Greg Restall

FIRST DEGREE ENTAILMENT, SYMMETRY AND PARADOX

Abstract. Here is a puzzle, which I learned from Terence Parsons in his "True Contradictions" [8]. *First Degree Entailment* (FDE) is a logic which allows for truth value gaps as well as truth value gluts. If you are agnostic between assigning paradoxical sentences gaps and gluts (and there seems to be no very good reason to prefer gaps over gluts or gluts over gaps if you're happy with FDE), then this looks no different, in effect, from assigning them a *gap* value? After all, on both views you end up with a theory that doesn't commit you to the paradoxical sentence or its negation. How is the FDE theory any different from the theory with gaps alone?

In this paper, I will present a clear answer to this puzzle — an answer that explains how being agnostic between gaps and gluts is a genuinely different position than admitting gaps alone, by using the formal notion of a *bi*-theory, and showing that while such positions might agree on what is to be *accepted*, they differ on what is to be *rejected*.

Keywords: first degree entailment; paradox; symmetry; models; theories

First Degree Entailment (FDE) is a logic well suited to fixed point solutions to the paradoxes. Perhaps it is *too* well suited, because it allows paradoxical sentences to be evaluated in two distinct ways: Paradoxical sentences can be assigned the value n (neither true nor false: $\{\}$) or b (both true and false — or $\{0, 1\}$) equally well. Are two possible values better than one? And more importantly, is agnosticism between *which* value to assign a paradoxical sentence like the liar — a stance Terence Parsons calls "agnostaletheism" — any different from assigning it the truth value n instead of b? After all, on either stance, neither the liar sentence nor its negation are to be accepted. In this note, I explore the symmetry

that is available in FDE, and I show how agnostaletheism may be clearly distinguished from the view according to which paradoxes are simply neither true nor false.

1. First Degree Entailment and relational evaluations

First Degree Entailment (FDE) is a simple and elegant logic, well suited to many different applications. It can be defined and understood in a number of different ways, but for our purposes it suits to introduce it as the generalisation of classical two-valued logic according to which evaluations are no longer functions assigning each sentence of a language a truth value from $\{0, 1\}$, but relations to those truth values. Relaxing the constraint that evaluations be Boolean functions means that sentences can be *neither* true nor false (the evaluation fails to relate the sentence to either 0 or 1) or *both* true and false (the evaluation relates the sentence to both truth values). This generalisation allows us to interpret the suite of connectives and quantifiers of predicate logic in a straightforward manner, generalising the traditional evaluation conditions due to Boole and Tarski as follows.[1]

Given a non-empty domain D of objects, an FDE-model for a language consists of a multi-sorted relation ρ defined as follows: For each n-place predicate F, ρ_F relates n-tuples of objects from D to the truth values $0, 1$. For each constant c in the language, ρ_c selects a unique object from D. An assignment α of values to the variables is a function from those variables to the domain D. Given an assignment α and the interpretation ρ we define the semantic value $[\![t]\!]_{\rho,\alpha}$ of a term t to be given by ρ_t if t is a name and $\alpha(t)$ if t is a variable. Then, relative to each assignment α we define the relation ρ_α which matches formulas in the language to truth values as follows:

$$
\begin{aligned}
(Ft_1 \cdots t_n)\rho_\alpha i &\quad \text{iff} \quad \langle [\![t_1]\!]_{\rho,\alpha}, \ldots, [\![t_n]\!]_{\rho,\alpha} \rangle \rho_F i \\
(A \wedge B)\rho_\alpha 1 &\quad \text{iff} \quad A\rho_\alpha 1 \text{ and } B\rho_\alpha 1 \\
(A \wedge B)\rho_\alpha 0 &\quad \text{iff} \quad A\rho_\alpha 0 \text{ or } B\rho_\alpha 0 \\
(A \vee B)\rho_\alpha 1 &\quad \text{iff} \quad A\rho_\alpha 1 \text{ or } B\rho_\alpha 1
\end{aligned}
$$

---

[1] The idea of a relational evaluation goes back to work by J. Michael Dunn, in the 1970s [4], and has been made use of in some influential papers by Nuel Belnap [1, 2]. My presentation follow's Graham Priest's treatment in *An Introduction to Non-Classical Logics* [9], since this is in wide use.

$$
\begin{array}{rcl}
(A \lor B)\rho_\alpha 0 & \text{iff} & A\rho_\alpha 0 \text{ and } B\rho_\alpha 0 \\
\neg A\rho_\alpha 1 & \text{iff} & A\rho_\alpha 0 \\
\neg A\rho_\alpha 0 & \text{iff} & A\rho_\alpha 1 \\
(\forall x)A\rho_\alpha 1 & \text{iff} & A\rho_{\alpha[x:=d]} 1 \text{ for every } d \text{ in } D \\
(\forall x)A\rho_\alpha 0 & \text{iff} & A\rho_{\alpha[x:=d]} 0 \text{ for some } d \text{ in } D \\
(\exists x)A\rho_\alpha 1 & \text{iff} & A\rho_{\alpha[x:=d]} 1 \text{ for some } d \text{ in } D \\
(\exists x)A\rho_\alpha 0 & \text{iff} & A\rho_{\alpha[x:=d]} 0 \text{ for every } d \text{ in } D
\end{array}
$$

The only deviation from classical first order predicate logic is that we allow for truth value gaps (ρ may fail to relate a given formula to a truth value) or gluts (ρ may relate a given formula to both truth values). Indeed, the possibilities of gaps and of gluts are, in a sense, *separable* or *modular*. It is quite straightforward to show that if a given interpretation ρ is a *partial function* on the basic vocabulary of a language — if it never over-assigns values to the extension of any predicate in that language — then it remains so over every sentence in that language. Sentences can be assigned gaps and not gluts. Similarly, if an interpretation is *decisive* over the basic vocabulary of some language — it never under-assigns values to the extensions of any predicate in that language — then it remains so over every sentence of that language. These sentences can be assigned gluts and not gaps. If an evaluation is *sharp* (if it allows for neither gaps nor gluts in the interpretation of any predicate), then it remains so over the whole language.

Relational evaluations are a very natural model for FDE. They show it to be an elementary generalisation of classical logic, allowing for gaps between truth values and over-assignment of those values. The interpretation of the connectives and the quantifiers remains as classical as in two-valued logic, except for the generalisation to allow for gaps and gluts between the two semantic values.

2. First Degree Entailment and four values

We can also see FDE in another light, not as a logic allowing for gaps and gluts between two truth values, but as a logic allowing for *four* semantic values. For clarity, we will write these four values: t, b, n, and f. We can translate between the two-valued and four-valued languages as

follows. Given any relational valuation ρ we define a matching functional valuation v_ρ like this:

$$v_\rho(A) = t \text{ iff } A\rho1 \text{ but not } A\rho0 \quad v_\rho(A) = b \text{ iff } A\rho1 \text{ and } A\rho0$$

$$v_\rho(A) = f \text{ iff } A\rho0 \text{ but not } A\rho1 \quad v_\rho(A) = n \text{ iff neither } A\rho1 \text{ nor } A\rho0$$

It follows then, that

$$A\rho1 \text{ iff } v_\rho(A) = t \text{ or } v_\rho(A) = b \quad A\rho0 \text{ iff } v_\rho(A) = f \text{ or } v_\rho(A) = b$$

Relational evaluations that are partial functions can be understood as functional evaluations taking semantic values from t, n, f — and the evaluation clauses in this case give us Kleene's strong three valued logic, K3. Evaluation relations that are decisive, allowing for no gaps, can be understood as taking semantic values from t, b, f — and the evaluation clauses in this case give us Priest's logic of paradox LP. In what follows, I will move between functional and relational vocabulary as seems appropriate.

3. First Degree Entailment and sequents

There are many different ways we can use FDE evaluations to analyse truth and consequence in the language of first order logic. One important notion goes like this: An interpretation ρ is said to be a *counterexample* to the sequent $X \succ Y$ if and only if ρ relates each member of X to 1 while it relates no member of Y to 1.[2] In other words, an interpretation provides a counterexample to a sequent if it shows some way that the sequent fails to preserve truth. Given some set $\mathcal{M}$ of evaluations, a sequent is said to be $\mathcal{M}$-valid if it has no counterexamples in the set $\mathcal{M}$. We reserve the term 'FDE-valid' for those sequents which have no counterexamples at all. Similarly, a sequent is said to be K3-valid if it has no counterexamples among partial function evaluations, and it sequent is said to be LP-valid if it has no counterexamples among decisive valuations.

All this is very well known in the literature on non-classical logics — see, for example Priest's *An Introduction to Non-Classical Logic* [9, Chapter 8] for more detail on the behaviour of these logical systems.

---

[2] For this paper it suffices to take sequents to be pairs $X \succ Y$ of sets of formulas. We allow these sets to be infinite. I follow Humberstone [6] in using '$\succ$' as a sequent separator, as a sequent with a counterexample remains a sequent. I will reserve the assertion sign '$\vdash$' for other uses.

The FDE-valid sequents include all of distributive lattice logic, with a de Morgan negation. For example, sequents such as these

$$A \vee (A \wedge B) \succ A \qquad A \succ A \wedge (A \vee B)$$

$$A \wedge (B \vee C) \succ (A \wedge B) \vee C$$

$$\neg(A \wedge B) \succ \neg A \vee \neg B \qquad \neg(A \vee B) \succ \neg A \wedge \neg B$$

$$\neg A \vee \neg B \succ \neg(A \wedge B) \qquad \neg A \wedge \neg B \succ \neg(A \vee B)$$

$$A \succ \neg\neg A \qquad \neg\neg A \succ A$$

are FDE valid. The next sequents are not valid in FDE, but they are valid in K3:

$$A, \neg A \succ \qquad A \vee B, \neg A \succ B$$

In both cases, an FDE interpretation which relates A to both 0 and 1 (but which fails to relate B to 1) serves as a counterexample. Similarly, the following sequents are *not* valid in FDE, but they are valid in LP:

$$\succ A, \neg A \qquad B \succ A \wedge B, \neg A$$

In both cases, an FDE interpretation which relates A to neither 0 nor 1 (but which relates B to 1) serves as a counterexample.

4. First Degree Entailment, theories and bitheories

From sequents we move to theories. The usual definition has it that a *theory* may be defined as a set of sentences closed under a logical consequence relation. So, given some collection $\mathcal{M}$ of interpretations, T is an $\mathcal{M}$-theory if and only if whenever the sequent $T \succ A$ (where A is a single formula) is $\mathcal{M}$-valid, then A is a member of T. $\mathcal{M}$-theories contain their own $\mathcal{M}$-consequences. We can think of theories as representing what is held to be true according to a certain stance — a consequences of what is held true is also (implicitly) held true. Elsewhere [11, 13], I have argued that in logics like FDE we have good reason to consider not only what is held true, but what is held *un*true. Sequents give us a straightforward vocabulary for describing this.

We may say that the disjoint pair $\langle T, U \rangle$ is an $\mathcal{M}$-*bitheory* if and only if whenever the sequent $T \succ A, U$ (where A is a single formula) is $\mathcal{M}$-valid, then A is a member of T, and whenever $T, A \succ U$ is $\mathcal{M}$-valid, then A is a member of U. Bitheories are *balanced*, in that $\langle T, U \rangle$ is a pair, consisting of what is (according to that bitheory) held true (to be

related to 1) on the one hand and what is held untrue (to be unrelated to 1), on the other.

Suppose $\mathcal{M}' \subseteq \mathcal{M}$ is another set of interpretations. If we define $T_{\mathcal{M}'}$ to be the set of all sentences true (related to 1) in all $\mathcal{M}'$-interpretations and likewise $U_{\mathcal{M}'}$ to be the set of all sentences untrue (not related to 1) in those interpretations, then $\langle T_{\mathcal{M}'}, U_{\mathcal{M}'} \rangle$ is clearly an $\mathcal{M}$-bitheory. Indeed, if $\mathcal{M}'$ is a singleton set, consisting of one interpretation, then the bitheory $\langle T_{\mathcal{M}'}, U_{\mathcal{M}'} \rangle$ is a partition of the language, deciding every formula to be either true or untrue. If the set $\mathcal{M}'$ is larger, containing interpretations which give a sentence A different truth verdicts, then the corresponding bitheory will no longer be a partition. If one interpretation judges A to be true, another judges it untrue, then A will neither feature in the left set nor the right set.

5. First Degree Entailment and truth

The puzzle under consideration in this paper arises from the behaviour of paradoxical sentences in FDE. The finer details of the paradoxes are not important to us, but regardless, we will consider a specific concrete case, the paradoxes of truth. We will consider a transparent truth predicate, so let us focus on first order languages in which we have a one-place predicate T for truth. Since the truth predicate is a *predicate*, it will apply to objects in the domain. To allow for fixed points (sentences which ascribe truth or falsity to sentences in the language, including themselves) we supply the language with quotation names for all sentences in that very same language. So, for each sentence A we have a name $\ulcorner A \urcorner$ among the singular terms of the language.[3] Fixed point constructions for truth in the manner of Kripke, Brady, Woodruff and Gilmore [3, 5, 7, 16] generate FDE-interpretations for a language in which the sentence A and the sentence $T\ulcorner A \urcorner$ are assigned the same semantic values. We will call such interpretations FDE^T interpretations. The construction method for FDE^T-interpretations assigns the extension of T in stages, keeping the rest of the evaluation as given, including the denotation for constants. The details of the proof are not important to us, but one essential idea is useful: the notion of *preservation* between evaluations. For two evaluations ρ and ρ', we have $\rho \sqsubseteq \rho'$ if and only if whenever ρ relates an atomic

---

[3] This could be achieved by way of a Gödel numbering, where $\ulcorner A \urcorner$ is the numeral of the Gödel number of the sentence A, but quotation need not use that device.

formula to a given truth value 0 or 1, so does ρ'. It is a straightforward induction on the complexity of formulas that this then extends to all of the formulas in the language: for any formula A, if $A\rho 0$ then $A\rho' 0$ too, and if $A\rho 1$ then $A\rho' 1$ too. The evaluations ρ and ρ' may still differ, because ρ might leave a *gap* where ρ' fills in a value, 0 or 1, or where ρ assigned only one value, ρ' might assign *both*.

The only requirement on quotation names for this fixed point construction to succeed is that quotation names for different sentences are different. This means that the construction will work whatever we take the denotation of other constants to be. So, let's consider a language with a countable supply of constants $\lambda, \lambda_1, \lambda_2, \ldots$ whose denotation can be freely set however we please.

So FDE^T is the set of relational FDE evaluations for this language in which T is a fixed point — that is, for any sentence A, that sentence receives the same evaluation as $T\ulcorner A\urcorner$. FDE^T can also be considered as a theory (or bitheory), if we wish to consider what holds (and fails to hold) in all such evaluations. We can do the same for K3^T and LP^T, when we restrict our attention to evaluations in which there are no truth value gluts or gaps respectively. Kripke's original construction shows us how to make K3^T evaluations, and the construction generalises to LP^T and FDE^T straightforwardly.

Now, to consider the behaviour of the paradoxical sentences, let's fix the referent of the term λ to be the same as the referent of the quotation name $\ulcorner \neg T\lambda \urcorner$, containing the term λ itself. It follows then that $T\lambda$ has the same value as $T\ulcorner \neg T\lambda \urcorner$ (since λ and $\ulcorner \neg T\lambda \urcorner$ have the same referent), which has the same value as $\neg T\lambda$ (since truth is transparent). In other words, λ denotes a liar sentence, which says of itself that it's not true. That is, the sentence $\neg T\lambda$ (and its mate, $T\lambda$) must be assigned the value b or n, since it is a fixed point for negation. The fixed point construction allows us to generate interpretations for the truth predicate in which sentences like $\neg T\lambda$ have the value n, and interpretations where those sentences have the value b — in fact, one can make the fixed point construction purely in K3 or in LP — and there are also mixed models in which some paradoxical sentences have the value n and others the value b.

So, if we take FDE^T to be an adequate *logic* of truth, then it seems as if we should be agnostic about whether a liar sentence like $\neg T\lambda$ has value n or b, unless we can find some consideration which breaks the tie between them. This position was named "agnostaletheism", by Terence Parsons [8].

Now perhaps there *is* a tie-breaking consideration between gaps and gluts. If we were to be agnostic between assigning $\neg T\lambda$ the value b and the value n, this looks a lot like assigning the value n and not b. After all, according to both theories, we don't assert $T\lambda$ and we don't assert its negation. This is the puzzling question: Is there an instability in FDE^T? Does FDE^T collapse into K3^T?

6. Symmetry in FDE theories

The profound symmetry between gaps and gluts in first degree entailment is manifest in the behaviour of the Routley star — a function on evaluations — introduced by Richard and Valerie Routley in the 1970s [14]. Given an evaluation ρ, we can define its *dual* evaluation ρ^* as follows — for each n-place predicate F:

- $\langle d_1, \ldots, d_n \rangle \rho_F^* 1$ holds iff $\langle d_1, \ldots, d_n \rangle \rho_F 0$ doesn't hold.
- $\langle d_1, \ldots, d_n \rangle \rho_F^* 0$ holds iff $\langle d_1, \ldots, d_n \rangle \rho_F 1$ doesn't hold.

In other words, an atomic formula is *true* according to ρ^* if and only if to ρ it is not false, and it is *false* according to ρ^* if and only if to ρ it is not true. This means that atomic formulas which are t by ρ's lights are also t by ρ^*'s, and similarly for f. But a formula that is n according to ρ is b to ρ^*, and a formula that is b according to ρ is n to ρ^*. The dual evaluation turns gaps into gluts, and gluts into gaps, for atomic formulas. This fact generalises to all of the formulas in the language of FDE.

FACT 1. *For any formula A in the language of* FDE *and for any relational evaluation ρ,*

- *$A\rho^*1$ holds iff $A\rho 0$ doesn't hold.*
- *$A\rho^*0$ holds iff $A\rho 1$ doesn't hold.*

This fact is established by a simple induction on the complexity of the formula A. The crucial feature of the connectives that makes this proof work is the balance between the positive and negative conditions in an evaluation ρ. For example, with conjunction we have

- $(A \wedge B)\rho_\alpha 1$ iff $A\rho_\alpha 1$ and $B\rho_\alpha 1$.
- $(A \wedge B)\rho_\alpha 0$ iff $A\rho_\alpha 0$ or $B\rho_\alpha 0$.

So we can proceed as follows (assuming that the fact holds for the simpler formulas A and B), $(A \wedge B)\rho_\alpha^* 1$ iff $A\rho_\alpha^* 1$ and $B\rho_\alpha^* 1$ iff $A\rho_\alpha 0$ doesn't hold and $B\rho_\alpha 0$ doesn't hold, iff neither $A\rho_\alpha 0$ nor $B\rho_\alpha 0$ hold, iff $A \wedge B\rho_\alpha 0$ doesn't hold. We have appealed to the parallel between these two clauses:

- $(A \wedge B)\rho_\alpha 1$ holds iff $A\rho_\alpha 1$ and $B\rho_\alpha 1$ don't hold.
- $(A \wedge B)\rho_\alpha 0$ doesn't hold iff $A\rho_\alpha 0$ and $B\rho_\alpha 0$ don't hold.

In the same way, for example, with the existential quantifier:

- $(\exists x)A\rho_\alpha 1$ holds iff $A\rho_{\alpha[x:=d]}1$ holds for some d in D,
- $(\exists x)A\rho_\alpha 0$ doesn't hold iff $A\rho_{\alpha[x:=d]}0$ doesn't hold for some d in D,

and the same form of argument applies. What holds for the existential quantifier and conjunction holds for the other connectives and quantifier of first degree entailment.

Excursus. This argument would fail if we had connectives or quantifiers in our language whose truth and falsity conditions are less well matched. For example, we could have a connective which is conjunctive with regard to truth and disjunctive with regard to falsity:

- $(A \times B)\rho_\alpha 1$ iff $A\rho_\alpha 1$ and $B\rho_\alpha 1$.
- $(A \times B)\rho_\alpha 0$ iff $A\rho_\alpha 0$ and $B\rho_\alpha 0$.

Given a evaluation ρ which relates the atomic formulas p to 1 only and q to 0 only, ρ^* does the same. According to both ρ and ρ^*, $p \times q$ is related to neither 1 nor 0, breaking the symmetry between gaps and gluts.   ◇

The Routley star sends relational evaluations to relational evaluations. It does not send theories to theories. It is natural to define the star of a set of sentences as follows: For any set S of formulas, $A \in S^*$ if and only if $\neg A \notin S$. However, the dual T^* of a theory T is not always itself a theory. Take, for example, the FDE-theory $(\neg p \vee \neg q)\!\downarrow$ consisting of all FDE-consequences of $\neg p \vee \neg q$ (it is the theory consisting of every sentence made true by every evaluation ρ where either $p\rho 0$ or $q\rho 0$). In particular, we have $\neg p \vee \neg q \in (\neg p \vee \neg q)\!\downarrow$ but $\neg p \notin (\neg p \vee \neg q)\!\downarrow$ and $\neg q \notin (\neg p \vee \neg q)\!\downarrow$. Now consider the dual set $(\neg p \vee \neg q)\!\downarrow^*$. This is not a theory, because $p \in (\neg p \vee \neg q)\!\downarrow^*$ (since $\neg p \notin (\neg p \vee \neg q)\!\downarrow$)) and $q \in (\neg p \vee \neg q)\!\downarrow^*$ (since $\neg q \notin (\neg p \vee \neg q)\!\downarrow$)) but the conjunction is not in the set: $p \wedge q \notin (\neg p \vee \neg q)\!\downarrow^*$ (since $\neg p \vee \neg q \in (\neg p \vee \neg q)\!\downarrow$) ensures that $\neg(p \wedge q) \in (\neg p \vee \neg q)\!\downarrow$ too).

However, it is straightforward to show the following fact, relating the Routley star and *bi*-theories.

FACT 2. *For any $\mathcal{M}$-bitheory $\langle T, U \rangle$, the pair $\langle \overline{U^*}, \overline{T^*} \rangle$ is an $\mathcal{M}^*$-bitheory, where $\overline{U^*}$ and $\overline{T^*}$ are the sets of sentences of the language not in U^* and T^* respectively, and where $\mathcal{M}^* = \{\rho^* : \rho \in \mathcal{M}\}$.*

Here is why: The interpretation ρ is a counterexample to $T \succ U$ has a counterexample iff ρ^* is a counterexample to $\neg U \succ \neg T$. It follows that $\overline{U^*} \succ A, \overline{T^*}$ fails at ρ^* iff $\neg \overline{T^*}, \neg A \succ \neg \overline{U^*}$ fails at ρ, but that means

$T, \neg A \succ U$ fails at ρ^*. So, $\overline{U^*} \succ A, \overline{T^*}$ holds in $\mathcal{M}^*$ iff $T, \neg A \succ U$ holds in $\mathcal{M}$. So, since $\langle T, U \rangle$ is an $\mathcal{M}$-bitheory, we have $\neg A \in U$, which means $A \in \overline{U^*}$ as desired. The case for $\overline{U^*}, A \succ \overline{T^*}$ to $A \in \overline{T^*}$ is dual.

Armed with these facts concerning the Routley star, we can attend to the behaviour of our theories (and bitheories) with gaps and gluts.

7. Two kinds of incompleteness

Theories in FDE can be incomplete in two distinct ways. Consider the FDE-theory consisting of every sentence true in those evaluations which relate p to 1 and relate q to neither 0 nor 1, and which relate r to either 1 or 0. This set of sentences contains p and it doesn't contain $\neg p$. It holds p to be *true*. However, it is incomplete concerning q and r — the theory doesn't contain q or $\neg q$, and it also doesn't contain r or $\neg r$. However, the theory has settled q to be neither true nor false. (In all of the evaluations, q receives the value n.) On the other hand, the value of r is unsettled. In some evaluations, r is true, in others it is false. In this way, FDE allows for two different kinds of incompleteness.

Now consider theories like $\text{K}3^T$ and FDE^T. Recall, FDE^T is given by all FDE evaluations for which $T^\ulcorner A^\urcorner$ and A receive the same value, and $\text{K}3^T$ is given by all K3 valuations with the same property. If we focus on the *theories* determined by each class of valuations, we see that a liar sentence like $T\lambda$ is incomplete in both theories. In $\text{K}3^T$, it is because in any such valuation, $T\lambda$ receives the value n — it is never true. In FDE^T it is because in any such valuation, $T\lambda$ either receives the value n or the value b. In some valuations it is true (those where it is b) and in others, it fails to be true. Again, the theory is incomplete concerning $T\lambda$.

Is there any way to distinguish these theories or distinguish this incompleteness?

In one sense, the answer will be *no*. The following fact contains the core of the reason:

FACT 3. *For any* K3 *evaluation* ρ, *the theory determined by* ρ *and the* FDE *theory determined by the two evaluations* ρ *and* ρ^* *are identical.*

It is easy to see that $\rho \sqsubseteq \rho^*$ in the case where ρ is a K3 evaluation. It follows that the truths according to ρ are exactly the truths according to both ρ and ρ^*.

This fact *generalises*. Consider an evaluation ρ, which may involve both gaps and gluts. We can define the evaluation ρ^n, which assigns n

to any atomic formula assigned either n or b by ρ, and which leaves t and f fixed. It is straightforward to see that $\rho^n \sqsubseteq \rho$. We can also define the evaluation ρ^b, which assigns b to any atomic formula assigned either n or b by ρ, and which leaves t and f fixed. In this case, we have $\rho \sqsubseteq \rho^b$. So, in general any FDE evaluation ρ is sandwiched between a K3 evaluation and an LP evaluation like so: $\rho^n \sqsubseteq \rho \sqsubseteq \rho^b$.

The generalisation of our previous fact can now be stated:

FACT 4. *For any* FDE *evaluation* ρ, *the* K3 *theory determined by* ρ^n *and the* FDE *theory determined by the two evaluations* ρ *and* ρ^n *are identical.*

The proof is as before: Now $\rho^n \sqsubseteq \rho$, so it follows that the truths according to ρ^n are exactly the truths according to both ρ^n and ρ. Now, the operation of sending all gaps and gluts either to *gaps* or to *gluts* does not disturb the logic of truth.

FACT 5. *If* ρ *is an* FDE^T *evaluation, then so are* ρ^n *and* ρ^b.

The only way that ρ^n could fail to be an an FDE^T evaluation is if for some formula A, the values in ρ^n of A and $T\ulcorner A \urcorner$ differ. But if values of two formulas differ in ρ^n, they also differ in ρ. (The same holds for ρ^b, too.)

Now we can state our general fact, concerning truth theories in FDE and K3. The basic idea is that the theories are identical, since theories that take the paradoxical sentences to be n and those that are agnostic between n and b take the same claims to be true. This is fair enough as far as it goes, but stated in this bald way, it does not go very far at all. The theories FDE^T and K3^T obviously have the same theorems — they both have no theorems. The silent evaluation which sends absolutely every every formula to n is a K3^T (and hence, FDE^T) evaluation, and this shows that both K3^T and FDE^T have no theorems at all. So, merely showing that K3^T and FDE^T share theorems does not say very much at all. We can do much better than this.

Suppose we have a set $\mathcal{M}$ of evaluations, such that whenever $\rho \in \mathcal{M}$ we also have $\rho^n \in \mathcal{M}$. Let $\mathcal{M}^n$ be the set of K3 evaluations in $\mathcal{M}$ — so $\mathcal{M}^n$ is $\{\rho^n : \rho \in \mathcal{M}\}$. We have the following result:

FACT 6. *The theory* $T_\mathcal{M}$ *of sentences true in all evaluations in* $\mathcal{M}$ *is identical to the theory* $T_{\mathcal{M}^n}$, *of sentences in all evaluations in* $\mathcal{M}^n$.

Clearly $T_{\mathcal{M}} \subseteq T_{\mathcal{M}}^n$. To show the converse, suppose the formula A is not in $T_{\mathcal{M}}$. So, it fails to be true on some evaluation $\rho \in \mathcal{M}$. It also fails in ρ^n, which is in $\mathcal{M}^n$.

So, for example, if we have some K3 valuation ρ for a language without the truth predicate, and we consider the set $\mathcal{M}$ of all FDE^T evaluations, extending ρ with a truth predicate. Here, grounded T-sentences will receive values as determined by the underlying valuation ρ, while other sentences will vary among all four values, t, f, n and b, constrained only by the requirement that A and $T\ulcorner A \urcorner$ agree in value. The set $\mathcal{M}^n$ is the subset of such evaluations in which the T-sentences receive the values t, f or n, not b. Our fact tells us the theories of $\mathcal{M}$ and $\mathcal{M}^n$ are indistinguishable. At the level of theories, we cannot distinguish between paradoxical sentences determinately receiving a gap value, and agnosticism between gaps and gluts.

Thankfully, we don't need to remain at the level of theories. The sets $\mathcal{M}$ and $\mathcal{M}^n$ determine the same set of theorems, but they determine different sets of *co*theorems. While they rule in the same sentences, they rule *out* different sentences. The liar sentence $\neg T\lambda$ is true in some valuations in $\mathcal{M}$ (those that assign it the value b) while it is true in no valuations in $\mathcal{M}^n$. In all valuations in $\mathcal{M}^n$ a liar sentence must receive the value n, so it is true in no valuation at all. The untruths of $\mathcal{M}$ differ from the untruths of $\mathcal{M}^n$. So if we attend to bitheories, the symmetry between gaps and gluts is completely restored. For our facts concerning gaps, we have matching facts concerning gluts.

FACT 7. *For any* FDE *evaluation* ρ, *the* LP *cotheory determined by* ρ^b *(the formulas U_{ρ^b} untrue in ρ^b) and the* FDE *cotheory determined by the two evaluations ρ and ρ^b (the formulas $U_{\{\rho, \rho^b\}}$) are identical.*

FACT 8. *If $\mathcal{M}$ is a set of valuations where for every ρ in $\mathcal{M}$ the valuation ρ^b is also in $\mathcal{M}$, then the cotheory $U_{\mathcal{M}}$ of sentences untrue in all evaluations in $\mathcal{M}$ is identical to the cotheory $U_{\mathcal{M}^b}$, of sentences untrue in all evaluations in $\mathcal{M}^b$.*

Symmetry is regained, and we can distinguish between agnostaletheism concerning paradoxical sentences and those views which assign them a gap, or assign them a glut. Glut views are distinguished from agnostaletheism as *theories* — they hold different sentences to be true, while gap views are distinguished from agnostaletheism as *cotheories* — they hold different sentences to be untrue.

This is all well and good when it comes to formally distinguishing these three views of paradoxical sentences. However, the puzzle wasn't just a puzzle about the formal development of these views. It is also a puzzle concerning what it is to hold those views, and this issue remains, even if we reject the model theory and the technical devices of theories, cotheories and bitheories.

8. Assertion and denial in FDE, K3 and LP

To answer the puzzle in those terms, we should say something about the speech acts of assertion and denial, or the cognitive states of accepting and rejecting. These are the practical analogues of the theoretical and abstract notions of theory and cotheory. To connect talk of accepting and rejecting (or assertion and denial) with logical notions, we need some kind of bridge principle. A principle I have endorsed elsewhere [10,11,13] goes like this:

BRIDGE PRINCIPLE 1. If the sequent $X \succ Y$ is valid, then don't accept (or assert) every member of X and reject (or deny) every member of Y.

To constrain what you accept and reject in line with such a bridge principle is to maintain a kind of coherence in your cognitive state. Since $A \lor B \succ A, B$ is valid, you would not accept the disjunction $A \lor B$ and reject both disjuncts A and B. If (as LP would have it) $\succ C \lor \neg C$ is valid, you would not reject that instance of the law of the excluded middle. If (as K3 would have it) $D \land \neg D \succ$ is valid, you would not accept that contradiction.

With this bridge principle at hand, we can distinguish the agnostaletheist (who uses a range of FDE^T valuations to define validity), the K3-theorist (who restricts her attention to $K3^T$ valuations) and the LP-theorist (who restricts his attention to LP^T valuations). The K3 theorist will not accept any contradiction. Contradictions are never true in any evaluation of theirs. The LP theorist will never reject any excluded middle. Excluded middle disjunctions are never untrue in their evaluations. The FDE theorist, on the other hand, can reject excluded middles and accept contradictions. That concerns validity and the first bridge principle, which amounts to a kind of coherence (or consistency) principle.

To accept a contingent theory, or better, the bitheory $\langle T, U \rangle$ is to constrain your acceptings and rejectings further.

BRIDGE PRINCIPLE 2. To accept a bitheory $\langle T, U \rangle$ is to accept each member of T and to reject each member of U.

This constraint is compatible with the first bridge principle if the pair $\langle T, U \rangle$ is indeed a bitheory. In that case, the sequent $T \succ U$ is not valid (if it were, then each formula A would be in T, since $T \succ A, U$ is valid, and in U, since $T, A \succ U$ is valid, but that is impossible, since T and U are, by definition, disjoint), so there is no issue with accepting all of T and rejecting all of U.

If we consider our three different views of the truth predicate (1) FDE^T allowing both gaps and gluts, (2) K3^T allowing only gaps, and (3) LP^T allowing only gluts as determining bitheories, we can see the difference in our acceptings and rejectings if we adopt BRIDGE PRINCIPLE 2 for each bitheory in turn. If we accept K3^T, we reject all contradictions, even those involving the liar sentence $T\lambda \wedge \neg T\lambda$. If we accept LP^T, we accept all excluded middles, including the excluded middle involving the liar: $T\lambda \vee \neg T\lambda$. But the agnostalethic position, accepting FDE^T, commits us to neither: we are free to accept the contradiction $T\lambda \wedge \neg T\lambda$ or to reject the disjunction $T\lambda \vee \neg T\lambda$.

So, an agnostaletheist and a gap theorist indeed agree on what to accept, but they disagree on what is to be rejected. In a similar way, an agnostaletheist and an glut theorist agree on what to reject, but they disagree on what to accept. Keeping the symmetry between accepting and rejecting in view, we have parity between gaps and gluts, and the agnostalethic position can be distinguished from its two neighbours.[4]

---

[4] Thanks to an anonymous referee for helpful comments and suggestions. One suggestion worth further exploration is other examples of the kind of duality between K3 and LP explored here. A clear case is the duality between intuitionistic logic and dual-intuitionistic logic [12, 15]. Intuitionistic logic and classical logic share counter-theorems but differ in theorems. Classical logic and dual-intuitionistic logic share theorems but differ in counter-theorems. It would be instructive to further explore this relationship as a part of a general theory of dualities between consequence relations.

References

[1] Nuel Belnap, "How a computer should think", pages 30–55 in G. Ryle (ed.), *Contemporary Aspects of Philosophy*, Oriel Press, 1976.

[2] Nuel Belnap, "A useful four-valued logic", pages 8–37, chapter 2, in J. M. Dunn and G. Epstein (eds.), *Modern Uses of Multiple-Valued Logic*, D. Reidel, 1977. DOI: 10.1007/978-94-010-1161-7_2

[3] Ross T. Brady, "The simple consistency of a set theory based on the logic CSQ", *Notre Dame Journal of Formal Logic*, 24, 4 (1983): 431–449. DOI: 10.1305/ndjfl/1093870447

[4] J. Michael Dunn, "Intuitive semantics for first-degree entailments and 'coupled trees'", *Philosophical Studies*, 29, 3 (1976): 149–168. DOI: 10.1007/BF00373152

[5] Paul C. Gilmore, "The consistency of partial set theory without extensionality", pages 147–153 in *Axiomatic Set Theory*, volume 13 of *Proceedings of Symposia in Pure Mathematics*, Providence, Rhode Island, 1974, American Mathematical Society. DOI: 10.1090/pspum/013.2/0360271

[6] Lloyd Humberstone, *The Connectives*, The MIT Press, 2011.

[7] Saul Kripke, "Outline of a theory of truth", *The Journal of Philosophy*, 72, 19 (1975): 690–716. DOI: 10.2307/2024634

[8] Terence Parsons, "True contradictions", *Canadian Journal of Philosophy*, 20, 3 (1990): 335–354. DOI: 10.1080/00455091.1990.10716495

[9] Graham Priest, *An Introduction to Non-Classical Logic: From if to is*, Cambridge University Press, Cambridge, 2008. DOI: 10.1017/CBO9780511801174

[10] Greg Restall, "Multiple conclusions", pages 189–205 in Petr Hájek, Luis Valdés-Villanueva, and Dag Westerståhl (eds.), *Logic, Methodology and Philosophy of Science: Proceedings of the Twelfth International Congress*, KCL Publications, 2005. http://consequently.org/writing/multipleconclusions

[11] Greg Restall, "Assertion, denial and non-classical theories", chapter 6 in K. Tanaka, F. Berto, E. Mares, and F. Paoli (eds.), *Paraconsistency: Logic and Applications*, volume 26 of *Logic, Epistemology and the Unity of Science*, Springer 2013. DOI: 10.1007/978-94-007-4438-7_6

[12] Greg Restall, "Pluralism and proofs", *Erkenntnis*, 79, 2 (2014): 279–291. DOI: 10.1007/s10670-013-9477-9

[13] Greg Restall, "Assertion, denial, accepting, rejecting, aymmetry and paradox", pages 310–321 in Colin R. Caret and Ole T. Hjortland (eds.), *Foundations of Logical Consequence*, Oxford University Press, 2015. DOI: 10.1093/acprof:oso/9780198715696.003.0011

[14] Richard Routley and Valerie Routley, "Semantics of first degree entailment", *Noûs*, 6, 4 (1972): 335–359. DOI: 10.2307/2214309

[15] Igor Urbas, "Dual-intuitionistic logic", *Notre Dame Journal of Formal Logic*, 37, 3 (1996): 440–451. DOI: 10.1305/ndjfl/1039886520

[16] Peter W. Woodruff, "Paradox, truth and logic. Part I: Paradox and truth", *Journal of Philosophical Logic*, 13, 2 (1984): 213–232. DOI: 10.1007/BF00453022

GREG RESTALL
Philosophy Department
The University of Melbourne
restall@unimelb.edu.au