Non-Classical Approaches to the Brandenburger-Keisler Paradox

Can BAŞKENT

Department of Computer Science, University of Bath, England

can@canbaskent.net   www.canbaskent.net/logic

26th International Conference on Game Theory, Stony Brook
July 20-25, 2015

Outlook of the Talk

- The Brandenburger - Keisler Paradox
- Non-well-founded set theoretic approach
- Paraconsistent approach
- Redefining the Paradox

The Paradox

The Brandenburg-Keisler paradox (BK paradox) is a two-person self-referential paradox in epistemic game theory (Brandenburger & Keisler, 2006).

Theorem (The Paradox)

The following configuration of beliefs is impossible:

> *Ann believes that Bob assumes that Ann believes that Bob's assumption is wrong.*

The paradox appears if you ask whether "Ann believes that Bob's assumption is wrong".

Notice that this is essentially a 2-person Russell's Paradox.

Why is it a Paradox?

The theorem uses two underlying, hidden assumptions:

- it uses ZF(C) set theory with *well-founded* sets,
- it uses classical logic which does not allow inconsistencies.

What would happen, if we change these assumptions?

Why is it a Paradox?

The theorem uses two underlying, hidden assumptions:

- it uses ZF(C) set theory with *well-founded* sets,
- it uses classical logic which does not allow inconsistencies.

What would happen, if we change these assumptions?

Model

Brandenburger and Keisler use belief sets to represent the players' beliefs.

The model (U^a, U^b, R^a, R^b) that they consider is called a *belief structure* where $R^a \subseteq U^a \times U^b$ and $R^b \subseteq U^b \times U^a$.

The expression $R^a(x, y)$ represents that in state x, Ann believes that the state y is possible for Bob, and similarly for $R^b(y, x)$. We will put $R^a(x) = \{y : R^a(x, y)\}$, and similarly for $R^b(y)$.

At a state x, we say Ann believes $P \subseteq U^b$ if $R^a(x) \subseteq P$.

Semantics

A modal logical semantics for the interactive belief structures can be given.

We use two modalities $\Box$ and $\heartsuit$ for the belief and assumption operators respectively with the following semantics.

$$x \models \Box^{ab}\varphi \quad \textit{iff} \quad \forall y \in U^b . R^a(x, y) \textbf{ implies } y \models \varphi$$
$$x \models \heartsuit^{ab}\varphi \quad \textit{iff} \quad \forall y \in U^b . R^a(x, y) \textbf{ iff } y \models \varphi$$

Note the bi-implication in the definition of the assumption modality!

Completeness

A belief structure (U^a, U^b, R^a, R^b) is called *assumption complete* with respect to a set of predicates Π on U^a and U^b if for every predicate $P \in \Pi$ on U^b, there is a state $x \in U^a$ such that x assumes P, and for every predicate $Q \in \Pi$ on U^a, there is a state $y \in U^b$ such that y assumes Q.

We will use special propositions $\mathbf{U}^a$ and $\mathbf{U}^b$ with the following meaning: $w \models \mathbf{U}^a$ if $w \in U^a$, and similarly for $\mathbf{U}^b$. Namely, $\mathbf{U}^a$ is true at each state for player Ann, and $\mathbf{U}^b$ for player Bob.

Incompleteness

Brandenburger and Keisler showed that no belief model is complete for its (classical) first-order language.

Therefore, “not every description of belief can be represented” with belief structures (Brandenburger & Keisler, 2006).

Incompleteness

The incompleteness of the belief structures is due to the *holes* in the model. A model, then, has a hole at φ if either $\mathbf{U}^b \wedge \varphi$ is satisfiable but $\heartsuit^{ab}\varphi$ is not, or $\mathbf{U}^a \wedge \varphi$ is satisfiable but $\heartsuit^{ba}\varphi$ is not.

Namely, φ is true for *b*, but cannot be assumed by *a* (or vice versa).

A big hole is then defined by using the belief modality $\Box$ instead of the assumption modality $\heartsuit$.

Theorem

Modal Version, (Brandenburger & Keisler, 2006)

There is either a hole at $\mathbf{U}^a$, a hole at $\mathbf{U}^b$, a big hole at one of the formulas

$$\heartsuit^{ba}\mathbf{U}^a, \qquad \Box^{ab}\heartsuit^{ba}\mathbf{U}^a, \qquad \Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$$

a hole at the formula $\mathbf{U}^a \wedge \mathbf{D}$, or a big hole at the formula $\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D})$. Thus, there is no complete interactive frame for the set of modal formulas built from $\mathbf{U}^a$, $\mathbf{U}^b$, and $\mathbf{D}$.

A model, then, has a hole at φ if either $\mathbf{U}^b \wedge \varphi$ is satisfiable but $\heartsuit^{ab}\varphi$ is not, or $\mathbf{U}^a \wedge \varphi$ is satisfiable but $\heartsuit^{ba}\varphi$ is not. A big hole is defined by using $\Box$ instead of $\heartsuit$.

Non-well-foundedness

- The Brandenburger - Keisler Paradox
- Non-well-founded set theoretic approach
- Paraconsistent approach

Concept

Non-well-founded set theory is a theory of sets where the axiom of foundation is replaced by the *anti-foundation axiom* which is due to Mirimanoff (Mirimanoff, 1917). Then, decades later, it was formulated by Aczel within graph theory, and this motivates our approach here (Aczel, 1988).

In non-well-founded (NWF, henceforth) set theory, we can have true statements such as '$x \in x$', and such statements present interesting properties in game theory. NWF theories are natural candidates to represent circularity (Barwise & Moss, 1996).

Concept

NWF set theory is not immune to the problems that the classical set theory suffers from. For example, note that Russell's paradox is **not** solved in NWF setting, and moreover the subset relation stays the same in NWF theory (Moss, 2009).

Therefore, we may not expect the BK paradox to *disappear* in NWF setting. Yet, NWF set theory will give us many other tools in game theory.

NWF Type Spaces

It seems to me that the basic reason why the theory of games with incomplete information has made so little progress so far lies in the fact that these games give rise, or at least appear to give rise, to an infinite regress in reciprocal expectations on the part of the players. In such a game player 1's strategy choice will depend on what he expects (or believes) to be player 2's payoff function U_2, as the latter will be an important determinant of player 2's behavior in the game. But his strategy choice will also depend on what he expects to be player 2's first-order expectation about his own payoff function U_1. Indeed player 1's strategy choice will also depend on what he expects to be player 2's second-order expectation - that is, on what player 1 thinks that player 2 thinks that player 1 thinks about player 2's payoff function U_2... and so on ad infinitum.

(Harsanyi, 1967)

NWF Type Spaces

> *Nevertheless, one may continue to argue that a state of the world should indeed be a circular, self-referential object: A state represents a situation of human uncertainty, in which a player considers what other players may think in other situations, and in particular about what they may think there about the current situation. According to such a view, one would seek a formulation where states of the world are indeed self-referring mathematical entities.*

(Heifetz, 1996)

Definition

What we call a non-well-founded model is a tuple $M = (W, V)$ where W is a non-empty non-well-founded set (*hyperset*), and V is a valuation. We will use the symbol $\models^+$ to represent the semantical consequence relation in a NWF model based on (Gerbrandy, 1999).

$$\begin{array}{lll} M, w \models^+ \Box^{ij}\varphi & \text{iff} & M, w \models^+ \mathbf{U}^i \ \wedge \\ & & \forall v \in w.(M, v \models^+ \mathbf{U}^j \to M, v \models^+ \varphi) \\ M, w \models^+ \heartsuit^{ij}\varphi & \text{iff} & M, w \models^+ \mathbf{U}^i \ \wedge \\ & & \forall v \in w.(M, v \models^+ \mathbf{U}^j \leftrightarrow M, v \models^+ \varphi) \end{array}$$

Counter-model

Consider the following NWF counter-model M. Let $W = \{w, u, v, t, y\}$ where $U^a = \{w, u\}$, and $U^b = \{v, t, y\}$. Put $w = \{v, t\}$, $v = \{u, w\}$, $u = \{t\}$, $y = \{u\}$.

Then, M satisfies the formulas given in the Main Theorem of BK.

First, M has no holes at $\mathbf{U}^a$ and $\mathbf{U}^b$ as the first is assumed at v, and the latter is assumed at w. Therefore, $v \models^+ \heartsuit^{ba}\mathbf{U}^a$. Moreover, it has no big holes, thus w believes $\heartsuit^{ba}\mathbf{U}^a$ giving $w \models^+ \Box^{ab}\heartsuit^{ba}\mathbf{U}^a$. Similarly, v believes $\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$ yielding $v \models^+ \Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$.

Counter-model

We have to be careful here!

The counter-model does not establish that NWF belief models are complete.

It establishes the fact that they do not have the same holes as the classical belief models.

The answer requires **category theory**.

Paraconsistency

- The Brandenburger - Keisler Paradox
- Non-well-founded set theoretic approach
- Paraconsistent approach

What is a Topology?

Definition

The structure $\langle S, \sigma \rangle$ is called a topological space if it satisfies the following conditions.

1. $S \in \sigma$ and $\emptyset \in \sigma$
2. σ is closed under finite unions and arbitrary intersections

Collection σ is called a topology, and its elements are called *closed* sets.

Problem of Negation

We stipulate that extensions of *any* propositional variables to be a closed set (Mortensen, 2000), to get a paraconsistent system.

Negation can be difficult as the complement of a closed set is not generally a closed set, thus may not be the extension of a formula in the language.

For this reason, we will need to use a new negation that returns the closed complement (closure of the complement) of a given set.

Topological Belief Models

The language for the logic of topological belief models is given as follows.

$$\varphi := p \mid \sim\varphi \mid \varphi \wedge \varphi \mid \Box_a \mid \Box_b \mid \boxplus_a \mid \boxplus_b$$

where p is a propositional variable, $\sim$ is the paraconsistent topological negation symbol which we have defined earlier, and $\Box_i$ and $\boxplus_i$ are the belief and assumption operators for player i, respectively.

Topological Belief Models

For the agents *a* and *b*, we have a corresponding non-empty type space A and B, and define closed set topologies τ_A and τ_B on A and B.

To connect τ_A and τ_B, we introduce $t_A \subseteq A \times B$, and $t_B \subseteq B \times A$.

We then call the structure $F = (A, B, \tau_A, \tau_B, t_A, t_B)$ a paraconsistent topological belief model.

A state $x \in A$ *believes* $\varphi \subseteq B$ if $\{y : t_A(x, y)\} \subseteq \varphi$. Furthermore, a state $x \in A$ *assumes* φ if $\{y : t_A(x, y)\} = \varphi$. Notice that in this definition, we identify logical formulas with their extensions.

Semantics

For $x \in A$, $y \in B$, the semantics of the modalities are given as follows with a modal valuation attached to F.

$$
\begin{array}{lll}
x \models \Box_a \varphi & \textit{iff} & \exists Y \in \tau_B \textit{ with } t_A(x, Y) \to \forall y \in Y . y \models \varphi \\
x \models \boxplus_a \varphi & \textit{iff} & \exists Y \in \tau_B \textit{ with } t_A(x, Y) \leftrightarrow \forall y \in Y . y \models \varphi \\
y \models \Box_b \varphi & \textit{iff} & \exists X \in \tau_A \textit{ with } t_B(y, X) \to \forall x \in X . x \models \varphi \\
y \models \boxplus_b \varphi & \textit{iff} & \exists X \in \tau_A \textit{ with } t_B(y, X) \leftrightarrow \forall x \in X . x \models \varphi
\end{array}
$$

The Result

Theorem ((Başkent, 2015))

The BK sentence is satisfiable in some paraconsistent topological belief models.

Namely, we can construct a state which satisfies the BK sentence - push the *holes* that create the inconsistencies to the boundaries.

Yablo's Original Paradox

Yablo's Paradox, on the other hand, is a *non*-self referential paradox unlike the Brandenburger - Keisler paradox (Yablo, 1993). Yablo considers the following sequence of sentences.

$$S_1 : \forall k > 1, S_k \text{ is untrue},$$
$$S_2 : \forall k > 2, S_k \text{ is untrue},$$
$$S_3 : \forall k > 3, S_k \text{ is untrue},$$
$$\vdots$$

Yablo shows that every sentence S_n is untrue. Then, "the sentences *subsequent* [his emphasis] to any given S_n are all untrue, whence S_n is true after all!" [ibid]. Yablo's paradox can be viewed as a non-self-referential liar's paradox.

Yablo-like Reformulation of the Paradox

Consider the following sequence of assumptions where numerals represent game theoretical agents.

A_1 : 1 believes that $\forall k > 1$, k's assumption is untrue,
A_2 : 2 believes that $\forall k > 2$, k's assumption is untrue,
A_3 : 3 believes that $\forall k > 3$, k's assumption is untrue,

$\vdots$

Now, for a contradiction, assume A_n is true for some n. Therefore, n believes that $\forall k > n$, k's assumption is untrue. In particular, $n + 1$'s assumption is untrue. Then, $n + 1$ believes that $\forall k > n + 1$, k's assumption is true, which contradicts the initial assumption that A_n is true. The choice of n was arbitrary, so each A_n in the sequence is untrue.

Thanks for your attention!

Talk slides and the papers (Başkent, 2015) are available at

www.CanBaskent.net/Logic

References I

ACZEL, PETER. 1988.
Non-well-founded Sets.
Vol. 14.
CSLI Lecture Notes.

BARWISE, JON, & MOSS, LAWRENCE S. 1996.
Vicious Circles.
Center for Study of Language and Information.

BAŞKENT, CAN. 2015.
Some Non-Classical Approaches to the Brandenburger-Keisler paradox.
Logic Journal of the IGPL, **23**(4), 533–552.

BRANDENBURGER, ADAM, & KEISLER, H. JEROME. 2006.
An Impossibility Theorem on Beliefs in Games.
Studia Logica, **84**, 211–240.

References II

GERBRANDY, JELLE. 1999.

Bisimulations on Planet Kripke.

Ph.D. thesis, Institute of Logic, Language and Computation; Universiteit van Amsterdam.

HARSANYI, J. C. 1967.

Games with incomplete information played by 'Bayesian' players: I. The basic model.

Management Science, **14**(3), 159–182.

HEIFETZ, AVIAD. 1996.

Non-well-founded Type Spaces.

Games and Economic Behavior, **16**, 202–217.

MIRIMANOFF, DMITRY. 1917.

Les antimonies de Russell et de Burali-Forti et le probleme fondamental de la theorie des ensembles.

L'Enseignement Mathématique, **19**(1), 37–52.

References III

MORTENSEN, CHRIS. 2000.

Topological Seperation Principles and Logical Theories.

Synthese, **125**(1-2), 169–178.

MOSS, LAWRENCE S. 2009.

Non-wellfounded Set Theory.

In: ZALTA, EDWARD N. (ed), *The Stanford Encyclopedia of Philosophy*, fall 2009 edn.

YABLO, STEPHEN. 1993.

Paradox without Self-Reference.

Analysis, **53**(4), 251–2.