Knowledge in Flux: Modeling the Dynamics of Epistemic States,* Cambridge (MA): The MIT Press.

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

Konolige, K. (1988), "On the relation between default and autoepistemic logic". *Artificial Intelligence* **35**: 343-382.

Lenzen, W. (1980), *Glauben, Wissen und Wahrscheinlichkeit*. Wien and New York: Springer.

Levi, I. (1991), "Consequentialism and Sequential Choice", in: M. Bacharach and S. Hurley (eds.), *Foundations of Decision Theory*. Cambridge, MA: Blackwell.

Moore, R.C. (1985), "Semantical considerations on nonmonotonic logic". *Artificial Intelligence* **25**.

Samet, D. (1996), "Hypothetical Knowledge and Games with Perfect Information," *Games and Economic B ehavior* 17, 230-251.

Schick, F. (1979), "Self-knowledge, uncertainty, and choice", *The British Journal of*

Philosophy **30**.

Vilks, A. (1997), “A Player’s Reasoning Process as a Sequence of Propositional Calculi”, in: M. Bacharach, et al. (eds.), *Epistemic Logic and the Theory of Games and Decisions*. Boston: Kluwer.

Vilks, A. (1998), “Knowledge of the Game, Relative Rationality, and Backwards Induction without Counterfactuals”, unpublished.