Logic Journal of IGPL Advance Access published March 29, 2015

Some non-classical approaches to the Brandenburger–Keisler paradox

CAN BAŞKENT*, *Department of Computer Science, University of Bath, England.*

Abstract

In this article, we discuss a well-known self-referential paradox in epistemic game theory, the Brandenburger–Keisler paradox. We approach the paradox from two different perspectives, non-well-founded set theory and paraconsistent logic, and provide models in which the paradox is solved.

Keywords: Brandenburger–Keisler paradox, paraconsistent logic, epistemic game theory, topological models of modal logic, non-well-founded sets.

1 Introduction and motivation

1.1 Introduction

In this article, we consider a well-known epistemic game theoretical paradox, and provide various alternative models in which the paradoxical statement becomes satisfiable. For this task, we resort to various non-classical frameworks, and reformulate the paradoxical statement in them. By achieving this, first we provide a richer toolkit that can be used in epistemic game theoretical formalisms, and secondly imply that the choice of classical and traditional models in epistemic game theory seems rather arbitrary.

The Brandenburg–Keisler paradox ('BK paradox', henceforth) is a two-person self-referential paradox in epistemic game theory [6]. In short, for players Ann and Bob, the BK paradox arises when we consider the following statement 'Ann believes that Bob assumes that Ann believes that Bob's assumption is wrong', and ask the question if 'Ann believes that Bob's assumption is wrong', where Bob's assumption is the sentence that 'Ann believes that Bob's assumption is wrong.' In this case, we have two possible answers to the question.

If the answer to the above question is a 'yes', then Ann *does* believe that Bob's assumption is wrong, which means that she believes that the statement 'Bob's assumption is wrong' is wrong. Therefore, Ann believes that Bob's assumption is correct. But, initially she believed that this assumption was wrong. This creates a contradiction. On the other hand, if the answer is 'no', then she *does not* believe that Bob's assumption is wrong, which means that Ann believes that Bob's assumption is correct. However, this contradicts the assumption that 'Ann believes that Bob's assumption is wrong.' This is a contradiction, too. Both possible answers to the question create a contradiction. Thus, we obtain a paradox. The BK paradox, as the above reasoning demonstrates, can be seen as a two-person Russell's paradox.

From a logical perspective, there are two main reasons as to why the BK argument turns out to be paradoxical. First, the limitations of set theory present some restrictions on the mathematical models that are used to describe self-referentiality and circularity in formal languages. In other words, sets are assumed to be well-ordered due to ZF(C) set theory that admits the *axiom of foundation*. It can

*E-mail: can@canbaskent.net, www.canbaskent.net/logic

doi:10.1093/jigpal/jzv001

be deduced from this axiom that no set can be an element of itself. On the other hand, in non-well-founded set theory, the axiom of foundation is replaced by the *anti-foundation axiom* that leads to, among many other things, generating sets that are members of themselves [2, 24]. Therefore, we claim that adopting non-well-founded set theory suggests a new approach to the paradox, and game theory in general. The power of non-well-founded set theory comes from its genuine methods to deal with circularity [4, 26].

Secondly, the limitations of classical logic introduce various metaphysical assumptions to the mathematical models. What makes the BK argument a logical paradox is the *principium contradictionis*, which suggests that contradictions are impossible. Paraconsistent (and dialetheic) logics disagree with this assumption, and argue that *some* contradictions can be true, and contradictions do not necessarily trivialize the logic [7, 9, 28, 30]. The BK paradox is essentially a self-referential paradox, and similar to any other paradox of the same kind, it can be analysed from a category theoretical or algebraic point of view [1, 38]. Moreover, paraconsistent logics present some strong algebraic and category theoretical structures. In this work, we make the connection between self-referentiality and paraconsistency clearer by presenting some (counter-)models based on our non-classical approach.

Apart from pursuing the subject for its own merits, we implicitly underline in this article that the original BK paper gives no argument as to why game theoretical agents need to be represented by a theory that relies on classical logic and ZFC set theory, and whether this formal framework is sufficient to express interactive and complex epistemic situations in games. Briefly, in this article, we first show that adopting the non-well-founded set theory makes a significant change in the structure of the paradox. We achieve this by constructing non-well-founded counter-models for the BK argument. Secondly, by resorting to paraconsistent logic, we also demonstrate that the BK argument *can* be satisfied in some situations.

1.2 The paradox

The BK paradox can be considered as a game theoretical two-person version of Russell's paradox. Let us call the players Ann and Bob with associated type space sets U^a and U^b respectively. Now, consider the following statement that we call the 'BK sentence':

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

A two-person Russell-like paradox arises if one asks the question whether *Ann believes that Bob's assumption is wrong*. In both cases, as we explained earlier, we obtain a contradiction. Thus, the BK sentence is impossible. This is a strong paradox, namely it is not possible to dissolve it by using some other formalizations within the limits of classical modal logic and set theory.

Brandenburger and Keisler used belief sets to represent players' beliefs. The model (U^a, U^b, R^a, R^b) that they consider is called a *belief model* 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 *vice versa* 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 $R^a(x) \subseteq P$. Now, a semantics for the interactive belief structures can be given. We use two different operators $\square$ and $\heartsuit$ which stand for belief and assumption operators, respectively, with the following semantics.

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

In addition to the above semantics for the belief and assumption operators, a modal definition for these operators can be given [6]. First, define an interactive belief frame as the structure (W,P,U^a,U^b) with a binary relation $P \subseteq W \times W$, and disjoint sets U^a and U^b such that (U^a,U^b,P^a,P^b) is a belief model with $U^a \cup U^b = W$, $P^a = P \cap U^a \times U^b$, and $P^b = P \cap U^b \times U^a$. Now, for a given valuation function that assigns propositional variables to subsets of W, the semantics of the belief and assumption modalities are given as follows.

$$x \models \Box^{ab}\varphi \quad \text{iff} \quad w \models \mathbf{U}^a \wedge \forall y (P(x,y) \wedge y \models \mathbf{U}^b \text{ implies } y \models \varphi)$$
$$x \models \heartsuit^{ab}\varphi \quad \text{iff} \quad w \models \mathbf{U}^a \wedge \forall y (P(x,y) \wedge y \models \mathbf{U}^b \text{ iff } y \models \varphi)$$

The semantics of the assumption modality as above was originally given by Brandenburger and Keisler. However, as Pacuit underlined, a similar modality was also discussed by Humberstone not in a game theoretical setting [16, 27].

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.

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 [6]. 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 $\heartsuit$.

In the original paper, the authors make use of the following lemma before identifying the holes in the system. First, let us define a special propositional symbol $\mathbf{D}$ with the following valuation $D = \{w \in W : (\forall z \in W)[P(w,z) \rightarrow \neg P(z,w)]\}$.

Lemma 1.1 ([6])
1. If $\heartsuit^{ab}\mathbf{U}^b$ is satisfiable, then $\Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a \rightarrow \mathbf{D}$ is valid.
2. $\neg\Box^{ab}\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D})$ is valid.

Based on this lemma, the authors observe that there is no complete belief models. Here, we give the theorem in two forms.

Theorem 1.2 ([6])
- First-order version: 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 that is complete for a language $\mathcal{L}$ that contains the tautologically true formulas and formulas (i)–(v).
- 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 $\mathbf{D}$.

Thus, in this model, there are some descriptions of beliefs that cannot be represented, including the BK sentence. The model relies on sets that are assumed to admit well-ordering in a classical logical framework. This observation is our starting point in this article.

1.3 Related literature and motivation

Due to its considerable impact on the various branches of game theory and logic, the BK paradox has gained an increasing interest in the literature. Now, we present a brief survey of the work that influenced the original BK paradox paper, and the work that is influenced by it.

A general framework for self-referential paradoxes was discussed earlier by Yanofsky in 2003 [38]. In his paper, Yanofsky used Lawvere's category theoretical arguments in some well-known mathematical arguments such as Cantor's diagonalization, Russell's paradox and Gödel's incompleteness theorems. Lawvere, on the other hand, discussed self-referential paradoxes in Cartesian closed categories in his early paper that appeared in 1969 [18]. Most recently, Abramsky and Zvesper used Lawvere's arguments to analyse the BK paradox in a category theoretical framework [1].

Pacuit approached the paradox from a modal logical perspective and presented a detailed investigation of the paradox in neighbourhood models and in hybrid systems [27]. Neighbourhood models are used to represent modal logics weaker than **K**, and can be considered as weak versions of topological semantics [8]. This argument was then extended to assumption-incompleteness in modal logics [39]. Mariotti *et al*., on the other hand, used compact belief models to represent interactive belief structures in a topological framework with further topological restrictions [23].

*

As we mentioned earlier, we approach the paradox from two different *non-classical* perspectives. However, before presenting the technical results, it is important to argue as to why and how such methods can be applicable to epistemic game theory.

1.3.1 Non-well-founded set theory in games

Non-well-founded (NWF) set theory is a mathematical theory of sets where the axiom of foundation is replaced by the *anti-foundation axiom* which is due to Mirimanoff [24]. Decades later, the axiom was re-formulated by Aczel within the domain of graph theory [2]. In NWF set theory, we can have true statements such as '$x \in x$', and such statements present interesting properties in game theory. NWF theories, in this respect, are natural candidates to represent circularity [4].

To the best of our knowledge, Lismont introduced non-well-founded type spaces to show the existence of universal belief spaces [21]. Then, Heifetz used NWF sets to represent type spaces and obtained rather sophisticated results [14]. He mapped a given belief space to its NWF version, and then proved that in the NWF version, epimorphisms become equalities. Harsanyi also noted earlier that circularity might be needed to express infinitary hierarchy of beliefs.

> 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*.
> [13]

Note that Harsanyi's concern for infinite regress or circularity is related to the epistemics of the game. However, some other ontological concerns can also be raised about the type spaces, and the way we define the states in the type spaces. In this respect, Heifetz motivated his approach, which is related to our perspective here, by arguing that NWF type spaces can be used 'once states of nature and types would no longer be associated with states of the world, but constitute *their very definition*.' [14, (emphasis in original)]. This is, indeed, a prolific approach to Harsanyi type spaces to represent uncertainty. Here is Heifetz on the very same issue.

> 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.
> [14, p. 204].

Notice that the paradoxical BK sentence appears in situations where the aforementioned belief interaction between the players assumes a central role. Yet, the classical model which is based on a well-founded set theory does not seem rich enough to express the level of interactiveness that the epistemic games require. However, we should not over-play our hand. It is important to note that NWF set theory is not immune to some of the problems that the classical set theory suffers from. For example, Russell's paradox is not solved in a NWF setting, and the subset relation stays the same in NWF theory [26]. The reason is quite straight-forward. As Heifetz also noted, 'Russell's paradox applies to the collection of all sets which do not contain themselves, not to the collection of sets which *do* contain themselves' [14, (emphasis in original)]. Therefore, we do not expect NWF set theory to completely solve or avoid the BK paradox. We will briefly revisit this issue when we discuss category theoretical tools.

The opponent of the use of NWF set theory might argue that it must be taken with a grain of salt. In this manner, Gerbrandy noted the following:

> ... [T]here are many 'more' Kripke models than there are possibilities of knowledge structures: each possibly corresponds [to] a whole class of bisimilar, but structurally different, models. In other words, a semantics for modal logic in the form of Kripke models has a finer structure than a semantics in terms of non-well-founded sets.
> [11]

In this work, we will observe how using an NWF background set theory can effect the BK paradox.

1.3.2 Paraconsistency in games

Paraconsistent logics allow us to construct belief models that are *inconsistency-friendly*. In paraconsistent models, contradictions do not necessarily trivialize the theory. Surprising enough, such methods have not been widely employed in game theory. Additionally, paraconsistent logics can be formalized by using a variety of mathematical methods including modal, topological, algebraic and category theoretical techniques. Some of such techniques are familiar to game theorists, yet they have not been extended to build a connection between non-classical logics and game theory. The lack of the aforementioned connection is one of our motivations in this article.

Also, notice that paraconsistent logics have originally been developed to analyse paradoxical situations, and to develop logical frameworks that can admit paradoxes without leading to trivialities [7, 9, 29]. Therefore, investigating the BK paradox within a paraconsistent framework seems only natural.

An interesting application of paraconsistency has been dialogical and discursive logics [17, 31, 32]. Similar to Hintikkan game semantics, dialogical logics present a pragmatical semantics for a variety of logics. The game element in such systems is not very strong and central. Akin to Hintikka's game semantics, the moves are specified syntactically, and the elements of rationality, strategies (mixed or pure) and epistemics are not directly evident in dialogical games. Nevertheless, they take the very first step to introduce the notion of paraconsistency in game-like situations. The literature on the connection between logic and games is rich [34]. Yet, the relation between *non-classical logics* and games is not widely studied. This article is an attempt to fill this gap.

2 Non-Well Founded set theoretical approach

We start with defining belief models using NWF sets. What we call an NWF model is a tuple $M=(W,V)$ where W is a non-empty NWF set (*hyperset*, for short), and V is a valuation assigning propositional variables to the elements of W. The semantics of (basic) modal logic in the NWF setting is given as follows where we use the symbol $\models^+$ to represent the satisfaction relation in a NWF model [11].

$$M,w\models^+\Diamond\varphi \quad \text{iff} \quad \exists v\in w \text{ such that } M,v\models^+\varphi$$
$$M,w\models^+\Box\varphi \quad \text{iff} \quad \forall v.v\in w \text{ implies } M,v\models^+\varphi$$

Based on this definition, we can now give a non-standard semantics for the belief and assumption modalities $\Box^{ij}$ and $\heartsuit^{ij}$, respectively, for $i,j\in\{a,b\}$.

$$M,w\models^+\Box^{ij}\varphi \quad \text{iff} \quad M,w\models^+\mathbf{U}^i\wedge \forall v(v\in w\wedge M,v\models^+\mathbf{U}^j \text{ implies } M,v\models^+\varphi)$$
$$M,w\models^+\heartsuit^{ij}\varphi \quad \text{iff} \quad M,w\models^+\mathbf{U}^i\wedge \forall v(v\in w\wedge M,v\models^+\mathbf{U}^j \text{ iff } M,v\models^+\varphi)$$

Several comments on the NWF semantics are in order here. First, notice that this definition of NWF semantics for belief and assumption modalities depends on the earlier modal definition of those operators given in [6, 11]. Secondly, belief or assumption of a formula φ at a state w is defined in terms of the truth of φ at the states that constitutes w, including possibly w itself. Therefore, these definitions address the philosophical and foundational points that Heifetz made about the uncertainty in type spaces. We call a belief state $w\in W$ a *Quine state* if $w=\{w\}$. Similarly, a state w is called an *urelement* if it is not the empty set, and it can be a member of a set but cannot have members. Finally, we call a set A *transitive* if $a\in A$ and $b\in a$, then $b\in A$.

For example, consider the model $W=\{w,v\}$ with $V(p)=\{w\}$ and $V(q)=\{v\}$ with a language with two propositional variables for simplicity. Let us assume that both w and v are Quine states. What does it mean to say that the player a assumes p at a Quine state w in NFW belief models? The following theorem establishes how belief and assumption operators work in Quine states.

THEOREM 2.1
Let $M=(W,V)$ be an NWF belief model with disjoint type spaces U^a and U^b respectively for two players a and b. If $w\in U^i$ is a Quine state or an urelement belief state for $i\in\{a,b\}$, then i assumes φ at w if and only if $M,w\not\models^+\varphi$. Moreover, i believes in any formula ψ at w.

PROOF. Let us start with considering the Quine states.

Without loss of generality, let $w\in U^a$ where w is a Quine state. Suppose $w\models^+\heartsuit^{ab}\varphi$. Since $w\in U^a$, $w\models^+\mathbf{U}^a$. Since, $w\in w$, and U^a and U^b are disjoint, we have $w\not\models^+\mathbf{U}^b$. Therefore, since $w\models^+\heartsuit^{ab}\varphi$, we conclude $w\not\models^+\varphi$.

For the converse direction, suppose that $w\not\models^+\varphi$. Since $w\in w$, and $w\not\models^+\mathbf{U}^b$ the biconditional is satisfied. Moreover, since $w\in U^a$, $w\models+\mathbf{U}^a$. Therefore, $w\models^+\heartsuit^{ab}\varphi$.

Similarly, now, without loss of generality, let $w\in U^a$ be an urelement. Then, $w\models^+\mathbf{U}^a$. For the left-to-right direction, note that since there is no $v\in w$, the conditional is vacuously satisfied. For the right-to-left direction, suppose $w\not\models^+\varphi$. Since, $w\notin w$ and $w\not\models^+\mathbf{U}^b$, the biconditional is satisfied again.

The proof for the belief operator follows immediately from the definitions for Quine states and urelements as assumption implies belief. ∎

Notice that the above proof heavily depends on the fact that the type spaces for the players are assumed to be disjoint. Let us now see how belief models change once we allow the intersection of NWF type spaces.

THEOREM 2.2
Let $M=(W,V)$ be an NWF belief model for two players a and b where U^a and U^b are not necessarily disjoint. For a Quine state w and different $i,j\in\{a,b\}$, $w\models^+\heartsuit^{ij}\top$ if and only if $w\in U^a\cap U^b$. In other words, Quine states with true assumptions belong to the both players.

PROOF. From left-to-right direction, assume without loss of generality that the Quine state w is in U^a. So, $w\models^+\mathbf{U}^a$. Suppose $w\models^+\heartsuit^{ab}\top$. Now, we will show that $w\in U^b$.

By the definition of the assumption modality, and the fact that $\top$ is everywhere true, we observe that $w\models^+\mathbf{U}^b$. Thus, $w\in U^b$ as well which gives $w\in U^a\cap U^b$.

From right-to-left direction, assume that $w\in U^a\cap U^b$, where w is a Quine state. Thus, $w\models^+\mathbf{U}^a$ and $w\models^+\mathbf{U}^b$. Thus, $\forall v(v\in w\wedge v\models^+\mathbf{U}^b)$ is satisfied. Also, by definition, $\top$ is satisfied everywhere, hence $v\models^+\top$. Therefore, by definition, we have $w\models^+\heartsuit^{ab}\top$. By a symmetric argument, one can easily show $w\models^+\heartsuit^{ba}\top$. Thus, we conclude that for a Quine state w and different $i,j\in\{a,b\}$, $w\models^+\heartsuit^{ij}\top$.

This concludes the proof. ∎

A game-theoretical implication of Theorem 2.2 is worth mentioning. Notice that Quine states correspond to the states that are reflexive. In other words, at a Quine state w, player i considers w possible for player j. Thus, such a state w is forced to be in the intersection of the type spaces.

On the other hand, intersecting type spaces do not seem to create a problem for belief models. To overcome this issue, one can introduce a *turn* function from the space of the belief model to the set of players assigning states to players. The functional definition of this construction necessitates that every state should be assigned to a unique player. Therefore, the game can determine whose turn it is at Quine atoms. Additionally, urelements, since they cannot have elements, are end states in games. At such states, players do not consider any states possible for the other players.

Now, based on the NWF semantics we gave earlier, it is not difficult to see that the following formulas discussed in the original paper are still valid as before if we maintain the assumption of the disjointness of type spaces.

$$\Box^{ab}\mathbf{U}^b \leftrightarrow \mathbf{U}^a,\ \ \Box^{ba}\mathbf{U}^a \leftrightarrow \mathbf{U}^b,\ \ \Box^{ab}\mathbf{U}^a \leftrightarrow \bot,\ \ \Box^{ba}\mathbf{U}^b \leftrightarrow \bot$$

Furthermore, the following formulas are not valid as before.

$$\Box^{ab}\mathbf{U}^b \to \mathbf{U}^b,\ \ \Box^{ab}\mathbf{U}^b \to \Box^{ba}\Box^{ab}\mathbf{U}^b,\ \ \Box^{ab}\mathbf{U}^b \to \Box^{ab}\Box^{ab}\mathbf{U}^b$$

However, for the sake of the completeness of our arguments, let us, for the moment, allow that type spaces may not be disjoint.

Consider an NWF belief model (W,V) where $w=\{w\}$ with $U^a=U^b=W$. In such a model $\Box^{ab}\mathbf{U}^a \leftrightarrow \bot$ fails, but $\Box^{ab}\mathbf{U}^a \leftrightarrow \top$ is satisfied. Similar observations can be made for $\Box^{ba}\mathbf{U}^b \leftrightarrow \bot$ and $\Box^{ba}\mathbf{U}^b \leftrightarrow \top$. Similarly, all $\Box^{ab}\mathbf{U}^b \to \mathbf{U}^b$, $\Box^{ab}\mathbf{U}^b \to \Box^{ba}\Box^{ab}\mathbf{U}^b$, and $\Box^{ab}\mathbf{U}^b \to \Box^{ab}\Box^{ab}\mathbf{U}^b$ are satisfied in the aforementioned NWF model.

Now, our aim is to construct an NWF belief model in which the Lemma 1.1 fails. For our purposes, however, we still maintain the assumption that the type spaces be disjoint as in the original BK paper.

As a first step, we redefine the diagonal set in the NWF setting. Recall that, in the standard case, diagonal set D is defined with respect to the accessibility relation P which we defined earlier. In NWF case, we will use membership relation for that purpose. We put $D^+ := \{w \in W : \forall v \in W.(v \in w \to w \notin v)\}$. We also define $\mathbf{D}^+$ as the propositional symbol with the valuation set D^+.

Now, we observe how the NWF models make a difference in the context of the BK paradox. Notice that BK argument relies on two lemmas which we have mentioned earlier in Lemma 1.1. Now, we present counter-models to Lemma 1.1 in NWF theory.

Proposition 2.3
In an 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 $W=\{w,v\}$ with $w=\{v\}$, $v=\{w\}$ where $U^a=\{w\}$ and $U^b=\{v\}$. To maintain the disjointness of the types, assume that neither w nor v is transitive.

Then, $w \models^+ \heartsuit^{ab}\mathbf{U}^b$ since all states in b's type space is assumed by a at w. Similarly, $v \models^+ \heartsuit^{ba}\mathbf{U}^a$ as all states in a's type space is assumed by b at v. Then, $w \models^+ \Box^{ab}\heartsuit^{ba}\mathbf{U}^a$. Continuing this way, we conclude, $w \models^+ \Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$.

However, by design, $w \not\models^+ \mathbf{D}^+$ since $v \in w$ and $w \in v$. Thus, the formula $\Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a \wedge \neg\mathbf{D}^+$ is satisfiable as well. ■

Proposition 2.4
The formula $\Box^{ab}\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D}^+)$ is satisfiable in some NWF belief model.

Proof. Take a non-transitive model (W,V) with $W=\{w,v,u,t\}$ where $w=\{v,w\}$, $v=\{u\}$, and $u=\{t\}$ where $u \notin t$. Let $U^a=\{w,u\}$, and $U^b=\{v,t\}$. Now, observe that the formula $\mathbf{U}^a \wedge \mathbf{D}^+$ is satisfiable only at u (as $w \in w$, w does not satisfy $\mathbf{U}^a \wedge \mathbf{D}^+$). Now, $v \models^+ \heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D}^+)$. Finally, $w \models^+ \Box^{ab}\heartsuit^{ba}\mathbf{U}^a \wedge \mathbf{D}^+$. Note that even if $w \in w$ as $w \notin U^b$, by definition of the box modality, w satisfies $\Box^{ab}\heartsuit^{ba}\mathbf{U}^a \wedge \mathbf{D}^+$. ■

Therefore, the Lemma 1.1 is refuted in NWF belief models. Notice that Lemma 1.1 is central in Brandenburger and Keisler's proof of the incompleteness of belief structures. We now construct a counter-model for Theorem 1.2 in the NWF setting using hypersets.

Consider the following counter-model. Let $W=\{w,v,u,t,r,s,x,y,z\}$ with $w=\{v,w\}$, $v=\{u\}$, $u=\{t\}$ $(u \notin t)$, $r=\{v,t,s,y\}$, $s=\{w,u,r,x,z\}$, $x=\{x,s\}$, $y=\{y,x\}$, $z=\{z,y\}$ where $U^a=\{w,u,r,x,z\}$ and

$U^b = \{v, t, s, y\}$. Now, step by step, we observe the following.

- $\mathbf{U} \wedge \mathbf{D}^+$ is satisfied only at u, since we have $w \in w$, $r \in s \wedge s \in r$, $x \in x$ and $z \in z$
- No hole at $\mathbf{U}^a$ as $s \models^+ \heartsuit^{ba}\mathbf{U}^a$
- No hole at $\mathbf{U}^b$ as $r \models^+ \heartsuit^{ab}\mathbf{U}^b$
- No big hole at $\heartsuit^{ba}\mathbf{U}^a$ as $x \models^+ \Box^{ab}\heartsuit^{ba}\mathbf{U}^a$
- No big hole at $\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$ as $y \models^+ \Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$
- No big hole at $\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$ as $z \models^+ \Box^{ab}\Box^{ba}\Box^{ab}\heartsuit^{ba}\mathbf{U}^a$
- No hole at $\mathbf{U}^a \wedge \mathbf{D}^+$ as $v \models^+ \heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D}^+)$
- No big hole at $\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D}^+)$ as $w \models^+ \Box^{ab}\heartsuit^{ba}(\mathbf{U}^a \wedge \mathbf{D}^+)$

The crucial point in the semantical evaluation of big holes is the fact that the antecedent of the conditional in the definition of the box modality is not satisfied if some elements of the current states are not in the desired type space. Therefore, the box modality is still satisfied if the current state has some elements from the same type space. This helped us to construct the counter-model.

This counter-model shows that Theorem 1.2 in its stated form 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 does, however, establish the fact that they do not have the same holes as the standard belief models. We will revisit this question later on, and give an answer from a category theoretical perspective.

3 Paraconsistent approach

Paraconsistency is the umbrella term for logical systems where *ex contradictione quodlibet* fails. Namely, in paraconsistent logics, for some formulas φ, ψ, we have $\varphi, \neg\varphi \nvdash \psi$. The semantical equivalence of this proof theoretical rule is that some contradictory statements are satisfiable in paraconsistent models. Endorsing paraconsistent logic does not necessarily entail that *all* contradictions are true. It simply means that the existence of contradictions does not trivialize the model while absurdity ($\bot$) always leads to trivial theories. Thus, in paraconsistent logics, there are some contradictions which are not absurd.

Paraconsistent logics can be captured by using several algebraic, topological and category theoretical formalisms. We will approach paraconsistency from such directions, and analyse the BK paradox from within these formalisms.

3.1 Algebraic and category theoretical approach

A recent work on the BK paradox shows the general pattern of self-referential paradoxical cases, and gives some positive results including a fixed-point theorem [1]. In this section, we instantiate the fixed-point results of the aforementioned reference to some other mathematical structures that are inherently inconsistency friendly. This will allow us to build a counter-model for the BK paradox.

There has been offered a variety of different logical and algebraic formalisms to represent paraconsistent logics [29]. First, we discuss co-Heyting algebras as they provide a broader perspective for paraconsistency.

Definition 3.1
Let L be a bounded distributive lattice. If there is 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 $\backslash : L \times L \to L$ such that

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

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

Some immediate examples of co-Heyting algebras are the closed subsets of a given topological space, and subtopoi of a given topos [5, 19, 25, 29].

The operators $\Rightarrow$ and $\backslash$ give rise to two different negations. The *intuitionistic negation* $\dot{\neg}$ is defined as $\dot{\neg}\varphi \equiv \varphi \to \mathbf{0}$ while the *paraconsistent negation* $\sim$ as $\sim\varphi \equiv \mathbf{1} \backslash \varphi$ where $\mathbf{0}$ and $\mathbf{1}$ are the bottom and the top elements of the lattice respectively. Therefore, $\dot{\neg}\varphi$ is the largest element disjoint from φ, and $\sim\varphi$ is the smallest element whose join with φ gives the top element $\mathbf{1}$ [33]. In a Boolean algebra both intuitionistic and paraconsistent negations coincide, and give the usual Boolean negation where we interpret $\varphi \Rightarrow \psi$ as $\neg\varphi \vee \psi$, and $\varphi \backslash \psi$ as $\varphi \wedge \neg\psi$ with the Boolean negation sign $\neg$. Algebraic structures such as co-Heyting algebras can be approached from a category theoretical point of view, which was used to analyse the BK paradox. Such results are based on Lawvere's celebrated theorem of 1969. Before discussing Lawvere's argument, we need to define *weakly point surjective* maps.

Definition 3.2
An arrow $f : A \times A \to B$ is called *weakly point surjective* if for every $p : A \to B$, there is an $x : \mathbf{1} \to A$ such that for all $y : \mathbf{1} \to A$ where $\mathbf{1}$ is the terminal object, we have $p \circ y = f \circ \langle x, y \rangle : \mathbf{1} \to B$. In this case, we say, p is represented by x.

We still need some basic concepts from the category theory. Let us mention them here for the completeness of our treatment. A category is *Cartesian closed* (CCC for short, henceforth), if it has a terminal object, and admits products and exponentiation. A set X is said to have the *fixed-point property* for a function f, if there is an element $x \in X$ such that $f(x) = x$. Category theoretically, an object X is said to have the fixed-point property if and only if for every endomorphism $f : X \to X$, there is $x : \mathbf{1} \to X$ with $xf = x$ [18].

Theorem 3.3 ([18], Lawvere's Lemma)
In any Cartesian closed category, if there exists an object A and a weakly point-surjective morphism $g : A \to Y^A$, then Y has the fixed-point property for g.

It was observed that CCC condition can be relaxed, and Lawvere's Theorem works for categories that have only finite products [1].[1]

What is the connection then between Theorem 3.3 and the BK paradox? Abramsky and Zvesper showed that it is possible to reduce Lawvere's Lemma to the BK paradox and *vice versa* [1]. Let us briefly mention the argument here.

First of all, the authors observe that under the assumptions of interactive belief models, every unary operator admits a fixed point. They define the predicate $p(x_0)$ as follows:

$$p(x_0) = \exists y.(R^a(x_0, y) \wedge R^b(y, x_0))$$

[1]This point was already made by Lawvere and Schanuel in *Conceptual Mathematics*. Thanks to Noson Yanofsky for pointing this out.

Notice that this predicate stands for the diagonal set we have defined earlier. Based on this formulation and under the assumptions of interactive belief models, they define the following:

$$q(x) = \exists y.(R^a(x,y) \wedge R^b(y,x))$$
$$p(x) = O(q(x))$$

Then, Abramsky and Zvesper observe that $q(x_0)$ is the fixed point of the unary O operator, which can be verified given the above statements: Substitute x_0 for x in $q(x)$ to obtain $\exists y.(R^a(x_0,y) \wedge R^b(y,x_0))$, which is $p(x_0)$. On the other hand, as $p(x) = O(q(x))$, we observe that $p(x_0) = O(q(x_0))$. Thus, $O(q(x_0)) = q(x_0)$ yielding that $q(x_0)$ is a fixed-point for the unary operator O. Here, the authors note that the BK sentence is obtained if O is taken as the classical negation operator as $\neg q(x_0) = q(x_0)$ is impossible in classical Boolean logic. Now, what needs to be done is to find a weakly point surjective mapping that acts as the fixed point $q(x_0)$ on a CCC. The authors achieve it in a categorical logic by reformulating the weakly point surjective mappings, and then showing that fixed points still exist, generating the BK sentence. In other words, Abramsky and Zvesper reduced the Lawvere's Theorem (Theorem 3.3) to the following.

Theorem 3.4
[1] Given the following two assumptions for all predicates p on U^a, every unary operator O has a fixed point.

- Assumption 1: $R^a(x_0) \subseteq \{y : R^b(y) = \{x : p(x)\}\}$
- Assumption 2: $\exists y.R^a(x_0,y)$

This approach relies on Theorem 3.3 and requires a CCC which gives us some more freedom to go beyond classical logic. Now, we will inquire if the same category theoretical qualities carry over to some other logics, particularly to paraconsistent logic. Let us start by investigating the category theoretical properties of co-Heyting algebras and the category of hypersets. First, recall that the category of Heyting algebras is a CCC. A canonical example of a Heyting algebra is the set of opens in a topological space [3]. The objects of such a category will be the open sets. The unique morphisms in that category exists from O to O' if $O \subseteq O'$. What about the co-Heyting algebras? We now state the following dual statement.

Proposition 3.5
Co-Heyting algebras are Cartesian closed categories.

Proof. Let $(L, \setminus)$ be a co-Heyting algebra. First, observe that the element **0** is the terminal element (dual of the **1** in Heyting algebra). Secondly, for x,y in L, the product exists and is defined as $x \wedge y$ as $x \wedge y \leq x$, and $x \wedge y \leq y$. Moreover, $x \leq y$ and $x \leq z$ imply that $x \leq y \wedge z$. Thirdly, the exponent x^y is defined as $x \wedge \neg y$. Notice that we also write this as $x \setminus y$. The evaluation of subtraction is $x \leq (x \setminus y) \vee y$. Unravelling the definition, we observe that the definition is sound:

$$x \leq (x \setminus y) \vee y = (x \wedge \neg y) \vee y = (x \vee y) \wedge (\neg y \vee y) = x \vee y$$

as we always have $x \leq x \vee y$. Thus, the terminal object, finite products and exponents exist in $(L, \setminus)$ rendering it a CCC. ∎

Example 3.6
As we have mentioned, the co-Heyting algebra of the closed sets in a topology is a well-known example of a CCC. Similar to the arguments that show that open set topologies are CCC, we can observe that closed set topologies are CCC as well.

Given two objects C_1, C_2, we define the unique arrow from C_1 to C_2, if $C_1 \supseteq C_2$. The product is the union of C_1 and C_2 as the finite union of closed sets exists in a topology. The exponent $C_1^{C_2}$ is then defined as $\mathsf{Clo}(\overline{C_1} \cap C_2)$ where $\overline{C_1}$ is the complement of C_1.

Now, we have the following result based on Theorems 3.3 and 3.4.

THEOREM 3.7
In a co-Heyting algebra, if there is an object A and a weakly point-surjective morphism $g : A \to Y^A$, then Y has the fixed-point property. Therefore, there exists a co-Heyting algebraic model with a satisfiable BK sentence.

PROOF. As we already observed, co-Heyting algebras are CCCs. Take one: $(L, \backslash)$. Let L be the object, and take $\backslash$ as the morphism. We first show that the join mapping $\vee : L \times L \to \{0,1\}$ is weakly point-surjective, where the join is defined as usual: $a \backslash b := a \vee {\sim} b$ [3]. Now, for any proposition $p : L \to \{0,1\}$, there is a negation $x : \mathbf{1} \to \{0,1\}$ which is defined as the smallest element whose join with p gives $\mathbf{1}$. Denote it as ${\sim}p$. Then, we observe that for any other proposition $y : L \to \{0,1\}$, we have $y \backslash p = g(x,y)$ which gives $y \backslash p = y \vee {\sim} p$. The function g here is from $A \times A$ to Y, so it can easily be re-written as a mapping from A to Y^A. Therefore, g is a weakly point-surjective morphism. Now, Theorem 3.3 applies. Thus, taken as a CCC, co-Heyting algebras admits fixed points on $\{0,1\}$.

By Theorem 3.4, we know that unary operators admit fixed points. Take the unary operator $\sim$ as we defined above on the set $\{0,1\}$ (which is the set Y in the statement of the theorem). It is now possible to define the boundary ∂ of a formula p as follows: $\partial(p) = p \wedge {\sim} p$. As ${\sim} p := 1 \backslash p$, and by De Morgan's Laws, we observe that for all predicates p, $\partial(p)$ is satisfied, producing $\mathbf{1}$ due to the following reasoning.

$$\begin{aligned}\partial(p) &= p \wedge {\sim} p \\ &= \mathbf{1} \backslash ((\mathbf{1} \backslash p) \backslash (\mathbf{1} \backslash p)) \\ &= \mathbf{1} \backslash \mathbf{1} \\ &= \mathbf{1}\end{aligned}$$

So, by Theorem 3.4, for all predicates p, we have a unary operator $\sim$ with a fixed point that satisfies the contradictory statement $p \wedge {\sim} p$. Take p as the BK statement to conclude the proof. ∎

Now, we show that a similar approach works for NWF sets. In our earlier discussion, we presented some counter-models for the classical BK sentence. However, we have not concluded that NWF models are complete. We need Lawvere's Lemma to show that NWF belief models cannot be complete. Consider the category **AFA** of hypersets with total maps between them.[2] Category **AFA** admits a final object $\mathbf{1} = \{\emptyset\}$. Moreover, it also admits exponentiation and products in the usual sense, making it a CCC. Thus, Lawvere's Lemma applies.

COROLLARY 3.8
There exists an impossible BK sentence in non-well-founded interactive belief structures.

PROOF. The argument is very similar to the classical case, so we present a sketch of the proof. As we argued earlier, with the classical negation and a CCC, Lawvere's Theorem shows the existence of fixed points. The above argument shows that **AFA** is a CCC, thus Theorem 3.3 applies. Together with Theorem 3.4 and with the fact that the negation remains classical in **AFA**, we conclude that an impossible BK sentence exists in NWF interactive belief structures by taking the unary operator as the classical Boolean negation, similar to the argumentation above. ∎

[2] Thanks to Florian Lengyel for pointing this out.

The above corollary translates to the fact that in NWF models with hypersets, the BK sentence still exists as a fixed point. As we already argued, the holes in NWF models are not the same as the ones in classical belief models. This is not surprising as Russell's paradox is not solved with hypersets.

3.2 Topological approach

In the previous section, we observed algebraically that it is possible to have a contradictory BK sentence satisfiable in some inconsistency-friendly systems. Now, by using topological models, we construct the belief models that were shown to be possible by the algebraic methods. In our construction, we will make use of relational representation of belief models which in turn produce belief and assumption modalities. We will then interpret those modalities over paraconsistent topological models. This result is important as it *constructs* a topological belief model with a *satisfiable* BK sentence.

Topological structures play an essential role in various paraconsistent logics and algebraic structures. In an early paper, Lawvere pointed out the role of boundary operator in co-Heyting algebras [19]. In a similar fashion, boundaries play a central role to give topological semantics for paraconsistent logics [5, 12, 25]. In the original BK paper where the paradox is first introduced, the authors discussed several complete models including topologically complete models where their topological space is a compact metrizable space satisfying several further conditions [6]. Our approach, however, does not depend on the topological qualities of the space *per se*, but rather depends on the topological semantics we build on it.

In the topological semantics for the classical modal logic, topological interior Int and closure Clo operators are identified with $\Box$ and $\Diamond$ modalities respectively. Then, the extension $[\cdot]$ of a modal formula $\Box\varphi$ is given as follows $[\Box\varphi] := \mathsf{Int}([\varphi])$. In the classical setting, in general, open or closed sets are produced by the modal operators. Thus, the extensions of Booleans are not necessarily topologically open or closed in the classical case. At this stage, we can take one step further, and stipulate that the extensions of propositional variables to be closed sets. This stipulation works well with conjunction and disjunction as the finite intersection (and respectively, the union) of closed sets is closed. However, it is not straightforward for negation as the complement of a closed set is not necessarily closed, but open. Therefore, we define a special negation, the paraconsistent negation, $\sim$ as the *closure of the complement*. Then, we obtain a co-Heyting algebra of closed sets as we have observed in Example 3.6. In this setting, inconsistent theories are the ones that include the formulas that are true at the boundaries [5, 25].[3]

The following is a step by step construction of the BK sentence in a paraconsistent topological setting. We will call these belief models *paraconsistent topological interactive belief models*.

For the agents a and b, we take corresponding non-empty type spaces A and B, and define closed set topologies τ_A and τ_B on A and B, respectively. Furthermore, 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 call the structure $F = (A, B, \tau_A, \tau_B, t_A, t_B, V)$ a paraconsistent topological interactive belief model with a valuation V. Here, the set A represents the possible epistemic states of the player a in which he/she holds beliefs about player b, or about b's beliefs etc., and vice versa for the set B and the player a, and the topologies represent those beliefs. For instance, for player

[3] Dually, if we stipulate that the extension of the propositional variables to be open sets, we obtain a model of intuitionistic logic with intuitionistic negation. Thus, the duality of intuitionistic and paraconsistent logics is rather clear if the topological semantics is adopted.

a at the state $x \in A$, t_A returns a closed set in $Y \in \tau_B \subseteq \wp(B)$. In this case, we write $t_A(x, Y)$ which means that at state x, player a believes that the states y in $Y \in \tau_B$ are possible for the player b, i.e. $t_A(x,y)$ for all $y \in Y$. Moreover, 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.

The modal language which we use has two modalities representing the beliefs of each agent. Akin to some earlier modal semantics for the paradox, we give a topological semantics for the BK argument in paraconsistent topological interactive belief models [6, 27]. Let us first give the formal language which we use. The language for our belief models is given as follows.

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

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

The constant $\top$ is true everywhere, and we discussed the semantics of the negation already. Also, the semantics for the conjunction is as usual. For $x \in A, y \in B$, the semantics of the modalities is given as follows with a modal valuation attached to F.

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

We define the dual modalities $\Diamond_a$ and $\Diamond_b$ as usual with the paraconsistent negation: $\Diamond_i \varphi := \sim\Box_i \sim\varphi$ for $i \in \{a, b\}$.

Now, we have sufficient tools to represent the BK sentence in our paraconsistent topological belief structure with respect to a state x_0:

$$x_0 \models \Box_a \boxplus_b \varphi \wedge \Diamond_a \top$$

Let us analyse this formula in paraconsistent topological belief models. Notice that the second conjunct guarantees that for the given $x_0 \in A$, there exists a corresponding set $Y \in \tau_B$ with $t_A(x, Y)$. On the other hand, the first conjunct deserves closer attention:

$$\begin{array}{lll} x_0 \models \Box_a \boxplus_b \varphi & \text{iff} & \exists Y \in \tau_B \text{ with } t_A(x_0, Y) \text{ implies } \forall y \in Y.\ y \models \boxplus_b \varphi \\ & \text{iff} & \exists Y \in \tau_B \text{ with } t_A(x_0, Y) \text{ implies} \\ & & [\forall y \in Y, \exists X \in \tau_A \text{ with } t_B(y, X) \text{ iff } \forall x \in X.\ x \models \varphi] \end{array}$$

Based on the above observation, let us show that $x_0 \models \Box_a \boxplus_b \varphi \wedge \Diamond_a \top$ can be satisfied for an inconsistent φ. For simplicity, let φ be $p \wedge \sim p$ for a propositional variable p. Denote the extension of φ with X_0, so $|p| = X_0$. Pick $x_0 \in \partial X_0$ where $\partial(\cdot)$ operator denotes the boundary of a set $\partial(\cdot) = \mathsf{Clo}(\cdot) - \mathsf{Int}(\cdot)$. By the assumptions of our framework, X_0 is closed. Moreover, by simple topology ∂X_0 is closed as well. By the second conjunct of the formula in question (i.e. $\Diamond_a \top$), we know that some $Y \in \tau_B$ exists such that $t_A(x_0, Y)$. Now, for all y in Y, we make an additional supposition and associate y with ∂X_0 giving $t_B(y, \partial X_0)$. We know that for all $x \in \partial X_0$, we have $x \models p$ as $\partial X_0 \subseteq X_0$. Moreover, for all $x \in \partial X_0$, $x \models \sim p$. The reason for that is the following. By definition, $|\sim p| = \mathsf{Clo}(\overline{X_0})$ where $\overline{X_0}$ is the set theoretical complement of X_0. Thus, $\partial(X_0) \subseteq \mathsf{Clo}(\overline{X_0})$, which gives $\partial(X_0) \subseteq |\sim p|$. Thus, we conclude that $x_0 \models \Box_a \boxplus_b (p \wedge \sim p) \wedge \Diamond_a \top$ for some *carefully selected* x_0.

In this construction, we have several suppositions. First, we pick the actual state from the boundary of the extension of some proposition (ground or modal). Secondly, we associate the epistemic accessibility of the second player to the same boundary set. Namely, a's beliefs about b includes his/her current state.

Now, the BK paradox appears when one substitutes φ with the following diagonal formula (whose extension is a closed set by definition of the closed set topology), hence breaking the aforementioned circularity:

$$D(x) = \forall y.[t_A(x,y) \to \ \sim t_B(y,x)]$$

The set $D(x)$, as before, can be represented with a special propositional symbol $\mathbf{D}$, as in Section 1.2. In this case, the formula $\mathbf{D}$ will be satisfied at the state x by the set $D(x)$. The BK impossibility theorem asserts that, under the seriality condition, there is no such x_0 satisfying the following formula $x_0 \models \Box_a \boxplus_b \mathbf{D}$.

$$\begin{aligned} x_0 \models \Box_a \boxplus_b \mathbf{D} \quad \text{iff} \quad & \exists Y \in \tau_B \text{ with } t_A(x_0, Y) \text{ implies} \\ & [\forall y \in Y, \exists X \in \tau_A \text{ with } t_B(y, X) \text{ iff} \\ & \forall x \in X.\ x \models \forall y'.(t_A(x,y') \to \ \sim t_B(y',x))] \end{aligned}$$

Motivated by our earlier discussion, let us analyse the logical statement in question. Let X_0 satisfy the statement $t_A(x,y')$ for all $y' \in Y$ and $x \in X_0$ for some Y. Then, $\partial X_0 \subseteq X_0$ will satisfy the same formula. Similarly, let $\sim X_0$ satisfy $\sim t_B(y',x)$ for all $y' \in Y$ and $x \in X_0$. Then, by the similar argument, $\partial(\sim X_0)$ satisfy the same formula. Since $\partial(X_0) = \partial(\sim X_0)$, we observe that any $x_0 \in \partial X_0$ satisfy $t_A(x,y')$ and $\sim t_B(y',x)$ with the aforementioned quantification. Thus, such an x_0 satisfies $\Box_a \boxplus_b D(x)$. Therefore, the states at the boundary of some closed set *satisfy* the BK sentence in paraconsistent topological belief structures. Thus, this is a counter-model for the BK sentence in the paraconsistent topological belief models.

Theorem 3.9
The BK sentence is satisfiable in some paraconsistent topological belief models.

Proof. See the above discussion for the proof which gives a model that satisfies the BK sentence. The argument is very similar to the algebraic proof that shows the existence of a satisfiable BK sentence in co-Heyting algebras.

Let φ denote the BK sentence. Then, $\Box_a \boxplus_b \varphi \wedge \Diamond_a \top$ denotes the BK sentence. Let the extension of φ be X_0 where X_0 satisfies the statement $t_A(x,y')$ for all $y' \in Y$ and $x \in X_0$ for some Y. Pick an arbitrary $x_0 \in \partial X_0$. Then, as we already have observed, $x_0 \models \Box_a \boxplus_b \varphi \wedge \Diamond_a \top$.

This shows the existence of a state x_0 and a model which satisfies the BK sentence. ∎

3.3 Product topologies

In the previous section, we introduced t_A and t_B to represent the belief interaction between the players. However, topological models provide us with a variety of tools to combine topological spaces to model various interactions between modalities and agents/players [10].

In this section, we make use of product topologies to represent belief interaction among the players. The novelty of this approach is not only to economize on the notation, but also to present a more natural way to represent the belief interaction. For our purposes here, we only consider two-player games as our results can easily be generalized to n-player. Here, we resort to a variety of formalisms presented in some recent works [35, 36].

Definition 3.10
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 interactive belief model for two agents is given as $(A \times B, \tau_A \times \tau_B, V)$ where V is the valuation function.

In this framework, we assume that the topologies are *full* on their sets—namely $\bigcup \tau_A = A$, and likewise for B. In other words, we do not want any non-expressibility results just because the given topologies do not cover such states. If the topologies are not full, we can reduce the given space to a subset of it on which the topologies are full without losing any epistemic expressibility.[4]

In this framework, if player a believes proposition $P \subseteq B$ at state $x \in A$, we stipulate that there is a closed set $X \in \tau_A$ such that $x \in X$ and a closed set $Y \in \tau_B$ with $Y \subseteq P$, all implying $X \times Y \in \tau_A \times \tau_B$. Player a assumes P if $Y = P$, and likewise for player b. Similar to the previous section, we make use of paraconsistent topological structures with closed sets and paraconsistent negation.

Here, we consider the same syntax of the formulas, and for $x \in A$ and $y \in B$, we give the semantics of the modalities as follows as the semantics of the Booleans are identical to the paraconsistent topological belief models which we discussed earlier.

$$(x,y) \models \Box_a \varphi \quad \text{iff} \quad \exists(\{x\}, Y) \in \tau_A \times \tau_B \text{ implies } \forall (x,y) \in (\{x\}, Y).(x,y) \models \varphi$$
$$(x,y) \models \boxplus_a \varphi \quad \text{iff} \quad \exists(\{x\}, Y) \in \tau_A \times \tau_B \text{ iff } \forall (x,y) \in (\{x\}, Y).(x,y) \models \varphi$$
$$(x,y) \models \Box_b \varphi \quad \text{iff} \quad \exists(X, \{y\}) \in \tau_A \times \tau_B \text{ implies } \forall (x,y) \in (X, \{y\}).(x,y) \models \varphi$$
$$(x,y) \models \boxplus_b \varphi \quad \text{iff} \quad \exists(X, \{y\}) \in \tau_A \times \tau_B \text{ iff } \forall (x,y) \in (X, \{y\}).(x,y) \models \varphi$$

Notice that, now, product topological paraconsistent interactive belief models do not have the functions t_A or t_B unlike the models we discussed in the previous section as the topological product operator $\times$ is defined to represent the interaction between the agents. This is how we economize in the notation.

Given a set $S \subseteq A \times B$, we say that S is *horizontally closed* if for any $(x,y) \in S$, there exists a closed set X with $x \in X \in \tau_A$ and $X \times \{y\} \subseteq S$. Similarly, S is *vertically closed* if for any $(x,y) \in S$, there exists a closed set Y with $y \in Y \in \tau_B$, and $\{x\} \times Y \subseteq S$ [35, 36]. In this framework, player a at $x \in A$ is said to believe a set $Y \subseteq B$ if $\{x\} \times Y$ is vertically closed. We can also define assumption-complete structures in product topologies.

Definition 3.11
For a given language $\mathcal{L}$ for our belief model, let $\mathcal{L}^a$ and $\mathcal{L}^b$ be the families of all subsets of A and B, respectively. Then, we observe that by assumption-completeness, we require every non-empty set $Y \in \mathcal{L}^b$ is assumed by some $x \in A$, and similarly, every non-empty set $X \in \mathcal{L}^a$ is assumed by some $y \in B$.

It is now possible to characterize assumption-complete paraconsistent topological belief models. Given type spaces A and B, we construct the closed-set topologies on respective type spaces τ_A and τ_B where *each* subset of A and B are in τ_A and τ_B, respectively. Therefore, it is easy to see that $A \times B$ is vertically and horizontally closed for any $S \subseteq A \times B$. For any $S \subseteq A \times B$, take an arbitrary (x,y) from S. Since, the singleton $\{x\}$ is in τ_A by construction and is closed, we observe that for all $y \in B$, we have $\{x\} \times \{y\} \subseteq S$ as (x,y) was picked from S. Therefore, we conclude that S is horizontally closed. By the same reasoning, S is vertically closed, too.

Moreover, under these conditions, product topological belief models are assumption-complete. Let $\mathcal{L}^a$ and $\mathcal{L}^b$ be the families of all subsets of A and B, respectively. Take them as τ_A and τ_B,

[4]This observation is a basic modal logical fact which suggests that adding states to a model which are not accessible from any other state by the modal operators does not change the expressive strength of the model, and two such models effectively satisfy the same formulas.

respectively. Then, without loss of generality, take an arbitrary $Y \in \tau_B$. We need to show that Y is assumed by some $x \in A$. However, by construction of the topologies τ_A and τ_B, for any $Y \subseteq B$ and any $X \subseteq A$, we have $X \times Y \in \tau_A \times \tau_B$. Thus, Y is assumed by all $x \in A$ (as $\{x\} \in \tau_A$). The proof used a set from τ_B. The same argumentation also works for a set X from τ_A.

We can summarize these observations in the following proposition.

Proposition 3.12
Given type spaces A and B, construct the closed set topologies on respective type spaces τ_A and τ_B where each subset of A and B are in τ_A and τ_B, respectively. Then,

- $A \times B$ is vertically and horizontally closed for any $S \subseteq A \times B$.
- The structure $(A \times B, \tau_A \times \tau_B, V)$ is assumption-complete.

We can relax some of these conditions. We define *weak assumption-completeness* for a topological belief structure if every set $S \in A \times B$ is both horizontally and vertically closed. In weak-assumption-complete models, we lack the topological liberty that came with the topological spaces that contain each and every subset of the given type space. Weak-assumption completeness can be considered a weakening of this topological assumption. Then, by definition, we observe the following.

Theorem 3.13
Let $M = (A \times B, \tau_A \times \tau_B, V)$ be a product topological paraconsistent interactive belief model. If M is horizontally and vertically closed, then it is weak assumption-complete.

Proof. The proof follows from the definitions, so we will briefly sketch it here. Let $M = (A \times B, \tau_A \times \tau_B, V)$ be a product topological paraconsistent interactive belief model which is both horizontally and vertically closed. To obtain a contradiction, assume that M is not weak assumption-complete. Thus, without loss of generality assume that M is horizontally closed but not vertically closed. This creates a contradiction as M was assumed to be both horizontally and vertically closed. ■

4 Conclusion and future work

In this article, we implicitly argued that the choice of traditional and classical logical tools to express a self-referential paradox in games is not well-justified. We achieved this by presenting two non-classical frameworks to formalize impossible beliefs in games. Non-well-founded sets, among many other things, enabled us to use a larger collection of models to represent self-referentiality. Paraconsistent logic, on the other hand, gave us additional tools to construct inconsistency-friendly models. This directly relates to the possibility of expressing inconsistent *knowledge* in epistemic games. With the help of such notions, it is possible that we can express (at least descriptively) various inconsistent game theoretical situations including the mistakes committed by rational agents, inconsistent signals and moves. Such long-term and far-reaching goals of our research programme cannot be achieved by the standard tools of epistemic game theory and classical modal logic. Also, what we did not discuss at all in the current work are the implications of the non-classical models and paraconsistent logics on the traditional notion of game theoretical rationality. For instance, it is worthwhile to pursue what NWF game models entail in terms of the rationality of the players or how the traditional utilitarian notion of rationality can be defined against a paraconsistent logical background theory.

The methodology which we presented in this work can also be applied to semantical games including discursive and dialogical games [17, 20, 31]. Conceptually, this raises the idea of discussing epistemics in semantical games. Therefore, a further research direction is to combine the formalisms of this work with that of dialogical logic to analyse various issues and paradoxes in epistemic game theory.

Even if the current paper discusses only a self-referential game theoretical paradox, some *non*-self-referential paradoxes interest philosophers and logicians. Yablo's paradox is a novel approach to paradoxes and dialetheism [37]. It seems appealing and perhaps challenging to search for epistemic game theoretical situations which can be expressed by a two-person formulation of Yablo's paradox. Thus, non-self-referential paradoxes may suggest even a broader approach to epistemic games and paraconsistency.

Finally, note that, from a logical perspective, paraconsistency has its dual intuitionistic form where the belief sets of the players may be *incomplete* or *paracomplete*. Hintikka and Sandu's Independence-Friendly logic (IF logic) can be an excellent tool to analyse the dualities of the BK paradox [15, 22].

Acknowledgements

I am more than grateful to Adam Brandenburger, Florian Lengyel, Rohit Parikh, Graham Priest and Noson Yanofsky for their help and encouragement. Two anonymous referees provided very valuable and detailed feedback to improve this article. I am thankful to them.

References

[1] S. Abramsky and J. Zvesper. From lawvere to brandenburger-keisler: interactive forms of diagonalization and self-reference. CoRR, abs/1006.0992, 2010.
[2] P. Aczel. *Non-well-founded Sets*, volume 14. CSLI Lecture Notes, 1988.
[3] S. Awodey. *Category Theory*. Oxford University Press, 2006.
[4] J. Barwise and L. S. Moss. *Vicious Circles*. Center for Study of Language and Information, 1996.
[5] C. Başkent. Some topological properties of paraconsistent models. *Synthese*, **190**, 4023–4040, 2013.
[6] A. Brandenburger and H. Jerome Keisler. An impossibility theorem on beliefs in games. *Studia Logica* **84**, 211–240, 2006.
[7] W. A. Carnielli, M. E. Coniglio and J. Marcos. Logics of formal inconsistency. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, vol. 14, pp. 15–107. Springer, 2007.
[8] B. Chellas. *Modal Logic*. Cambridge University Press, 1980.
[9] N. C. A. da Costa, D. Krause and O. Bueno. Paraconsistent logics and paraconsistency. In D. Jacquette, editor, *Philosophy of Logic*, vol. 5, pp. 655–781. Elsevier, 2007.
[10] D. M. Gabbay, A. Kurucz, F. Wolter and M. Zakharyaschev. *Many Dimensional Modal Logics: Theory and Applications*, vol. 145 of *Studies in Logic and the Foundations of Mathematics*. Elsevier, 2003.
[11] J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, Institute of Logic, Language and Computation; Universiteit van Amsterdam, 1999.
[12] N. D. Goodman. The logic of contradiction. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, **27**, 119–126, 1981.
[13] J. C. Harsanyi. Games with incomplete information played by 'bayesian' players: I. the basic model. *Management Science*, **14**, 159–182, 1967.

[14] A. Heifetz. Non-well-founded type spaces. *Games and Economic Behavior*, **16**, 202–217, 1996.
[15] J. Hintikka and G. Sandu. Information independence as a semantical phenomenon. In J. E. Fenstad, I. T. Frolov and R. Hilpinen, editors, *Logic, Methodology and Philosophy of Science VIII*, pp. 571–589. Elsevier, 1989.
[16] I. L. Humberstone. The modal logic of 'all and only'. *Notre Dame Journal of Formal Logic*, **28**, 177–188, 1987.
[17] S. Jaśkowski. A propositional calculus for inconsistent deductive systems. *Logic and Logical Philosophy*, **7**, 35–56, (translated from the 1948 original) 1999.
[18] F. W. Lawvere. Diagonal arguments and Cartesian closed categories. In A. Dold and B. Eckmann, editors, *Category Theory, Homology Theory and their Applications II*, vol. 92 of *Lecture Notes in Mathematics*, pp. 134–145. Springer, 1969.
[19] F. W. Lawvere. Intrinsic co-heyting boundaries and the leibniz rule in certain toposes. In A. Carboni, M. Pedicchio and G. Rosolini, editors, *Category Theory*, vol. 1488 of *Lecture Notes in Mathematics*, pp. 279–281. Springer, 1991.
[20] H.-J. Lebbink, C. L. M. Witteman and J.-J. Meyer. Dialogue games for inconsistent and biased information. *Electronic Notes in Theoretical Computer Science*, **85**, 134–151, 2004.
[21] L. Lismont. Une classe universelle anti-fondée pour les croyances. *Comptes rendus de l'Académie des sciences, Série 1, Mathématique*, **315**, 1117–1120, 1992.
[22] A. L. Mann, G. Sandu and M. Sevenster. *Independence-Friendly Logic*. Cambridge University Press, 2011.
[23] T. Mariotti, M. Meier and M. Piccione. Hierarchies of beliefs for compact possibility models. *Journal of Mathematical Economics*, **41**, 303–324, 2005.
[24] D. Mirimanoff. Les antimonies de russell et de burali-forti et le probleme fondamental de la theorie des ensembles. *L'Enseignement Mathématique*, **19**, 37–52, 1917.
[25] C. Mortensen. Topological seperation principles and logical theories. *Synthese*, **125**, 169–178, 2000.
[26] L. S. Moss. Non-wellfounded set theory. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/fall2009/entries/nonwellfounded-set-theory, fall 2009 edition, 2009.
[27] E. Pacuit. Understanding the brandenburger-keisler belief paradox. *Studia Logica*, **86**, 435–454, 2007.
[28] G. Priest. What is so bad about contradictions? *Journal of Philosophy*, **95**, 410–426, 1998.
[29] G. Priest. Paraconsistent logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, vol. 6, pp. 287–393. Kluwer, 2002.
[30] G. Priest. *In Contradiction*, 2nd edn. Oxford University Press, 2006.
[31] S. Rahman and W. A. Carnielli. The dialogical approach to paraconsistency. *Synthese*, **125**, 201–231, 2000.
[32] S. Rahman and T. Tulenheimo. From games to dialogues and back. In O. Maher, A. Pietarinen and T. Tulenheimo, editors, *Games: Unifying Logic, Language and Philosophy*, pp. 153–208. Springer, 2009.
[33] G. E. Reyes and H. Zolgaghari. Bi-heyting algebras, toposes and modalities. *Journal of Philosophical Logic*, **25**, 25–43, 1996.
[34] J. van Benthem. *Logic in Games*. MIT Press, 2014.
[35] J. van Benthem, G. Bezhanishvili, B. ten Cate and D. Sarenac. Modal logics for product topologies. *Studia Logica*, **84**, 375–399, 2006.
[36] J. van Benthem and D. Sarenac. The geometry of knowledge. In *Aspects of Universal Logic*, vol. 17 of *Travaux Logic*, pp. 1–31, 2004.

[37] S. Yablo. Paradox without self-reference. *Analysis*, **53**, 251–252, 1993.

[38] N. S. Yanofsky. A universal approach to self-referential paradoxes, incompleteness and fixed points. *The Bulletin of Symbolic Logic*, **9**, 362–386, 2003.

[39] J. Zvesper and E. Pacuit. A note on assumption-completeness in modal logic. In G. Bonanno, B. Löwe and W. van der Hoek, editors, *Proceedings of the 8th International Conference on Logic and the Foundations of Game and Decision Theory*, pp. 190–206. LOFT '08, 2010.

Received 1 October 2013