Some Non-Classical Approaches to the Brandenburger-Keisler Paradox

Can BAŞKENT

The Graduate Center of the City University of New York
cbaskent@gc.cuny.edu www.canbaskent.net

KGB Seminar
The Graduate Center of the City University of New York
April 8^{th}, 2011

Outlook of the Talk

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

The Paradox

The Brandenburg-Keisler paradox ('BK paradox', henceforth) is a two-person self-referential paradox in epistemic game theory (Brandenburger & Keisler, 2006).
The following configuration of beliefs is impossible:

Paradox
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.

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

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

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.

Completeness

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

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

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.

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

Two Lemmas

In the original paper, the authors make use of two lemmas before identifying the holes in the system.
These lemmas are important for us as we will challenge them in the next section.
First, let us define a special propositional symbol **D** with the following valuation
$D = \{w \in W : (\forall z \in W)[P(w,z) \to \neg P(z,w)]\}$.

Lemma

1. If $\heartsuit^{ab}\mathbf{U}^b$ is satisfiable, then $\Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a \to \mathbf{D}$ is valid.
2. $\neg\Box^{ab}\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D})$ is valid.

Theorem

First-Order Version (Brandenburger & Keisler, 2006)

Every belief model M has either a hole at U^a, a hole at U^b, a big hole at one of the formulas
(i) $\forall x.P^b(y,x)$ (ii) x believes $\forall x.P^b(y,x)$
(iii) y believes [x believes $\forall x.P^b(y,x)$],
a hole at the formula (iv) $D(x)$,
or a big hole at the formula (v) y assumes $D(x)$
Thus, there is no belief model which is complete for a language $\mathcal{L}$ which contains the tautologically true formulas and formulas (i)-(v).

Theorem

Modal Version

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 all modal formulas built from $\mathbf{U}^a$, $\mathbf{U}^b$, and $\mathbf{D}$.

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

On the other hand, 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.

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*, for short), 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).

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

Lemmas

Define $D^+ = \{w \in W : \forall v \in W.(v \in w \rightarrow w \notin v)\}$.
We define the propositional variable $\mathbf{D}^+$ as the propositional variable with the valuation set D^+.

Lemmas

Lemma

In a NWF belief structure, if $\heartsuit^{ab}\mathbf{U}^b$ is satisfiable, then the formula $\Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a \wedge \neg\mathbf{D}^+$ is also satisfiable.

Proof

Let $W = \{w, v, u, t, z\}$ with $w = \{v\}$, $v = \{u, w\}$, $u = \{t\}$, $t = \{z\}$ where $U^a = \{w, u, z\}$, and $U^b = \{v, t\}$. To maintain the disjointness of the types, assume that the sets w, v, u, t, z are not transitive. Then, $w \models^+ \heartsuit^{ab}\mathbf{U}^b$. Moreover, we have $w \models^+ \Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$. But, by design, $w \not\models^+ \mathbf{D}^+$.

Lemmas

Lemma

The formula $\Box^{ab}\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D}^+)$ is satisfiable in some NWF belief structures.

Proof

Take $M = (W, V)$ with $W = \{w, v, u, t\}$ where $w = \{v\}$, $v = \{u\}$, $u = \{t\}$ with $u \notin t$. Let $U^a = \{w, u\}$ and $U^b = \{v, t\}$. Then, it is easy to see that $M, w \models \Box^{ab}\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D}^+)$. $\Box$

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 Theorem.
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

The state u also satisfies D, and it is assumed by y, thus y assumes $D(u)$. This counter-model shows that Theorem does not hold in NWF belief structures.

Yet, we have to be careful here. Our counter model does not establish the fact that NWF belief models are complete. It establishes the fact that they do not have the same holes as the standard belief models. We will get back to this question later on, and give an answer from category theoretical point of view.

An Application

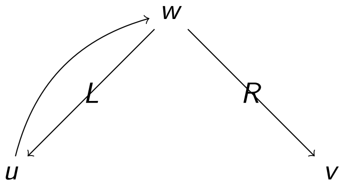
Theorem

For every (labeled) rooted directed connected graph, there corresponds to a unique two-player NWF belief structure up to the permutation of type spaces, and the order of players.

Therefore, we can use any connected graphs (not only trees, but also connected graphs with loops) to represent games in extensive form (up to the natural conditions in the theorem).

An Example

Consider the following labeled, connected directed graph.



The two-player NWF belief structure of this game is as follows. Put $W = \{w, v, u\}$ where $w = \{u, v\}$ and $u = \{w\}$. Assume that $U^a = \{w\}$, $U^b = \{u, v\}$ (or any other combination of type spaces). Therefore, this graph corresponds to the game where Bob can *reset* the game if Alice plays L at w.

Definition

The well-studied notion of deductive explosion describes the situation where any formula can be deduced from an inconsistent set of formulae, i.e. for all formulae φ and ψ, we have $\{\varphi, \neg\varphi\} \vdash \psi$, where $\vdash$ denotes the classical logical consequence relation.

In this respect, both "classical" and intuitionistic logics are known to be explosive. Paraconsistent logic, on the other hand, is the umbrella term for logical systems where the logical consequence relation $\vdash$ is not explosive (Priest, 2002).

Motivation for Paraconsistency

Motivation for paraconsistency is usually this: we may be in a situation where our theory/information is inconsistent, but we still would like to make inference **sensibly**.
There are several class of situations where paraconsistency could be thought of a natural approach.

- Computer databases
- Scientific theories
- Law
- Counterfactuals
- Various human behavior

(Priest, 2002)

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.

Paraconsistent Topological Semantics

Use of topological semantics for paraconsistent logic is not new.
To our knowledge, the earliest work discussing the connection between inconsistency and topology goes back to Goodman (Goodman, 1981)[1].
Namely, in basic modal logic, only modal formulas produce topological objects.
If we stipulate that:
extension of *any* propositional variable to be a closed set (Mortensen, 2000), we get a paraconsistent system.

[1] Thanks to Chris Mortensen for pointing this work out.

Problem of Negation

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 symbol $\sim$ that returns the closed complement (closure of the complement) of a given set.

Self-Reference

Recently, a category theoretical approach has been presented (Abramsky & Zvesper, 2010).
They focus on the fixed points and extend their analysis to category theory.
Lawvere's Theorem says that if $g : X \to V^X$ is surjective, then every function $f : V \to V$ has a fixed point (Lawvere, 1969).
BK paradox occurs if f plays the role of a Boolean negation.

Conditions

Lawvere's Theorem says that if $g : X \to V^X$ is surjective, then every function $f : V \to V$ has a fixed point (Lawvere, 1969). There is an important restriction:

- X should be cartesian closed (actually, should only admit exponents)

Usually people consider the category of sets **Set**.

Co-Heyting: definitions

Let L be a bounded distributive lattice. If there is defined a binary operation $\Rightarrow: L \times L \to L$ such that for all $x, y, z \in L$,

$$x \leq (y \Rightarrow z) \text{ iff } (x \wedge y) \leq z,$$

then we call $(L, \Rightarrow)$ a Heyting algebra.
Dually, if we have a binary operation $\setminus : L \times L \to L$ such that

$$(y \setminus z) \leq x \text{ iff } y \leq (x \vee z),$$

then we call $(L, \setminus)$ a co-Heyting algebra.
We call $\Rightarrow$ implication, $\setminus$ subtraction.

Co-Heyting: definitions

In Boolean algebras, Heyting and co-Heying algebras give two different operations. We interpret $x \Rightarrow y$ as $\neg x \vee y$, and $x \setminus y$ as $x \wedge \neg y$.

In other words, a co-Heyting algebra is a generalization of a Boolean algebra that allows a generalization in which *principium contradictionis* is relaxed.

Closed set topologies are co-Heyting algebras. The topological paraconsistent negation $\sim$ is defined as $\sim\varphi \equiv \mathbf{1} \setminus \varphi$ where $\mathbf{1}$ is the top element of the lattice.

Paraconsistent BK Paradox

Therefore, even if we have paraconsistent framework. we will have fixed points.
How:

- Take a co-Heyting algebra - which is a natural candidate for paraconsistency.
- Observe that it admits exponents: $x^y \equiv x \wedge \neg y$.
- Thus, Lawvere's Theorem applies.
- It will still have fixed points: instead of the Boolean negation, take co-Heyting negation as the unary operator.

NWF Categories

Category of hypersets is also CCC .
Thus, Lawvere theorem also applies.
Therefore, we will have "different" fixed points, BK sentences in NWF setting.

Conclusion

We have observed that

- In NWF setting, BK models may have different fixed-points
- In paraconsistent setting, BK paradox can be modelled

Conclusion

What is in the paper, but not in this short talk?

- Two different ways to construct topological models: paraconsistent, and product models
- A counter-model where BK sentence fails

Thanks for your attention!

Talk slides and the papers are available at:

www.CanBaskent.net

References I

Abramsky, Samson, & Zvesper, Jonathan. 2010.
From Lawvere to Brandenburger-Keisler: Interactive forms of diagonalization and self-reference.
CoRR, abs/1006.0992.

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.

Brandenburger, Adam, & Keisler, H. Jerome. 2006.
An Impossibility Theorem on Beliefs in Games.
Studia Logica, **84**, 211–240.

Gerbrandy, Jelle. 1999.
Bisimulations on Planet Kripke.
Ph.D. thesis, Institute of Logic, Language and Computation; Universiteit van Amsterdam.

Goodman, Nicolas D. 1981.
The Logic of Contradiction.
Zeitschrift für Mathematische Logik und Grundlagen der Mathematik, **27**(8-10), 119–126.

References II

LAWVERE, F. WILLIAM. 1969.
Diagonal Arguments and Cartesian Closed Categories.
Pages 134–145 of: DOLD, A., & ECKMANN, B. (eds), *Category Theory, Homology Theory and their Applications II*.
Lecture Notes in Mathematics, vol. 92.
Springer.

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.

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.

PRIEST, GRAHAM. 2002.
Paraconsistent Logic.
Pages 287–393 of: GABBAY, DOV, & GUENTHNER, F. (eds), *Handbook of Philosophical Logic*, vol. 6.
Kluwer.