Bridging the two plans in the semantics for relevant logic

Takuro Onishi

1 Introduction

There have been two "plans" in the semantics for relevant logic. One is the tradition of Belnap-Dunn style four-valued semantics [8, 1, 2][1], which is called the *American plan*. The other is the *Australian plan* that is two-valued relational (frame) semantics, also known as Routley-Meyer semantics [24, 25, 23]. The present paper considers how these two plans are related to each other and concludes that the Australian plan can be regarded as a developed form of the American plan.

I start with R. Routley's 1984 paper "American plan completed," in which he examined how to extend the American plan to accommodate full relevant implication [22]. His attitude in the paper is indeed interesting. While he condemns the extended American plan for being hopelessly complicated, he also suggests that it might perhaps provide an explanatory basis for the star function employed in the Australian plan, which has been often criticized as unintuitive. In this paper I try to make clear the import of this suggestion referring to the recent discussion on negative modality.

2 Routley's American plan

This section quickly reviews the two plans in the semantics of relevant logic, and introduces Routley's extended American plan.

---

[1]In addition to these original works, I refer to [10, 19, 11, 14] for equivalent formulations of FDE and helpful references.

2.1 The two plans

The American plan is a four-valued semantics for FDE (First Degree Entailment) that assigns to each formula one of the four values, *true, false, neither true nor false*, and *both true and false*. It is equivalently formulated as a semantics in which a valuation assigns truth and falsity separately to each formula, and hence there are four cases (true but not false, false but not true, neither, and both) corresponding to the above four values.

Given a valuation over atomic formulas, truth and falsity of complex formulas are determined by *classical* semantic clauses, especially for negation, we have:

$$\sim A \text{ is true} \iff A \text{ is false}$$
$$\sim A \text{ is false} \iff A \text{ is true}$$

So there is a sense in which the meaning of negation (and the other connectives, conjunction and disjunction) is the same as that in classical logic. The only difference is that while in classical logic truth and falsity are mutually exclusive and jointly exhaustive, they are not in FDE.

The Australian plan is a two-valued frame semantics in which truth and falsity (i.e. non-truth) are relativized to states (worlds or situations). The main feature is the semantic clauses for implication and negation :

$$x \models A \to B \iff \forall y, z : Rxyz \,\&\, y \models A \Rightarrow z \models B;$$
$$x \models \sim B \iff x^* \not\models A,$$

where R is a ternary relation, $*$ a function on a frame, and $x \models A$ means that A is true at a state x.

The ternary relation is a very powerful tool to represent various principles concerning relevant implication as corresponding to properties of it, which is impossible in the American plan. On the other hand, the intuitiveness of the classical semantic clauses is lost in the Australian plan. In particular, the so-called *star function* or *Routley star* $*$ that models relevant negation is often criticized for its lack of intuitive meaning [5, 26].

One may think that it would be great if we could put together the two plans to obtain a semantics that has both ternary relation for implication and the intuitive classical clauses for negation, avoiding the use of Routley star. This is what Routley attempted in the 1984 paper. Let us look at where it lead him.

2.2 Four-valued relational semantics

We call it the *extended American plan*. It is a four-valued frame semantics in which each model is equipped with two ternary relations, R and S, and a twin valuation $\left\langle \models^+, \models^- \right\rangle$ that determines truth and falsity of formulas

separately. The additional ternary relation S is for the falsity condition of implication. Truth and falsity are relativized to states, but the semantic clauses for conjunction, disjunction, and importantly, negation preserve classicality, which means we use no star function.

Below are some definitions and basic facts concerning the semantics based on Routley's presentation, some details are changed though.

Definition 1 (Language). We work with the set L of formulas built from the connectives $\wedge, \vee, \sim, \rightarrow$ over the set PV of propositional variables.

Definition 2 (Frames and models). A structure $\langle 0, 0^*, U, V, K, R, S \rangle$, where K is a non-empty set of states, $0 \in U \subseteq K$, $0^* \in V \subseteq K$, and $R, S \subseteq K^3$, is a **B-frame** if

- $\exists u \in U : Ruxy \iff \exists v \in V : Sxvy$

  - We write $x \leq y$ if either one of these holds.

- $x \leq x^2$;

- $Rxyz \, \& \, x' \leq x \Rightarrow Rx'yz$; $Syxz \, \& \, x \leq x' \Rightarrow Syx'z$.

Let $\models^+, \models^- \subseteq K \times PV$ be *hereditary* valuations on a **B**-frame $\mathcal{F} = \langle 0, 0^*, U, V, K, R, S \rangle$ such that for any $x, y \in K$ and any propositional variable p,

$$x \models^+ p \, \& \, x \leq y \Rightarrow y \models^+ p; \quad x \models^- p \, \& \, x \leq y \Rightarrow y \models^- p.$$

We call $\left\langle \mathcal{F}, \left\langle \models^+, \models^- \right\rangle \right\rangle$ a **B-model**.

Given a **B**-model $\mathcal{M} = \left\langle 0, 0^*, U, V, K, R, S, \left\langle \models^+, \models^- \right\rangle \right\rangle$, the twin valuation $\left\langle \models^+, \models^- \right\rangle$ is extended to arbitrary formulas as follows:

$$
\begin{aligned}
x \models^+ A \wedge B &\iff x \models^+ A \, \& \, x \models^+ B \\
x \models^- A \wedge B &\iff x \models^- A \text{ or } x \models^- B \\
x \models^+ A \vee B &\iff x \models^+ A \text{ or } x \models^+ B \\
x \models^- A \vee B &\iff x \models^- A \, \& \, x \models^- B \\
x \models^+ \sim A &\iff x \models^- A \\
x \models^- \sim A &\iff x \models^+ A \\
x \models^+ A \rightarrow B &\iff \forall y, z : Rxyz \, \& \, y \models^+ A \Rightarrow z \models^+ B \\
x \models^- A \rightarrow B &\iff \exists y, z : Syxz \, \& \, y \models^- B \, \& \, z \not\models^- A.
\end{aligned}
$$

We write like $\mathcal{M}, x \models A$ to make explicit which model is in question.

---

[2]In what follows we assume that the binary relation $\leq$ so defined is always a partial order, namely that it is not only reflexive but also transitive and anti-symmetrical. As usual, this does not affect the logic.

The (twin) heredity property of valuations is extended to all formulas.

Proposition 3. For any formula $A \in L$ and any states x, y in a **B**-model,

$$x \models^{+} A \,\&\, x \leq y \Rightarrow y \models^{+} A; \quad x \models^{-} A \,\&\, x \leq y \Rightarrow y \models^{-} A.$$

Definition 4 (Entailments and validities)**.** We define different notions of entailment and validity at different level. Let $\mathcal{F}$ denote an arbitrary **B**-frame $\langle 0, 0^{*}, U, V, K, R, S \rangle$, and $\mathcal{M}$ an arbitrary **B**-model on $\mathcal{F}$.

$$A \models^{\mathcal{M}^{+}} B \iff \forall x \in K : \mathcal{M}, x \models^{+} A \Rightarrow \mathcal{M}, x \models^{+} B$$

$$A \models^{\mathcal{M}^{-}} B \iff \forall x \in K : \mathcal{M}, x \models^{-} B \Rightarrow \mathcal{M}, x \models^{-} A$$

$$A \models^{\mathcal{M}} B \iff A \models^{\mathcal{M}^{+}} B \,\&\, A \models^{\mathcal{M}^{-}} B$$

$$A \models^{\mathcal{F}^{+}} B \iff A \models^{\mathcal{M}^{+}} B \text{ for any } \mathcal{M} \text{ on } \mathcal{F}$$

$$A \models^{\mathcal{F}^{-}} B \iff A \models^{\mathcal{M}^{-}} B \text{ for any } \mathcal{M} \text{ on } \mathcal{F}$$

$$A \models^{\mathcal{F}} B \iff A \models^{\mathcal{F}^{+}} B \,\&\, A \models^{\mathcal{F}^{-}} B$$

$$A \models^{+} B \iff A \models^{\mathcal{F}^{+}} B \text{ for any } \mathcal{F}$$

$$A \models^{-} B \iff A \models^{\mathcal{F}^{-}} B \text{ for any } \mathcal{F}$$

$$A \models B \iff A \models^{+} B \,\&\, A \models^{-} B.$$

We say that *A t-entails* (*f-entails*, or *2-entails*) *B* in a model $\mathcal{M}$ when $A \models^{\mathcal{M}^{+}} B$ ($A \models^{\mathcal{M}^{-}} B$ or $A \models^{\mathcal{M}} B$ resp.). Similarly for entailment in a frame and entailment *simpliciter* (i.e. $\models^{+}, \models^{-}$ and $\models$).

A formula *A* is said to be *t-valid* (*f-valid*) *in a model* $\mathcal{M}$ if $0 \models^{+} A$ ($0^{*} \not\models^{-} A$) in $\mathcal{M}$. If *A* is both *t*-valid and *f*-valid in $\mathcal{M}$, it is *2-valid* in $\mathcal{M}$. Validity in a frame and validity *simpliciter* are defined in an obvious way.

Proposition 5. For any **B**-model $\mathcal{M}$,

$$A \models^{\mathcal{M}^{+}} B \iff 0 \models^{+} A \to B$$

$$A \models^{\mathcal{M}^{-}} B \iff 0^{*} \not\models^{-} A \to B.$$

Theorem 6 (Routley 1984)**.** A formula *A* is 2-valid if and only if it is provable in the axiom system of **B**.

2.3 Extensions

Routley's extended American plan captures the basic relevant logic **B** employing two ternary relations but without the star function. The next step is to represent various extensions of **B** in terms of structural properties of frames.

In the usual two-valued framework, we have the following kind of correspondence. For example, *EAssertion* $(A \to (A \to B)) \to B$ is valid in a frame if and only if the frame satisfies the commutativity of R:

$$Rxyz \Longrightarrow Ryxz.$$

It is one of the great features of the Australian plan that it enables us to develop a correspondence theory between axiom schemes and frame properties. It is highly desirable that this should be carried over to the extended American plan which aims to unify the two plans.

There is no problem with *positive* extensions, extensions with axiom schemes not involving negation. All that we have to do is to add the dual version of the original frame properties. For EAssertion and commutativity, we add the commutativity of S:

$$Sxyz \Longrightarrow Sxzy,$$

which corresponds to f-validity of EAssertion, while $Rxyz \Rightarrow Ryxz$ is now the counterpart of its t-validity.

Things are not this easy with *negative* extensions of **B**, extensions with axiom schemes involving negation. We focus on the axiom form of Contraposition since Routley takes it as "critical, as well as representative" [22, p.145]. In the Australian plan, the frame condition that corresponds to Contraposition are concisely formulated appealing to the star function:

$$(A \to B) \to (\sim B \to \sim A) \qquad \text{corresponds} \qquad Rxyz \Longrightarrow Rxz^*y^*.$$

How can we force Contraposition in our extended American plan without the star? Routley considered a several options and decided to *double* the semantic clauses for implication. That is, he revised them as follows:

$$x \models^+ A \to B \iff \forall y, z : Rxyz \Rightarrow \begin{cases} y \models^+ A \Rightarrow z \models^+ B & \text{and} \\ z \models^- B \Rightarrow y \models^- A \end{cases}$$

$$x \models^- A \to B \iff \exists y, z : Syxz \Rightarrow \begin{cases} y \models^- B \,\&\, z \not\models^- A & \text{and} \\ z \not\models^+ B \,\&\, y \models^+ A \end{cases}$$

The idea is to make Contraposition built in the meaning of implication, which sounds nicely parallel to the definition of 2-entailment we have seen in the previous section.

Unfortunately this causes awful complication. Especially for nested implication schemes, even when considering only t-validity, we would have to take into account the falsity of (embedded) implication, and hence some interaction between the dual ternary relations R and S. Worse, there is a case where a structural property of R and S looks helpless however complicated we make it: Consider proving t-validity of EAssertion under

5

the doubled semantic clauses (and some frame property). By proposition 5, we would show that there is no state x in any model on it such that $x \models^+ A$ and $x \not\models^+ (A \to B) \to B$. Suppose otherwise. Then for some y, z such that $Rxyz$, either (1) $y \models^+ A \to B$ and $z \not\models^+ A$ or (2) $z \models^- B$ and $y \not\models^- A \to B$. Now, in the latter case, how could we derive a contradiction with the assumption $x \models^+ A$? One will immediately see that any structural property of R and S would not be enough. We need some kind of *cross-over between truth and falsity* ($\models^+$ and $\models^-$).

Confronted with this problem, Routley gave up a star-free semantics. He introduced a *star-like relation* or *star-imitation*, which looks like the star function but is not supposed to be a function. We are going to give a closer look to the relation in the next section, but before that, let us review his attempt and consider how we should proceed hereafter.

2.4   Changing the strategy

The original goal was to build a star-free semantics that is powerful and flexible enough to develop a correspondence theory. Now it has turned out that the semantics is far more complicated than the Australian plan, and that the use of star-like thing is unavoidable anyway. Routley concludes the paper saying:

> The American plan is, in the end, when duly enriched with the Australian mateship component [star-like relation], little more than a very circuitous and cumbersome way of avoiding the much more intuitive Australian plan. [22, p.157]

This does not necessarily mean the victory of the Australian plan over the American plan. Indeed, though Routley seems to think that the Australian plan including the star function is "much more intuitive," many will disagree. Unless the Routley star is made sense, the Australian plan won't be accepted as an appropriate semantics for relevant logic either.

Here I'd like to take notice of a suggestion Routley himself made in the same paper:

> So perhaps a major virtue of the American approach is that it provides another approach to and explanation of——what it was set up partly in opposition to, and was supposed to be getting away from——the star negation rule and its role in the Australian plan. [22, p.136]

He is suggesting a slightly different strategy, that is, a strategy of using the American plan as conceptual resource to *explain* the star function, rather than to *avoid* it. Unfortunately he did not fully make clear in the paper how the strategy will develop. Hence it is the aim of the present paper. In what follows I try to make clear how the American plan works as an explanatory basis for the star function.

6

3 Star-like relation(s)

In this section, I introduce Routley's star-like relation and discuss how it is explained on the basis of the American plan.

3.1 Explanatory task

The star-like relation Routley employed is a binary relation M on a frame such that for any state x and y, if xMy holds then we have

$$x \models^- A \iff y \not\models^+ A \tag{M}$$

for any formula A[3]. As you may notice, the equivalence is similar to the star-negation rule

$$x \models\, \sim A \iff x^* \not\models A.$$

However, x's M-correlate y may not be the function of x. M is only assumed to be symmetric, which gives an effect similar to that brought by $x = x^{**}$.

In fact, a relation seems to be sufficient to formulate frame properties that correspond to negative extensions. E.g., Contraposition $(A \to B) \to (\sim B \to \sim A)$ corresponds to the frame property

$$\begin{cases} Rxyz \Rightarrow \exists y', z' : yMy' \,\&\, zMz' \,\&\, Rxz'y' \text{ and} \\ Syxz \Rightarrow \exists y', z' : yMy' \,\&\, zMz' \,\&\, Sz'xy'. \end{cases}$$

Reductio $(A \to \sim A) \to \sim A$ is represented by

$$\begin{cases} \exists x' : xMx' \,\&\, Rxx'x \text{ and} \\ \exists x' : xMx' \,\&\, Sxxx' \end{cases}$$

These are just a relational version of our familiar conditions $Rxyz \Rightarrow Rxz^*y^*$ and Rxx^*x. And it is not likely that there is an axiom schemes that requires a function rather than a relation.

Then, we should say that our explanatory task is twofold. First, of course we are obliged to explain the meaning of the star-like relation, that is, what it is for two state to stand in the relation. In addition to that, we need an explanation of the functionhood of the star-function. Why is it supposed to be a function in the standard Routley-Meyer semantics?

---

[3]Strictly speaking, it is first assumed that any valuation satisfies the condition for propositional parameters, and then it is proved that the equivalence extends to arbitrary formulas, appealing to several conditions imposed on R, S and M.

3.2 Compatibility and exhaustiveness

The story that I am going to tell about the meaning of the star-like relation is like this. As we will see shortly, thanks to the recent study on *negation as modal operator*, we are in a much better position concerning this explanatory task than Routley was when he wrote the paper [9, 21, 13, 15]. But the explanatory strategy of the negation-as-modality approach faces some fundamental objection. Then we will see that, on the framework of the extended American plan, we can employ the useful concepts from the negation-as-modality approach while avoiding the objection. I will argue that this is what Routley suggests as the virtue of the American plan.

First, let us decompose the equivalence (*M*). If xMy, we have the following two implications:

$$x \models^- A \Longrightarrow y \not\models^+ A \tag{1}$$
$$x \not\models^- A \Longrightarrow y \models^+ A. \tag{2}$$

(1) is equivalent to "it is excluded that A is false at x while it is true at y." In other words, the information contained x and y never clash each other. In this sense, x and y can be said to be *compatible* with each other. The converse direction, (2), is equivalent to "either A is false at x or A is true at y." Thus, x and y jointly cover the whole logical possibilities, truth and falsity. They are jointly *exhaustive*. Therefore, for two states to stand in the star-like relation M is for them to be mutually compatible and jointly exhaustive. The star-like relation is the conjunction of compatibility and exhaustiveness.

Some might say that compatibility and exhaustiveness are, properly speaking, relations between propositions or properties, not between states. Here is an example from logic that will motivate our notion of compatibility and exhaustiveness.

A class of models defined by some properties determines (in)validity of formulas. Each can be seen as a state that contains information about validity and a different class of models will carry different information from another. Take as an example the Kripke-style model theory for modal logic. Let x be the class of all Kripke models (with no constraint on accessibility relation). Then we can write, for example, $x \models^- \Box A \supset A$, in the sense that x contains a counter model to the axiom scheme. On the other hand, if y is the class of models in which a binary accessibility relation is reflexive, then we have $y \not\models^- \Box A \supset A$. Similarly, take Hilbert style axiom systems. Each of them may also be regarded as a state. We write $z \models^+ A$ if the system z proves A. By adding or dropping axioms we have different states. Now if a proof system y is sound with respect to a class x of models, then (1) holds. Soundness is compatibility between models and proof systems. Conversely, exhaustiveness (2) means completeness in the world of logic.

Models and proof systems form a system of information states which enable us to make a certain range of inferences. If A has a counter model in a class x of models and a proof system y is compatible with (sound with respect to) y, one can infer that A is not provable in y, etc. We can find such systems in various areas and fields in which different bodies of information are somehow related so that we can infer what is true (or false) in one of them from what is true (or false) in another. I believe that accessibility relations deployed in frame semantics *in general*, including binary relation for necessity and possibility and ternary relation for relevant implication as well as compatibility and exhaustiveness, can be understood in this way, i.e., as inferential relationships between information states. This is not a place to develop a general theory, so I just refer to the notion of situated inference due to Mares [13] and the conception of ternary relation as information channel by Restall [20] as pointing to the direction I am thinking about. I shall go forward assuming that the meaning of star-like relation has been explained enough.

3.3 The explanatory virtue of the American plan

The star-like relation M is the conjunction of compatibility and exhaustiveness. Let us divide it explicitly and work with models that have two binary relations C (compatibility) and E (exhaustiveness) such that

$$xCy \,\&\, x \models^- A \Longrightarrow y \not\models^+ A \qquad\qquad \text{(Com)}$$
$$xEy \,\&\, x \not\models^- A \Longrightarrow y \models^+ A. \qquad\qquad \text{(Exh)}$$

They are assumed to be symmetric as in the case of M.

Given binary relations on a frame, we may introduce two *negative modal operators*, $\rhd$ and $\blacktriangleright$, with the semantic clauses:

$$x \models^+ \rhd A \iff \forall y : xCy \Longrightarrow y \not\models^+ A \qquad\qquad (\rhd)$$
$$x \models^+ \blacktriangleright A \iff \exists y : xEy \,\&\, y \not\models^+ A. \qquad\qquad (\blacktriangleright)$$

These are justifiably called *negative* modality since they support contraposition (If $A \models B$ then $\rhd B \models \rhd A$ and $\blacktriangleright B \models \blacktriangleright A$) and some inferences characteristic to negation. $\rhd$ is *impossibility* operator and $\blacktriangleright$ is understood as *unnecessity*. There is now much work on negative modality of this kind, which are usually based on two-valued semantics: the truth conditions are

$$x \models \rhd A \iff \forall y : xCy \Longrightarrow y \not\models A \qquad\qquad (\rhd_2)$$
$$x \models \blacktriangleright A \iff \exists y : xEy \,\&\, y \not\models A. \qquad\qquad (\blacktriangleright_2)$$

Kosta Došen [7] initiated the investigation into negative modality parallel to that into positive modality (necessity and possibility), aiming to

elaborate a correspondence theory between axiom schemes and frame properties. His base logic was intuitionistic logic. In the relevant tradition, logicians like Dunn, Restall and Mares have been trying to make sense of star-negation in terms of negative modality [9, 21, 13]. According to the *star postulate* they found, the star state is characterized as "maximally compatible state." It seems to be reasonable to say that the star function is now fully made sense by the star postulate. But here we consider an objection that the negative modality approach *in general* must answer.

The negative modality approach aims to explain their target negations as negative modality, and, in turn, negative modality in terms of binary relation. In other words, binary relation is assumed to be conceptually prior to negative modality and negation. De and Omori argue that this conceptual priority is objectionable:

> If there is any kind of explanation as to why these worlds should be incompatible with ours, we can only see that it must ultimately appeal to negation. [6]

The above ($\triangleright_2$) tells us that the negation of A (or impossibility of A) holds at a state x if and only if A fails at any state that is compatible with x. But then, what is it for two states to be compatible with each other? The only answer available to us seems to be that two states are compatible if truth of a proposition at one of them excludes the *negation* of the proposition at another. This is just another way of reading ($\triangleright_2$) and we are forced to use the notion of negation to explain the notion of compatibility. Thus, the strategy to explain negation in terms of binary relation is circular, unless the notion of compatibility is given some characterization independent from ($\triangleright_2$), which seems unavailable. As far as I know, we have no principle concerning negation and (in)compatibility other than the above ($\triangleright_2$) on the framework of *two-valued* semantics (similarly for $\blacktriangleright_2$ and exhaustiveness).

It might be possible to find such an independent characterization and resolve the worry of circularity. But it does not matter to us anyway. I mean, in the extended American plan, there is no need to care about the worry at all, for we started with four-valued semantics where negation is already defined by the classical clauses. We don't have to use (Com) and (Exh) as an explanation of negation and falsity. (Com) and (Exh) can be used simply as an explanation of C and E in terms of primitive notions of truth and falsity (or negation).

I claim this is the virtue of the American plan that Routley suggested. If you are a member of the Australian camp trying to make sense of the unintuitive star-negation in terms of binary relation on a two-valued frame, you will have a trouble concerning conceptual priority. In contrast, if you start under the American plan, then negation is simply defined by the intuitive classical clauses and there is no obstacle to introduce compatibility

and exhaustiveness relations on a frame, which will be useful to make sense of the Routley star. In this way, the American plan works as an explanatory basis for the star function [4].

4 The star function

We have finished the first task of explaining the meaning of star-like relation based on the American plan: the star-like relation is the conjunction of compatibility and exhaustiveness. In this section I shall make clear why the star-function is supposed to be a function.

As I mentioned above, the star function is now understood through the star postulate, formally written as

$$\forall x \exists y : xCy \,\&\, \forall z : xCz \Rightarrow z \leq y, \tag{S}$$

That is, each state has a mate state that is maximally compatible with it. On a frame satisfying this constraint, we can define the star function by setting x^* to be the maximally compatible state y. But I will discuss the functionhood of the star in light of the dualistic version of this postulate, the *dualist star postulate*,

$$\forall x \exists y : xCy \,\&\, xEy \,\&\, (\forall w : xCw \Rightarrow w \leq y) \,\&\, (\forall w : xEw \Rightarrow y \leq w), \tag{DS}$$

which is essentially the conjunction of the uniqueness and existence conditions for the functionhood of the star [5].

Before going into the discussion, note that to suppose the star to be a function, or equivalently to assume the (dualist) star postulate, does not change the logic. This is seen from the canonical model construction that Routley presented in the 1984 paper. For each prime theory a, which is an element of the canonical model, he defined the *opposite* of a to be

$$a^* := \{B \mid \,\sim B \notin a\}.$$

---

[4]There is an approach to semantics of negation and logical constants in general which is based on the notion of incompatibility between *propositions* (sentences, statements, and the like) [3, 4, 17, 18]. A similar objection would be raised concerning the conceptual priority to this approach too, and the proponents might answer to it by invoking an explanation of incompatibility in terms of preclusion relation between commitment and entitlement to propositions: roughly speaking, propositions A and B are incompatible if anyone who is committed to A is thereby precluded from counting as entitled to B (and vice versa) [4, p.120]. This may satisfy the doubt of circularity, or may not. The notion of preclusion might still appear to involve some notion of negation. I am neutral on this point. The point of the present argument is just that we don't need to care about this if we start with the American plan.

[5]I think the dualist version is better than the original one in several aspects. And it is one of its merits that, as we will see, the dualist version allows us to analyze the uniqueness and existence conditions separately. For other merits of the dualist approach, see [15, 16].

as a useful "auxiliary notion." Indeed this behaves just as the star. Thus, the star function is *implicitly* there in the canonical model, which implies the assumption of functionhood of the star does not affect the logic as the set of valid formulas and entailments. In the following, I will examine what kind of change is introduced to the entire structure of models by *explicitly* assuming functionhooood of the star.

4.1 Models with star-like relations

First, let me define a model equipped with compatibility and exhaustiveness formally.

Definition 7 (B^{CE}-frames and models)**.** Let $\mathcal{F} = \langle 0, 0^*, U, V, K, R, S \rangle$ be a **B**-frame, and $C, E \subseteq K^2$. A structure $\langle \mathcal{F}, C, E \rangle$ is called a **$\mathbf{B}^{CE}$-frame** if

1. $xCy \Rightarrow yCx$; $xEy \Rightarrow yEx$;

2. $xCy \,\&\, x' \leq x \Rightarrow x'Cy'$; $xEy \,\&\, x \leq x' \Rightarrow x'Ey$;

3. $\exists x' : Szx'y \,\&\, xCx' \Rightarrow \exists y', z' : yEy' \,\&\, zCz' \,\&\, Rxy'z'$;

4. $\exists x' : Rx'yz \,\&\, xEx' \Rightarrow \exists y', z' : yCy' \,\&\, zEz' \,\&\, Sz'xy'$.

If a hereditary twin valuation $\left\langle \models^+, \models^- \right\rangle$ on a **$\mathbf{B}^{CE}$**-frame satisfies the compatibility and exhaustiveness conditions, i.e.

$$xCy \,\&\, x \models^- p \Longrightarrow y \not\models^+ p \qquad\qquad \text{(Com)}$$
$$xEy \,\&\, x \not\models^- p \Longrightarrow y \models^+ p. \qquad\qquad \text{(Exh)}$$

for any propositional parameter p, we call $\left\langle \mathcal{F}, C, E, \left\langle \models^+, \models^- \right\rangle \right\rangle$ a **$\mathbf{B}^{CE}$**-**model**. The notions of validity and entailments at different levels are defined as before.

We add to the language two unary operators $\triangleright$ and $\blacktriangleright$ for which the semantic rules are as follows:

$$x \models^+ \triangleright A \iff \forall y : xCy \Longrightarrow y \not\models^+ A$$
$$x \models^+ \blacktriangleright A \iff \exists y : xEy \,\&\, y \not\models^+ A$$
$$x \models^- \triangleright A \iff \exists y : xEy \,\&\, y \not\models^- A$$
$$x \models^- \blacktriangleright A \iff \forall y : xCy \Longrightarrow y \not\models^- A.$$

These negative modalities are not ultimately necessary for our purpose, but they serve to make things clear.

The three conditions on valuations, heredity, compatibility and exhaustiveness, can be extended to arbitrary formulas as expected:

Proposition 8. For any formula A and any states x and y in a $\mathbf{B}^{CE}$-model,

$$x \leq y \,\&\, x \models^+ A \Longrightarrow y \models^+ A$$
$$x \leq y \,\&\, x \models^- A \Longrightarrow y \models^- A$$
$$xCy \,\&\, x \models^- A \Longrightarrow y \not\models^+ A \qquad\qquad \text{(Com)}$$
$$xEy \,\&\, x \not\models^- A \Longrightarrow y \models^+ A. \qquad\qquad \text{(Exh)}$$

Proof. Routine. The condition 2 is for the heredity of $\rhd$ and $\blacktriangleright$, and 3 and 4 for the case of implication in the inductive proof of (Com) and (Exh) respectively. If other intensional connectives like fusion are added, further conditions are required. □

4.2 Dualist star postulate

Now we shall consider what happens if we add the *dualist star postulate*:

$$\forall x \exists y : xCy \,\&\, xEy \,\&\, (\forall w : xCw \Rightarrow w \leq y) \,\&\, (\forall w : xEw \Rightarrow y \leq w), \quad \text{(DS)}$$

which claims the existence of "maximally compatible" and "minimally exhaustive" state y for each state x. As opposed to the original star postulate as a constraint on C, this is dualistic in the sense that it features E as well as C. Note that (DS) is equivalent to the conjunction of the following two conditions (modulo the transitivity of $\leq$):

$$\forall x, y, z : xCy \,\&\, xEz \Longrightarrow y \leq z, \qquad\qquad \text{(DS1)}$$
$$\forall x \exists y : xCy \,\&\, xEy. \qquad\qquad \text{(DS2)}$$

Recall that Routley's star-like relation M is the conjunction of compatibility (C) and exhaustiveness (E). Then clearly (DS2) is the existence condition for the functionhood of M, and it should be also obvious that (DS1) is the uniqueness condition: xMy and xMz imply $y = z$ under (DS1) (modulo the anti-symmetry of $\leq$). Let us look at them separately.

First, it seems rather reasonable to assume (DS1):

Proposition 9. In a $\mathbf{B}^{CE}$-model, we can assume (DS1) without changing the logic.

Proof. If xCy and xEz, then, thanks to (Com) and (Exh), we have

$$y \models^+ A \Rightarrow x \not\models^- A \Rightarrow z \models^+ A;$$
$$y \models^- A \Rightarrow x \not\models^+ A \Rightarrow z \models^- A$$

for any formula A. Considering that the ordering relation in the canonical model is defined by subset relation between the sets of true formulas and false formulas in different states, we can assume $y \leq z$ here. □

As shown in the proof, (DS1) can be seen as a natural by-product of (Com) and (Exh), so I assume that every $\mathbf{B}^{CE}$-model satisfies (DS1) hereafter. In contrast, it is not immediately clear where (DS2) comes from. I will discuss the origin of the existence condition in the next section. Before that, let us see what happens if we further add (DS2). Indeed the change caused is quite big. The following is an application of the result of [15]:

Proposition 10 (Cf. [15]). On a $\mathbf{B}^{CE}$-model with (DS1) and (DS2), we can define a function $*$ such that:

$$x \leq y \Longrightarrow y^* \leq x^*; \quad x^{**} = x;$$
$$x \models^+ \sim A \iff x^* \not\models^+ A \iff x \models^+ \vartriangleright A \iff x \models^+ \blacktriangleright A;$$
$$x \models^- \sim A \iff x^* \not\models^- A \iff x \models^- \vartriangleright A \iff x \models^- \blacktriangleright A;$$
$$Rxyz \iff Sz^*x^*y^*.$$

Proof. Immediate. See the condition 3 and 4 for the last item. □

And more importantly, under (DS2), the model is essentially two-valued. Once having defined the star-function on a $\mathbf{B}^{CE}$-model as above, we can just forget the falsity valuation $\models^-$ and the structure of $0^*, V$ and S. The result is a standard two-valued $\mathbf{B}$-model with Routley star. Conversely, given a two-valued $\mathbf{B}$-model with the star, by defining

$$x \models^+ p \iff x \models p$$
$$x \models^- p \iff x \models \sim p$$
$$Sxyz \iff Ry^*z^*x^*$$
$$xCy \iff xEy \iff y = x^*$$

etc., we get a four-valued $\mathbf{B}^{CE}$-model with (DS2). This construction constitutes an isomorphism between the class of two-valued $\mathbf{B}$-model with Routley star and that of four-valued $\mathbf{B}^{CE}$-models with (DS2). And it is immediate that (the $\{\wedge, \vee, \rightarrow, \sim\}$-fragment of) the logic defined by the class of $\mathbf{B}^{CE}$-model is $\mathbf{B}$ (Note that $\vartriangleright$ and $\blacktriangleright$ collapse to $\sim$).

Explicitly assuming the functionhood of star, the whole framework of four-valued semantics collapses to the two-valued one. In other words, we have seen how the Australian plan generates from the American plan: the Australian plan is the extended American plan with (DS2). However, as I noted, the origin of (DS2) is unclear. How can such an assumption that causes a big change be motivated?

4.3 Contraposition

I shall now turn to contraposition. It may sound a bit abrupt, but it will shed a light on the structure of the extended American plan. Here we are

concerned with rule-contraposition ($A \models B$ implies $\sim B \models \sim A$) rather than formula-contraposition ($(A \to B) \to (\sim B \to \sim A)$).

Let us first observe that validity (entailment preservation) of rule-contraposition is *built in* the definition of entailment in Routley's American plan since it is defined as 2-entailment, truth preservation *and* backward falsity preservation: $A \models B$ is trivially equivalent to $\sim B \models \sim A$. But a closer look readily reveals that the bi-directional preservation is not necessary.

Proposition 11. With the class of $\mathbf{B}^{CE}$-frames, we have

$$A \models^+ B \iff A \models^- B \quad (\iff \sim B \models^+ \sim A).$$

Recall that $\models^+$ and $\models^-$ here denote truth preservation and backward falsity preservation in all frames respectively. Also note that the same holds good for the class of $\mathbf{B}$-frames since a $\mathbf{B}$-frame is just a special case of $\mathbf{B}^{CE}$-frame where $C = E = \emptyset$.

Proof. It suffices to show any counter-model against $A \models^+ B$ can be turned into one against $A \models^- B$.

Let $\mathcal{F} = \left\langle 0, 0^*, U, V, K, R, S, C, E, \models^+, \models^- \right\rangle$ be a $\mathbf{B}^{CE}$-model such that for some $w \in K$,

$$w \models^+ A \text{ but } w \not\models^+ B.$$

We define a dualized model $\mathcal{F}_d = \left\langle 0_d, 0_d^*, U_d, V_d, K_d, R_d, S_d, C_d, E_d, \models_d^+, \models_d^- \right\rangle$ to be such that:

$$0_d = 0^*, \ 0_d^* = 0, \ U_d = V, \ V_d = U, \ K_d = K,$$

$$
\begin{array}{ll}
R_d xyz \iff Szxy & S_d xyz \iff Ryzx \\
x C_d y \iff x E y & x E_d y \iff x C y \\
x \models_d^+ p \iff x \not\models^- p & x \models_d^- p \iff x \not\models^+ p.
\end{array}
$$

It is immediately seen that $\mathcal{F}_d$ is a $\mathbf{B}^{CE}$-model. Note that we have

$$x \leq_d y \iff y \leq x$$

for the ordering relation $\leq_d$ defined by $x \leq_d y \iff R_d 0_d xy$. The rest of the conditions for $\mathbf{B}^{CE}$-model follow from this.

Then, as expected, we can easily prove by induction that for any formula A and any $x \in K$,

$$x \models_d^+ C \iff x \not\models^- C \qquad x \models_d^- C \iff x \not\models^+ C.$$

Here are a few examples.

$$x \models_d^+ C \to D \iff \forall y, z : R_d xyz \,\&\, y \models_d^+ C \Longrightarrow z \models_d^+ D$$
$$\iff \forall y, z : Szxy \,\&\, y \not\models^- C \Longrightarrow z \not\models^- D$$
$$\iff x \not\models^- C \to D.$$
$$x \models_d^+ \rhd C \iff \forall y : xC_d y \Longrightarrow y \not\models_d^+ C$$
$$\iff \forall y : xEy \Longrightarrow y \models^- C$$
$$\iff x \not\models^- \rhd C.$$

Thus, applying this to the above assumption, we can conclude that

$$w \not\models_d^- A \text{ but } w \models_d^- B,$$

which implies that $\mathcal{F}_d$ is a counter-model against $A \models^- B$. $\qquad\square$

So, why do we need 2-entailment if the set of entailment is characterized by truth preservation alone? I don't know how Routley would answer. I only see that we need 2-entailment at the level of single model or frame. Indeed, t-entailment *in a model* and f-entailment *in a model* do not necessarily coincide:

Proposition 12. There is a $\mathbf{B}^{CE}$-model $\mathcal{M}$ in which $A \models_{\mathcal{M}}^+ B$ but $A \not\models_{\mathcal{M}}^- B$.

Proof. Consider a two-element $\mathbf{B}^{CE}$-frame $\langle 0, 0^*, \{0\}, \{0^*\}, \{0, 0^*\}, R, S, C, E \rangle$, where

$$R = \{\langle 0,0,0 \rangle, \langle 0,0^*,0^* \rangle\}, S = \{\langle 0,0^*,0 \rangle, \langle 0^*,0^*,0^* \rangle, \langle 0,0,0^* \rangle\}.$$

and $C = E = \varnothing$. Define a valuation on this by

$$0 \models^- p, 0 \not\models^- q, 0^* \not\models^- p,$$

and arbitrarily for the other cases. This satisfies the conditions for valuation trivially since $\leq$ is just the identity relation and C and E are empty in this frame. Now since $0 \models^+ p \to p$ and $0^* \models^+ p \to p$, we have $q \models^+ p \to p$. On the other hand, $0 \not\models^- q$ and $0 \models^- p \to p$ imply that $q \not\models^- p \to p$. $\qquad\square$

Thus, the rationale of the notion of 2-entailment seems to consist in validating rule-contraposition at the level of single model. I am quite indifferent to whether it should be or not, since it holds at the level of entailment *simpliciter* anyway. I observed this just because it will serve to clarify the role that the star function or the star postulate (DS2) plays.

The above contrast between two levels may be understood like this: while the structure of the entire class of $\mathbf{B}^{CE}$-models is rich enough to provide the dual model for each model, a single model may not be that

much. Even when you have a structure that gives a counter example to a *t*-entailment in a model, it may not be possible to find the dual structure within the same model that will refute *f*-entailment, and vice versa.

And what the star function or (DS2) provides is that dualizable structure within a single model or frame. Recall that in Proposition 10 we had $Rxyz \Longleftrightarrow Sz^*x^*y^*$ in a $\mathbf{B}^{CE}$-model with (DS2), and also that we defined $R_d xyz \Longleftrightarrow Szxy$ in the proof of Proposition 11, where we constructed the dualized model. Also compare the equivalence $x \models^+ {\sim} A \Longleftrightarrow x^* \not\models^+ A$ in Proposition 10 and $x \models^-_d C \Longleftrightarrow x \not\models^+ C$ in Proposition 11. The star function makes possible the dualizing operation within a single model, which is otherwise available only at the level of the entire class of models. As a result, in a $\mathbf{B}^{CE}$-model with (DS2), *t*-entailment in a model and *f*-entailment in a model become equivalent, and hence rule-contraposition is validated at the level of single model without appealing to 2-entailment.

I don't have any strong argument that we *should* introduce such a dualizer at the level of single model. But it should be noted that the dualizable structure itself is *already there* at the level of the class of models, so it is not foreign even to the friend of the (extended) American plan. The star function only internalizes the existing structure down to the level of single model. And it should be added that the dualizer does not change the logic. As we saw above, whereas (DS2) collapses the four-valued structure into the essentially two-valued one, the logic remains to be $\mathbf{B}$. Then, why not?

5   Conclusion

I have told a story about how the Australian plan is developed from the (extended) American plan. We started with Routley's four-valued frame semantics. Since it has falsity as a primitive notion, we can explain and introduce the star-like relations without falling into the circularity that De and Omori allege to find in the two-valued framework of the negation-as-modality approach. I claimed that this is what Routley suggested as the virtue of the American plan. It indeed provides an explanatory basis to make sense of the star function, which it is originally purported to avoid.

Then I analyzed the final step towards the Australian plan. We saw that the dualist star postulate on the binary relations of compatibility and exhaustiveness, especially its half (DS2), makes them a function, and in the present case, collapses the four-valued structure into the essentially two-valued one. An examination of rule-contraposition at different levels showed that the step is understood as internalization of dualizable structure that can be found in the class of models. We may perhaps say that this is the *ontological virtue* of the American plan. The structure of the star function as a dualizer is not an invention of the Australian plan. It is essentially there from the beginning, and the star just makes it available within

a single model.

And moreover, the final step does not change the logic. So as far as its significance is clearly understood as I have explained, we may say that the step is harmless. Thus, the (extended) American plan provides an explanatory and ontological basis for the Australian plan. The latter is just a developed form of the former[6].

References

[1] BELNAP, N. How a computer should think. In *Contemporary Aspects of Philosophy*, G. Ryle, Ed. Oriel Press, 1977, pp. 30–55.

[2] BELNAP, N. D. A useful four-valued logic. In *Modern Uses of Multiple-valued Logic*, J. Dunn and G. Epstein, Eds. Springer, 1977, pp. 5–37.

[3] BRANDOM, R. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1998.

[4] BRANDOM, R. *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford University Press, 2008.

[5] COPELAND, J. On When a Semantics is not a Semantics: some reasons for disliking the Routley-Meyer semantics for relevance logic. *Journal of Philosophical Logic 8* (1979), 399–413.

[6] DE, M., AND OMORI, H. There is more to negation than modality. *Journal of Philosophical Logic 47*, 2 (Apr 2018), 281–299.

[7] DOŠEN, K. Negation as a modal operator. *Reports on Mathematical Logic 20* (1986), 15–27.

---

[6]Although we have focused on **B**, it is clear that, for any positive extensions of **B** Routley considered in the paper, the step of adding (DS2) and making the star function definable does not change the logic. So the Australian plan for an extension of the kind can also be taken as just a developed form of the (extended) American plan for *that* extension. On the other hand, as we saw, some negative extensions do not have a decent formulation in the extended American plan. Then, there is a sense in which the Australian plan for, e.g. **R**, which has the axiom of Contraposition, is *not* a developed form of the American plan for it, since the latter does not exist (I am not denying the possibility of "four-valued" semantics for **R** in any sense. See [12]). But there is another sense in which it *is*, since **R** is an extension of **B** for which the Australian plan is a developed form of the American plan.

[8] DUNN, J. M. Intuitive semantics for first-degree entailments and 'coupled trees'. *Philosophical studies 29*, 3 (1976), 149–168.

[9] DUNN, J. M. A comparative study of various model-theoretic treatments of negation: a history of formal negation. In *What is Negation?*, D. M. Gabbay and H. Wansing, Eds. Springer, 1999, pp. 23–51.

[10] DUNN, J. M., AND RESTALL, G. Relevance logic. In *Handbook of Philosophical Logic, Vol.6*, D. Gabbay and F. Guenthner, Eds. Kluwer Academic Publishers, 2002, pp. 1–128.

[11] HUMBERSTONE, L. Power matrices and Dunn–Belnap semantics: Reflections on a remark of Graham Priest. *The Australasian Journal of Logic 11*, 1 (2014).

[12] MARES, E. D. "Four-valued" semantics for the relevant logic R. *Journal of Philosophical Logic 33*, 3 (2004), 327–341.

[13] MARES, E. D. *Relevant Logic: A Philosophical Interpretation*. Cambridge University Press, 2004.

[14] OMORI, H., AND WANSING, H. 40 years of FDE: An introductory overview. *Studia Logica 105*, 6 (2017), 1021–1049.

[15] ONISHI, T. Substructural negations. *The Australasian Journal of Logic 12*, 4 (2015), 177–203.

[16] ONISHI, T. Understanding negation implicationally in the relevant logic R. *Studia Logica 104*, 6 (2016), 1267–1285.

[17] PEREGRIN, J. Logic as based on incompatibility. *The Logica Yearbook* (2010), 157–168.

[18] PEREGRIN, J. Logic reduced to bare (proof-theoretical) bones. *Journal of Logic, Language and Information 24*, 2 (2015), 193–209.

[19] PRIEST, G. *An Introduction to Non-classical Logic: From If to Is, 2nd Edition*. Cambridge University Press, 2008.

[20] RESTALL, G. Information flow and relevant logics. *Logic, language and computation* (1996), 1–14.

[21] RESTALL, G. Negation in relevant logics (how I stopped worrying and learned to love the Routley star). In *What is Negation?*, D. M. Gabbay and H. Wansing, Eds. Springer, 1999, pp. 53–76.

[22] ROUTLEY, R. The American plan completed: Alternative classical-style semantics, without stars, for relevant and paraconsistent logics. *Studia Logica 43*, 1-2 (1984), 131–158.

[23] ROUTLEY, R., AND MEYER, R. K. The semantics of entailment. In *Truth, Syntax and Modality*, H. Leblanc, Ed. North Holland, 1972, pp. 199–243.

[24] ROUTLEY, R., AND MEYER, R. K. The semantics of entailment II. *Journal of Philosophical Logic 1* (1972), 53–73.

[25] ROUTLEY, R., AND ROUTLEY, V. The semantics of first degree entailment. *Noûs 6*, 4 (1972), 335–359.

[26] VAN BENTHEM, J. What is dialectical logic? *Erkenntnis 14*, 3 (1979), 333–347.