On the logical foundations of game theory
(Abstract)

Arnis Vilks
Handelshochschule Leipzig

1. Introduction

During the last 15 years there has been growing interest among game theorists in epistemic conditions for game-theoretic solution concepts. Most of the work in this area has more or less explicitly employed some version of Kripke-style epistemic logic. Actually, most game theorists do not work with the syntactic formulations of epistemic logics, but instead with information-structures. An information structure, however, can be viewed as a Kripke model, and as the relation between Kripke models and normal modal systems are well known to logicians, I do not need to go into that here.

Instead, I will use the weakest normal system K (in fact, the multi-agent version thereof) to explain a problem with this kind of logic which I believe can only be adequately resolved by moving beyond "normal" modal systems. In the second part of the paper I will therefore suggest an epistemic logic which resolves the problem in what I believe to be a satisfactory way.

2. The logic K_Γ

For a given finite extensive game of perfect information (PI) Γ, we define the logic K_Γ as follows: atomic formulas are: move formulas a, b, c, ... one for each move of Γ, and preference formulas $a \succ_i b$ ... where a, b are any moves, and i is a player of Γ. Wffs are made up from these atomic formulas in the usual way by applications of negation, conjunction, and belief operators B_i. The axioms of K_Γ are the usual ones of multi-agent K plus Γ-specific axioms describing the rules of the game Γ and the preferences of the players according to the payoff function of Γ. For the simplest nontrivial example of a PI game, Γ_0, which has just one player 1, who has to choose between moves a and b, whereof he prefers the former, the Γ-specific axioms are $(a \vee b) \wedge \neg(a \wedge b)$ and $a \succ_1 b$. The rules of inference are modus ponens and epistemization (which may be applied to all the axioms including the Γ-specific ones).

3. The problem of self-knowledge of rationality and options

Within K_Γ, we give a sufficient condition for the backward induction play of Γ which can be shown to be weaker than the one of Aumann (1995). In this abstract, we explain our condition only for the one-player example Γ_0 (described above), which suffices to explain the problem we seek to solve in this paper.

As the player may have false beliefs in $K_{\Gamma 0,}$ his choice of *b* – contrary to his preference – may be due to his belief that *a* is not possible. This motivates a condition we call *relative* rationality:

(RR) $\quad \neg B_1 \neg a \Rightarrow \neg b$

As $a \succ_1 b$ is an axiom of $\Gamma_{0,}$ this says that the player will not take action b if he considers the preferred action possible. Clearly, RR $\Rightarrow$ a does not hold in $K_{\Gamma 0.}$ However, it seems natural to add the assumption that the player does consider *a* possible. We call this assumption Possibility of Backward Induction moves:

(PBI) $\quad \neg B_1 \neg a$

For our simple example, RR $\wedge$ PBI $\Rightarrow$ *a* is trivially a theorem of Γ_0. (For the general case, an analogous, but more elaborate theorem holds.) However, a problem arises from the fact that it seems natural and in line with the usual informal assumptions of game theory to assume of *all* moves that they are considered possible, and that there is mutual (or even common) belief in rationality and the structure of the game. Clearly, B_1(RR $\wedge \neg B_1 \neg a \wedge \neg B_1 \neg b$) is inconsistent in Γ_0: The player can infer from what he believes that what he considers possible will not be the case.

2. The logic L_Γ

To resolve the above problem we suggest an epistemic logic which has a sequence of belief operators B^0, B^1, B^2, … for each player, corresponding to the temporal sequence of the player's states of belief. Limiting ourselves (in this abstract) to the one-player case again, we consider the axiomatic system (for which we also provide a belief-set semantics, similar to the autoepistemic logics of Moore, 1985, and Konolige, 1988) with the following axiom schemes:

(A1) φ , whenever φ is a propositional calculus tautology or a Γ-specific axiom;
(A2) $B^t(\varphi)$, whenever φ is a propositional calculus tautology or a Γ-specific axiom;
(A3) $B^t(\varphi) \wedge B^t(\psi) \Rightarrow B^t(\varphi \wedge \psi)$;
(A4) $B^t(\varphi) \Rightarrow B^{t+1}(\varphi)$;
(A5) $B^t(\varphi) \Rightarrow B^{t+1}(B^t(\varphi))$;
(A6) $\neg B^t(\varphi) \Rightarrow B^{t+1}(\neg B^t(\varphi))$;

the sole rule of inference being modus ponens. Among other properties of this logic L_Γ , we show that a delayed version of the epistemization rule holds.

3. A Solution to the Problem

Within L_Γ, the problem explained above can be easily resolved: Writing (RR^0) for $\neg B^0 \neg a \Rightarrow \neg b$, one can verify that $\neg B^0 \neg a \wedge \neg B^0 \neg b \wedge B^0(RR^0)$ is consistent, and so is $B^1(RR^0 \wedge \neg B^0 \neg a \wedge \neg B^0 \neg b)$. These formulas can be naturally taken to describe a situation where *initially* the player considers both options possible and himself to be rational, *and then*, on reflection, recognizes that he will not take b, while remembering that he initially considered both options possible.

A multi-agent version of L_Γ can be used to reconstruct both the backward induction argument and that it may fail if the players have insufficient knowledge about each other's reasoning processes.

References

Aumann, R. (1995), "Backward induction and common knowledge of Rationality", Games and Economic Behavior 8: 6-19.

Moore, R. (1985), "Semantical considerations on nonmonotonic logic", Artificial intelligence 25: 75-94

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

Vilks, A. (1999), "Knowledge of the game, relative rationality, and backwards induction without counterfactuals", HHL Working Paper 25.

ERROR: undefined
OFFENDING COMMAND: --------------CC9F83EE4AE40EC38962993C--

STACK: