The Central Theorems of Kripke's Theory of Truth

We start with an interpreted language $\mathfrak{L}$, and we add a new predicate "Tr," which is supposed to represent the set of true sentences of the enlarged language. We need to determine how the new predicate is to behave. In classical semantics, without truth-value gaps, this means assigning a set E as the extension of "Tr," so that, if the closed term τ denotes something in E, $\mathrm{Tr}(\tau)$ will be true, whereas, if τ denotes something outside E, $\mathrm{Tr}(\tau)$ will be false and $\neg\mathrm{Tr}(\tau)$ will be true. Because of the Liar Paradox, it isn't possible to pick the set E so that E = the set of sentences that are made true by assigning E as the extension of "Tr." In other words, we can't arrange things so that $\mathrm{Tr}([^{+}S'])$ is true or false according as S is true or false.

If we allow truth-value gaps, we shall need to assign to "Tr" an *extension* E, consisting of things that definitely satisfy the predicate, and an *antiextension* A, consisting of things that are definitely fail to satisfy it. E and A cannot overlap, but there may be things that aren't in either of them. If $\mathrm{Den}(\tau)$ is in E, $\mathrm{Tr}(\tau)$ will be true, whereas if $\mathrm{Den}(\tau)$ is in A, $\mathrm{Tr}(\tau)$ will be false. If $\mathrm{Den}(\tau)$ is in neither E nor A, $\mathrm{Tr}(\tau)$ will be neither true nor false. The central theorem underlying Kripke's theory of truth (which was proven also by Robert Martin and Peter Woodruff) tells is that we can find a pair of nonoverlapping sets E and A with the following properties:

> E = the set of sentences that are made true by setting the extension of "Tr" equal to E and setting the antiextension equal to A.
>
> A = the union of the set of nonsentences with the set of sentences that are made false by setting the extension of "Tr" equal to E and the antiextension equal to A.

In other words, $\mathrm{Tr}([^{+}S'])$ will be true, false, or unsettled according as S is true, false, or unsettled. We have thus arranged things so that $\mathrm{Tr}([^{+}S'])$ and S always have the same semantic status, even though we haven't arranged things so that sentences of the form

$$\mathrm{Tr}([^{+}S']) \leftrightarrow S$$

are all true.

We'll prove the theorem for $\mathfrak{L}$ the language of arithmetic, but we're using the language of arithmetic as a stand-in for a vast range of other languages we might equally well have used. In particular, the language of arithmetic lacks truth-value gaps, so that any truth value gaps we encounter occur because of the indeterminacy of "Tr." It doesn't have to be this way. The construction works just as well if the base language has truth-value gaps.

I. The Classical Situation

The language $\mathfrak{L}^{Tr}$ is obtained from the language of arithmetic $\mathfrak{L}$ by adjoining the new predicate "Tr." Thus the *terms* of $\mathfrak{L}^{Tr}$ are specified by the following stipulation:

> "0" is a term.
> Each of the variables "x," "x'," "x"," "x'"," "x""," "x""'," and so on, is a term.

[These are the variables for official purposes. In practices, the notation is cumbersome, so I'll ask other letters from the end of the alphabet, with or without numerical subscripts, to play the role of variables.]

> If τ is a term, $S\tau$ is a term.
> If τ and ρ are terms, $(\tau + \rho)$ is a term.
> If τ and ρ are terms $(\tau \times \rho)$ is a term.
> Nothing is a term unless it is required to be by the five clauses above.

A *closed term* is a term with no variables. The *denotation* of closed terms is defined inductively, as follows (for any closed terms τ and ρ):

> Den("0") = 0.
> Den($S\tau$) = Den(τ) + 1
> Den((τ + ρ)) = Den(τ) + Den(ρ).
> Den(($\tau \times \rho$)) = Den(τ) • Den(ρ).

The *formulas* of $\mathfrak{L}^{Tr}$ are specified as follows:

> If τ and ρ are terms, then
>
> $\tau = \rho$
>
> $\tau < \rho$
>
> and
>
> Tr(τ)
>
> are formulas.
>
> If P and Q are formulas, so are
>
> $(P \wedge Q)$

$\neg$ P

and

$(\forall v)$P,

for each variable v.

Nothing is a formula unless it's required to be by the clauses above.

In the past, we've included other symbols in what we were calling the language of arithmetic: the disjunction sign "$\vee$," the conditional "$\rightarrow$," the biconditional "$\leftrightarrow$," and the existential quantifier "$\exists$." We're leaving them out here, but it's no real loss, since we can treat them as defined:

$(P \vee Q) =_{Def} \neg(\neg P \wedge \neg Q)$

$(P \rightarrow Q) =_{Def} \neg(P \wedge \neg Q)$

$(P \leftrightarrow Q) =_{Def} (\neg(P \wedge \neg Q) \wedge \neg(\neg P \wedge Q))$

$(\exists v)P =_{Def} \neg(\forall v)\neg P$

The reason for leaving them out of the official language is that the proofs we are about to give are very long and tedious, leaving out unneeded symbols makes the proofs not quite as bad, by avoiding repetitions.

An occurrence of a variable v in a formula is *bound* if it occurs within some subformula that begins with "$(\forall v)$." (We count a formula as a subformula of itself.) If not bound, *free*. A formula with no free variables is a *sentence*. It is a sentence that can say something that is either true or false.

Let $\mathbb{N}$ be the standard model of $\mathfrak{L}$, the model whose domain is the natural numbers, in which "0" denotes the number 0 and "+" stands for addition and "×" for multiplication. We extend $\mathbb{N}$ to a *classical model* $(\mathbb{N}, E)$ of $\mathfrak{L}^{Tr}$ by designating a set E of natural numbers to serve as the extension of "Tr." We assume that we have fixed a Gödel numbering, so that we'll talk about sentences being members of E. We specify the conditions under which a sentence is *true* in $(\mathbb{N},E)$, as follows:

$\tau = \rho$ is true in $(\mathbb{N},E)$ iff $Den(\tau) = Den(\rho)$.

$\tau < \rho$ is true in $(\mathbb{N},E)$ iff $Den(\tau) < Den(\rho)$.

$Tr(\tau)$ is true in $(\mathbb{N},E)$ iff $Den(\tau) \in E$.

$(P \wedge Q)$ is true in $(\mathbb{N},E)$ iff P and Q are both true in $(\mathbb{N},E)$.

$\neg P$ is true in $(\mathbb{N},E)$ iff P isn't true in $(\mathbb{N},E)$.

$(\forall v)P(v)$ is true in $(\mathbb{N},E)$ iff, for each closed term τ, $P(\tau)$ is true in $(\mathbb{N},E)$.

Here "$P(\tau)$" means the result of substituting τ for each free occurrence of "v" in "P(v)."

A sentence is said to be *false* in $(\mathbb{N},E)$ iff it's negation is true in $(\mathbb{N},E)$. Consequently,

$\tau = \rho$ is false in $(\mathbb{N},E)$ iff $Den(\tau) \neq Den(\rho)$.

$\tau < \rho$ is false in $(\mathbb{N},E)$ iff $Den(\tau) \not< Den(\rho)$.

$Tr(\tau)$ if false in $(\mathbb{N},E)$ iff $Den(\tau) \notin E$.

$(P \wedge Q)$ is false in $(\mathbb{N},E)$ iff either P or Q is false in $(\mathbb{N},E)$.

$\neg P$ is false in $(\mathbb{N},E)$ iff P isn't false in $(\mathbb{N},E)$ iff P is true in $(\mathbb{N},E)$.

$(\forall v)P(v)$ is false in $(\mathbb{N},E)$ iff, for some closed term τ, $P(\tau)$ is false in $(\mathbb{N},E)$.

The defined terms "$\vee$," "$\rightarrow$," "$\leftrightarrow$," and "$\exists$" behave just the way you'd expect:

A disjunction $(P \vee Q)$ is true in $(\mathbb{N},E)$ iff one or both disjuncts are true in $(\mathbb{N},E)$.

A disjunction is false in $(\mathbb{N},E)$ iff both its disjuncts are false in $(\mathbb{N},E)$.

A conditional $(P \rightarrow Q)$ is true in $(\mathbb{N},E)$ iff its antecedent P is false in $(\mathbb{N},E)$ or its consequent Q is trui in $(\mathbb{N},E)$.

A conditional is false in $(\mathbb{N},E)$ iff its antecedent is true in $(\mathbb{N},E)$ and its consequent is false in $(\mathbb{N},E)$.

A biconditional $(P \leftrightarrow Q)$ is true in $(\mathbb{N},E)$ iff its components are either both true or both false in $(\mathbb{N},E)$.

A biconditional $(P \leftrightarrow Q)$ is false in $(\mathbb{N},E)$ iff one of its components is true in $(\mathbb{N},E)$, and the other is false in $(\mathbb{N},E)$.

An existential sentence $(\exists v)P(v)$ is true in $(\mathbb{N},E)$ iff, for some closed term τ, $P(\tau)$ is true in $(\mathbb{N},E)$.

$(\exists v)P(v)$ is false in $(\mathbb{N},E)$ iff, for every closed term τ, $P(\tau)$ is false in $(\mathbb{N},E)$.

We want to see that, whatever set E we pick as the extension of "Tr," we'll never have

$E = \{\text{sentences true in } (\mathbb{N},E)\}$.

We'll use the Gödel Self-Referential Lemma. In its general form, the lemma tells us that, for any formula $Q(v_0,v_1,v_2,...,v_n)$, there is a formula $P(v_1,v_2,...,v_n)$ so that the biconditional

$$(\forall v_1)(\forall v_2)...(\forall v_n)(Q([^+P(v_1,v_2,...,v_n)'],v_1,v_2,...,v_n) \leftrightarrow P(v_1,v_2,...,v_n))$$

is a theorem of Peano Arithmetic (PA). In case n = 0, this tells us that, for any formula Q(v), we can find a sentence P so that

$$(Q([^+P']) \leftrightarrow P)$$

is a theorem of PA. Thus we can find a sentence L so that

$$(\neg Tr([^+L']) \leftrightarrow L)$$

is a theorem of PA. L is the formal equivalent of Eubulides' "This statement is not true." For any set of numbers E, $(\neg Tr([^+L']) \leftrightarrow L)$ is true in $(\mathbb{N},E)$. Consequently,

$L \in \{\text{sentences true in } (\mathbb{N},E)\}$
iff L is true in $(\mathbb{N},E)$
iff $\neg Tr([^+L'])$ is true in $(\mathbb{N},E)$
iff $Tr([^+L'])$ isn't true in $(\mathbb{N},E)$
iff $Den([^+L']) \notin E$
iff $L \notin E$

So we cannot have $E = \{\text{sentences true in } (\mathbb{N},E)\}$.

II. Truth-value Gaps

We get truth-value gaps when we allow "Tr" to be partially undefined. We assign to "Tr" a pair (E,A), with $E \cap A = \varnothing$. The extension E is intended to consist of things that are definitely true, while the antiextension A contains things that are definitely not true. We stipulate what it is for a sentence to be true or false in $(\mathbb{N},(E,A))$:

$\tau = \rho$ is true in $(\mathbb{N},(E,A))$ iff $Den(\tau) = Den(\rho)$.
$\tau = \rho$ is false in $(\mathbb{N},(E,A))$ iff $Den(\tau) \neq Den(\rho)$.

$\tau < \rho$ is true in $(\mathbb{N},(E,A))$ iff $Den(\tau) < Den(\rho)$.
$\tau < \rho$ is false in $(\mathbb{N},(E,A))$ iff $Den(\tau) \not< Den(\rho)$.
$Tr(\tau)$ is true in $(\mathbb{N},(E,A))$ iff $Den(\tau) \in E$.
$Tr(\tau)$ is false in $(\mathbb{N},(E,A))$ iff $Den(\tau) \in A$.
$(P \wedge Q)$ is true in $(\mathbb{N},(E,A))$ iff P and Q are both true in $(\mathbb{N},(E,A))$.
$(P \wedge Q)$ is false in $(\mathbb{N},(E,A))$ iff one or both of P and Q are false in $(\mathbb{N},(E,A))$.
$\neg P$ is true in $(\mathbb{N},(E,A))$ iff P is false in $(\mathbb{N},(E,A))$.
$\neg P$ is false in $(\mathbb{N},(E,A))$ iff P is true in $(\mathbb{N},(E,A))$.
$(\forall v)P(v)$ is true in $(\mathbb{N},(E,A))$ iff, for each closed term τ, $P(\tau)$ is true in $(\mathbb{N},(E,A))$.
$(\forall v)P(v)$ is false in $(\mathbb{N},(E,A))$ iff, for some closed term τ, $P(\tau)$ is false in $(\mathbb{N},(E,A))$.

In the presence of truth-value gaps, the Gödel Self-Referential Lemma takes a slightly different form. Given a formula $Q(v_0,v_1,v_2,...,v_n)$, there is a formula $P(v_1,v_2,...,v_n)$ such that, for any closed terms τ_1, τ_2,..., τ_n and any nonoverlapping sets E and A, $Q(\ulcorner {}^{+}P(v_1,v_2,...,v_n)' \urcorner,\tau_1,\tau_2,...,\tau_n)$ is true, false, or unsettled under $(\mathbb{N},(E,A))$ according as $P(\tau_1,\tau_2,...,\tau_n)$ is true, false, or unsettled under $(\mathbb{N},(E,A))$.

III. The Smallest Fixed Point

A *fixed point* is a pair (E,A) with

E = {sentences true in $(\mathbb{N},(E,A))$}
A = {nonsentences} $\cup$ {sentences false in $(\mathbb{N},(E,A))$}

We intend to show that there is a fixed point (E_∞,A_∞). (E_∞,A_∞) is the *smallest* fixed point, in the following sense: If (E,A) is another fixed point, we have $E_\infty \subseteq E$ and $A_\infty \subseteq A$.

If you look at Kripke's paper, you will see that he constructs E_∞ and A_∞ simultaneously. The argument he gives uses infinite ordinal numbers, and we haven't talked about infinite ordinal numbers, so I'd like to give a more direct, but also more laborious, argument here. It will simplify matters if we begin by constructing E_∞, without worrying about A_∞. We construct E_∞ by including within it all the sentences that have to be true, without fretting over how to make A_∞ include all the sentences that have to be false. Once we're done, we can obtain A_∞ as {nonsentences} $\cup$ {sentences S: $\neg$ S $\in E_\infty$}. Constructing E_∞ by itself simplifies matters, but it doesn't simplify matters very much, since falsity conditions are just truth conditions for negations. Thus, instead of having to deal separately with the falsity conditions for atomic sentences, conjunctions, negations, and universal sentences, we deal with the truth conditions for negated atomic sentences, negated conjunctions, negated negations, and negated universal sentences, in addition to having to deal with the truth conditions for atomic sentences, conjunctions, and universal sentences. So the definition of E_∞ is still unpleasantly complicated.

E_∞ is the smallest collection of sentences that satisfies the following twelve conditions:

i) If $Den(\tau) = Den(\rho)$, then $\tau = \rho$ is in E_∞.
ii) If $Den(\tau) \neq Den(\rho)$, then $\neg\, \tau = \rho$ is in E_∞.
iii) If $Den(\tau) < Den(\rho)$, then $\tau < \rho$ is in E_∞.
iv) If $Den(\tau) \not< Den(\rho)$, then $\neg\, \tau < \rho$ is in E_∞.
v) If $Den(\tau) \in E_\infty$, then $Tr(\tau)$ is in E_∞.
vi) If $Den(\tau)$ isn't a sentence, then $\neg Tr(\tau)$ is in E_∞.
vii) If $Den(\tau)$ is a sentence whose negation is in E_∞, then $\neg\, Tr(\tau)$ is in E_∞.
viii) If P and Q are in E_∞, then $(P \wedge Q)$ is in E_∞.
ix) If either $\neg P$ or $\neg Q$ is in E_∞, $\neg(P \wedge Q)$ is in E_∞.
x) If P is in E_∞, then $\neg\, \neg\, P$ is in E_∞.
xi) If $F(\tau)$ is in E_∞, for each closed term τ, then $(\forall v)F(v)$ is in E_∞.
xii) If $\neg F(\tau)$ is in E_∞, for some closed term τ, then $\neg(\forall v)F(v)$ is in E_∞.

When I say that E_∞ is the "smallest" set that meets that twelve conditions, I mean that E_∞ is a set that meets the twelve conditions that is included in every other set that meets the twelve conditions. To see that there have to be such a set, notice that there is a set that meets the twelve conditions, namely, the set of all sentences. Moreover, the intersection of all sets of sentences that meet the twelve conditions meets the twelve conditions. That intersection is E_∞.

The definition is set us in such a way that any sentence that is in E_∞ has been put into E_∞ by one of the twelve conditions. Moreover, for any sentence in E_∞, there is only one of the twelve conditions that could have put it there, and you can tell which condition that is by the syntactic form of the sentence.

Let A_∞ = {nonsentences} $\cup$ {sentences S: $\neg S \in E_\infty$}. Our first task is to show that E_∞ and A_∞ are disjoint. We'll do this by showing that $(E_\infty \sim A_\infty)$ satisfies the twelve conditions. Since E_∞ is the smallest set that meets the twelve conditions, if $(E_\infty \sim A_\infty)$ is also a set that meets the twelve conditions, then E_∞ must be included in $(E_\infty \sim A_\infty)$, which means that no member of E_∞ is in A_∞. So we have to verify the twelve conditions.

i) If $Den(\tau) = Den(\rho)$, then clause i) puts $\tau = \rho$ in E_∞, and none of the clauses puts $\neg\, \tau = \rho$ into E_∞. So $\tau = \rho$ isn't in A_∞.

ii) If $Den(\tau) \neq Den(\rho)$, then clause ii) puts $\neg\, \tau = \rho$ into E_∞. None of the clauses puts $\tau = \rho$ into E_∞. The only way $\neg\, \neg\, \tau = \rho$ could be in E_∞ would be for it to be put there by clause x), and clause x) would only put $\neg\, \neg\, \tau = \rho$ into E_∞ if $\tau = \rho$ were in E_∞. So $\neg\, \neg\, \tau = \rho$ isn't in E_∞, and $\neg\, \tau = \rho$ isn't in A_∞.

iii) Similar to i).

iv) Similar to ii).

v) Suppose that Den(τ) is in $E_\infty \sim A_\infty$. Then, because E_∞ satisfies clause v), Tr(τ) is in E_∞. Also, because Den(τ) is in E_∞, Den(τ) must be a sentence, and so $\neg$Tr(τ) isn't put into E_∞ by clause vi). The negation of Den(τ) isn't in E_∞, and so $\neg$Tr(τ) wasn't put into E_∞ by clause vii). So $\neg$Tr(τ) isn't in E_∞ at all; hence Tr(τ) isn't in A_∞.

vi) If Den(τ) isn't a sentence, then $\neg$Tr(τ) is in E_∞, because E_∞ satisfies clause vi). Tr(τ) isn't in E_∞, because it isn't put into E_∞ by clause v). So $\neg \neg$ Tr(τ) isn't put into E_∞ by clause viii). So $\neg \neg$ Tr(τ) isn't in E_∞, and $\neg$Tr(τ) isn't in A_∞.

vii) If Den(τ) is a sentence whose negation is in $E_\infty \sim A_\infty$, then $\neg$Tr(τ) is put into E_∞ by clause vii). Because the negation of Den(τ) isn't in A_∞, the negation of the negation of Den(τ) isn't in E_∞. So Den(τ) isn't in E_∞. So clause v) doesn't put Tr(τ) into E_∞. So Tr(τ) isn't in E_∞, which means that clause x) doesn't put $\neg \neg$ Tr(τ) into E_∞. So $\neg$Tr(τ) isn't in A_∞.

viii) If P and Q are both in $E_\infty \sim A_\infty$, then (P $\wedge$ Q) is in E_∞. Neither $\neg$P nor $\neg$Q is in A_∞, so clause ix) doesn't put $\neg$(P $\wedge$ Q) in E_∞. So (P $\wedge$ Q) isn't in A_∞.

ix) If $\neg$P is in $E_\infty \sim A_\infty$, then $\neg$(P $\wedge$ Q) is in E_∞. Because $\neg$P isn't in A_∞, $\neg \neg$ P isn't in E_∞. So P isn't in E_∞. So clause viii) doesn't put (P $\wedge$ Q) into E_∞. So clause x) doesn't put $\neg \neg$(P $\wedge$ Q) into E_∞. So $\neg$(P $\wedge$ Q) isn't in A_∞. The argument that, if $\neg$Q is in $E_\infty \sim A_\infty$, then $\neg$(P $\wedge$ Q) is in $E_\infty \sim A_\infty$ is similar.

x) If P is in $E_\infty \sim A_\infty$, then $\neg \neg$ P is in E_∞, because E_∞ satisfies clause x). Because $\neg$P isn't in E_∞, $\neg \neg \neg$P isn't in E_∞. So $\neg \neg$ P isn't in A_∞.

xi) Suppose that, for each closed term τ, F(τ) is in $E_\infty \sim A_\infty$. Then $(\forall v)$F(v) is in E_∞. There is no closed term τ such that $\neg$F(τ) is in E_∞. So clause xii) didn't put $\neg(\forall v)$F(v) into E_∞. So $\neg(\forall v)$F(v) isn't in E_∞, and $(\forall v)$F(v) isn't in A_∞.

xii) Suppose that, for some closed term τ, $\neg$F(τ) is in $E_\infty \sim A_\infty$. Because E_∞ satisfies clause xii), $\neg(\forall v)$F(v) must be in E_∞. Since $\neg$F(τ) isn't in A_∞,$\neg \neg$F(τ) must not be in E_∞. So F(τ) must not be in E_∞. This means that clause xi) can't put $(\forall v)$F(v) into E_∞, and so clause x) can't put $\neg \neg (\forall v)$F(v) into E_∞. So $\neg(\forall v)$F(v) isn't in A_∞.

This shows that E_∞ and A_∞ are disjoint. We now need to show that (E_∞,A_∞) is a fixed point. We do this by proving, by induction on the complexity of P, that, for any sentence P, we have:

$P \in E_\infty$ iff P is true in $(\mathbb{N},(E_\infty,A_\infty))$
$P \in A_\infty$ iff P is false in $(\mathbb{N},(E_\infty,A_\infty))$

That is, we assume, as inductive hypothesis, that these conditions hold for all sentences simpler than P, and show that they hold for P. There are six cases:

Case 1. P has the form $\tau = \rho$. Then

$P \in E_\infty$
iff P is put into E_∞ by clause i)
iff $Den(\tau) = Den(\rho)$
iff $\tau = \rho$ is true in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is true in $(\mathbb{N},(E_\infty,A_\infty))$.

This verifies the first clause. For the second, we have:

$P \in A_\infty$
iff $\neg P \in E_\infty$
iff $\neg P$ is put into E_∞ by clause ii)
iff $Den(\tau) \neq Den(\rho)$
iff $\tau = \rho$ is false in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is false in $(\mathbb{N},(E_\infty,A_\infty))$

Case 2. P has the form $\tau < \rho$. Similar.

Case 3. P has the form $Tr(\tau)$. We have

$P \in E_\infty$
iff P is put into E_∞ by clause ii)
iff $Den(\tau)$ is in E_∞
iff $Tr(\tau)$ is true in $(\mathbb{N}, (E_\infty,A_\infty))$
iff P is true in $(\mathbb{N},(E_\infty,A_\infty))$

We get the second clause as follows:

$P \in A_\infty$
iff $\neg P \in E_\infty$
iff $\neg P$ is put into E_∞ either by clause vi) or clause vii)
iff either $Den(\tau)$ isn't a sentence or $Den(\tau)$ is a sentence whose negation is in E_∞
iff $Den(\tau) \in \{\text{nonsentences}\} \cup \{\text{sentences S: } \neg S \in E_\infty\}$
iff $Den(\tau) \in A_\infty$
iff $Tr(\tau)$ is false in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is false in $(\mathbb{N},(E_\infty,A_\infty))$

Case 4. P has the form $(Q \wedge R)$. We have:

$P \in E_\infty$
iff P is put into E_∞ by clause viii)
iff Q and R are both in E_∞
iff Q and R are both true in $(\mathbb{N},(E_\infty,A_\infty))$ (by inductive hypothesis)
iff $(Q \wedge R)$ is true in $(\mathbb{N},(E_\infty,A_\infty))$

iff P is true in $(\mathbb{N},(E_\infty,A_\infty))$

For falsity, the argument goes like this:

$P \in A_\infty$
iff $\neg P \in E_\infty$
iff $\neg P$ is put into E_∞ by clause ix)
iff either $\neg Q$ or $\neg R$ is in E_∞
iff either Q or R is false in $(\mathbb{N},(E_\infty,A_\infty))$ (by inductive hypothesis)
iff $(Q \wedge R)$ is false in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is false in $(\mathbb{N},(E_\infty,A_\infty))$

Case 5. P has the form $\neg Q$. For truth, the argument is as follows:

$P \in E_\infty$
iff $Q \in A_\infty$
iff Q is false in $(\mathbb{N},(E_\infty,A_\infty))$ (by inductive hypothesis)
iff $\neg Q$ is true in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is true in $(\mathbb{N},(E_\infty,A_\infty))$

Here's the story for falsity:

$P \in A_\infty$
iff $\neg P \in E_\infty$
iff $\neg\ \neg Q \in E_\infty$
iff $\neg\ \neg$ Q is placed into E_∞ by clause x)
iff $Q \in E_\infty$
iff Q is true in $(\mathbb{N},(E_\infty,A_\infty))$ (by inductive hypothesis)
iff $\neg Q$ is false in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is false in $(\mathbb{N},(E_\infty,A_\infty))$

Case 6. P has the form $(\forall v)F(v)$.

$P \in E_\infty$
iff clause xi) puts P into E_∞
iff, for every closed term τ, $F(\tau)$ is in E_∞
iff, for every closed term τ, $F(\tau)$ is true in $(\mathbb{N},(E_\infty,A_\infty))$ (by inductive hypothesis)
iff $(\forall v)F(v)$ is true in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is true in $(\mathbb{N}.(E_\infty,A_\infty))$

For falsity,

$P \in A_\infty$
iff $\neg P \in E_\infty$
iff $\neg P$ is put into E_∞ by clause xii)
iff, for some closed term τ, $\neg F(\tau)$ is in E_∞
iff, for some closed term τ, $F(\tau)$ is in A_∞
iff, for some closed term τ, $F(\tau)$ is false in $(\mathbb{N},(E_\infty,A_\infty))$ (by inductive hypothesis)
iff, $(\forall v)F(v)$ is false in $(\mathbb{N},(E_\infty,A_\infty))$
iff P is false in $(\mathbb{N},(E_\infty,A_\infty))$

This shows that (E_∞,A_∞) is a fixed point. It is easy to check that, if (E,A) is another fixed point, E satisfies conditions i) - xii). Consequently, $E_\infty \subseteq E$, and (E_∞,A_∞) is the *smallest* fixed point.

IV. Other Fixed Points

There are many fixed points other than the smallest. To ger other fixed points, we use the following:

> **Theorem**. Given nonoverlapping sets D_0 and B_0, with $D_0 \subseteq$ {sentences true in $(\mathbb{N},(D_0,B_0))$} and $B_0 \subseteq$ {nonsentences} $\cup$ {sentences false in $(\mathbb{N},(D_0,B_0))$}. There exists a fixed point (D_∞,B_∞) with $D_0 \subseteq D_\infty$ and $B_0 \subseteq B_\infty$.

The pair (D_∞,B_∞) we construct will, in fact, be the smallest fixed point that extends (D_0,B_0). That is, it is a fixed point extending (D_0,B_0) that is included in every other such fixed point.

The construction of (D_∞,B_∞) will resemble the construction of (E_∞,A_∞), with the small changes required to ensure that D_0 is in the extension and B_0 in the antiextension. D_∞ will be the smallest set of sentences that meets the following eleven conditions:

a) If $Den(\tau) = Den(\rho)$, then $\tau = \rho \in D_\infty$.
b) If $Den(\tau) \neq Den(\rho)$, then $\neg\, \tau = \rho \in D_\infty$.
c) If $Den(\tau) < Den(\rho)$, then $\tau < \rho \in D_\infty$.
d) If $Den(\tau) \not< Den(\rho)$, then $\neg\, \tau < \rho \in D_\infty$.
e) If $Den(\tau) \in D_\infty$ or $Den(\tau) \in D_0$, then $Tr(\tau) \in D_\infty$.
f) If $Den(\tau)$ isn't a sentence or $Den(\tau)$ is a sentence whose negation is in D_∞ or $Den(\tau) \in B_0$, then $\neg Tr(\tau) \in D_\infty$.
g) If P and Q are in D_∞, $(P \wedge Q)$ is in D_∞.
h) If either $\neg P$ or $\neg Q$ is in D_∞, $\neg(P \wedge Q)$ is in D_∞.
i) If P is in D_∞, so is $\neg\, \neg\, P$.
j) If, for every closed term τ, $F(\tau)$ is in D_∞, $(\forall v)F(v)$ is in D_∞;
k) If, for some closed term τ, $\neg F(\tau)$ is in D_∞, $\neg(\forall v)F(v)$ is in D_∞.

B_∞ is defined to be {nonsentences} $\cup$ {sentences S: $\neg S \in D_\infty$}.

The proof that (D_∞,B_∞) is a fixed point is long and tedious, and it doesn't involve any new ideas not already found in the proof that (E_∞,A_∞) is a fixed point, so I won't give it here. Here I'll give only a sketch. The hard part will be to show that D_∞ and B_∞ are disjoint. The proof is tricky, just because, if you don't prove things in just the right order, the proof gets bogged down.

> **Lemma 1.** All the sentences true in $(\mathbb{N},(D_0,B_0))$ are in D_∞, and all the sentences false in $(\mathbb{N},(D_0,B_0))$ are in B_∞.

The proof is by induction on the complexity of sentences. Once we have Lemma 1, we'll know that $D_0 \subseteq D_\infty$ and $B_0 \subseteq B_\infty$.

> **Lemma 2.** No member of D_∞ is false in $(\mathbb{N},(D_0,B_0))$.

We show this by proving that $D_\infty \sim$ {sentences false in $(\mathbb{N},(D_0,B_0))$} satisfies conditions a) through k). We use both Lemma 1 and Lemma 2 to prove:

> **Lemma 3.** D_∞ and B_∞ are disjoint.

The proof of Lemma 3 consists in showing that $D_\infty \sim B_\infty$ satisfies conditions a) through k). Once we have Lemma 3, an induction on the complexity of sentences proves that, for any sentence P, we have:

> P $\in D_\infty$ iff P is true in $(\mathbb{N},(D_\infty,B_\infty))$
> P $\in B_\infty$ iff P is false in $(\mathbb{N},(D_\infty,B_\infty))$

Our smallest fixed point, (E_∞,A_∞), is what we get in the special case in which $D_0 = \varnothing$ and $B_0 = \varnothing$. There are other fixed points too. For example, take T to be a *Truthteller*, so that T with T semantically equivalent to $\mathrm{Tr}([^+T'])$. In the smallest fixed point, T is neither true nor false. However, in the fixed point obtained by setting $D_0 = \{T\}$ and $B_0 = \varnothing$, T will be true. In the fixed point obtained by setting $D_0 = \varnothing$ and $B_0 = \{T\}$, T is false.

Let T_1, T_2, and T_3 be three different Truthtellers. Then setting $D_0 = \{T_1)$ and $B_0 = \{T_2\}$ will give us a fixed point in which T_1 is true, T_2 false, and T_3 unsettled. On the other hand, if L is a Liar sentence – L is semantically equivalent to $\neg\mathrm{Tr}([^+L'])$ – L is unsettled in every fixed point.

How do we get our three Truthtellers? Let V(x,y) abbreviate the following formula:

> $(\exists z)$(z is the result of substituting [y] for free occurrences of "y" in the formula whose Gödel number is x $\wedge$ Tr(z))

Use the Self-Referential Lemma to find a formula U(y) that is semantically equivalent to $V([^+U(y)'],y)$, and let our three Truthtellers be U([1]), U([2]), and U([3]). This procedure gives us infinitely many Truthtellers. If R and S are subsets of our infinite set of Truthtellers with R $\cap$ S

= ∅, we can find a fixed point in which all the members of R are true, all the members of S are false, and the rest of our Truthtellers are all undecided.

V. The Classical Logic Version

The version of the Kripke construction we've been looking at so far does a good job at untangling complicated self-reference, but it has a hard time with simple generalizations. For example, the apparently harmless statement "$(\forall x)(Tr(x) \rightarrow Tr(x))$," which says, "Every true sentence is a true sentence," is shoved into the gap between truth and falsity.

Kripke gave an alternative construction that avoids this kind of problem. The new construction (which is based upon an idea of Bas van Fraassen) Treats a sentence P as true in $(\mathbb{N},(E,A))$ iff P is true in every classical model $(\mathbb{N},S)$ with $E \subseteq S \subseteq \mathbb{N} \sim A$. The proof that there are fixed points continues to work with this alternative way of handling truth-value gaps.. The alternative construction treats all the sentences regarded as valid by classical logic as true, and it treats every classical consequence of true sentences as true. It has the odd feature that a disjunction can be true without either disjunct being true An example is $(L \vee \neg L)$, where L is the Liar sentence. So we give up the very simple and natural picture of how the truth and falsity of a complex sentence is grounded in the truth or falsity of its simple components, but in return we get to count as true harmless generalizations like "All true sentences are true sentences," that aren't implicated in the Liar paradox.

VI. References

Kripke, Saul A. :"Outline of a Theory of Truth." *Journal of Philosophy* 72 (1975): 690-716. Reprinted in Martin, pp. 53-81.

McGee, Vann. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth.* Indianapolis: Hackett, 1991.

Martin, Robert L,, ed. *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press, 1984.

Martin, Robert L., and Peter W. Woodruff. "On Representing 'True-in-L' in L." *Philosophia* 5 (1975): 213-17. Reprinted in Martin, pp. 47-51.