Classical Extensions of Kripke-Feferman
Constraint Satisfaction and Alethic Paradox

Adam In Tae Gerard *

TBD University
August 16, 2024†

Abstract

Summary of the article.

*I am deeply indebted to David Ripley, Graham Priest, and Solomon Feferman for their original comments and objections. Also, to JC Beall, Bradd Hart, and Jason Winning for their advice. Last (but not least), Valia Allori, Will Robinson, Matt Babb, Richard Han, and Renee Jorgensen for clarificatory questions and reviews.

†This paper represents the culmination of several inspiring discussions begun over a decade ago and ones had with many luminaries across many fields: physics, math, logic, and philosophy. Feferman and Kripke are Rolf Schock Prize laureates - the equivalent of the Fields Medal, Wolf Prize, or Nobel Prize for philosophy.

Contents

"What is truth?" - John 18:38[1]

1 Introduction and Historical Survey

Few have denied that *Truth* **should not** be a proper object, concept, or area of study within the broader activities called *Science*, *Religion*, *Maths*, *Philosophy*, and *Law*. Indeed, such a denial appears to be a nearly futile task since the very act of denying *Truth* as a concept seems to invoke it in the first place.

And yet despite all that (all the profound and sacred texts attesting to revelation, the dizzying libraries of arcane and ancient tomes filled with deepest understandings, and all the accrued knowledge from across the Age of Enlightment, the digital era, and the Scientific Revolution), equal to its near ubiquity as a central subject to these most essential human activities, its apparent simplicity, and the unanimously espoused oft-foundational importance of *Truth* as a pillar upon which all such domains rest, is its resistance to analysis and the stubborn lack of satisfactory solution to the many puzzles surrounding its mysterious inner-workings.

Increasingly, teams of scholars have begun amassing technical tools (automated theorem proving, theology, bleeding edge mathematics, the best empirical science from across the spectrum) and hurling logical proofs like a Battle Royale Hobbesian Free For All in an increasing attempt to dislodge incumbent and entrenched positions and parties!

1.1 Storied History

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.

Definition 1.1 (Tarski's 1933 **Definition of Truth**). [2] For all x, $True(x)$

[1] Here is a humble attempt to reply to this scornful question.

[2] Semantic Conception of Truth, 1933

if and only if $\phi(x)$

1.2 The Science of Truth

Many Philosophers have taken up claims similar to Frege that *"Logic is the science of the most general laws of truth" - Frege Logik 1897*[3]

From the standpoint of modern science (Linguistics, that most recent science to fork from Natural Philosophy) there's an implicit and overt mandate that Linguistics should investigate the totality of linguistic phenomena. [4] In the same way that there's a Science of Living Things, Spacetime, etc. so have many great minds turned to *Truth* (itself) as a topic of scientific inquiry.

The greatest scientists of centuries last place *Truth* at the center of their scientific inquiries. Faraday did (*Truth* is central to both his views on Religion and Science). So too Einstein, etc. [5]

So it should be unsurprising that defects in our understanding about *Truth*, have proven quite vexing and moreover because of its seeming centrality to all the other branches of Science! [6].

1.3 Computer Science

Generally speaking: it's a kind of lingering bug (going all the way back to 300 BC) in our repository of knowledge.

Winnowing our field of vision a bit, most programmers are probably familiar with the following *Data Types* (commonly basic, primitive) within many widespread programming languages:

- *bool*
- *Boolean*
- *boolean*

Remark. We observe that *Boolean Algebra*'s are *Zero-Order* (and can't talk about *True* or *False* sentences directly)[7] (Indeed, *Reflection API's* are supplemental to core *Data Types.*)

[3] TODO Bib. pp. 330 https://d-nb.info/1247441296/34

[4] Go forth and study *Univeral Grammar* and universally!

[5] TODO finish this Bib.

[6] Refer to: https://www.thoughtscript.io/papers/000000000002

[7] More below.

But even at the level of core language *Boolean Algebra* representations (e.g. - Java wrapping *Reference Types*) there are non-trivial side-effects and representation scenarios already. Programmatic implementations of the *Liar Paradox* (or at least their close approximations) can be easily constructed even in most modern languages:

```
// Implementation from Reddit Thread
bool ThisStatementIs(bool x) {
    return ThisStatementIs(!x);
}
```

8

Scenarios of *Infinite Recursion* (above) differ from the pure mathematical contexts of the Liar Paradox (although they overlap and the Computer Science aspects of Alethic Paradox are interesting in their own right and deserve further attention and interest). One key difference: Programming and Programming Languages allow for inconsistencies to a degree that purely mathematical languages do not.[9] For example, return Types might be incompatible, some value that's supposed to be something isn't, or some asynchronous call doesn't return (in time or just at all) causing data to be missing and none of those concepts in purely mathematical languages alone.

Remark. We also observe that while the Boolean Type exists in most programming languages, we observe that the Liar Paradox emerges only in languages with Predication (first or higher order, not zero). In other words, the Liar Paradox cannot be formulated using Boolean objects alone. (only its approximations)

Remark. Undecidability - more precisely, the Liar Paradox (and other Semantic Paradoxes) are formally undecidable since the Liar Paradox lacks stable fixed points (it oscillates between *Truth Values*). As such, no determinate answer terminates the recursion chain (resulting in the infinite recursion). It's pretty cool to see this happen empirically and in real-time (programmatically)!

Also, consider the following well-known and humorous JavaScript quirks:

```
[] == ![]; // -> true
```

[8]https://www.thoughtscript.io/blog/000000000109.html

[9]Programming is in some sense mathematical - I mean the difference between *ZFC Set Theory* and say Java

```
true == ![]; // -> false
false == ![]; // -> true
```

Remark. This has less to do with Alethic Paradox in these cases but demonstrates quirks in how *Falsy* values are handled (showing that *Truth* concepts are still a bit surprising even in mature languages like ECMA).

`https://github.com/denysdovhan/wtfjs?tab=readme-ov-file#true-is-not-equal--but`

Natural Language Processing - *Truth* is of increasing relevance to Computer Science.[10]

People talk about *Truth* (itself) quite frequently. But how does *Truth* work? How do we model it correctly?

2 Logic

2.1 Definitions and Concepts

Definition 2.1 (**Soundness**)**.** An *axiom system* S is **sound** just in case each sentence s that is *provable* in system S is *True*.

Remark. If axiom system S has only tautologies as axioms and has *modus ponens* as its only rule of inference then, axiom system S is **sound**.

Definition 2.2 (**Completeness**)**.** An axiom system S is *complete* just in case each sentence s that is *True* is *provable* in system S.

Remark. By proving that a *complete* system M can be proven in S, one can show that S is also *complete*.

Definition 2.3 (**Validity**)**.** TODO

[11]

Definition 2.4 (**Logic**)**.** Defined.

1 A *logic* is a *language*, a *semantics* to interpret that language and a *proof system*.

[10] A scant search on the topic brings up many relevant articles: TODO

[11] Add definitions from my class handout: https://www.thoughtscript.io/papers/000000000001

2 A formal *language* is an *alphabet* and a *grammar*.

3 An *alphabet* is comprises a set of *logical symbols* and a set of *non-logical symbols*.

4 A *grammar* is a set of syntactic formation rules.

5 A *semantics* provides an interpretation of and the truth conditions for expressions of the language.

6 A *proof system* is a set of axioms and/or inference rules for making deductions within the language.

Definition 2.5. Conventions.

1 We shall assume the standard conventions for parenthetical dropping, precedence, quotation, and uniform substitution.

2 'Logical operator' shall be used interchangeably with 'logical connective'.

3 'Scheme' shall be used interchangeably with 'schema'.

4 'Proof system' shall be used interchangeably with 'calculus'.

5 'Grammar' shall be used interchangeably with 'syntax'.

6 'Model Theory' shall be used interchangeably with 'semantics'.

7 A variety of symbols will be deployed to denote meta-variables.

8 Arity is the number of arguments that a function or predicate can take.

Remark.

- Some might find it puzzling that we're arguing about Truth using Classical Logic when the very issue of Classical Logic is at hand! This is usually explained by appeal to the fact that we're implicitly using Tarski's *Meta-Language* and *Object-Language* distinction.

- The *Meta-Language* in question here is Classical and the target *Object-Languages* are the touted Classical or Non-Classical solutions being discussed.

- (Note this is standard practice in Compute Science - consider that the very first Java compiler developed by Sun Microsystems was written in C, Ruby's implementation langauge is C, and so on - though more archaic the original concept originates in Tarski's distinction between Meta- and Object- Languages.)

- Also, Deduction nevertheless holds in systems of 3 Truth Values and Non-Classical systems (locally or globally) so schematic arguments (that rely on Modus Ponens, Validity, and the like would remain invariants between the systems).

2.2 Classical Logic

Zero-Order

Deals with Propositions and Compounds of Propositions (Expressions) but not their internal contents (Predicates).

It is well-known that Boolean Algebra is isomorphic to Classical Logic and many systems have been shown to be as such:

- Boolean Algebra - Set Theoretic - Boole
- Venn Diagrams - Pictorial and Set-Theoretic - Venn
- Classical Logic - Modern Symbolic Logic
- Syllogistic Square - Aristotle
- Tableaux
- Lukasiewicz Simple Axiomatization
- Nicod's Scheffer Stroke Axiomatization
- Laws of Form - Spencer-Brown
- Some Apparently Classical Logics Internal to Category Theory - Univalent Foundations and Homotopy Type Theory

Note the numerous ways that people have defined systems that are taken and shown to be equivalent to each other (involving definitions, starting points, "building blocks" - ontologies or constituents of the theories if you

prefer a less loaded expression, etc.). That should be heartening to those who see alternative ways to "tinker" with Classicality.
[12]

2.3 From Zero- to First-Order Logic

This is the relevant level. Akin to Prolog (and not just the data type Boolean alone).

Of somewhat great importance to the discussion below is the idea that it's standard practice to add identity (supplement FOL with identity) after the fact.
[13]

3 Predicates, Techniques, and Truth-Tellers

It's worth discussing a few central debates and conventions. Namely, why *Truth* should be conceived as a *Predicate*.

3.1 Axioms as the Definition of a Concept

An influential way of viewing *Concepts* is to see them as essentially characterized (the totality of their content captured and expressed) by way of strictly and unambiguously formalized axioms. This approach has definitely informed the Truth Definition discussion (indeed why Tarski calls it that. e.g. - that the **T-Scheme** is the definition of and best complete description of *Truth*).

If a formal language is precise and unambiguous then its inferential relations can be wholly and completely articulated, revealed, or described (or so the argument goes).

Axioms aren't just stipulated rules, in Classical Logic:

- They are *Sound* (*True*)
- Not just *True* but provably *Sound* (in the Proof System)
- They can be proven without any Premises

[12] Add definitions from my class handout: https://www.thoughtscript.io/papers/000000000001
[13] Add definitions from my class handout: https://www.thoughtscript.io/papers/000000000001

3.2 Truth as a Metalinguistic Predicate

One rather unpopular view has it that *Truth* should not be cast as a *Predicate* (metalinguistic or otherwise) at all. That it should be understood as a Sentential Operator. From our prior discussion above, it should be obvious that's not a trivial distinction. Sentential Operators are introduced at the pre-grammatical level (before we recursively define our Well-Formed Formulae and indeed from which we construct them - they are the building blocks of grammatically valid Expressions within a Logic). *Predicates* by contrast are defined at the very end by recourse to a Model and Domain of Discourse.

This is not a trivial distinction either. Nothing "backs" or "grounds" the Sentential Operators save for Truth Tables. *Predicates* by contrast are defined by the Truth-Functional mapping between the *Predicate* symbol and an Extension.

Remark. Observe that: $@ := \neg\neg$ is not the same as $T(S) \leftrightarrow \neg\neg T(S)$

In the former case, the addition of the *Sentential Operator* called *Truth* (@) is just the logical equivalent to *Double-Negation Elimination.* Consider the well-known property that one can define any logical connective out of any of the rest plus negation.

Remark. Observe that:

- $p \rightarrow q \equiv \neg(p \wedge \neg q)$
- $p \rightarrow q \equiv \neg p \vee q$
- $\neg(p \wedge \neg q) \equiv \neg p \vee q$

At the level of operators (which logical connectives are), these are equivalences. So, truth would just be negated negation:

Moreover, we observe the strict definitional equivalence on the left and the biconditional association on the right. If we did define a *Sentential Operator* by way of the right-hand approach it would just be a relationship between *Predicates.*

The argument against such a view is relatively straightforward. By *reductio*, consider if we did take *Truth*, itself, to be a *Sentential Operator*. Among the many absurdities (including how we reconcile *Truth Values*, Formal Interpretations, and the like - and modal concepts too!) is the following:

P1 If $@ := \neg\neg$

P2 Then *Truth* is equivalent to *Double-Negation Elimination.*

⊢ This would also imply that any logic for which *Double-Negation Elimination* is *False* (or invalid) along with its entailments (such as the *Law of the Excluded Middle*) would not have any truth concept whatsoever.

⊢ It would imply that all Intuitionist and Constructive logics would not only be wrong but impossible. (And, that their inventors were not only wrong but technically unskilled at even the most basic notions within logic.)

[14]

Does the the prosententialist accept something weaker - e.g. keeping fundamental *Truth Values* (something that *Truth-Predicate* theorists do as well)?

Remark. Mapping $@p$ to an interpretation function:

- $I(p) = T \equiv @p$
- $@p := I(p) = T$

But this just equates the *Sentential Operator* with the mapping of the interpretation function (which is essentially an extension) - that is to say, this appears to equate the *Sentential Operator* with something like a *Predicate.* So, either such an operator is really just a *Predicate* (and, that tracks the assignment of the underlying truth-value T), or it's just shorthand for *Double-Negation.*

Note too that the introduction of Liar Paradox isn't due to the Predicativity, Fixed Point, or Self-Referential properties of the *Truth-Predicate* (itself). It's due to the Name-Forming mapping. The Liar Paradox can be trivially introduced given any Name-Forming Operator, **T-Scheme**, and Negation. Consider: $S := \neg @S$ and $@S \leftrightarrow S$. (Note that the Name Forming Operator is Function-Predicate notation and so itself comprises a valid compound expression in FOL.)

[14] Add: https://www.thoughtscript.io/blog/000000000107

3.3 Sentence Constructions

The well-known technique called **Curry-Howard Correspondance** establishes the formal way that Sentences, Symbols, and their Proofs arise from more primitive items within a Language.

Here and below I'll use the convention $\langle , \rangle$ to denote the familiar **Gödel Numbering** technique [15]. More precisely:

Definition 3.1. (Sentence Name) The Name of a Sentence (e.g. - a Variable Name in Computer Science) P for a Sentence S shall be written: $P := S$.

Remark. We observe that P and S can be used interchangeably, intersubstituted, and are logically equivalent to each other.

Definition 3.2. (Name-Forming Operator) $\langle S \rangle$ represents the mapping of some Proposition or Expression S to it's Name. $\langle S \rangle \equiv P := S$ returning P (P is the Name for the Expression).

Definition 3.3 (**Diagonalization**)**.** A technique that:

- Visually depicts the assignment of Sentence Names to Expressions.
- Associates the Fixed Point of a Sentence containing S as a sub-expression so that S is its own name.

This should come as no surprise since it forms the historical and mathematical basis for Variable Naming, Memory Addressing, and Value Assignment within programming languages.

So, on the view most commonly held, Truth is a *Predicate* that attaches to the Name of a Sentence. (And, it's Metalinguistic precisely because of this relationship. It attaches to the Name of Sentence not the Sentence itself.)

We say of the Sentence "It's rainy Cats and Dogs" that it is True (or False):

- $isTrue($ "It's rainy Cats and Dogs" $)$
- $P :=$ "It's rainy Cats and Dogs"
- $isTrue(P)$
- $T(P)$

[15] Following JC Beall and David Ripley.

- $T(\langle$ "It's rainy Cats and Dogs" $\rangle)$

Consequently, we'll use the notation: $T(\langle S \rangle)$

Remark. If $P :=$"It's rainy Cats and Dogs", then $T(\langle$ "It's rainy Cats and Dogs" $\rangle)$ is synonymous with $T(P)$.

These two formal techniques justify how we can define a Sentence within a Formal Language (and assign such a Sentence a Name).

3.4 Adding Predicates and Cohen Forcing

Cohen Forcing justifies the move to expand the Models under question. This is a famed and highly acclaimed technique from Set Theory.

Additionally, provided a Domain remains the same, it can be reasonably argued that regimenting or divvying the Domain (into extensions) adds or detracts nothing from the Domain (just the Extensions themselves).

3.5 Truth Tellers

Controversial and less problematic than their paradoxical kin are so-called *Truth-Tellers* - expressions that assert *Truth* of themselves (rather than *Untruth* as in the case of the *Liar Sentence*).

Definition 3.4 (Truth Teller)**.** Defined:

- Like the *Liar Sentence* but expressing *Truth* of itself.
- Constructed via Fixed-Point Diagonalization like the *Liar Sentence.*
- *Self-Referential.*

Examples:

- $S := T(S)$
- $P := S \wedge (S) isTrue$ (alternative notion sometimes encountered in the literature)
- $P := S \wedge T(P)$
- "X + 1 = 4 and I am telling the truth" (a compound *Truth-Teller*)

- “Everything I’ll tell you will be accurate (true)”, . . . , “told you I was right” (a *Truth-Teller Sequence*)

Consider the fourth example: the left conjunct can enter into **T-Scheme** but the right quite plausibly reports the conclusion of the inference. we might think it gets a *Truth Value* for the conjunct and does not enter into the truth inference itself (again).

Philosophers and Logicians have divided on whether *Truth-Tellers* should be allowed or declared to be something akin to “syntactic outlaws”. Arguments for disfavoring *Truth-Tellers* typically stem from two directions: arguments against the acceptability of **Self-Reference** and arguments from **Semantic Opacity**. [16]

But *Truth-Tellers* should not be precluded on syntactic or semantic grounds, are completely acceptable expressions, and moreover should not be excluded merely on the basis of being collateral damage when blocked for *Liar Paradox* reasons.

Consider that *Self-Reference* is indispensable to *Reflection* and *Introspection* in programming. Consider the ECMA *this* keyword and enforced $self$ convention (first constructor parameter) of Python: “This has important effects with subclassing. For another example, *Reflect.get()* allows you to run a getter with a custom *this* value, while property accessors always use the current object as the *this* value.”

https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/F

Consider the following from JavaScript:

```
const X = function() {
    return !!this == true
}
const Y = function() {
    return this == true
}

console.log(X()) // true
console.log(Y()) // false

const Z = console.log(this)
```

[16]TODO Bib.

```
/*
    //...
    //console.log(X())
    //console.log(Y())

    const Z = console.log(this)

    //...
*/
```

From a Mathematical standpoint, P can also be S anywhere there are fixed points allowed (nothing outright prevents this given a **Gödel Numbering Assignment** although we do typically construct WFF themselves without allowing Self-reference).

4 Alethic Paradox

Note that the following is a tautology:

P	Q	(P	↔	Q)	→	(P	→	(P	∧	Q))
T	T	T	T	T	T	T	T	T	T	T
T	F	T	F	F	T	T	F	T	F	F
F	T	F	F	T	T	F	T	F	F	T
F	F	F	T	F	T	F	T	F	F	F

And it follows that so too is:

P	(P	↔	¬	P)	→	(P	→	(P	∧	¬	P))
T	T	F	F	T	T	T	F	T	F	F	T
F	F	F	T	F	T	F	T	F	F	T	F

Remark. To prove $(P \leftrightarrow \neg P)$ is therefore to prove contradiction (from the above).

4.1 Alethic Paradox Defined

I provide the first formal definition for *Alethic Paradox*[17] to help recast the traditional debate about the *Liar Sentence*to the general family of *Seman-*

[17]At least that I'm aware of, anywhere.

tic Paradoxes that are relevant to our present discussion. With regard to **Howard-Curry Correspondance**:

Definition 4.1 (**Alethic Paradox**)**.** For any sentence S : The shortest proof resulting in Contradiction that requires the use of **T-Scheme** (**F-Schema**, or our other Alethic inferences including proven biconditionals involving the *Truth Predicate*).

While nearly all discussions on *Alethic Paradox* have fixated on the **Liar Sentence**, we observe that there are several species of Liar-like expressions, sequences, or constructions that give rise to the same kind of phenomena (a contradiction results when we combine those linguistic items with **T-Scheme**). (For example Rossi [18].) Many have been identified thus far:

4.2 Boolean Compounds

A **Boolean Compound** takes the form
$\ulcorner S := \neg T(S) \vee \bot \urcorner$

Proof. Construct $S := \neg T(S) \vee \bot$
TODO □

4.3 Curry Sentences

(The) Haskell Curry [19] spent some time on the Liar Paradox and its ilk:

Proof. Construct $S := T(S) \rightarrow \bot$
TODO □

4.4 Liar Sentences

Proof. Construct $S := \neg T(S)$
TODO □

[18] A UNIFIED THEORY OF TRUTH AND PARADOX pp. 7
[19] From which Currying, the Programming technique and language come from

4.5 Revenge Sentences

Beall calls scenarios where a purported solution is shown to have a contradiction within that arises from the purported solution itself, *Revenge*. Consider the imposed constraint that we preclude all *Self-Referential* sentences (that impredicativity is the problem). We formally parse this as:

- Let $Bugger(S)$ stand for *S is Self-Referential.* ($C(S)$ originally)
- Such sentences are considered *un-True* and/or *un-False*.
- More formally, a *Self-Referential* sentence S is *un-True*: $Bugger(S) \rightarrow \neg T(S)$
- $Constraint(S)$ is then defined in such a way so that $Constraint(S) \rightarrow \neg Bugger(S)$

[20]

Call the above *Naive Revenge* (since it lacks any other formal considerations - the near ubiquity of such proposals is a sad state of affairs discussed in the literature).

It's easy to show that such *Naive Revenge* proposals easily result in contradiction within any Classical (Bivalent, Closed, and Total) system governed by **T-Scheme**.

Proof. Construct $S := C(S)$.

- By **Curry-Howard Correspondance**, this yields logical equivalences $S \leftrightarrow C(S)$.
- From, $C(S) \rightarrow \neg T(S)$ and $S \leftrightarrow C(S)$: $S \rightarrow \neg T(S)$.
- By **T-Scheme**: $S \leftrightarrow T(S)$ ($CAPTURE$ in the original *Prolegomenon to future revenge pp. 1*)

Therefore, we obtain $T(S) \leftrightarrow \neg T(S)$. □

[20]TODO - Beall - Prolegomenon to future revenge pp. 3

4.6 Liar Cycles

Proof. Construct two sentences $S := \neg T(Q)$ and $Q := T(S)$.

- By **Curry-Howard Correspondance**, this yields logical equivalences $S \leftrightarrow \neg T(Q)$ and $Q \leftrightarrow T(S)$.
- $T(Q) \leftrightarrow Q$ follows by **T-Scheme** and Q.
- $T(Q) \leftrightarrow T(S)$ by the transitivity of the biconditionals.
- By **T-Scheme** we derive $T(S) \leftrightarrow S$.
- We derive $T(S) \leftrightarrow \neg T(Q)$ by the transitivity of the biconditionals.
- $T(Q) \leftrightarrow \neg T(Q)$ follows.

□

4.7 Infinite Liar Cycles

TODO

Proof. TODO □

4.8 Visser-Yablo Sequences

TODO

Proof. TODO □

4.9 McGee Sentences

TODO

Proof. TODO □

4.10 Diagnosing Alethic Paradox

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

Definition 4.2 (Triad)**.** Defined.

- **Language**: For any language with sufficient syntactic and semantic expressiveness 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 (using the traditional Classical Logic).

From **Language** we acquire a language, such as $LT*$, 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.

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. [21]

As Alfred Tarski observed, natural languages appear to be semantically closed. 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 middle13, the

[21] TODO here and below

law of noncontradiction[22], taking on a third truth-value[23], or by embracing a paraconsistent logic.

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 reasoning. However, the bulk of our mathematical and scientific theories remain formulated using classical logic. What does the failure of classical logic mean for Philosophy - how arguments and debates have been decided based on classical validities in the metalogic, etc?

The third reason is that normally when we are faced with a disconfirming case we reject or revise the theory 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. 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. It is prudent to note that it is not the aim of this paper to insist that classical logic ought to

[22] TODO

[23] TODO

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. This is otherwise a fairly original approach - almost wide open as the available path.

4.11 Desiderata

How do we rule on what theory's best? By defining decision criteria (Desiderata). The following Desiderata have been universally championed as good decision criteria:[24]

Definition 4.3 (Acceptance Criteria)**.** Desiderata:

- Formal (Technical) Proof of Soundness, Consistency
- A philosophical Explanation.
- Colyvan's *Uniform Solution*: Ideally, a solution that solves all the Semantic Paradoxes.
- Must get all the Alethic Paradoxes (not just the *Liar Sentence* and including Revenge Paradox).
- Minimize the number of intuitive theses constituting **Triad** that we reject since they are all seemingly plausible.

4.12 Formal and Explanatory

Distinguish between Philosophical Explanations and formal technical solutions. Philosophical Explanations provide a complete explanation of the technical machinery. Some Explanations assert that the Technical Machinery is at fault (like Tarski - that our technical machinery was inadequate to distinguish between the levels of artificial or possibly natural language). The

[24]Add references to this later

Technical Implementation may shed some light on particular philosophical explanations but might be seen as potentially compatible with two or more Explanations that jointly explain the phenomena.

Priest helpfully recommends the *Inclosure Schema* to characterize Semantic Paradoxes but refrains from advocating that as *Bugger* criteria. We might follow him and distinguish between:

- Features characteristic of Alethic and/or Semantic Paradox.
- Criteria used to solve Alethic and/or Semantic Paradox. (The Technical Solution)

One might wonder how the above are connected (arguably they are connected by way of a so-called "philosophical" solution). Does the former entail the latter? Priest identifies the Inclosure Schema but offers P3 as the actual solution. So, we must be careful to distinguish between these moves. Above, we see most attempts that give rise Revenge in fact attempt to equate the two moves.

Some problems:

- It's oft asserted that no proposal gets both for all alethic paradox consistently. (key being all and consistently)
- Numerous Explanations are offered but which prescind from any technical implementation. While interesting they offer no insight into technical problems with paradox.
- Moreover, no Philosophical Explanation has provided anything like rigorous Formal Proof to validate the criteria of their accuracy. Indeed Beall asserts that *Bugger* and Revenge Paradox are widespread.
- Historically, this means almost every solution that's been offered so far gives rise to Revenge Paradox.

5 Kripke-Feferman

5.1 Kripke

Explored gappy Truth Theories and 3-Valued systems starting with an analysis of Fixed Points.

5.2 Feferman

Axiomatized Kripke's Theory into a Kleene and Classical systems.

Call the Classical Bivalent axiomatization: **KF**.

5.3 Beall and Ripley

A convenient way of framing the debate around **T-Scheme**:

Definition 5.1 (**CAPTURE**)**.** Defined.

- The rule going from S to $T(\langle S \rangle)$
- Conditional subrule of the **T-Scheme** biconditional.

[25]

Definition 5.2 (**RELEASE**)**.** Defined.

- The rule going from $T(\langle S \rangle)$ to S.
- Conditional subrule of the **T-Scheme** biconditional.

[26]

Definition 5.3 (**Truth Transparency**)**.** The principle that S and $T(\langle S \rangle)$ are always and everywhere intersubstitutable.

[27]

Adding to these:

Definition 5.4 (**Truth Eliminability**)**.** Defined.

- In rewriting $T(\langle S \rangle)$ in the lexiographical form S (**Truth Transparency**) S must contain content that doesn't predicate *Truth*.
- **Truth Transparency** requires that $T(\langle S \rangle)$ can be rewritten in a form where no T appears (where *Truth* is not predicated).

[28]

Definition 5.5 (**Truth Opacity**)**.** When a sentence S cannot be rewritten (via **Truth Transparency**) without a T appearing (where *Truth* is not predicated).

[25] TODO - Reasoning with Truth - Beall and Ripley
[26] TODO - Reasoning with Truth - Beall and Ripley
[27] TODO - Reasoning with Truth - Beall and Ripley
[28] TODO - Reasoning with Truth - Beall and Ripley

6 Classical Extensions of Kripke-Feferman

6.1 Constraint Satisfaction

A paradigm from *Computer Science*, *Electrical Engineering*, *Logic*, etc.

Motivated by developments with modern Logic Gates (XOR, NAND). A constraint or condition can be fundamental.

Results from Mathematical Logic show the isomorphism between deeply complex operators like Nicod's *Sheffer Stroke Axiomatization* and Classical Logic that might traditionally represented as conditional inferences.[29]

6.2 A Finite Algorithm

Motivated by and identifies **Truth Opacity** and yet **Truth Opacity** is not itself *Bugger* - this is a separation of concerns that most touted solutions don't respect. Formally, this assists in blocking the material move from *Bugger* to *Untruth*.

Definition 6.1 (Truth Eliminability Algorithm). Sketch.[30]

- Introspects a sentence S.
- If after resolution it contains a T or an F (Predicate) stop and place S (or its name) into the Complementary Extension of $Constraint(\ldots)$.
- Then resolve all Sentence Names to their underlying expressions (the inverse function of the Name Forming Operator above), check step two again.
- Otherwise place S (or its name) into the Extension of $Constraint(\ldots)$.

[31] [32]

That's the first step which is a sorting operation. Then we populate the Extension of $Constraint(\ldots)$ second.

Remark. Lingering questions:

[29] Nicod A Reduction in the number of the Primitive Propositions of Logic

[30] The original formulation is found here: https://www.thoughtscript.io/papers/000000000002

[31] TODO - Reasoning with Truth - Beall and Ripley

[32] https://github.com/Thoughtscript/truth-eliminability-sorting-algorithm

- There's a question about whether the name has to be put into the extension. (The predicate $Constraint(\ldots)$ might be a kind of Set-Theoretic Predicate - e.g. Set Builder Style: $s \in Extension(Constraint)$ (alternatively, $s \in C$).[33] It's not clear a lot turns on this but it might have some impact as mentioned below. Is $Constraint(\ldots)$ a non-diagonalied Sentential Operator?)

- If $s \in C$ is used here **KFG** adds nothing beyond the underlying Set Theoretic Machinery (which all parties accept is being added - e.g. Peano or Robinson Arithmetic which are both formulated using the Theory of Sets), adjustments to the Interpretation Function (justified by appeal the standard practice of defining different Truth Tables and so on), and the Truth Predicate.

- Note it doesn't have to fully resolve each sentence - Visser-Yablo is sorted in O(1) time.

These are Recursive Functions (standard recursively defined functions) and so are Computable Recursive Functions.[34] We shall take to referring to this two-step construction as such (but not much hinges on whether an Algorithm or not).

It does an O(1) lookup for most problem sentences: Liar, Visser-Yablo, etc. to determine whether they contain a *Truth-Predicate*. So no looping. Decideable. Fast.

Accounts for indirect reference (referring to a sentence without using a *Truth-Predicate* until later in the reference chain). Note that the finite algo I provide is decidable in a finite sequence of steps - it might very well be a shorthand way to define Rossi's Satisfaction condition above.

Infinitary Extensions to Hypercomputation. (In any Model in which we could decide the Truth of an infinite sequence of sentences we'd presumably be in a Hypercomputation context: where N+1 is decideable in N time.)

So it gets Visser-Yablo Sequence.

Remark. We observe that the Algorithm is a sorting operation for Truth Opacity. It's a Classification or labeling operation.

[33] ZFC so no diagonalization

[34] TODO

6.3 Simplified Classical Axioms

Per the prior section, Kripke and Feferman explored gappy, non-binary, etc. logical systems.

Here I extract what I take to be the primary move made and simplify the axioms for *Truth* as follows:

Definition 6.2 (**Kripke-Feferman-Gerard**)**.** Classical **T-Scheme** with Constraint Satisfaction:

- $Constraint(S) \rightarrow (T(\langle S \rangle) \leftrightarrow S)$
- $Constraint(S) \rightarrow (F(\langle S \rangle) \leftrightarrow \neg T(\langle S \rangle))$
- F and T Predicates are decoupled.
- $I(S)$ and $T(\langle S \rangle)$ are decoupled.
- There are consistent Models for the Liar Sentence such that it is "is neither true nor false", "is false", "is true and false", and "is true".

[35]

6.4 Truth Transparency

A known consequence of any **KF**-like is that the following are permissible Interpretations (using the notation and conventions above):

Remark. Given a Non-Opaque Sentence S (sorted into the appropriate extension), an interpretation function I, and a *False-Predicate* F and using $\perp$, $\top$:

- $I(S) = \top, I(\neg S) = \perp, I(T(\langle S \rangle)) = \top, I(\neg T(\langle S \rangle)) = \perp, I(F(\langle S \rangle) = \perp, I(\neg F(\langle S \rangle) = \top$
- $I(S) = \top, I(\neg S) = \perp, I(T(\langle S \rangle)) = \perp, I(\neg T(\langle S \rangle)) = \top, I(F(\langle S \rangle) = \top, I(\neg F(\langle S \rangle) = \perp$ (in the original)

[35] TBD - Satisfies Tarski's Undefinability Theorem since **T-Scheme** will fail for $T(\langle \ldots \rangle))$ - also **KFG** should always be $\top$, $\top$ (holds of non-Truth Opaque Sentences); or $\perp$, $\perp$ (will never be $\top$ or derivable as $\top$ for previously Contradicatory Truth Opaque Sentences); or $\perp$, $\top$ for Truth Tellers in terms of the Material Conditional - will be Valid.

- $I(S) = \bot, I(\neg S) = \top, I(T(\langle S \rangle)) = \top, I(\neg T(\langle S \rangle)) = \bot, I(F(\langle S \rangle) = \bot, I(\neg F(\langle S \rangle) = \top$ (in the original)
- $I(S) = \bot, I(\neg S) = \top, I(T(\langle S \rangle)) = \bot, I(\neg T(\langle S \rangle)) = \top, I(F(\langle S \rangle) = \top, I(\neg F(\langle S \rangle) = \bot$
- ...

[36]

Remark. Alternative Set Mapping notion that's perhaps a bit easier to read:

- $I : \{S, T(\langle S \rangle), \neg F(\langle S \rangle)\} \to \{\top\}, \{\neg S, \neg T(\langle S \rangle), F(\langle S \rangle)\} \to \{\bot\}$
- $I : \{S, \neg T(\langle S \rangle), F(\langle S \rangle)\} \to \{\top\}, \{\neg S, T(\langle S \rangle), \neg F(\langle S \rangle)\} \to \{\bot\}$
- $I : \{\neg S, T(\langle S \rangle), \neg F(\langle S \rangle)\} \to \{\top\}, \{S, \neg T(\langle S \rangle), F(\langle S \rangle)\} \to \{\bot\}$
- $I : \{\neg S, \neg T(\langle S \rangle), F(\langle S \rangle)\} \to \{\top\}, \{S, T(\langle S \rangle), \neg F(\langle S \rangle)\} \to \{\bot\}$
- ...

Suppose they are consistent (as they may appear on the face of it to be), then *Truth* (and *False*) Predicates can (the operative notion here) "disconnect" (speaking loosely) from the underlying Sentence $S's$ *Truth Value* (against **Truth Transparency**) in some models.

This gives rise to an apparently long-standing criticism of **KF** - that it has wonky and/or unacceptable Models:

Definition 6.3 (KF Wonkiness)**.** If **KF** is Consistent, then there exists some Model M (a Structure or Interpretation in which every sentence S of **KF** is *True*) such that $M \models S$ and yet $M \models \neg T(\langle S \rangle)$ (and there are others too like too.)

So, not only is *KF Wonkiness* potentially problematic from an aesthetic standpoint, it may be problematic from a **Truth Transparency** standpoint: **Truth Transparency** would hold but not consistently (of course!):

- S and $T(\langle S \rangle)$ must be intersubstitutable (**Truth Transparency**).

[36] $\bot$, F, 0 can all be used instead as *Truth Values* - they are notational preferences. We use $\bot$ to avoid confusion. Also, if we allow Constraint restriction of the **F-Scheme** we arrive at the original consistent interpretations: https://www.thoughtscript.io/papers/000000000002

- But, because of *KF Wonkiness*, in such a Structure, we'd arrive at a contradiction immediately.

This appears to mean that **Truth Transparency** can't be obtained in **KF**. Let's also take a look at **Truth Transparency** from another standpoint. Consider some of more limited scenarios that might hold or obtain locally (within a single Structure or Model, not generally):

- Whenever S is in the system L then so too is $T(\langle S \rangle)$ and vice-versa (but not via entailment nor substitution alone).

- Anywhere S is *True* so too is $T(\langle S \rangle)$ and vice-versa.

- So, anywhere S is and $T(\langle S \rangle)$ is so too are the expressions: $S \to T(\langle S \rangle)$, $T(\langle S \rangle) \to S$, and $T(\langle S \rangle) \leftrightarrow S$ (not as general schemas but the precise expressions alone).

Also, consider that of **Truth Transparency** I think it quite reasonable for us to expect it to hold of itself:

- So, whenever S is in the system L then so too is $T(\langle S \rangle)$ and vice-versa (substitution).

- We can substitute into **Truth Transparency** (the line above) the contents of S and S itself. That'd be a natural of **Truth Transparency** applied to itself (it is presumably a *True* or accurate description of an essential property of the correct theory of Truth).

- We can substitue into **Truth Transparency** through other means so we don't have arrive at that from a Meta-Logical property (e.g. - a property of a logic) direction.

So, I'll call the following provisionally **Weaker Truth Transparency** and note that it is seemingly entailed by **Truth Transparency** (from the considerations above, *a fortiori*, and so on).

Definition 6.4 (**Weaker Truth Transparency**)**.** The idea here is given by the following:

- Whenever S is in the system L then so too is the content of S and vice-versa.

- Whenever S is in the system L and the contents of S are, they are both evaluated to be the same.
- The idea here is that we move the "transparency" of Truth down "one level" to look at the Propositional Interpretation and guarantee it lines up with the content.
- Arguably (Beall), an ideal system that's fully consistent would have it that **Truth Transparency** would be intimately connected **Weak Truth Transparency** (via entailment, substitution, proof, etc.) and both would obtain.
- Perhaps, this close connection (revealed by entailment, substitution, proof, etc.) is the very "heart" of **Truth Transparency** though - that the content of an expression isn't masked by a particular lexicographical form. (Indeed, a key motivation why Analytic philosophy placed/places so much emphasis on revealing the Logical Form of a Sentence - that the content or "Truest" nature of an expression is revealed via another means.)
- Can **Weaker Truth Transparency** then potentially meet these objectives without being connected to **Truth Transparency** (via entailment, substitution, proof, etc.) - e.g. in a system where **Truth Transparency** fails? The idea here is that **Truth Transparency** is desireable because of **Weaker Truth Transparency** and substituion.

This approach doesn't hold generally in **KFG** (it requires the underlying *Truth* of S for example - and we see above there are consistent Models of S where S isn't evaluated *True*) and it gets very close to **Truth Transparency** (if not to it).

The hope is that **KFG** can support **Truth Transparency** or the spirit of it (and I think **Weaker Truth Transparency** is a good way to achieve that - especially since it is so intimately inferentially to **Truth Transparency** itself).

6.5 Optional Extensions

Given the range of Interpretations (above) we might want to impose constraints to guarantee the uniformity of our Models (say to meet **Truth**

Transparency), to do so in an entirely optional way, and/or to narrow the Interpretations to find a specific Model or Model Class (a typically requisite kind ofproof to show the legitimacy of a Theory). And, the hope too is that this also addresses an apparently long-standing criticism of **KF** - that it has wonky and/or unacceptable Models (Wonkiness above).

Remark. For example, given a Non-Opaque Sentence S (sorted into the appropriate extension), an interpretation function I, and a *False-Predicate* F we might require of the truth-assignment:

- $I(S) = \top$ only if $I(T(\langle S \rangle)) = \top$
- $I(S) = \top$ only if $I(F(\langle S \rangle)) = \bot$
- ...

While **KFG** doesn't necessarily rule on a specific Model being correct, we can impose constraints on the Model to harmonize certain interpretations w.r.t. **Weak Truth Transparency**. I call this **Truth Normalization** and it's fully optional.

Definition 6.5 (Truth Normalization)**.** Given the definition D of an Interpretation Function I with Domain dom_I and Codomain cod_I^1 such that: $I : dom_I \rightarrow_D cod_I^1$. Truth Normalization involves:

- A definition $D*$ that narrows the Codomain of I to a subset:
- $I : dom_I \rightarrow_D^* cod_I^2$
- $cod_I^2 \subseteq cod_I^1$

How legitimate is this kind of move? Here are several arguments: First if we're (Classical) Constructivists we're done already. Proof by Construction. (I'm partly jesting here.)

Second, if additional argument is needed then by appeal and with careful attention to the fact Sentences and their Interpretations are constructed recursively, sequentially, and in a step-by-step manner already. Per the above, authors divide in their notation, techniques, and approach to the Interpretation Function - e.g. Atomic Propositions are first assigned *Truth Values*, then Complex Expressions are (and ditto for Well-Formed Formulae recursive constructions). The move here is seen solely as a constraint on the Interpretation Function itself. So, the moves above can be seen as a slight modification and within the realm of standard practice to the otherwise standard-fare Interpretation rules that **do not** modify any other Truth Assignment. It's a supplemental technique. [37] (Consider Project Euler 209, XOR, and NAND Truth Tables, and the like. Such alternative definitions are the kind of D^* definition above.)

Additionally, the definition narrows the Codomain of a Function to a subset. If we're justified in mapping the Domain to the Codomain in the first place it follows that there's a natural mapping already from the Domain to the subset defined within the first definition (with no other alterations or additions). So, *a fortiori*. We might strengthen that argument with a follow-up and ask why the first Definition is viable in the first place (given the second)? Standard practice? Thus, the argument can be run back the other way and there's no reason to prefer one over the other on this line of reasoning.

Lastly, the move can be justified as narrowing a range of Structures to find a satisfactory Model (a mapping from Structures to the appropriate Model Class or Model). In this case, the supplemental rules/technique is less a substantive ruling on how the Interpretation Function should behave and more about identifying the properties of desirable models for Truth - e.g. Models in which Weak **Truth Transparency** holds. This flexibility should equip the proponent of the view with a range of moves they can take without committing much to other theses (one can take the Model-Theoretic View or challenge the very notion of the Interpretation Function itself deeming a viable area of fertile inquiry that is far as I know hasn't been studied that much).

Regarding the relationship between such Normalized interpretations to **T-Scheme**, for **Opaque** expressions. The goal is for and slightly reprising the above:

- Opaque Sentences should not enter into the **T-Scheme** or other Alethic Inferences.
- Given **Truth Normalization** - the following will incidentally hold: S and $T(\langle S \rangle)$ (but this is not derived from the **T-Scheme** but by Truth Assignment). This would be **Weak Truth Transparency**.

[37] https://www.thoughtscript.io/papers/000000000001 pp. 6 - Model Theory B and C w.r.t to Sentential and Propositional Truth Assginments.

- In lieu of (strong) **Weak Truth Transparency**, **Weak Truth Transparency** is obtained and we may make contingent inferences about S to its logical form and contents.

From the list of permissible interpretations far above - we can whittle down the list to those that are consistent or specific Models within the Model Class:

Proof. Observe that in a **Weak Truth Transparency** scenario where S is the Liar:

- S and $\neg T(S)$ both hold (say in the manner above).
- And so too do then specific conditionals $S \rightarrow \neg T(S)$, $\neg T(S) \rightarrow S$, and $\neg T(S) \leftrightarrow S$.
- But in this interpretation: $\neg T(S) \leftrightarrow S$ holds. (Note that: $T(S)$ does not have to receive a contradictory *Truth Value* from the same Interpretation.)
- It meets criteria weaker than **Truth Transparency** since $I(S) = \top$ and $\neg T(S)$ both hold (since S and $\neg T(S)$ are interchangeable per section on Names above). Indeed, Weaker Truth Transparency (above) might be taken as the crux of the motivation for **Truth Transparency**.

□

[38]

Consider the following Interpretation:

Proof. Where S is the Liar:

- S and $T(S)$ both hold (say in the manner above).
- And so too do then specific conditionals $S \rightarrow T(S)$, $T(S) \rightarrow S$, and $T(S) \leftrightarrow S$.
- One might be tempted to run the Liar through the proven and resultant specific biconditional: $T(S) \leftrightarrow S$ arriving at: $T(S) \leftrightarrow \neg T(S)$

[38] This is the original base case example used before FYI: https://www.thoughtscript.io/papers/000000000002

- But in this interpretation: $\neg T(S)$ is *False* so that'd be a failure of the biconditional.

□

We'd obviously want to have our **Truth Normalization** select for the Model Class.

7 Tautology, Revenge Paradox, and Consistency Proof

7.1 Argument from Tautology

Let $\vdash P$ denote $P's$ being a *Tautology* (following standard conventions from *Proof Theory*).

A few brief definitions to get everyone up to speed:

Definition 7.1 (Tautology)**.** A *Tautology*:

- Is never $False$. (e.g. - is *Necessarily True*)
- Is $True$ under every *Structure* or $Model$.
- Can be demonstrably proven to be $True$ using $Truth - Tables$.

Remark. We write $\vdash P$ when P is a *Tautology.* (Using the Proof Theoretic notation described previously above. This notates that P can be proven without any additional or supplemental Premises.)

Definition 7.2 (Weakening)**.** A well-known *Tautology*: $\vdash (Q \to (P \to Q))$

Proof. [39] **Weakening** can be trivially checked by verifying its *Truth-Table*:

P	Q	Q	P	Q	→	(P	→	Q)
T	T	T	T	T	T	T	T	T
T	F	F	T	F	T	T	F	F
F	T	T	F	T	T	F	T	T
F	F	F	F	F	T	F	T	F

□

[39] https://mrieppel.net/prog/truthtable.html

Corollary 7.0.1. It obviously follows (in the *Metalogic*): $(\vdash Q) \Rightarrow (\vdash P \rightarrow Q)$ (This is helpful to more fully elucidate the *Proof-Theoretic* inferences that can be *Validly* made.)

Proof. From the above (either within the Deductive Metalogic or just by simple Substitution):

- Substitute $Constraint(S)$ for P: $\vdash (Q \rightarrow (Constraint(S) \rightarrow Q))$ $[P/Constraint(S)]$
- Substitute **T-Scheme** for Q: $\vdash$ (**T-Scheme** $\rightarrow$ ($Constraint(S)$ $\rightarrow$ **T-Scheme**)) [Q/**T-Scheme**]

□

The move to **KF** is therefore justified not only as a consequence of the aforementioned reasoning but on the basis of **Weakening**. If one holds **T-Scheme** then their view entails **KFG** within context (or Language) in which both are being considered (as a matter of elementary Classical Logic).

A slight variant:

Proof. If P is a Tautology $(\vdash P)$ then so too is $Q \rightarrow P$ a Tautology $(\vdash Q \rightarrow P)$. By the Truth Table above - a consequent can never be True and the Conditional False. □

Per the above, the argument here turns on **SCHEME**. The idea that the **T-Scheme** is Analytic (a Tautology) is widespread and nearly universal.

Thus, from the two arguments above: to deny the move to **KF** is to deny one or more of the following:

- Deny the truth of the **T-Scheme**.
- Deny the analyticity of the **T-Scheme**.
- Deny **Classical Logic**.

Furthermore, there's another argument that can be run here. Any other solution to the Liar Paradox will likely overgenerate a solution since **KFG** is entailed by their view (at least in any scenario where the views are compared).

One last observation here:

Remark.

- If it is the case that some theory / solution TH selects for just the set of sentences comprising Alethic Paradox, then **KFG** satisfies TH.
- From before, a technical implementation can satisfy multiple explanations.
- We also observe that proposals like Propositional Depth, Semantic Instability, Graph-Theoretic Cycles, etc. are all tracked by the **Truth Eliminability Algorithm**.
- Given the **Argument from Tautology** above, this fact as much of a surprise (since by the above TH overgenerates).

7.2 Revenge Immunity

Beall has characterized **Revenge (of the Liar Sentence) Paradox** as fundamentally involving *Bugger* - e.g. that the very criteria alleging to "fix the problem" itself generate further Alethic anarchy. Bacon also takes a similar position alleging that no Classical Solution (and especially *Restrictionist* ones) can prove their own "Healthiness"[40] and that they generate Revenge Paradox.

A few definitional items:

Definition 7.3 (Healthiness and Bugger)**.** Bacon and Beall:

- It seems that Bacon's $H(\dots)$ is Beall's $\neg Bugger(\dots)$
- $H(\dots) \leftrightarrow \neg Bugger(\dots)$
- $\neg H(\dots) \leftrightarrow Bugger(\dots)$
- $H(\dots)$ would be akin to a sentence S falling into the Extension of $Constraint(\dots)$ (setting aside questions around $s \in C$ here, etc. more below).
- For simplicity's sake we'll stick with $Constraint(\dots)$ as the relevant $H(\dots)$ (and discuss the tenability of this association below).

[40]Can the Classical Logician Avoid the Revenge Paradoxes? - Bacon pp. 8-9

Rossi and Murzi draw out and draw careful attention to the following line of reasoning implicitly found in Bacon:[41]

Remark. Bacon asserts that the following holds of any *Restriction* of **T-Scheme**:

- A sentence S can be constructed: $H(\langle S \rangle) \rightarrow \neg T(\langle S \rangle)$ via the Diagonal Lemma (with biconditional omitted).
- And that for every classical theory TH that interprets Robinson's Arithmetic that $\vdash_T S \wedge \neg H(\langle S \rangle)$ holds. "That is, restricting TS to healthy sentences doesn't prevent T from proving of one of its theorems that it is unhealthy." [42]
- Furthermore that there exists some sentence $H(\langle S \rangle)$ such that: $H(\langle S \rangle) \rightarrow T(\langle S \rangle) \wedge \neg T(\langle S \rangle)$.

The complete argument reprised with slightly modified notation[43]:

Looking more closely at the full argument on page 8 with the following assumptions:

Definition 7.4 (SRT)**.** Sententially Restricted T-Schema. $H(\langle S \rangle) \rightarrow (T(\langle S \rangle) \leftrightarrow S)$

Proof. Given the following assumptions:

- (i) TH either does not contain some axiom of classical logic, does not contain some axiom of Peano arithmetic, or is not closed under the classical rules of inference.
- (ii) TH does not contain every instance of the schema $H(\langle S \rangle) \rightarrow (T(\langle S \rangle) \leftrightarrow S)$
- In order to show this we assume that neither (i) nor (ii) hold and construct a sentence, S, such that:
- (1) S is a theorem of TH and
- and (2) $\neg H(\langle S \rangle)$ is also a theorem of TH

[41] Reflection principles and the Liar in context pp. 12
[42] Reflection principles and the Liar in context pp. 12
[43] Can the Classical Logician Avoid the Revenge Paradoxes? - Bacon pp. 8-9

Derive the following:

P1 $S \leftrightarrow (H(\langle S\rangle) \to \neg T(\langle S\rangle))$ (Diagonal lemma).

P2 $H(\langle S\rangle) \to (T(\langle S\rangle) \leftrightarrow S)$ (by SRT).

P3 $H(\langle S\rangle) \to (T(\langle S\rangle) \to (H(\langle S\rangle) \to \neg T(\langle S\rangle)))$ (by 1, 2, and biconditional weakening).

P4 $H(\langle S\rangle) \to (\neg T(\langle S\rangle))$ (from 3 by the classical tautology below).

P5 $H(\langle S\rangle) \to (H(\langle S\rangle) \to \neg T(\langle S\rangle))$ (from 4).

P6 $H(\langle S\rangle) \to S$ (by 5 and transitivity of $\to$).

P7 $H(\langle S\rangle) \to T(\langle S\rangle))$ (by lines 6 and 2 and classical logic).

P8 $\neg H(\langle S\rangle)$ (by lines 4 and 7).

P9 S (by 1, 8 and classical logic).

□

Notes:

p	q	(p	→ (	q	→ (	p	→	¬	q)))	→ (	p	→	¬	q)
T	T	T	F	T	F	T	F	F	T	T	T	F	F	T
T	F	T	T	F	T	T	T	T	F	T	T	T	T	F
F	T	F	T	T	T	F	T	F	T	T	F	T	F	T
F	F	F	T	F	T	F	T	T	F	T	F	T	T	F

Diagonal Lemma. Let TH be a theory extending first-order arithmetic. For every formula $\phi(x)$ there is a sentence S such that $TH \vdash S \leftrightarrow \phi(\langle S\rangle)$. (Note that $\phi(\langle S\rangle)$ must be a Number Theoretic Function of the kind $\phi(1,2) = 3 = 1+2$ in the original proofs. That detail is often omitted and this finer point is discussed further below.) □

`https://plato.stanford.edu/entries/self-reference/`

Rossi and Murzi argue that Bacon's conclusion is too hasty.

Here are some standard ways in which argument might fail:

- One or more premises are not **Sound**.

- The argument is not **Valid**.
- The argument doesn't hold locally or is irrelevant - it might apply to some species of argument, a domain, or other phenomena but not the relevant one.
- A formal, technical, or other mistake is made.
- The assumptions are such that although the argument might be **Valid** and the numbered premises **Sound**, that the argument fails to apply to a particular theory or target. The conditions for getting "off the ground" (so to speak) aren't satisfied for a particular theory. Some specified criterion or assumption for the argument to be relevant, that must be satisfied, or that must hold doesn't.

Generally, my criticism stem from the following:

- Whether **KFG** requires Bacon's $H(\dots)$, Beall's $\neg Bugger(\dots)$, some other Meta-Linguistic Predicate (beyond the Truth Predicate) at all. If not Diagonalization isn't a problem.
- Even if **KFG** is committed to some Meta-Linguistic Predicate (beyond the Truth Predicate) after all, it's not clear such expressions are Theorems, that they need to be, or that expressions of **Truth Opacity** (or the lack of **Truth Opacity**) not being Theorems is a problem. They are provable but not Theorems - the Truth of such expressions is given the **Truth Eliminability Algorithm** (arguably a Contingent Truth Assignment). It is evident that such expressions are or are not in virtue of the Algorithm ("Constructive Proof" as it were, in the Meta-Logic prior to the Language being Formalized) but not required of the Proof Theory at the level of premise-less inference within the system itself.
- That the Restrictionist Conception Bacon takes aim at isn't actually exhaustive of the Concept - **KFG** has certain distinctive features unique to its approach that aren't of naive **Restrictionist** views.
- It's not clear that Bacon's argument isn't susceptible to problems itself.

First, addressing some additional higher criticisms:

- Bacon says : "Restricting [**T-Scheme**] to healthy sentences doesn't prevent TH from proving of one of its theorems that it is unhealthy".
- While that might be generally true of naive *Restrictionist* approaches, that result doesn't seem so damning w.r.t. **KFG**.
- The **Truth Eliminability Algorithm** itself doesn't fit the Schematic profile for being a *Bugger*, an *Unhealthy*, and so on (e.g. - an Algorithm determines the Contingent Truth that a Sentence is Truth Opaque or not, it may or may not require a Predicate, and the T-Schema/Thereoms need not be Truth Opaque nor not Truth Opaque - they all get Truth Values - expressions of Truth Opacity aren't Theorems.) It sequentially sorts sentences into two piles. And, sentences in both piles receive Truth Values. So there's no entailment here between this sorting, membership in one pile or the other, to *Untruth* although specific Truth Assignment might such a particular sentence receives a $\perp$.
- It would be far more damning, if it were the case that "Unhealthiness", $\neg H(\langle S \rangle)$, generally **entails** *Untruth* (or "Healthiness" for that matter).
- Since this is not the case we turn to the argument given for a single constructed diagonalized sentence.

Second, in considering the assumptions:

- I think it's unclear whether [P2] (or the assumptions) actually hit(s) its mark against **KFG** since the relevant Healthiness notion in question might be a Set Theoretic Inclusion Sub-Expression on S rather than its Name.
- If so, are there grounds to say the argument fails to match schematically?
- If that line of reasoning is correct, then subsequent moves wouldn't apply to **KFG** (although they'd probably still hit the mark against other naive **Restrictionist** views). That's also possibly non-trivial since the argument relies on Diagonalization.

- Here's the argument: "There is a formula that expresses the relation ϕ in the system TH, provided that the function $f(x)$ satisfies certain conditions — usually the condition is that it be a primitive recursive number-theoretic function."

- So, when we consider the formula: $(\phi \in C \to \neg T(\langle\phi\rangle))$ with one free variable it isn't clear it's a *primitive recursive number-theoretic function* - it's a Set Theoretic formula (we are in a theory with Peano Arithmetic after all) with a compounded function clause (and if it only contained a Truth Predicate, then I would agree it's diagonalizable). (I consider this approach a bit more fully below.)

- If the line of think above is correct, then there are grounds to say the argument fails to match schematically (and in a principled and reasoned way).

- If wrong we are fully licensed to diagonalize into the formula: $(\phi \in C \to \neg T(\langle\phi\rangle))$ like so: $S \leftrightarrow (s \in C \to \neg T(\langle S\rangle))$

  `https://www.jamesrmeyer.com/ffgit/diagonal_lemma`

Additionally:

- Recollect that Bacon assumes that *ii* is *False*. Obviously, an antecedent can be *False* and the consequent *True*. So *ii* can generally hold in **KFG** given the lack of entailment described above.

- Regarding the nature of $H(\langle S\rangle)$ expressions: no $H(\langle S\rangle)$ is a Theorem (that is proven without assumptions or premises) nor is $\neg H(\langle S\rangle)$ - they are contingent (they can be proven given other assumptions but never alone). A sentence for which $H(\langle S\rangle)$ holds will always receive the same Truth-Value for $H(\ldots)$ in any Interpretation however (although again these are not Theorems - they are Analytic if we want to use more familiar philosophical terminology).

Still even if no $H(\langle S\rangle)$ expression is theorem we might be able to show that contradictions emerge. (And, barring the success of the above attempt to sidestep the schematic appropriateness of any otherwise potential valid argument.)

Observe that $p \to \neg p$ is not a contradiction:

P	P	→	¬	P
T	T	F	F	T
F	F	T	T	F

So, Bacon's move from $H(\langle S \rangle) \rightarrow \neg H(\langle S \rangle)$ from lines **2** to **8** is not a contradiction. However, from lines **4**, **7**, and the following the above argument appears to show that there exists some sentence $H(\langle S \rangle)$ such that: $H(\langle S \rangle) \rightarrow T(\langle S \rangle) \wedge \neg T(\langle S \rangle)$:

P	Q	((P	→	¬	Q)	∧	(P	→	Q))	→	(P	→	(Q	∧	¬	Q))
T	T	T	F	F	T	F	T	T	T	T	T	F	T	F	F	T
T	F	T	T	T	F	F	T	F	F	T	T	F	F	F	T	F
F	T	F	T	F	T	T	F	T	T	T	F	T	T	F	F	T
F	F	F	T	T	F	T	F	T	F	T	F	T	F	F	T	F

To resist this contradiction we must reject seemingly reject **Classical** or read the conclusion as a *Reductio*. And, I think the prior considerations inform that latter approach. So our two considerations seem intimately interconnected. By rebutting one line of criticism we might be able to address both.

We might pause before concluding that. If one could should that some premise was unsound (weak subgoal) or contradictory (strong subgoal) itself then:

- A False antecedent and a False consequent makes a conditional True. (By the Truth Table for a Material Conditional.)
- If one can show either that: $H(\langle S \rangle)$ is not **Sound** or even better contradictory then we can show that even if the argument given is **Valid**, that it's not problem.
- Consider the prior discussion about $H(\langle S \rangle)$ being a Thereom. It's not.
- Provided that we can diagonalize into S the very sentence S is not "Healthy" (by the definition of the sorting algorithm and the established connection between $H(\dots)$ and $Constraint(\dots)$) since it itself contains a Truth Predicate. So, $\neg H(\langle S \rangle)$.
- But that's essential for the move to **P2** (drawing attention to the legitimacy of substitution into **SRT**).
- Before we celebrate showing the contradiction in the argument (that it contains an Unsound, Invalid, or Contradictory step in *its* assumptions), let's ensure this feature doesn't affect the theory as well.

- Furthermore, consider the following Structure within which such a diagonal sentence is constructed: $\neg S$, $T(\langle S \rangle)$, $S := (H(\langle S \rangle) \rightarrow \neg T(\langle S \rangle))$, and the $\neg H(\langle S \rangle)$ (since S would contain a Truth Predicate). Note that the left-side and right-side Truth Values of the diagonal biconditional don't match S. Beyond questions about the legitimate use of Diagonalization, it's not clear if the Schematic type can be

- So, there are several questions here if the argument starts from a contradiction (which it seems like it does when run against **KFG**): this might add additional reason to go the other route mentioned above (w.r.t to the Set Theoretic Inclusion Sub-Expression: $s \in C$), it might show a chink in the armor of diagonalization, and/or it can be a False biconditional (and therefore the argument's not **Sound**).

Now, lastly, regarding the conclusion that every **Restrictionist** theory TH cannot prove of its sentences S that they are $H(\langle S \rangle)$ (and whether that then entails $H(\langle S \rangle)$ is another step we might consider). Presumably, if it can't prove of all its sentences that they are $H(\langle S \rangle)$ (e.g. - generally or universally), that some would be *Healthy* and others not: that there'd exist some S, Q: $H(\langle S \rangle)$, and $\neg H(\langle Q \rangle)$. In fact, it might very well show the exact function of the sorting operation: to say of sentences that they are unhealthy or not (in his terminology) (e.g. - We expect the sorting operation (algorithm) to do exactly what the conclusion of the argument is: it says this sentence is **Truth Opaque** and this sentence is not)!

To reprise, the sorting operation sorts sentences into two piles whereby one pile contains all and only the **Truth Opaque** sentences (in the sense above) and the other the rest. So, the sort is identifying just those sentences that are "Unhealthy" and putting them together into an extension (and the rest into the complement). In other words, the conclusion of Bacon's argument can be understood as requiring that **KFG** be able to prove of itself that some of its sentences are **Truth Opaque** - the very thing the sorting operation is trying to do. (And in doing so the argument from *SRT* might very well strengthen one's preference for sorting and indirect reference (pun) algorithm in the first place.)

In fact, the very absence of this ability (e.g. - if the conclusion of the argument above weren't correct) it would seem to imply that the sorting operation would be impossible (or at least unsuccessful in some other unspecified way).

Furthermore, suppose his conclusion is correct, why is a sentence being *Unhealthy* so damning if *Untruth* is not entailed? Consider also that for every Sentence that is not **Truth Opaque**, it can be shown that they are "Healthy" per the above. The allegedly damning conclusion to the argument is tantamount to "That is, restricting TS to sentences that aren't Truth Opaque doesn't prevent T from proving of one of its theorems that it is Truth Opaque."

Remark. Consider sentence: $S{:}T(\langle S\rangle) \to T(\langle S\rangle)$. S is a Theorem and it's Truth Opaque. It gets a Truth-Value. It's True. It's not run through the **T-Scheme**.

Importantly and again, the Algorithm is a separation between the **Opacity** criteria and the **T-Scheme**. **Opacity** is not as it were "Unhealthy" in the sense of entailing *Untruth* although the very intent is for the language to talk the problematic sentences. The semantic criteria are decoupled from the syntactic criteria in this way (and "Unhealthiness" appears to require a tight connection to semantical *Truth Values* to really be a "problem"). And, indeed we would expect as such if the language really were semantically closed. (Between *Unhealthy* and full admission into the inferences of the **T-Scheme** is a constraint condition and unlike naive approaches which necessitate the Falsity or Untruth of a Sentence, this approach does not.)

Remark. Observe that the sentence: $T(\langle S\rangle) \to S$ isn't run through **T-Scheme** but can be proven (from other sentences), can be a Theorem, and can be of the form of a more general Axiom Schemata. **T-Scheme** and **KFG** aren't run through **T-Scheme** either. We decouple **T-Scheme** and any requirement that Theorems be run through **T-Scheme**.

See the Consistency Proof.

7.3 Consistency Proof

Given some sentence S:

P1 $P(S_1)$

P2 $(\forall n)(P(S_n) \to P(S_{n+1}))$

$\vdash$ $(\forall n)(P(S_n))$

44

Mathematical Consistency can be proven via Mathematical Induction (the standard approach to Consistency Proof). One proves that Consistency is obtained at each step of the sequence and that there's no change or variation at each further step. We can then be reasonably assured that the sequence in question is mutually Sound (Consistent).

Then by Mathematical Induction:

Proof.

- Iterated, non-diagonal, or concatenated expressions can be given a consistent interpretation if their base constituent expressions can. Consider the following base (non-iterative, non-diagonal) expressions and their consistent interpretations[45]

□

7.4 A Second Approach

This approach falls more on the **K** side of team **KF** (whereas the above perhaps leans more on the **F** side of things).

An alternate way to approach Truth was perhaps revealed above. Guided by our discussion around just what a Theory of Truth entails (e.g. - arriving at a "mechanics" and Formal treatment of Truth in the same way we approach Gravity, Physics, and other linguistic phenomena), we see a Classical Theory of Truth in the same way we think of theories in Abstract Algebra: Field Theory, Group Theory, and so on. Again, such an insight is motivated by the multitude of ways that Mathematicians have approached Classically isomorphic systems: Spencer-Brown, Venn, Boole, Frege, and so on (refer to the many ways people have constructed Classical systems in the first few sections). That is to say:

- We "equip" Classical Logic with additional Logical "machinery".
- We add to Classical Logic the extra stuff required to understand the Domain of Discourse at hand.
- This is built into Model Theory and Classical Logic - it's standard practice.

[44]This formal notation: https://plato.stanford.edu/entries/sorites-paradox/

[45]See Section 6.4 and https://www.thoughtscript.io/papers/000000000002 pp 11

- Provided that the Proof proves all the Propositions of Classical Logic, the underlying machinery is fair game - that Mathematics affords high levels of creativity when we engineer the systems (but require of them Classical Completeness).

From this vantage point we can see perhaps several additional approaches:

- What about the Interpretation Function? How do Interpretation Functions relate to each other in the case of Alethic Paradox (for instance it's standard practice to first define a Propositional Interpretation and then a Sentential Interpretation)?

- In addition to the challenges raised above (against those arguments that purport to show the untenability of all Classical solutions and *Restrictionist* solutions in particular) we might ask about several routes that have been discussed in other contexts that have not considered at all in this. For example: Truth Predicate Pluralism, Predicate Typing, and Quantification Restrictionism.

- Perhaps what Kripke and others have brought into question concerns the limitations of and justification for Diagonalization as a general technique. Ostensibly, it's unrestricted (and the very unrestrictedness of the Diagonalization has led to most of the Semantic Paradoxes).

Here's a sketch of some of the above to hopefully (eventually) do one better than the move far above (w.r.t. **Truth Transparency** - e.g. to move from **Weaker Truth Transparency** to **Truth Transparency** - perhaps they are the functionallty the same but perhaps they're not - we've got Consistency but also to get the rest of the way):

- We sequentially define a Predicate T^* that does not permit Fixed Points. A "Truth-Sayer" Predicate - it allows one Sentence to talk about another (but not itself).

- We define an Interpretation Function I^* to include standard Sentential Truth Interpretations and to handle the special case of Predicate T^*.

- We might futher define several such "Truth-Sayer" Predicates and organize them or relate them finitely by Type so that no two Predicates of the same typing can "Truth-Say" of each other. The idea here is justified from the line of thought that Liar Cycles exhibit not merely Chronological or Sequential order but are bookended by two distinctive types of Predicates - one that looks forward and one that looks backward - indicated by Tense - Tense modifies or specifies distinctive Truth Predication. (Obviously, this would take more to get off the ground.) Also, definitions are often interdefined. Why think Extensions aren't?

- We might also allow a weaker *Truth-Predicate* T that admits Fixed Points but no negation symbols at the Predicate level. This would be a "sub-Expression-Predicate" (that cannot itself be a WFF alone - it must be part of complex Sentence in the same way two-arity logical connective must) - taking the path of Truth Eliminability to indicate a failure to distinguish between kinds of Predicates and the roles they play within an Expression.

- Such might be justified from the view that we begin at Truth then define its opposites (in fact this is often standard practice: $\bot := \neg\top$). In this case, we would define negation symbols after formally defining the relationship between T and T^*.

- We then bind the Truth of T^* to T where T (requiring certain conditions of both to hold) holds and T^* would then analogous to a Sentential Truth Interpretation function in standard practice.

- **T-Scheme** is then restricted to T^* (implicitly per Type Quantification Restriction) or explicitly (use of T^* alone, or by Constraint Satisfaction, or Restrictionism above).

- Lastly, and perhaps most fruitfully we might look to the Name Forming Operator we've all been using. Why does Naming not come with more rules or constraints on use? In practice, people seem to be reluctant to use Names (naming Baptisms without regarding certain social practices - like kinship, friendship, and years of acquaintance as in the case of nicknames).

8 Philosophical Observations

For a philosophy and logic paper I've perhaps spent a surprising amount of time on topics and areas outside of philosophy far above. Here, I dive into some the Philosophical discussions and the Explanation, relevance, and so on.

8.1 Summary

This view:

- This solution simplifies Kripke-Feferman and combines the Constraint criteria with Truth Eliminability and a Definition.
- This solution draws inspiration from Beall's Neglected Deflationary Approach, Kripke Kleene 3-Valued Fixed Point Solution, and Feferman's Classical Axiomatization of the aforementioned.
- It locates the problem in Truth Opacity and the missing requirement that we should only allow sentences that are recursively constructed Truth expressions the same way we do already in Boolean Algebra/-Classical Logic with WFF.
- Sorites is only an apparent paradox that conflates numerical quantities with shapes, aggregations, and the configuration of quantified constituents.[46] - Separate argument
- The approach taken above (if ultimately successful) might also inform the Knowledability Paradox by providing a general technique. The sorting method might take either T or K.

  `https://plato.stanford.edu/entries/self-reference/#ConEpiPar`

Remark. If **KFG** is consistent then the set of sentences that are *True* and which can be proven so (although not necessarily as theorems) is not equivalent to other supposedly **Sound** and **Complete** solutions. (It proves more sentences than its alternatives.) On the view taken here, the other theories fail to sufficiently the relevant set of Truth-apt linguistic phenomena.

Regarding the distinction between an Explanation and a Technical Implementation. Regarding the Explanation:

- I provide the first formal definition of Alethic Paradox (narrowing down from the Inclosure Schema) to my knowledge.

[46] TODO - think this was raised by another person or two

- Transparency and Eliminability (see **RELEASE**). That **T-Scheme** necessitates being to remove Truth-Predication from a sentence/expression.
- Confusion about the scope of **T-Scheme**.
- Confusion about the order of recursive sentence constructions w.r.t. *Truth-Predicate*.

like Priest, there's a distinction between the two that's respected.

8.2 Alternative Nearly Classical Views

Regarding Rossi's Conversational Contextualism:[47]

- What's a Conversational Context? What delimits one from another (that seems essential to the view).
- Can't the Liar Paradox be written? And not require a conversation? Does Conversational Contextualism apply in those circumstances? Are spoken language contexts different than written ones? (Socrates appears to have thought so, probably many others)
- Appears to give rise to Revenge.
- What about Infinite Sequences like Visser-Yablo? How do we accommodate these within supposedly finite conversational language contexts?
- Deeper question perhaps: Does alethic Paradox bifurcate into Conversational and Monological scenarios (say following Rescher)?

I'm actually fairly sympathetic to this approach for the Liar Paradox. But one sticking question is simply that I don't think most people (save for Philosophers and a handful of Logicians) talk about the Liar Paradox that much (or at all) and almost none run through all the Inferences or Contextual moves asserted to occur by the above theory anecdotally. For instance, the Reddit threads above which only scantly discuss the topic. They don't run through the inferences at all.[48]

Regarding Rossi UTTP:

[47] REFLECTION PRINCIPLES AND THE LIAR IN CONTEXT

[48] Verify using these examples

- Describes the looping nature of dependency chains (akin to the explanations offered by Read and Englebretsen) [49]
- Partial technical implementation.
- Undecidable loop checking required (a known problem with differences between Depth First Search and Breadth First Search in C.S.) and *Truth-Tellers* (also loop but should be allowed per the above) (the above is O(1) even with Visser-Yablo)
- Models Classical Logic but seems to require opting for 3 *Truth Values*: $0, 1, 1/2$. Either this collapses into another Non-Classical solution or it invokes more complexity than Truth Normalization of the Interpretation Function (above) since it introduces much more baggage by comparison.[50]

Propositional Depth, Loops, Programmatic Recursion, Fixed Points, Impredicativity, Self-Reference, etc. Stephen Read, Lorenzo Rossi, and discussions in Computer Science. The above have all been similarly discussed (they appear to be all related). These appear to run afoul of Beall's *Bugger*-type criteria. The above mostly describes phenomena but doesn't explain why or how to uniformly address these features.

Regarding Ripley:

- Drops the Classical Rule Cut (deemed essential by mathematicians today [51])
- Substructural
- Purports to get Sorites and Alethic Paradox.

Regarding Zhen Zhao's **T* Restriction**:[52]

- Interesting combination of Tarskian approaches and Restrictionism that purports to unite a multitude of Truth Predicates into a singular Monistic view (about the Truth Predicate) while trading both the singularity and general validity of **T-Scheme**.

[49] TODO

[50] pp. 4 TODO

[51] TODO

[52] The Outline of A New Solution to The Liar Paradox 2012

- Inherits the Tarskian Ascending Hierarchy of Languages approach and constructs a top-level Language LT (maximum ordinal a or **Theorem 3.8**) from the assemblage preventing both the general validity of unrestricted **T-Scheme** and only **T-Scheme** for the Ordinal level allowed that holds within the Language.

- Uses the following constraints, where c^1_{x1} is a Constant in Language $L_0 : c^1_{x1}$, x is a Formula of L_0 and 1 means x is a Closed Formula.

- $(c^a_{x1}) \leftrightarrow x$ is the resultant **T-Scheme** and there is at least one defined for each Ascending Language (whose "level" is specified by its ordinal).

- Bacon's argument appears to hold since we're in LT at the maximum ordinal a, it doesn't seem that LT can prove theorems about itself (since it would necessitate moving to $a+1$ by the language construction which it can't since a is the maximum ordinal).

- There's another question, Zhao's approach seems to prove many things (indeed the entire Tarski-inspired assemblage, language construction, numerous **T-Schemes**, and so on) within some Meta-Logic. (And which one?) But that Meta-Logic cannot be some $a+1$, so either there really is another more general theory they're presupposing or theory theory's wrong or both.

- Despite those defects, it's an incredibly interesting and potentially overlooked approach in line with the **Restrictionist** approach offered above. A proof or two that they give shows that there are potential Models that can hold of a **Restricted T-Scheme**, and that would be compatible with **KFG**.

- **KFG** differs in the following respects: **KFG T-Scheme** is generally Valid, the approach is simplified by relaxing the **T-Scheme** by **Weakening** and not specifying a connection between **T-Scheme's** and the Language Ordinal Level. On both views **Original T-Scheme** is Contingent and holds only of some Models and not others. Zhao's motivation and inspiration seems inspired from **LANGUAGE** rather than **SCHEME** but they arrive **Restrictionism** nevertheless.

- Tarski's original approach is triply profligate: it requires an infinitude of languages (sequentialy by Object- and Meta- Language), Predicates (one at each Meta- Language), and therefore T-Schemes (one for each Truth Predicate). Zhao's proposal cuts this to a-many Langauges (used to construct the top-level Language) and a -many T-Schemes.

- **KFG** only requires one Language, one Truth Predicate, and one **T-Scheme**

Others:

- Only provide an explanation but no technical implementation and ultimately no formal proof of consistency.
- Misses phenomena like Visser-Yablo - only partially address Alethic Paradox.
- Attempts to provide a technical solution that ends up only being a partial solution (as Tarski himself admitted at the end of his legendary paper) or lacks an explanation.

[53]

8.3 Empirical Observations

People don't run the *Liar Sentence* through **T-Scheme**, they laugh about how strange a Sentence it is. This anecdotal data point supports the claim that **T-Scheme** should be constrained/secure/restricted by the "security checkpoint" metaphor.

We observe that solutions to the Liar Paradox have been thrown into the ring from four corners:

8.4 The Catuṣkoṭi

Per the above, **KF** (and **KFG**) give rise to multiple consistent Models some of which have been deemed counter-intutive or peculiar. Consider that something like (stressing "like") the following can all be consistent:

- "is neither true nor false"

[53] Rossi writes: ' It is often argued that any purported solution to the semantic paradoxes faces inevitable revenge problems, both in classical and in nonclassical settings" Reflection principles and the Liar in context - pp 12

- "is false"
- "is true and false"
- "is true"

The technical details are discussed above but the natural tendency has been to take that consequence of the view to show:

- A fundamental mismatch in *Truth Values*, Truth Assignments, and the Truth Predicate (violating principles like **Truth Transparency** that require their alignment).
- It's an unacceptable semantic or syntactic peculiarity that demonstrates the ad-hocness of the view.
- It masks some deeper contradiction that's lurking beyond the superficial formal machinery (perhaps at the Truth concept level itself).

First indeed like most at the time (apparently), I even balked at **KF** as a consequence of the above but I don't now think that so bad. In fact, upon closer inspection, there are some far more interesting things going on. For one, there's a surprising silver lining: the solution can accommodate the four divergent intuitions that we may say of the Liar (or any such sentence) it thereby both explaining the various positions held by those in the Liar Paradox debate and providing an ecumenical solution to Alethic Paradox. On that last notion, naturally, we would expect an accurate description of Truth to account for the wide range of philosophical positions that have been held by various Philosophers, Logicians, and Mathematicians and we indeed see that revealed above. I now count this as strong evidence in favor of the empirical accuracy of the view. All other theories just dismiss everyone else's as mere confusion. But that perhaps seems unlikely given the luminaries involved.

Indeed, we revisit the following found in JavaScript and we observe the close similarity to the mismatch (although in slightly different circumstances - *Falsy Type Coercion*):

```
[] == ![]; // -> true
true == ![]; // -> false
false == ![]; // -> true
```

In many Programming Langauges they've accepted such quirks and it seems to have little impact on the success or viability of those langauges.

Second, the normalization extensions above provide I think a way out (so the conclusion is too quick to dismiss **KF** on the basis of the above). **KFG** provides the machinery to create a simplified, consistent, Classical, and Truth Transparent solution.

Moreover, and perhaps of greater relevance is the ancient religious and philosophical notion called the *Catuṣkoṭi* whose expressed Truth potentialities appear to be precisely the above. [54] Although it is localized to the Alethic Paradoxes it is found nevertheless embedded into the rest of classical logic on this view. This is perhaps interesting in its own right since Eastern and Western logic have long been understood to be in opposition (Aristotle's Syllogistic Square vs. so-called Indian Logic). Here they appear to occupy subregions of the same logic.

So, the aesthetic preference (let's call it what it is) for the requirement that every S and $T(\langle S \rangle)$ pair align on their Truth Values (over and beyond the dictums of Classical Logic itself) might very well reflect some deeper cultural bias. Many Christian sects for instance challenged Aristotelian orthodoxy in the past [55] in addition to the oft-cited "Eastern Logics". And many more have argued for Classicality (indeed it was central to many Catholic Fathers of the Church). So there's Aesthetic relativism, subjectivism, or underdetermination (that there's no legitimate reason to accept one over the other beyond merelh arbitrary preference). If Classicality and Consistency are achieved then what's the blocker, really? Aesthetic preferences don't seem to be compelling grounds to reject the view.

Prescinding from the above, I know of no other such Syntactic or Semantic artifact of Religious significance that's been naturally "discovered" within our best natural theories. Logic plays a central in multiple religions: Logos in Christianity, the Catuskoti in Buddhism, and many so-called Indian Logics central to several religions. From a historical vantage point, the *Catuṣkoṭi* predates many religions, sects, and their attendant theologies (including many that endorsed Aristotelianism).

[54]TBD

[55]For example many Orthodox Christians "Florovsky, by denying the Law of Non-Contradiction, clearly falls into the trap of several twentieth-century Orthodox thinkers" Florovsky's logical relativism: a philosophical and theological analysis - Harry James Moore pp. 8ish

8.5 Restrictionism, Alethic and Logical Nihilism

Definition 8.1 (**Restrictionism**). That **T-Scheme** should be conditionalized. (There are constraints to move from left to right on the T-Scheme.)

From the vantage point of Philosophy, one commonly encounters the challenge that the failure of our theories and conceptions about *Truth* results in widespread Skepticism about Logic, Validities, and Reasoning.

Others have challenged the tenability of all truth-apt (Rationality and all-knowledge) projects in the face of worries surrounding the *Liar Paradox*.

Such are the purported implications of *Alethic Paradox* (which include the limitations of *Classical Logic* and hence, *Rationality*, *Religion*, *Science*, *Mathematics*, *Law*, and most of the deeply entrenched assumptions framing contemporary metaphysics).

[56]

[57]

Metaphors: Security Vestibule, Bending a Metal into a Stronger Shape, Logic Gate, Checkpoint

Restrictionism doesn't rule on Metaphysical questions. It's consistent with a range of philosophical proposals.

It offers a potentially easier out than rejecting Logic (itself) when replying to supposed failures of Logical Validities.

Inductive Argument:

P1 All prior supposedly universal Laws or Theories come to ultimately apply only to a specific region or domain (they are Conditionalized or have Ceteris Parabis Conditions).

P2 We have no reason to think our current or future Laws/Theories are in different in this respect.

$\vdash$ Therefore all current or future Laws/Theories will come to ultimately apply only to specific regions or domains.

Premise One is substantiated by recourse to the history of General and Special Relativity, Hyperbolic vs. Euclidean Geometry, The supposed universality of Newton's Theory of Gravity (and the retreat from those universalist

[56] Refer to: https://www.thoughtscript.io/blog/000000000025

[57] Add bibliography and such from the other draft

claims to the current quartet regime ("Branches of Physics") we have today: Quantum, Newtonian, Special Relativity/Quantum Field Theory, and General Relativity), and Black Holes.

Therefore, we have no good reason to think **T-Scheme** is universal (e.g. - we have good inductive reasons to support restrictionism).

Does this same argument affect **Classical**? Sure, partly jesting here, but only ast per Quine's Maxim of Minimal Mutilation to our beliefs.

8.6 Semantic Closure

Tarski argued for a distinction between Meta- and Object- languages as a consequence of the Liar Paradox. If we've arrived at a consistent solution, do we need such cumbersome machinery?

Can we complete Tarski's quest? [58]

8.7 Type Quantification

Can proponents of **T-Scheme** bake their cake and still eat it too?

Argument from Superficiality:

- **T-Scheme** masks implicit Type Quantification [59] restrictions (which explains the millennia of confusion and the apparent analyticity of the **T-Scheme**).
- The quantifier implicitly ranges over a set of sentences and this set is typically omitted when **T-Scheme** is presented or depicted.
- That set are just the sentences that satisfy $ConstraintS$ (or $\neg Bugger(S)$)

Almost any solution implicitly endorses the above. So this should not be so controversial. The intuitive of the **T-Scheme** is its apparent simplicity- an appeal to a particular lexicographical aesthetics. And, that can nevertheless coexist with **KFG**.

There is another interesting area that I think has remained unexplored to the *Restrictionist* as well (and again, I think this approach is one of only two *Restrictionist* approaches at least that I'm aware of). Perhaps the line

[58] As they themself alluded to at the end of his seminal paper...

[59] https://www.microsoft.com/en-us/research/video/the-structural-theory-of-pure-type-systems/

of thinking elsewhere in this article takes us down the path of questioning the type of Sentences and Expressions we consider Well Formed (following Tarski). Tarski rejects the Liar Sentence as completely syntactically invalid - it should even be formable as an expression (only by conflating the levels of artificial language) - and therefore blocked as Well Formed. (Call this Another Potential Solution.)

Can we take a weaker guided by the above? Suppose we allow the Liar Sentence to be formed but not Well Formed (in-between Tarski the approach described above)?

- A theory can be locally or globally restricted (**KFG** takes the local approach applying restriction solely to the **T-Schema**).[60]
- We can allow the Liar Sentence as an expression but ban it as a WFF. There is another question too about Classical Logic. Suppose we were to assign the Liar Sentence a Truth Value but ban it is a WFF. Would that violate **Completeness** or not? Allegedly under the global restriction approach the other laws, rules, theorems, axioms of Classical Logic would only range over WFF so possbily not.
- And we observe that Classical Logic (even pre-Truth Predicate) separates WFF from non-WFF - and we normally just say that those Expressions which don't receive Truth Values are part of our **Completeness** concerns. Classical Logic appears to restrict questions about two Truth Values solely to the WFF but as we zoom out to consider all Expressions we observe that some (many) aren't assigned anything.
- Or as Tarski did ban the Liar Sentence in both ways (as a valid expression and WFF).

9 Conclusion

I have introduced a simplified Classic variant of Kripke-Feferman, provided a Consistency proof by Mathematical Induction, a finite decision procedure for determinately classifying Sentences (and extension to hypercomputing/infinitary contexts), and offered the first formal definition for Alethic Paradox along the way.

[60]The original paper talks about the implicit assumptions going into just what Propositional Variables and WFF are.

I further, discuss the implications of this solution to Semantic Closure, Buddhism, so-called Indian Logics, and Empiricism.

This is important since everyone cares about Truth (Philosophers, Logicians, Scientists, Religions, etc.)!

The aim is to put the language science of truth on a better footing (like the rest of Lingusitics) and to help in many domains (NLP, Logic, Philosophy).

It's the first Classically Consistent view and one of the very first Restrictionist views opening a new path to reply to Non-Classical critics.

[61]

References

[1] Beall, J. A Neglected Deflationist Approach to the Liar. *Analysis*. **61**, 126-129 (2001)

[2] Beall, J. Prolegomenon to Future Revenge. *Revenge Of The Liar: New Essays On The Paradox*. pp. 1-30 (2007)

[3] Colyvan, M. Vagueness and Truth. *From Truth To Reality*. pp. 29-40 (2009)

[4] Feferman, S. Axioms for Determinateness and Truth. *Review Of Symbolic Logic*. **1**, 204-217 (2008)

[5] Field, H. Saving Truth From Paradox. (Oxford University Press,2008)

[6] Field, H. Solving the Paradoxes, Escaping Revenge. *Revenge Of The Liar: New Essays On The Paradox*. (2007)

[7] Halbach, V. How Not to State T-Sentences. *Analysis*. **66**, 276-280 (2006)

[8] Horwich, P. Truth. (Clarendon Press,1990)

[9] Kripke, S. Outline of a Theory of Truth. *Journal Of Philosophy*. **72**, 690-716 (1975)

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

[11] Priest, G. Logic of Paradox. *Journal Of Philosophical Logic*. **8**, 219-241 (1979)

[12] Quine, W. The Pursuit of Truth, 1St Ed.. (Mass.: Harvard University Press,1990)

[13] Read, S. Further Thoughts on Tarski's T-scheme and the Liar. *Unity, Truth, And The Liar: The Modern Relevance Of Medieval Solutions To The Liar Paradox*. pp. 205-225 (2008),

[61] TODO - fill this in, make sure the bib items are/remain relevant, and verify everything

[14] Rossi, L. A Unified Theory of Truth and Paradox. *Review Of Symbolic Logic*. **12**, 209-254 (2019)

[15] Tarski, A. The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy And Phenomenological Research*. **4**, 341-376 (1943)

[16] Yablo, S. Paradox Without Self-Reference. *Analysis*. **53**, 251-252 (1993)

[17] Bacon, A. Can the Classical Logician Avoid the Revenge Paradoxes?:. *Philosophical Review*. **124** pp. 299-352 (2015,7)