Truth Grounding and the Liar

Adam InTae Gerard[1]

0.0 Introduction and Aim

Alethic paradox has beguiled philosophers and logicians since classical antiquity. Consider the following sentence $\lambda =_{df} \lambda$ is not true. If λ is true, then λ is not true. If λ is not true, then λ is true. But, λ cannot be both true and not true. Call sentence λ the *liar sentence* and the paradox that results from λ the *liar paradox*. Standard proposals to remedy the liar paradox require either an ascending hierarchy of languages or the denial of classical logic. Most, perhaps all, solutions remain susceptible to *revenge* or fail to resolve associated antinomies like the Yablo paradox.

Some philosophers have come to dismiss the Liar paradox as an uninteresting or otherwise irrelevant problem. Such an attitude is not only vastly ignorant of the implications of alethic paradox (which include the limitations of classical logic and hence, rationality; mathematics; and most of the deeply entrenched assumptions framing contemporary metaphysics), such an attitude also ignores the rich history behind alethic paradox (and the semantic paradoxes in general) which have, for example, motivated most of the great mathematical and philosophical projects of the early 20th century including Russell and Whitehead's *Principia Mathematica*; the gradual development of ZFC set theory which has for the most part been accepted as the foundation for modern mathematics (one of the oft-called four pillars of mathematics); and Tarski's semantic conception of truth which led to the development of Model Theory (one of the other four pillars of mathematics) which in turn has guided computational semantics, formal linguistics, and most of the contemporary philosophical truth debate.

It is the aim of this short essay to combine the *neglected deflationary solution* proposed by JC Beall[2] with restrictions on axiom[3] **T-Scheme** in the manner of Solomon Feferman[4] and Saul Kripke[5] in order to secure a classical, sentential, standard *revenge*-immune solution to known alethic paradox.

1.0 Alethic Paradox

I assume familiarity with the canonical alethic paradoxes, namely: the liar paradox, Boolean compounds[6], Curry's paradox[7], Yablo's paradox[8] and paradoxes arising from liar cycles[9].

---

[1] Northern Illinois University. E: adam.gerard@gmail.com. W: www.adamintaegerard.com. I am deeply indebted to David Ripley, Graham Priest, Solomon Feferman, Matt Babb, and Richard Han for their comments and objections and to JC Beall, Bradd Hart, and Jason Winning for their advice.
[2] See Beall 2001.
[3] For ease I'll treat **T-Scheme** and **F-Schemes** as axioms.
[4] See Feferman 2008.
[5] See Kripke 1975.
[6] See Appendix I.
[7] See Appendix II.
[8] See Yablo 1993 and Appendix III.

Roughly, an alethic paradox is a paradox that pertains to our conception of truth - a paradox whose shortest proof requires the use of our alethic inferences. I begin with a suitable first-order language ***L*** supporting a standard sound and complete first-order axiom system ***FOL***. I extend ***L*** by introducing a truth-predicate T (with associated machinery[10]) and a truth grounded predicate * yielding language $\boldsymbol{L}_{T^*}$.

I now extend ***FOL*** with the unqualified axioms **T-Scheme** and **F-Scheme** yielding axiom system $\boldsymbol{FOL}_T$. $\boldsymbol{FOL}_T$ thus comprises the standard classical first-order axioms supplemented with the axioms of Peano Arithmetic (***PA***) and the following alethic axioms:

[T-Scheme] $Tf(S) \leftrightarrow S$ **[F-Scheme]** $Tf(\neg S) \leftrightarrow Ff(S)$

Here 'S', 'Q', ... range over unambiguous declarative sentences of the language under inquiry and 'f' denotes an adequate name-forming function sending sentences to sentence-names. It is easy to show that $\boldsymbol{FOL}_T$ is inconsistent. Consider the liar sentence λ: $\lambda \vdash_{\mathrm{FOLT}} Tf(\lambda) \wedge \neg Tf(\lambda)$.

> **Proof.** By diagonalization construct the sentence $\lambda =_{\mathrm{df}} \neg Tf(\lambda)$. This yields logical equivalence $\lambda \leftrightarrow \neg Tf(\lambda)$. By **T-Scheme** we acquire $Tf(\lambda) \leftrightarrow \lambda$. By the transitivity of the biconditionals we derive $\neg Tf(\lambda) \leftrightarrow Tf(\lambda)$. $Tf(\lambda) \wedge \neg Tf(\lambda)$ immediately follows. ■

It is the aim of this paper to improve on truth-theory $\boldsymbol{FOL}_T$ by restricting the axioms of our theory of truth to *truth grounded* sentences.

1.1 Diagnosing Alethic Paradox

Alethic paradox arises from the conjunction of the following three theses:

[Language] For any language with sufficient syntactic and semantic expressiveness[11] ***L***, truth-in-***L*** is defined within ***L***.

[Scheme] **T-Scheme** and **F-Scheme**.

[Classical] A theory of truth for natural language ought to be classically constrained.

From **language** we acquire a language, such as $\boldsymbol{L}_{T^*}$, with the syntactic resources and expressive machinery to support the construction of a sentence that contains a negated alethic predicate predicated of its own sentence-name. As we have seen, from **language**, **scheme**, and **classical** we obtain an alethic paradox.

---

[9] See Appendix IV.

[10] Namely, the language of ZFC set theory and a suitable name-forming operator.

[11] A language at least capable of constructing the ordinals, supporting set-theory and describing arithmetic operations.

We have a choice as to which of the three theses we reject. Each of the three theses are plausible and intuitive and we should do as little violence as possible to our native intuitions. It is therefore prudent to reject but one of the three theses if rejecting one resolves salient inconsistency. I will now turn to a variety of options available to that end.

1.2 An Available Route

On denying **language**, Tarskian solutions restrict truth-talk for sentences in a language ***L*** to a meta-language ***L'***. Tarskian solutions thereby rule out the liar paradox because there is no liar sentence. No sentence can predicate truth or untruth of itself because no language can define its own truth predicate. While attractive, such a solution comes at a significant cost. As Alfred Tarski observed, natural languages appear to be *semantically closed*.[12] If languages are indeed semantically closed, then Tarski's solution cannot be applied to natural language; a consequence that is deeply unsatisfying for the truth theorist who seeks to produce a theory of truth for a language like English. Granted that the Tarskian route is the most plausible route to denying **language**, it appears that to deny **language** is to preclude natural language from a satisfactory definition of truth. Indeed, most present theories of truth take **language** as a starting point.

Many logicians have taken alethic paradox as a sign that classical logic needs to be revised in order to accommodate the salient phenomena. Such a view requires the denial of **classical**. Non-classical solutions attempt to dissolve alethic paradox by rejecting the law of the excluded middle[13], the law of noncontradiction[14], taking on a third truth-value[15], or by embracing a paraconsistent logic[16].

There are four reasons why I aim to preserve **classical** in this paper. The first reason is that while there are numerous touted non-classical solutions, there are few viable classical contenders on the market. Providing a classical solution is therefore interesting in that it provides a novel route by which to deal with alethic paradox.

The second reason why I aim to preserve **classical** is that a classical solution is desirable. We desire theories that cohere with our best empirical and mathematical theories, most of which are classically constrained. It is true that non-classical and non-monotonic logics have been employed to deal with empirical phenomena in quantum mechanics and everyday

---

[12] A language ***L*** is semantically closed *just in case* ***L*** can talk about its own semantics and can apply semantic properties to its own constituent expressions. See section eight in Tarski 1944.
[13] Field (unpublished manuscript) is a paradigm example.
[14] See Priest 2006 part one.
[15] See Priest 1979.
[16] A logic in which the *principle of explosion* ("from a contradiction, anything follows") is not valid which requires abandoning one of (1), (2), (3) and one of (4), (5), (6).
(1) *Disjunction introduction*: $S \vdash S \vee Q$.
(2) *Disjunctive syllogism*: $S \vee Q, \neg S \vdash Q$.
(3) *Transitivity*: $\Gamma \vdash S; S \vdash Q \Rightarrow \Gamma \vdash Q$.
(4) *Reductio ad absurdum*: $S \rightarrow (Q \wedge \neg Q) \vdash \neg S$.
(5) *Rule of weakening*: $S \vdash Q \rightarrow S$.
(6) *Double negation elimination*: $\neg\neg S \vdash S$.

reasoning. However, the bulk of our mathematical and scientific theories remain formulated using classical logic.

The third reason is that normally when we are faced with a disconfirming case we reject or revise the theory (in this case $\boldsymbol{FOL}_T$) and not the logic itself. This applies equally well within the empirical sciences as it does within the mathematical. In the empirical sciences the falsification of a theory results in the development of a new theory often formulated using the same logic as its predecessor. The set-theoretic contradictions revealed in the first two decades of the 20th century were taken as signs that set-theory needed to be reformulated in a more precise and rigorous manner. Classical logic was retained in the development of a successor to naive set theory.

The fourth reason that I aim to preserve **classical** is that revising our alethic inferences is less damaging than revising our logical validities or laws. Per Quine's Principle of Minimal Mutilation, we should do as little harm as we can to our web of beliefs when confronted with belief revision.[17] Classical logic is central to philosophy, mathematics, and a good deal of the sciences. Revising classical logic thereby does greater violence to our overall web of belief than to say revise our theory of truth.

The view that classical logic ought to be retained whenever we may restrict our alethic inferences has been criticized as a dogma.[18] It is prudent to note that it is not the aim of this paper to insist that classical logic ought to be privileged over its non-classical brethren. Rather, the reasons motivating this approach are merely methodological and say nothing about the status of classical logic or its non-classical brethren. The proposal presented in this paper therefore trades **Scheme** securing **classical** and **language**.

There is a precedent of rejecting **Scheme**. For example, Stephan Read rejects **T-Scheme** and thereby **Scheme** from the claim that **T-Scheme** is incapable of supporting sentences whose truth-conditions can only be ascertained in a particular conversational context.[19] Consider the following natural language instance of the **T-Scheme**:

(**I**) 'I am here' is true *just in case* I am here.

Read claims that the biconditional immediately above fails as there is nothing to ensure that the named-sentence on the left pairs with the sentence on the right - there's ambiguity that **T-Scheme** fails to account for. For example, the statement on the left-hand could be uttered by person A while the indexical 'I' on the right-hand side could refer to another person, person B. But, in both cases, 'here' (another indexical) could be referring to the same location. Hence, **I** fails. I think that sentences containing indexicals can easily be reconciled with **T-Scheme** by requiring the logic undergirding our theory of truth to be one of unambiguous (including all satisfactorily disambiguated) sentences as per section 1.0 - in other words, all such situations in which the immediately above scenario would arise would thereby be excluded from alethic considerations anyway. Why is this permissible? Because

[17] See Quine 1990 p. 14.
[18] See Field 2008.
[19] See Read 2008.

truth-aptness, it seems plausible to hold, requires disambiguation at least insofar as we are constraining truth-aptitude by a two-value logic. Consider the following expression:

[*] Sarah beat Sally.

When considering the truth-aptness of this statement, we must first determine the relevant disambiguation of the verb 'beat'. Suppose in this case Sarah won a game against Sally but has never physically struck Sally. Upon disambiguating the relevant syntactic item 'beat', we are thus able to determine the truth or falsity of the sentences in question. This is not because the pre-disambiguated sentences carry a truth value that we proceed to discover. It is due to the fact that ambiguous sentences do not carry a truth value at all. At least not if an ambiguous sentence *S* does not bear the logical form or whose logical form does not entail $Q_1 \vee \ldots \vee Q_n$ where each member of $\{Q_1, \ldots, Q_n\}$ is understood to be a relevant disambiguation of *S*. Why? Because ambiguous sentences must either bear a logical form or have a logical form that entails the disjunction of its relevant disambiguations. Any else seems insensible. If an ambiguous sentence does not bear a logical form or have a logical form that entails the disjunction of its relevant disambiguations it's hard to see how we could reason from ambiguous sentences to their relevant disambiguations.

But if ambiguous sentences bear a logical form or have a logical form that entails the disjunction of its relevant disambiguations, then such sentences are not ambiguous at all (but for in name) since they give rise to clearly explicated well-formed expressions. In the first case, an "ambiguous" sentence is just a disjunction – though we do not know which disjunct is intended in utterance – which is not ambiguous syntactically or semantically (but rather epistemically and pragmatically). In the second case, if the sentence entails a disjunction, the logical form must itself be explicable in a format that is conducive to preserving truth-aptitude across inference. But no two-valued logical system can manage truth-aptitude without it giving rise to a well-formed and explicit expression in the first place. Here, I take the epistemicist position and maintain that sentences such as:

[**] The lampshade is not orange.

to have a definite truth-value and a definite meaning though the application conditions for the term 'orange' remain opaque in borderline cases. Thus, though I disagree with Read, I too will take to rejecting **Scheme** in favor of a restricted **T-Scheme**.

2.0 Restricting the T-Scheme

There are a number of restriction-based strategies that purport to prevent the threat of alethic paradox however, the literature on restriction-based solutions to alethic paradox is vast and I cannot survey them in detail here.

Nevertheless, it is prudent to note that restriction-based solutions to alethic paradox tend to take one of three forms: "restrict" the **T-Scheme** by pronouncing liar-like sentences syntactic outlaws by invoking a ramified theory of types or an ascending hierarchy of languages, take liar-like sentences to fail a criterion on truth-aptitude or meaning,[20] or by suitably revising **T-Scheme** in the manner explored by Saul Kripke[21] and Solomon Feferman.[22]

I take it that the essential feature of the Kripke-Feferman strategy is found in restricting the **T-Scheme** by a condition C requisite on each sentence S for which the truth-predicate is transparent.[23] More formally:

[Kripke-Feferman] $C(S) \rightarrow (Tf(S) \leftrightarrow S)$

Kripke-Feferman guarantees the transparent linguistic profile of the truth-predicate for all sentences that meet condition C. Condition C is to be interpreted as a point demarcating sentences that engender alethic paradox from those that do not[24]. Importantly, **Kripke-Feferman** permits the violation of **T-Scheme** without requiring the failure of bivalence. In the following section, $C(S)$ will be interpreted as the truth grounded predicate * predicated of $f(S)$ accompanied by supporting axioms and a semantics characterizing the linguistic profile of the truth grounded predicate *.

2.1 Truth Grounded Sentences

Most solutions to alethic paradox diverge in how they attempt to dissolve alethic paradox while nevertheless converging in their diagnosis of the root, source or cause of alethic paradox. Beall has suggested that the transparency of the truth-predicate holds the key to a deflationary resolution of alethic paradox.[25] On the deflationary account relevant to our present discussion, the truth-predicate is eliminable[26] from any sentence over which the **T-Scheme** holds and sentences of the form $\ulcorner Tf(S) \urcorner$ are always intersubstitutable in non-opaque contexts *salva veritate* with some sentence S in which the truth-predicate does not appear. Liar-like sentences are defective under such a deflationary account.[27] We observe that liar-like sentences resist truth-predicate elimination. Further, there are no sentences within which the truth-predicate does not appear that are intersubstitutable in non-opaque contexts *salva veritate* with any given liar-like sentence.

Similarly, Kripke motivates his famed fixed-point construction in part from the observation that liar-like sentences and sentences like $S_1 =_{df} Tf(S_1)$ fail to be grounded. Sentences like S_1 and λ are not grounded in a statement in which the truth-predicate does not appear.[28] The

---

[20] Requiring the failure of bivalence, the invocation of truth-gaps, or a commitment to propositions as the bearers of truth.
[21] See Kripke 1975.
[22] See Feferman 2008.
[23] To be precise Feferman treats 'C' as a meta-linguistic predicate. Feferman defines $Cf(S)$ as follows: $Cf(S) =_{df} Tf(S) \vee Ff(S)$. I will, however, treat '$C(S)$' as '$\neg * f(S)$'. See Feferman 2008.
[24] Non liar-like sentences may be precluded as well.
[25] See Beall 2001.
[26] See **2** below.
[27] Note that this does not wed me to a deflationist account of truth. Beall's suggestion is merely a conceptual starting point from which I construct a truth grounded predicate that is indifferent to the truth realism debate.
[28] See Kripke 1975.

remainders of both Kripke's *fixed-point* and *closed-off* constructions are set aside here for the aim of developing a classical theory of truth that conforms more closely to the framework of standard two-valued semantics.

While the solutions proposed by Beall and Kripke differ significantly in how they handle alethic paradox, they do seem to converge on the identification of a feature shared by liar-like sentences. First, both proposals identify semantic defects common to liar-like sentences. Second, both proposals indicate that liar-like sentences do not exhibit the semantic conditions requisite for substitution into **T-Scheme.** Third, both proposals suggest that the defects common to liar-like sentences are closely related to the semantic conditions under which sentences are truth-evaluable. I will take the proposals made by Beall and Kripke as my inspiration and will initially operate with the rough notion that:

[1] *S* is *truth grounded only if* the <u>truth-predicate</u> is eliminable from *S*.
[2] *S* is *truth grounded only if* there exists a sentence *Q* within which the <u>truth-predicate</u> does not appear having the same truth-conditions as *S*.

I note that my initial gesture at truth grounding is probably unsatisfactory as it precludes many tautologies from being counted as truth grounded. There are, after all, many tautologies for which the truth-predicate is not eliminable. Thankfully, I will explicitly define truth grounding in a mathematically rigorous way that avoids the pitfalls of my initial impression and which is neutral with respect to the truth realism debate. The conditions under which a sentence succeeds in meeting the truth grounded condition are precisely captured by way of the following procedure.

First, I write '$\Phi | S$' to denote that a sentence *S* contains Φ as a constituent expression. '*directly refers*' and '*indirectly refers*' are reserved as technical terms to be explicitly defined below. The relationship between 'directly refers' and 'indirectly refers' is given as follows:

[3] I write '*S* directly refers to *Q*' *when and only when* $f(Q) | S$ *or* whenever a constituent expression of *S* quantifies over *Q*.
[4] Indirect reference is recursively defined as follows:
- **[a]** If *S* directly refers to *Q* and *Q* directly refers to *R*, then *S* indirectly refers to R.
- **[b]** If *S* indirectly refers to *Q* and *Q* directly refers to *R*, then *S* indirectly refers to *R*.
- **[c]** If *S* indirectly refers to *Q* and *Q* indirectly refers to *R*, then *S* indirectly refers to *R*.
- **[d]** If *S* directly refers to *Q* and *Q* indirectly refers to *R*, then *S* indirectly refers to *R*.
- **[e]** There are no other instances of indirect reference.

3 captures the intuitive notion that a sentence *S* directly refers to a sentence *Q just in case S* contains $f(Q)$ as a constituent expression. **a-e** indicate the conditions under which a sentence *S* indirectly refers to a sentence *R*. For example, condition **a** states that *S* indirectly refers to a sentence *R* whenever sentence *S* contains a constituent expression $f(Q)$ and *Q* contains a constituent expression $f(R)$. Consider the following sequence of sentences:

(*S*) $Tf(Q)$
(*Q*) $Tf(R)$

(R) $P(\mathrm{x})$[29]

Sentence S directly refers to sentence Q and indirectly refers to sentence R. Sentence Q directly refers to sentence R and sentence R does not refer to any sentence. Note that liar-like sentences both directly and indirectly refer to themselves.

I have already alluded to the truth grounded predicate * and will now show how the extension of the predicate is to be constructed. The extension of the truth grounded predicate *, ext(*), is recursively constructed as follows:

[5] $f(\varphi) \in \mathrm{ext}(*)$ *whenever*:
- **[a]** φ is well-formed *and*
- **[b]** φ does not directly refer to a sentence containing a truth-predicate.
- **[c]** φ does not directly refer to a sentence containing a false-predicate.
- **[d]** φ does not indirectly refer to a sentence containing a truth-predicate.
- **[e]** φ does not indirectly refer to a sentence containing a false-predicate.

[6] *If* φ is a tautology, *then* $f(\varphi) \in \mathrm{ext}(*)$.
[7] *If* $f(\varphi) \in \mathrm{ext}(*)$, *then* $f(Tf(\varphi)) \in \mathrm{ext}(*)$.
[8] *If* $f(\varphi) \in \mathrm{ext}(*)$, *then* $f(Ff(\varphi)) \in \mathrm{ext}(*)$.
[9] *If* $f(\varphi) \in \mathrm{ext}(*)$, *then* $f(\neg\varphi) \in \mathrm{ext}(*)$.
[10] *If* $f(\varphi), f(\sigma) \in \mathrm{ext}(*)$, *then* $f(\varphi \wedge \sigma) \in \mathrm{ext}(*)$.
[11] *If* $f(\varphi), f(\sigma) \in \mathrm{ext}(*)$, *then* $f(\varphi \vee \sigma) \in \mathrm{ext}(*)$.[30]
[12] *If* $f(\varphi), f(\sigma) \in \mathrm{ext}(*)$, *then* $f(\varphi \rightarrow \sigma) \in \mathrm{ext}(*)$.[31]
[13] There are no other elements of ext(*).

Conditions **5**-**13** can be divided into two parts. The first part, conditions **5** and **6**, specify a privileged set of sentences that serve to ground sentences that contain an alethic predicate. The second part, conditions **7**-**13**, specify which of the sentences that contain an alethic predicate are truth grounded.

Each sentence satisfying condition **5** likewise satisfies conditions **1** and **2**. By stipulation we get tautologies for free by condition **6**. Sentences meeting conditions **7** through **9** predicate truth or falsity of satisfactory sentences meeting the aforementioned conditions guaranteeing that all sentences predicating truth or falsity of a sentence are ultimately truth grounded.

It turns out that λ is not truth-grounded - λ directly refers to itself, a sentence containing an alethic predicate. Truth-teller sentences, like S_1, Yablo's infinite sequences likewise fail to be truth-grounded, and so are the rest of the sentences that engender alethic paradox. The present proposal connects the semantic oddities identifying liar, truth-teller, and Yablo sentences with the conditions under which sentences are truth-grounded. It is therefore

---

[29] Let P and x be arbitrary constants.
[30] I include this here for ease. It is implicitly entailed by **9** and **10**. Any logical connective not negation is can be defined by any other logical connective and negation.
[31] I include this here for ease. It is implicitly entailed by **9** and **10**. Any logical connective not negation is can be defined by any other logical connective and negation.

prudent to note that the present solution unifies a solution to several related but distinct linguistic phenomena each problematic in their own right.

One objection to the proposal offered here is that normally when we think of truth-apt sentences we do not have to go through the seemingly elaborate process described in **1**-**12**.[32] The rebuttal to this is straightforward. Most people, when confronted with liar-like phenomena, do in fact take liar-like phenomena to be an outlier - a peculiar artifact of language - though they may not be able to express why. I believe that liar-like phenomena are implicitly calculated and distinguished by the aforementioned method. The brain is a startling complex device capable of vastly complex calculations in real-time, most completely unbeknownst to individual thinkers. Meta-cognition is a mechanism by which people are able to penetrate such processes, but most such processes remain completely opaque despite a lifetime of reflection.

By interpreting C of **Kripke-Feferman** as the truth grounded predicate * sentences like S_1 and λ are allowed to violate the unqualified **T-Scheme** while preserving classical logic. On the view proposed here, liar-like sentences are grammatically well-formed but fail a criterion on **T-Scheme**. We may of course, utilize liar-like sentences in our proof-theoretic reasoning. The proposal offered below accommodates the observation that we may utilize liar-like sentences in our proof-theoretic reasoning by retaining the truth-aptness of liar-like sentences while sequestering liar-like sentences to the periphery of truth-apt language.

2.2 Axioms and Model Theory

I extend a sound and complete first-order axiom system ***FOL*** with the axioms of ***PA*** along with the following axioms yielding consistent axiom system $\boldsymbol{FOL_{T^*}}$ supported by language $\boldsymbol{L_{T^*}}$:

[T1] $(^*f(S) \wedge Tf(S)) \rightarrow S$
[T2] $(^*f(S) \wedge S) \rightarrow Tf(S)$
[F1] $(^*f(S) \wedge Tf(S)) \rightarrow \neg Ff(S)$
[F2] $(^*f(S) \wedge \neg Ff(S)) \rightarrow Tf(S)$
[TT1] $(^*f(S) \wedge Tf(\neg S)) \rightarrow \neg Tf(S)$
[TT2] $(^*f(S) \wedge \neg Tf(S)) \rightarrow Tf(\neg S)$

Here, our unqualified alethic inferences are replaced by controlled alethic inferences **T1**-**TT2**. For example, instead of the left-hand implication of **T-Scheme**, $Tf(S) \rightarrow S$, I have the controlled inference **T1**. Note that because most sentences are truth grounded, **T1** behaves very much like $Tf(S) \rightarrow S$.[33] The same is true of each of **T2**-**TT2**. By adding a requisite condition into the antecedents of our axioms, liar-like sentences are precluding from entering into our alethic inferences thereby preventing alethic paradox. Axioms **T1-TT2** prove:

---

[32] I thank Matt Babb for this objection.
[33] Perhaps explaining, in part, why the truth grounded proposal has thus far eluded philosophers and logicians.

[*TS] $*f(S) \rightarrow (Tf(S) \leftrightarrow S)$[34] **Proof:** obvious[35] by [T1] and [T2] ■
[*TF1] $*f(S) \rightarrow (Tf(S) \leftrightarrow \neg Ff(S))$ **Proof:** obvious by [F1] and [F2] ■
[TF] $*f(S) \rightarrow (Tf(\neg S) \leftrightarrow Ff(S))$ **Proof:** obvious by [TT1] and [TT2] ■
[TT*] $*f(S) \rightarrow (Tf(\neg S) \leftrightarrow \neg Tf(S))$ **Proof:** obvious by [*TF1] and [TF] ■
[*FS] $*f(S) \rightarrow (Ff(S) \leftrightarrow \neg S)$ **Proof:** obvious by [*TS] and [*TF1] ■
[*TF2] $*f(S) \rightarrow (Tf(S) \vee Ff(S))$ **Proof:** obvious by [*TF1] and LEM ■

Here, each of our alethic inferences are restricted to sentences that are truth grounded.

Volker Halbach notes that almost all authors agree that banning sentences containing a truth-predicate from entering into our alethic inferences is simply too restrictive. [36] One advantage of the solution proposed in this paper over other restriction based strategies is that sentences containing a truth-predicate are allowed to enter into the above axioms, so long as they are truth grounded.

I will now show the consistency of the $\boldsymbol{FOL}_{T*}$ by finding an interpretation that satisfies all the axioms composing $\boldsymbol{FOL}_{T*}$. Let **M** be an $\boldsymbol{L}_{T*}$ structure supporting the axioms of ***PA***.[37] Let E_M be a standard first-order truth-evaluation under **M**.[38] **M** is model of $\boldsymbol{FOL}_{T*}$ whenever **M** is constrained by the following conditions **14**-**29**. Conditions **14**-**17** constrain interpretation I under **M**:

[14] ($f(S) \in$ ext(*) *and* $f(S) \in$ ext(T)) *only if* $f(S) \notin$ ext(F).
[15] ($f(S) \in$ ext(*) *and* $f(S) \in$ ext(F)) *only if* $f(S) \notin$ ext(T).
[16] ($f(S) \in$ ext(*) *and* $f(S) \in$ ext(T)) *only if* $f(\neg S) \notin$ ext(T).
[17] ($f(S) \in$ ext(*) *and* $f(S) \in$ ext(F)) *only if* $f(\neg S) \notin$ ext(F).

Constraints **18** and **21** limit the construction of ext(T) and ext(F) so that they conform closely to the usual way in which they are defined[39]:

[18] ($f(S) \in$ ext(*) *and* $E_M(S) = 1$) *only if* $f(S) \in$ ext(T).
[19] ($f(S) \in$ ext(*) *and* $f(S) \in$ ext(T)) *only if* $E_M(S) = 1$.
[20] ($f(S) \in$ ext(*) *and* $E_M(S) = 0$) *only if* $f(S) \in$ ext(F).

---

[34] Doesn't restricting **T-Scheme** run counter to the neglected deflationary solution in part motivating section 2.1? Not if deflationism is understood as the thesis that 'the truth-predicate does not refer to a substantive property' which I take to be the core deflationary commitment. Note that followers of Horwich are unlikely to be satisfied with the truth grounded solution given that it calls for a restriction of the **T-Scheme**. See Horwich 1998

[35] $(((P \wedge Q) \rightarrow R) \wedge ((P \wedge R) \rightarrow Q)) \rightarrow (P \rightarrow (Q \leftrightarrow R))$ is a tautology.

[36] See Halbach 2006 p. 276.

[37] Here, for ease, a structure M is treated as a triple: $\langle D, I, E_M \rangle$ where D is a domain, E_M a truth-evaluation function sending sentences to $\{0,1\}$, and I a standard interpretation function sending each element of the signature to an element of D (constants), a subset of D^n (n-ary relations), or a mapping from $D^n \rightarrow D$ (n-ary functions).

[38] i.e., the standard semantics for the logical connectives, predicates, relations, and quantifiers are preserved.

[39] Usually ext(F) is constructed from the set of sentence-names that name sentences which have been truth-evaluated 0. Ext(T) is usually constructed from the set of sentence-names that name sentences which have been assigned a truth-value of 1.

[21] ($f(S) \in \text{ext}(*)$ *and* $f(S) \in \text{ext}(F)$)*only if* $E_M(S) = 0$.

Conditions **22**-**29** impose constraints on E_M.

[22] $E_M(S) = 1$ *if and only if* $E_M(S) \neq 0$.
[23] $E_M(S) = 1$ *if and only if* $E_M(\neg S) \neq 1$.
[24] ($E_M(Tf(S)) = 1$ *and* $E_M(*f(S)) = 1$) *only if* $E_M(S) =1$.
[25] ($E_M(S) = 1$ *and* $E_M(*f(S)) = 1$) *only if* $E_M(Tf(S)) = 1$.
[26] ($E_M(*f(S)) = 1$ *and* $E_M(Tf(S)) = 1$) *only if* $E_M(Ff(S)) = 0$.
[27] ($E_M(*f(S))= 1$ *and* $E_M(Ff(S)) = 0$) *only if* $E_M(Tf(S)) = 1$.
[28] ($E_M(*f(S)) = 1$ *and* $E_M(Tf(\neg S)) = 1$) *only if* $E_M(Tf(S)) = 0$.
[29] ($E_M(*f(S))= 1$ *and* $E_M(Tf(S)) = 0$) *only if* $E_M(Tf(\neg S)) = 1$.

To show how conditions **14**-**29** work, I'll use two sentences $S =_{df} P(x)$ and $Q =_{df} R(x)$. To begin, I will assume an interpretation in which x is assigned to both the extension of P and the extension of R. So, $E_M(S) = 1 = E_M(Q)$. Next, I verify that S and Q are truth grounded using conditions **5**-**13**. They are, so $f(S), f(Q) \in \text{ext}(*)$. One ascertains that $E_M(*f(S)) = 1 = E_M(*f(Q))$. So, by condition **25**, $E_M(Tf(S)) = 1 = E_M(Tf(Q))$. Hence, $f(\neg S), f(\neg Q) \in \text{ext}(*)$ from condition **9**. By conditions **22** and **23** one acquires $E_M(\neg S) = 0 = E_M(\neg Q)$. So, by condition **20**, $f(\neg S), f(\neg Q) \in \text{ext}(F)$. So, $Ff(\neg S) = 1 = Ff(\neg Q)$. Conditions **14**-**29** have thus far produced the following string of true sentences:

[M1] $\{S, Q, *f(S), *f(Q), Tf(S), Tf(Q), *f(\neg S), *f(\neg Q), \mathrm{F}f(\neg S), \mathrm{F}f(\neg Q)\}$

Continuing, $Ff(S) = 0 = Ff(Q)$ by condition **26**,. So, $E_M(\neg Ff(S)) = 1 = E_M(\neg Ff(Q))$ by conditions **22** and **23**. One judges that $E_M(\neg S) = 0 = E_M(\neg Q)$ by **22** and **23**. So, $f(\neg S)$, $f(\neg Q) \in \text{ext}(F)$ from condition **20**. By condition **15**, $f(\neg S), f(\neg Q) \notin \text{ext}(T)$. Hence, $E_M(Tf(\neg S)) = 0 = E_M(Tf(\neg Q))$. Thus, $E_M(\neg Tf(\neg S)) = 1 = E_M(\neg Tf(\neg Q))$ by conditions **22** and **23**. Conditions **14**-**29** have produced the following string of true sentences:

[M2] **M1** $\cup \{\neg \mathrm{F}f(S), \neg \mathrm{F}f(Q), \neg Tf(\neg S), \neg Tf(\neg Q)\}$

Of course one could continue this process indefinitely but I think the above should suffice to show that conditions **14**-**29** constrain first-order structures in a way that produces consistent interpretations of $\boldsymbol{FOL}_{T*}$.

The consistency proof goes as follows. First, a logic is consistent with respect to a particular transformation by which each sentence or propositional form P is transformed into a sentence of form $P*$, if there is no sentence form P such that $\vdash P \wedge P*$.

First, $\boldsymbol{FOL}_{T*}$ is an extension of ***FOL***. A logic $L*$ is an extension of a logic L *just in case* sig(L) $\subset$ sig($L*$) and the theorems provable in L are a subset of the theorems in $L*$.[40] Second, one need only check the consistency of base (non-iterative, non-diagonal) expressions containing an alethic predicate or *, subsequent expressions that are iterations or concatenations of those aforementioned base expressions, and then diagonal alethic (or *) expressions in order to check the consistency of $\boldsymbol{FOL}_{T*}$. At each stage, T and * show consistent truth-values.

Consider that iterated, non-diagonal, or concatenated expressions can be given a consistent interpretation if their base constituent expressions can. Now, consider the following base (non-iterative, non-diagonal) expressions and their consistent interpretations:

$Tf(P)$	1	0
$Ff(P)$	0	1
$*f(P)$	1	1

It follows that (1) all iterations or concatenations composed of these base expressions can likewise be given consistent interpretations and (2) because there is nothing particularly unique about these base expressions (when compared to similar expression like $Tf(Q)$, $\neg*f(S)$, or any other well-formed alethic expression), all other similar base expressions can likewise be given consistent interpretations and hence, so too can their iterations and concatenations.

All iterations or concatenations of diagonal sentences can likewise be assigned consistent interpretations if their base diagonal sentences can. Consider the following diagonal base expression and its consistent interpretations:

$S =_{df} \neg Tf(S)$

$\neg Tf(S)$	0	1	1	0
S	0	1	0	1
$*f(S)$	0	0	0	0

It follows that (1) all iterations or concatenations of this base diagonal expressions can likewise be given consistent interpretations and (2) because there is nothing particularly unique about these base diagonal expressions, all other similar base expressions can likewise be given consistent interpretations and hence, so too can their iterations and concatenations. Thus, $\boldsymbol{FOL}_{T*}$ is consistent. ■[41]

2.3 Seeking Revenge

[40] As per section 1.0, where $\boldsymbol{FOL}_{T*}$ is constructed as an extension of ***FOL***. Hence, all theorems provable in ***FOL*** are provable in $\boldsymbol{FOL}_{T*}$ and the only expressions that appear in the latter and not the former are those that contain the alethic predicates and/or *.

[41] Gödel's Second Incompleteness Theorem, which states that no consistent system F capable of expressing arithmetic can prove its own consistency, becomes salient whenever a 'consistency' predicate is introduced into the language along with its governing inferences. No such predicate has yet been introduced. Hence, $\boldsymbol{FOL}_{T*}$ is provably consistent.

A *revenge* paradox is a semantic paradox formulated for a purported solution to alethic paradox. A standard way to induce *revenge* (call this *standard revenge*) is to introduce into the relevant language a predicate 'ς' where 'ς' is the condition that is used to dissolve alethic paradox such that for each liar sentence S: $\neg\varsigma(S)$. A *revenge* sentence is then constructed to the end of showing an inconsistency in the solution making use of condition ς. Revenge sentences often take the form $\ulcorner S =_{\mathrm{df}} \neg\varsigma(S) \vee Ff(S)\urcorner$ or $\ulcorner S =_{\mathrm{df}} \neg\varsigma(S\urcorner$. Beall correctly observes that *revenge* is induced whenever $\varsigma(S)$ entails $Ff(S)$.[42] Here, '$\neg\varsigma(S)$' is '$\neg{*f}(S)$'. We note, however, that neither ${*f}(S)$ nor $\neg{*f}(S)$ entail $\neg Tf(S)$ or $Ff(S)$.

Revenge paradox has a terrible way of popping up despite the best efforts of theorists to suppress it. Fortunately, $\boldsymbol{FOL}_{T*}$ resists the standard way of inducing *revenge*. It is easy to see why. First, the sentences $S_2 =_{\mathrm{df}} \neg{*f}(S_2)$ and $S_3 =_{\mathrm{df}} {*f}(S_3)$ have consistent interpretations.[43] Second, let us construct a second standard *revenge* sentence $S_4 =_{\mathrm{df}} \neg{*f}(S_4) \vee Ff(S_4)$. S_4, however, has consistent interpretations under $\boldsymbol{FOL}_{T*}$ - S_4 may be evaluated true. If there is a lurking inconsistency within $\boldsymbol{FOL}_{T*}$ it does not appear to be found in the standard way of inducing *revenge*.

2.4 Remarks on $\boldsymbol{FOL}_{T*}$

While $\boldsymbol{FOL}_{T*}$ preserves the transparency of the truth-predicate for *truth-grounded* sentences and preserves classical logic it does so at a cost. $\boldsymbol{FOL}_{T*}$ absorbs the semantic peculiarity of alethic paradox by conditionalizing **T-Scheme**. On some consistent interpretations λ is evaluated true and so is $Ff(\lambda)$. The logical truth of **T-Scheme** is traded for the preservation of classical logic.

There is another cost associated with the proposal that is worth mentioning. The following sets are consistent interpretations of $\boldsymbol{FOL}_{\mathbf{T}*}$ - they are consistent with constraints **14**-**29** imposed in section 2.2: $\{\lambda, \neg Tf(\lambda), \neg \mathrm{F}f(\lambda), ...\}$, $\{\lambda, \neg Tf(\lambda), \mathrm{F}f(\lambda), ...\}$, $\{\neg\lambda, Tf(\lambda), Ff(\lambda), ...\}$, and $\{\neg\lambda, Tf(\lambda), \neg Ff(\lambda), ...\}$. The first set assigns the liar sentence a truth-value of 1 but requires us to say of the liar that it is neither true nor false.[44] The second set assigns the liar sentence a truth-value of 1 but requires us to say of the liar that it is not true and false. The third set says that the liar is both true and false while assigning the liar sentence a truth-value of 0. The fourth set says of the liar that it true and not false while assigning the liar sentence a truth-value of 0. This feature is a well-known consequence of any classical solution utilizing **Kripke-Feferman**.

On the one hand, the solution can accommodate the four divergent intuitions that we may say of the liar that it "is neither true nor false", "is false", "is true and false", and "is true" thereby both explaining the various positions held by those in the liar debate and providing an ecumenical solution to alethic paradox. On the other hand, there is something distinctly

---

[42] See Beall 2007.

[43] Observe that S_2 is truth grounded. So, S_2 is false (but this does not lead to paradox - it is simply false). S_3 is true - S_3 is truth grounded.

[44] Note that this is neither the same as the conjunction $Tf(\lambda) \wedge \neg Tf(\lambda)$ nor $Ff(\lambda) \wedge \neg Ff(\lambda)$.

peculiar about the four interpretations.[45] For example, we cannot say of the liar that it is true when it is evaluated 1 and we can say of the liar that it is true when it is evaluated 0. On the matter, I think that the problem lies with liar-like sentences themselves. Liar-like sentences are outliers on the way we secure talk about truth. Observe that truth-grounded sentences, which comprise the "overwhelming majority" of sentences in $\boldsymbol{L}_{\Gamma^*}$, do not exhibit the phenomenon in question. It is the point of the truth-grounded solution to relegate liar-like sentences to the periphery of language. While our theory of truth holds for all sentences, truth-inferences are restricted to truth-grounded sentences of language $\boldsymbol{L}_{\Gamma^*}$

There is a further observation on the matter; just about any classical solution respecting **language** will exhibit some of the features present in the truth-grounded proposal. Standard two-valued semantics requires that whenever the liar sentence is evaluated 1 that it is impossible to say so and that whenever the liar is evaluated 0 that we must say it is true owing to the manner in which the liar sentence is itself constructed. We are thus led to an interesting trilemma: we can accept the quirkiness of classically constrained liar-like phenomena, we can reject T-Scheme and what is usually taken to be our conception of truth (and given that **Kripke-Feferman** is likely the next-best formulation of that conception[46], we still end up with a quirky classical solution), or we accept that classical logic is not the appropriate logic for truth-talk. This trilemma ultimately collapses into two distinct choices, to opt for classical logic or to deny classical logic. I note that the paper presupposes classical logic. The choice of which of these two options is left up to the reader though this author recommends the latter and takes the results of this paper to be a negative result.[47]

Works Cited

[1] Beall, JC. (2001), 'A neglected deflationist approach to the liar', in *Analysis* 61, pp. 126-129.

[2] Beall, JC. (2007), 'Prolegomenon to Future Revenge', in *Revenge of the Liar*, JC Beall (ed), Oxford: Oxford University Press.

[3] Feferman, S. (2008), 'Axioms for determinateness of truth', in *Review of Symbolic Logic* 1, pp. 204-217

[4] Field, H. (2008), *Saving Truth from Paradox*, Oxford: Oxford University Press.

[5] Field, H. (2006), 'Solving the paradoxes, escaping revenge' (unpublished manuscript). Available at ⟨http://as.nyu.edu/docs/IO/1158/revengetex3.pdf⟩

[45] It is not my project to determine which interpretation is *the* correct interpretation owing to the fact that I think that there are severe epistemic constraints on determining the truth-value of the liar sentence. For starters, it does not seem that we can discover the truth-value of the liar sentence empirically. Add to that what seems to be the consensus view that the liar sentence is neither a necessary truth nor a necessary falsehood.

[46] By definition, as it excludes exactly liar-like sentences and phenomena.

[47] A conclusion that most logicians seem to have accepted already but that few philosophers have paid any attention to. Noting here, that the results appear to be radically underdetermined.

[6] Halbach, V. (2006), 'How Not to State T-Sentences', in *Analysis* 66(268), pp. 276-280.

[7] Horwich, P. (1998), *Truth*, New York: Oxford University Press.

[8] Kripke, S. (1975), 'An Outline of a Theory of Truth', in *The Journal of Philosophy* Vol. 72(19), pp. 690-717.

[9] Priest, G. (1979), 'The Logic of Paradox', in the *Journal of Philosophical Logic* 8, pp. 219-241.

[10] Priest, G. (2006), *Doubt Truth to Be a Liar*, Oxford: Oxford University Press.

[11] Quine, W.V.O. (1990), *The Pursuit of Truth*, Cambridge, MA: Harvard University Press.

[12] Read, S. (2008), 'Further Thoughts on Tarski's T-scheme and the Liar', in *Logic, Epistemology, and the Unity of Science* Vol. 8(1), pp. 205-255.

[13] Tarski, A. (1944), 'The semantic conception of truth', in *Philosophy and Phenomenological Research* 4, pp. 13-47.

[14] Yablo, S. (1993), 'Paradox without Self-Reference', in *Analysis* Vol. 53(4), pp. 251-252.

Appendix

I. Boolean Compounds

A Boolean compound takes the form $\ulcorner S =_{\mathrm{df}} \neg Tf(S) \vee \bot \urcorner$ where $\bot$ is a necessary falsehood. Boolean compounds give rise to alethic paradox.

> **Proof.** Construct $S =_{\mathrm{df}} \neg Tf(S) \vee \bot$. This yields logical equivalence $S \leftrightarrow \neg Tf(S) \vee \bot$. By **T-Scheme** we acquire $Tf(S) \leftrightarrow S$. By the transitivity of the biconditionals $Tf(S) \leftrightarrow \neg Tf(S) \vee \bot$. By logical equivalence we derive $\neg Tf(S) \leftrightarrow (Tf(S) \wedge \neg\bot)$. $Tf(S) \leftrightarrow \bot$ and $\neg Tf(S) \rightarrow Tf(S)$ follow. ■

II. Curry Sentences

An alethic Curry sentence takes the form $\ulcorner S =_{\mathrm{df}} Tf(S) \rightarrow \bot \urcorner$ where $\bot$ is a necessary falsehood. Alethic Curry sentences are intimately related to Boolean compounds (they are logically equivalent) and like them give rise to alethic paradox.

> **Proof.** Construct $S =_{\mathrm{df}} Tf(S) \rightarrow \bot$. This yields logical equivalence $S \leftrightarrow (Tf(S) \rightarrow \bot)$. By **T-Scheme** we acquire $Tf(S) \leftrightarrow S$. By the transitivity of the biconditionals we derive $Tf(S) \leftrightarrow (Tf(S) \rightarrow \bot)$. By logical equivalence we acquire $\neg Tf(S) \leftrightarrow (Tf(S) \wedge \neg\bot)$. $Tf(S) \rightarrow \bot$ and $\neg Tf(S) \rightarrow Tf(S)$ follow. ■

III. Yablo Sequences

A Yablo sequence takes the form:

$S_{n0} =_{df} \forall x \forall n{>}0\ (\neg Tf(x_n))$
$S_{n+1} =_{df} \forall x \forall n{>}1\ (\neg Tf(x_n))$
$S_{n+2} =_{df} \forall x \forall n{>}2\ (\neg Tf(x_n))$
.
.
.

Here, we have an infinite sequence of sentences, each of which predicates untruth of every subsequent sentence. The proof of alethic paradox informally goes as follows:

> **Proof.** Choose an arbitrary sentence S_k in the sequence. If S_k is true, then each sentence in the sequence following S_k is not true. But then what sentence S_{k+1} says obtains. So, by **T-Scheme** S_{k+1} is true, contradicting the truth of S_k. If S_k is not true, then there is some sentence Q_j following S_k that is true. If Q_j is true, then every sentence following Q_j is not true. But then what sentence Q_{j+1} says obtains. So, by **T-Scheme** Q_{j+1} is true, contradicting Q_j. ■

IV. Liar Cycles

A liar cycle is usually presented as a pair of sentences S, Q such that $S =_{df} \neg Tf(Q)$ and $Q =_{df} Tf(S)$. Liar cycles engender alethic paradox.

> **Proof.** Construct two sentences $S =_{df} \neg Tf(Q)$ and $Q =_{df} Tf(S)$. This yields logical equivalences $S \leftrightarrow \neg Tf(Q)$ and $Q \leftrightarrow Tf(S)$. By **T-Scheme** we acquire $Tf(Q) \leftrightarrow Q$. $Tf(Q) \leftrightarrow Tf(S)$ by the transitivity of the biconditionals. By **T-Scheme** we get $Tf(S) \leftrightarrow S$. We derive $Tf(S) \leftrightarrow \neg Tf(Q)$ by the transitivity of the biconditionals. $Tf(Q) \leftrightarrow \neg Tf(Q)$ follows. ■