Dialetheism and a Game Theoretical Paradox

Can BAŞKENT

Institut d'Histoire et de Philosophie des Sciences et des Techniques

can@canbaskent.net   www.canbaskent.net

25 Years *In Contradiction* Conference

University of Glasgow

December 6-8, 2012

Outlook of the Talk

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

"Dialetheism is outrageous, at least to the spirit of contemporary philosophy. For this reason, many may be tempted to dismiss the idea out of hand. I ask only that they remember how many scientific theories that came to be accepted started their history wearing this mark, and give the idea a fair hearing. I am confident that if this be done the merits of the position will speak for themselves."

Graham Priest, *In Contradiction*, Preface to the First Edition

The Paradox

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

The following configuration of beliefs is impossible:

Theorem (The 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.

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

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

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 $\heartsuit$.

Two Lemmas

In the original paper, the authors give two lemmas before identifying the holes in the system.

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.

Main Theorem of BK

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 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 $\heartsuit$.

Some Remarks

- BK paradox is a game theoretical example of a self-referential paradox
- It is a simple step towards the possibility of paraconsistent games - a broader research program in progress
- It raises the possibility of discussing discursive/dialogical logics within game theory proper
- Provides an interesting take on Hintikka's interrogative theory - how to inquire about a paradoxical sentence?

Some Remarks

- BK paradox is a game theoretical example of a self-referential paradox
- It is a simple step towards the possibility of paraconsistent games - a broader research program in progress
- It raises the possibility of discussing discursive/dialogical logics within game theory proper
- Provides an interesting take on Hintikka's interrogative theory - how to inquire about a paradoxical sentence?

Some Remarks

- BK paradox is a game theoretical example of a self-referential paradox
- It is a simple step towards the possibility of paraconsistent games - a broader research program in progress
- It raises the possibility of discussing discursive/dialogical logics within game theory proper
- Provides an interesting take on Hintikka's interrogative theory - how to inquire about a paradoxical sentence?

Some Remarks

- BK paradox is a game theoretical example of a self-referential paradox
- It is a simple step towards the possibility of paraconsistent games - a broader research program in progress
- It raises the possibility of discussing discursive/dialogical logics within game theory proper
- Provides an interesting take on Hintikka's interrogative theory - how to inquire about a paradoxical sentence?

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

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.

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).

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

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 $U^b = \{v, t\}$. 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, $w \not\models^+ \mathbf{D}^+$. $\Box$

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 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!

Our 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.

We will get back to this question later on, and give an answer from category theoretical point of view.

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.

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).

In classical modal logic, only modal formulas produce topological objects.

However, if we stipulate that:

extension of *any* propositional variable to be a closed set (Mortensen, 2000), we get a paraconsistent system.

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.

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 respectively. Furthermore, in order to establish connection between τ_A and τ_B to represent belief interaction among the players, we introduce additional constructions $t_A \subseteq A \times B$, 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.

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

The Result

Theorem

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.

Product Topological Models

It is also possible to construct a topological model with topological products.

Definition

Let a, b be two players with corresponding type space A and B. Let τ_A and τ_B be the (paraconsistent) closed set topologies of respective type spaces. The product topological paraconsistent belief structure for two agents is given as $(A \times B, \tau_A \times \tau_B)$.

In this structure, we can get almost-complete game models.

Self-Reference

Recently, a category theoretical approach has been presented for the BK paradox (Abramsky & Zvesper, 2010).

They focused on the fixed points and extended their analysis to category theory. They started from Lawvere's theorem.

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, and V is a valuation. The *fixed-point* then refers to the paradoxical sentence. At a fixed-point, under a suitable negation, the sentence is equivalent to its negation.

Conditions

Lawvere’s result explains various paradoxes of self-reference; Liar, Halting Problem etc... (Yanofsky, 2003)

It also explains the BK paradox.

Then, what about paraconsistent BK paradox?

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

Namely, the category should admit exponents and products, and have a terminal object.

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

Is is possible to use Lawvere's Theorem with paraconsistent negation?

Namely, is there a paraconsistency-friendly cartesian closed category?

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

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

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

We call $\Rightarrow$ implication, $\setminus$ subtraction.

Co-Heyting: definitions

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

Closed set topologies which we discussed 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.

And Co-Heyting algebras are cartesian closed categories.

Paraconsistent BK Paradox

Therefore, even if we have paraconsistent framework, we will have fixed points with respect to negation - where a formula and its negation are both true.

How:

- Take a co-Heyting algebra that represents paraconsistent topological models
- Observe that it admits exponents: $x^y \equiv x \wedge \sim y$.
- Thus, Lawvere's Theorem applies.
- It will still have fixed points: instead of the Boolean negation, take the paraconsistent negation $\sim$ as the unary operator.
- Translate the fixed-points to BK paradoxical holes by Abramsky & Zvesper result.

What does it mean?

It means that in our paraconsistent model, the BK sentence is satisfiable.

Briefly, we can push the holes to the boundaries that satisfy the contradictory statements.

Or, we can obtain them by using Lawvere's Theorem as fixed-points.

NWF Categories

What about NWF models?

Category of hypersets is also cartesian closed.

Thus, Lawvere theorem also applies.

Therefore, we will have “different” fixed points, BK sentences in NWF setting.

Thanks for your attention!

Talk slides and the papers are available at

`www.CanBaskent.net/Logic`

and, thank you Graham for your pioneering work!

Thanks for your attention!

Talk slides and the papers are available at

`www.CanBaskent.net/Logic`

and, thank you Graham for your pioneering work!

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.

References II

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.

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.

References III

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.

References IV

YANOFSKY, NOSON S. 2003.

A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points.

The Bulletin of Symbolic Logic, **9**(3), 362–386.