Understanding the Brandenburger-Keisler Belief Paradox

Eric Pacuit
Institute of Logic, Language and Information
University of Amsterdam
epacuit@staff.science.uva.nl
staff.science.uva.nl/∼epacuit

January 31, 2006

In their textbook, Osborne and Rubinstein describe game theory as “a bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact” ([8] page 1). They go on to say that one of the basic assumptions of game theory is that when agents make decisions, they take into account “their knowledge or expectations of *other* decision-makers’ behavior (they *reason strategically*).” In other words, when agents involved in a multi-agent interactive situation are making decisions about what action to perform next, that decision is influenced by what actions they *expect* the other agents will perform. This assumption leads very naturally to questions about what agents believe about the other agents’ beliefs.

This observation has prompted a number of game theorists to propose that the basic models of game theory (extensive games forms and normal game forms) be extended to include a representation of the agents’ beliefs (see [1, 9, 3, 4, 7] for a discussion of the relevant literature). Essentially the idea is that when describing a strategic interactive situation part of that description should include the agents’ beliefs about the relevant ground (non-epistemic) facts, beliefs about the other agents’ beliefs about these ground facts, beliefs about the other agents’ beliefs about the other agents’ beliefs about these ground facts, and so on. Early on in 1967, John Harsanyi [6] developed an elegant formal model which can be used to represent the epistemic state of the agent in a game theoretic situation[1]. The

[1]Harsanayi’s original motivation was to study games of *incomplete* information, i.e., games

idea is that each agent could be any one of a number of different **types**, where a type is intended to represent an *infinite hierarchy of beliefs*, i.e., the agent's first-order beliefs about the strategies of the other agents, second-order beliefs about the other agents' first-order beliefs, third-order beliefs about the other agents' second-order beliefs, and so on. Thus the problem of adding beliefs to the basic models of game theory reduces to finding an appropriate collection of possible types for each agent.

Adam Brandenburger and H. J. Keisler have recently discovered a Russell-style paradox lurking in the background of the above discussion. In [5], they show that is the following configurations of beliefs is impossible: *Ann believes that Bob assumes*[2] *that Ann believes that Bob's assumption is wrong.* This suggests that it may not always be possible to find a type space to represent certain configurations of beliefs.

In [5] a modal logic interpretation of the paradox is proposed. The idea is to introduce two modal operators intended to represent the agents' *beliefs* and *assumptions*. The goal of this paper is to take this analysis further and study this paradox from the point of view of a modal logician. In particular, we show that the paradox can be seen as a theorem of an appropriate hybrid logic[3]. Furthermore, we propose a sound and complete axiomatization of a modal logic with belief and assumption modal operators, a question left open in [5].

References

[1] Aumann, R. Interactive epistemology I: knowledge. *International Journal of Game Theory 28* (1999), 263–300.

[2] Blackburn, P., and Seligman, J. Hybrid languages. *Journal of Logic, Language and Information 4* (1995), 251 – 272.

[3] Bonanno, G., and Battigalli, P. Recent results on belief, knowledge and the epistemic foundations of game theory. *Research in Economics 53*, 2 (June 1999), 149–225.

in which the agents are uncertain about the structure of the game.

[2]An assumption is a belief that implies all other beliefs. It is shown in [5] that it is crucial the statement be about "one particular belief of Bob and all of Ann's beliefs".

[3]Hybrid logic is a modal logic with distinguished propositional variables called *nominals* that are used to name each world in a Kripke structure. See [2] for more information.

[4] Brandenburger, A. The power of paradox: some recent developments in interactive epistomology. *International Journal of Game Theory* (2006 (forthcoming)).

[5] Brandenburger, A., and Keisler, H. An impossibility theorem on belies in games. forthcoming in *Studia Logica*, available at `pages.stern.nyu.edu/∼abranden/itbg072904.pdf`, July 2004.

[6] Harsanyi, J. C. Games with incompletete information played by bayesian players parts I-III. *Management Sciences 14* (1967).

[7] M.O.L. Bacharach, L. A. Gerard-Varet, P. M., and Shin, H. S., Eds. *Epistemic logic and the theory of games and decisions.* Theory and Decision Library, Kluwer Academic Publishers, 1997.

[8] Osborne, M., and Rubinstein, A. *A Course in Game Theory.* The MIT Press, 1994.

[9] Stalnaker, R. Belief revision in games: forward and backward induction. *Mathematical Social Sciences*, 36 (1998), 31 – 56.