A Theory of Truth that prefers falsehood

Melvin Fitting
mlflc@cunyvm.cuny.edu
Dept. Mathematics and Computer Science
Lehman College (CUNY), Bronx, NY 10468
Depts. Computer Science, Philosophy, Mathematics
Graduate Center (CUNY), 33 West 42nd Street, NYC, NY 10036 *

April 10, 1994

Abstract

We introduce a subclass of Kripke's fixed points in which falsehood is the preferred truth value. In all of these the truthteller evaluates to false, while the liar evaluates to undefined (or overdefined). The mathematical structure of this family of fixed points is investigated and is shown to have many nice features. It is noted that a similar class of fixed points, preferring truth, can also be studied. The notion of intrinsic is shown to relativize to these two subclasses. The mathematical ideas presented here originated in investigations of so-called *stable* models in the semantics of logic programming.

1 Introduction

Briefly stated, the job of a theory of truth is to assign truth values to sentences in a language allowing self-reference, in a way that respects intuition while avoiding paradox. Of course this can not be done in the framework of classical, two-valued logic because of liar sentences. Some generalization allowing *partial* truth assignments, or perhaps *contradictory* ones is needed. Kripke [13] (partly anticipated by [15]), provided a satisfactory mathematical mechanism using partiality. But for several reasons it did not eliminate the problem — primarily because it did not produce a unique candidate for a truth assignment, but rather a whole family of 'fixed points,' any of which is a good candidate. Some of these fixed points stand out as warranting special attention — most notably, the smallest one. But all of them are of interest, and any of them could serve as a truth assignment.

Given this multiplicity of possible truth assignments, it is of importance to find subclasses that are narrower in scope, ones that are characterized in natural ways. Kripke himself began this with his *intrinsic* fixed points. (We give proper definitions later, in order to keep things relatively self-contained.) In this paper we propose another natural subclass of interest (actually, we propose four related ones). It is a subclass having very nice mathematical properties, and is characterized by a condition having a creditable intuition. We put forward this subclass (or rather, these subclasses) for special attention.

*Research partly supported by NSF Grant CCR-9104015.

The liar sentence, asserting its own falsehood, acquires no classical truth value in any of Kripke's fixed points. This is as it should be — the sentence is paradoxical, after all. But the truth-teller, asserting its own truth, is true in some fixed points, false in some, and left without a classical truth value in still others. As a consequence, in both the smallest of Kripke's fixed points and in the biggest intrinsic fixed point the truth-teller lacks a classical truth value. In other words, in the most prominent of Kripke's fixed points the truth assignment treats the liar and the truth-teller alike, though for different reasons. The liar cannot be assigned a value; the truth-teller can be assigned anything, so a particular value is not determined. But the truth-teller is not contradictory, and there is no evidence to support its truth (other than itself, provided we assume its truth), so perhaps we should simply take it to be false.

We do not intend to argue in this paper for the correctness of the position sketched in the previous paragraph. We will say, however, that it is a plausible position that warrants investigation. In particular, what would the mathematics of such a position look like? This is what will be developed here: the theory of a subclass of Kripke's fixed points that, roughly, take falsehood as the preferred value whenever possible. This yields a class of fixed points having a very nice mathematical structure, with properties that are worth further investigation for this reason, as well as for their more 'philosophical' features.

The treatment just sketched can be dualized: one can take truth as the preferred value instead of falsehood. The resulting structure is essentially like that which results when falsehood is preferred, so we do not present the development explicitly. It is enough to mention the possibility, and leave its investigation to those who wish to chose truth over falsehood.

Finally, Kripke's notion of an intrinsic fixed point combines with either of the subclasses mentioned above. This is not the same thing as considering those members of the falsehood (or truthhood) preferring subclass that are intrinsic in Kripke's sense — rather, the definition itself is relativized. This yields two more natural subclasses of Kripke's fixed points, subclasses whose properties are largely open questions.

A few words about the origins of the development presented here. There is often a close relationship between theories of truth and the semantics of logic programming. Generally speaking, the same mathematical machinery is appropriate for both. Some years ago I showed how a development, much like Kripke's, could supply meanings for logic programs that made use of negation [4]. Other semantical approaches were developed as well, motivated by various concerns that need not be gone into here. One very successful such approach was the *stable model* semantics, due to Michael Gelfond and Vladimir Lifschitz [9], and independently to Kit Fine [3]. In [8] I gave an abstract, generalized treatment of this stable model semantics, using in part machinery developed by Stephen Yablo for application to the theory of truth [19]. The present paper can be seen as shifting the mechanism of stable model semantics for logic programming back to the philosophical arena, repaying the debt incurred to Kripke years ago. Since the term "stable" already has a meaning in philosophical logic, in work of Gupta and Herzberger [11, 12], the term *GLF-stable*, to credit Gelfond, Lifschitz, and Fine, will be used here.

The basic subject matters of this paper and of logic programming, as in [8], are quite different. Nonetheless, starting from different domains, the immediate goal is to define certain operators and prove certain monotonicity conditions for these operators. Once this is done the mathematical machinery to be applied is the same in either case, and is essentially algebraic in nature, deriving entirely from lattice theory. This should not be surprising since programming language semantics and natural language semantics both rely heavily on inductive definability notions. It is well-known, for instance, that Kripke's concept of intrinsic fixed point appeared independently in computer science [14]. The two communities have much to learn from each other.

2 Logic

Kripke carried out his development in [13] using three-valued logic — actually he considered several different versions: Kleene's weak, Kleene's strong, and supervaluations. Here we will use yet another, a four-valued logic credited to Dunn [2] and to Belnap [1]. This is a multiple-valued logic that contains Kleene's strong three-valued logic as a natural sub-logic (and in a certain sense Kleene's weak three-valued logic as well — see [7]). It has decided advantages. One is that, adding a fourth truth value (of *overdefined* or *inconsistent*) simplifies the mathematics by giving us a lattice, rather than a semi-lattice to work with. Another is that it adds a degree of symmetry between under-defined and over-defined that allows us to state and prove some striking theorems bounding the family of GLF-stable valuations. And finally, allowing overdefined as a truth value has a degree of naturalness to it. After all, a liar sentence can just as well be taken as having both classical truth values, as having neither. Other investigations, notably [17] and [18], have used the same logic. Still we hasten to point out that nothing critical depends on our choice of logic. We could have used Kleene's weak, or strong, three-valued logic instead. The situation with supervaluations remains to be investigated. In addition, any *distributive bilattice* could have been chosen (see [8]). The methodology is fairly robust.

Now for the details of the four-valued logic we have chosen to use. We denote the four values as *false*, *true*, $\bot$ and $\top$. Think of $\bot$ as undefined, or neither false nor true. Likewise think of $\top$ as overdefined, or both false and true. As Belnap observed, these values can be given two natural orders. One is on the *degree of information or knowledge* involved. In this ordering $\bot$ is least, strictly below both *false* and *true*, which are incomparable. These, in turn are strictly below $\top$. Intuitively, an increase in this ordering amounts to a gain of a classical truth value (possibly in addition to one already possessed), without loosing anything. We use $\leq_k$ for this ordering. In effect, when Kripke talks about a *least* fixed point for his truth revision operator, it is least with respect to this ordering.

In addition there is a second natural ordering of the four values, one in which an increase intuitively signifies an *increase in truth content, or a decrease in falsehood content*. We use $\leq_t$ for this ordering. The value *false* is least in this ordering. Moving from it to $\bot$ decreases the degree of falsehood (because $\bot$ is neither true nor false), while moving from *false* to $\top$ increases the degree of truth (because $\top$ is both true and false, while *false* is only false). So both $\bot$ and $\top$ are strictly above *false* in this ordering, though they are not comparable with each other. Finally, moving from $\bot$ to *true* increases the degree of truth, while moving from $\top$ to *true* decreases the degree of falseness, so both $\bot$ and $\top$ are below *true* in this sense.

Both orderings can be displayed simultaneously, in the double Hasse diagram of Figure 1. This way of displaying such things is due to Matt Ginsberg [10]; Belnap's logic is the simplest example of a *bilattice* in Ginsberg's sense.

Both orderings give $\mathcal{FOUR}$ the structure of a complete lattice. That is, all meets and joins exist, with respect to both orderings. We will use the following notation for meets and joins. For the $\leq_t$ ordering, binary meets are denoted $\wedge$, binary joins are denoted $\vee$. Arbitrary meets and joins are denoted $\bigwedge$ and $\bigvee$. It is important to note that, restricted to the substructure $\{\mathit{false}, \mathit{true}\}$, the operations $\wedge$ and $\vee$ are the usual conjunction and disjunction of classical logic. Likewise, restricted to $\{\mathit{false}, \bot, \mathit{true}\}$ they are the operations of Kleene's strong three-valued logic. Thus working with Belnap's system loses nothing that working with Kleene's would have obtained for us. Incidentally, the relationship between Kleene's and Belnap's logics is a special case of a general phenomenon — see [7] for a discussion of this point.

With respect to the $\leq_k$ ordering, we use $\otimes$ and $\oplus$ for binary meet and join, and $\prod$ and $\sum$

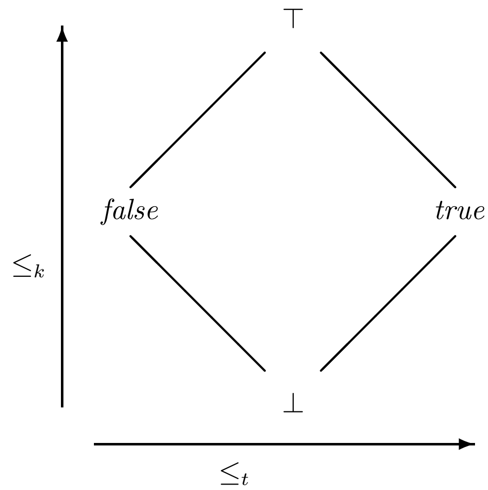

Figure 1: The Logic $\mathcal{FOUR}$

for the arbitrary versions. These operations are less familiar, doubtless. I have been calling $\otimes$ a *consensus* operation, because it produces the most information that two truth values can agree on. Likewise I have been calling $\oplus$ a *gullibility* operation — it accepts anything it's told.

The two sets of operations are not independent of each other: *all possible distributive laws hold.* For instance:

$$\begin{array}{rcl} x \land (y \lor z) & = & (x \land y) \lor (x \land z) \\ x \land (y \otimes z) & = & (x \land y) \otimes (x \land z) \\ x \oplus (y \otimes z) & = & (x \oplus y) \otimes (x \oplus z) \end{array}$$

and so on. All these play a significant role later on.

It is an easy consequence of the various distributive laws that *each of the operations is monotonic with respect to both orderings.* For example,

$$\begin{array}{rcl} x \leq_t y & \Rightarrow & x \otimes z \leq_t y \otimes z \\ x \leq_t y & \Rightarrow & x \land z \leq_t y \land z \\ x \leq_k y & \Rightarrow & x \land z \leq_k y \land z \end{array}$$

and so on. These are called the *interlacing conditions*. These too play a significant role below.

Finally, there is a natural negation operation: $\neg$*false* $=$ *true*; $\neg$*true* $=$ *false*; $\neg\top = \top$; and $\neg\bot = \bot$. This is a natural generalization of both classical negation and the negation of Kleene's logic. Negation reverses the $\leq_t$ ordering, while preserving the $\leq_k$ ordering. It is an easy consequence that the usual DeMorgan laws hold with respect to the operations $\land$ and $\lor$, but the operations $\otimes$ and $\oplus$ are self-dual.

Things can be summed up generally by saying $\mathcal{FOUR}$ is extraordinarily well-behaved. The rest is details.

3 Language and valuations

We need a language in which self-reference is possible. For this purpose we use Gödel numbering, though other mechanisms are possible. For the rest of this paper L is a first-order language with a one-place relation symbol T, intended to represent the truth predicate; 0 is a constant symbol; s is a one-place function symbol intended to represent successor (so we have representatives for the natural numbers); d^1, d^2, $d^3, \ldots$ are function symbols, with d^k being $k+1$-place (these are intended to represent substitution functions — more on this below); and otherwise, there can be constant, function, and relation symbols intended to represent various 'real world' things. This part is rather arbitrary. Propositional connectives are limited to $\wedge$, $\vee$, and $\neg$. Both existential and universal quantifiers are assumed present.

Since we have 0 and a successor function symbol in L, there is a numeral for every natural number. We denote the numeral representing n by $\overline{n}$. That is, $\overline{n} = s^n(0)$. Next, we assume that a Gödel numbering has been introduced, so that every sentence of the language has a number. We write $\ulcorner X \urcorner$ for the Gödel number of the sentence X. We assume the Gödel numbering is onto, so that for every number n there is some sentence X such that $\ulcorner X \urcorner = n$. This is not necessary, but it does simplify things mildly. We also assume that for each k we have an effective enumeration of the collection of all formulas with $x_1, \ldots, x_k$ as free variables. We write $\phi_n^k(x_1, \ldots, x_k)$ for the n-th formula in this enumeration. Finally, we assume that D^1, D^2, $D^3, \ldots$ are primitive recursive functions, with D^k being $k+1$-place, such that $D^k(n, m_1, \ldots, m_k)$ is the Gödel number of $\phi_n^k(\overline{m_1}, \ldots, \overline{m_k})$. D^k is the intended interpretation of d^k.

As far as models for the language L go there is an intended domain, containing the natural numbers and possibly other 'real world' objects. For simplicity we assume every member of the domain has a closed term of L intended to name it. Interpretations of all symbols except T are fixed. The symbol 0 is always interpreted by the number 0; the function symbol s is interpreted by the successor function; symbols intended to denote various 'real world' items are interpreted by those items; and so on. In particular, the function symbol d^k is always interpreted by the function D^k mentioned in the previous paragraph. Thus the only thing up for grabs is the interpretation of T, which we specify using the notion of a valuation.

Definition 3.1 A *valuation* is a mapping of closed formulas of the form $T(t)$ to the space of truth values $\mathcal{FOUR}$. We assume that if t is any closed term whose interpretation is not a natural number, $v(T(t)) = \bot$.

The action of a valuation is extended to all sentences in a straightforward way. If $R(t_1, \ldots, t_n)$ is any atomic sentence whose relation symbol is not T, $v(R(t_1, \ldots, t_n))$ is taken to be *true* or *false* according to whether the intended interpretation of R holds or does not hold of the intended interpretations of $t_1, \ldots, t_n$. Then, inductively, $v(X \wedge Y) = v(X) \wedge v(Y)$, where the $\wedge$ on the right is the meet operation of $\mathcal{FOUR}$ with respect to the $\leq_t$ ordering, and similarly for $\vee$ and $\neg$. $v((\forall x)\phi(x)) = \bigwedge\{v(\phi(t)) \mid \text{all closed terms } t\}$, where this is the arbitrary meet of $\mathcal{FOUR}$ with respect to $\leq_t$, and similarly for the existential quantifier.

The aim of a theory of truth is to produce a valuation that gives T a meaning as closely approximating its intended interpretation as a truth predicate as possible. A minimum requirement is that $T(\ulcorner X \urcorner)$ and X should have the same truth value for every sentence X.

Finally, a word on the substitution function. It is, of course, used to produce self-referential sentences in the standard way. We omit details at this point, but we assume it can be used in the usual way to create liar sentences, truth tellers, and so on. One caution, however. We write $X \leftrightarrow Y$ to indicate that X and Y have the same truth value (member of $\mathcal{FOUR}$) under every valuation. This is *not* a definable connective of our language. One sometimes sees a connective $\equiv$, introduced by: $X \equiv Y$ abbreviates $(\neg X \vee Y) \wedge (\neg Y \vee X)$. This does not have the behavior one might expect. For instance, $X \equiv X$ evaluates to $\bot$ if X itself does. The relation $\leftrightarrow$ is purely two-valued, and $X \leftrightarrow X$ is simply true. A liar sentence is a sentence X such that $X \leftrightarrow \neg T(\ulcorner X \urcorner)$.

4 Kripke's way of doing things

Kripke introduced a *truth revision operator*, which we will denote by Φ. It maps valuations to valuations, and the intention is that $\Phi(v)$ will make the truth predicate reflect the way that v sees the world.

Definition 4.1 Let v be a valuation. Then $\Phi(v)$ is the valuation such that, for each sentence X of L,

$$\Phi(v)(T(\ulcorner X \urcorner)) = v(X).$$

Kripke argued that a fixed point of Φ would be a good candidate for a valuation of L (where v is a fixed point if $\Phi(v) = v$). It is not immediately clear that Φ has any fixed points, so Kripke gave an argument that amounts to the following.

The underlying space of truth values, $\mathcal{FOUR}$, has two partial orderings, $\leq_t$ and $\leq_k$. These, in turn, induce orderings on the space of valuations, as follows.

Definition 4.2 We say $v_1 \leq_t v_2$ provided $v_1(A) \leq_t v_2(A)$ for each closed atomic formula A. Similarly $v_1 \leq_k v_2$ if $v_1(A) \leq_k v_2(A)$ for each closed atomic formula A.

It is not hard to check that the space of valuations is a complete lattice under both $\leq_t$ and $\leq_k$. For instance, the meet of v_1 and v_2 in the $\leq_t$ ordering exists, and is the valuation v_3 such that for each closed atomic formula A, $v_3(A) = v_1(A) \wedge v_2(A)$. And so on.

The orderings of valuations were defined using closed *atomic* formulas. This extends to non-atomic formulas well or badly, depending.

Proposition 4.3

1. *Suppose $v_1 \leq_k v_2$. Then $v_1(X) \leq_k v_2(X)$ for every sentence X.*
2. *Suppose $v_1 \leq_t v_2$. Then $v_1(X) \leq_t v_2(X)$ for every sentence X that does not contain negation.*

The verification of this Proposition is by a straightforward induction on formula complexity, and uses the interlacing conditions mentioned earlier. In a sense, it is the fact that the $\leq_t$ ordering behaves badly when negation is present that is the source of all difficulties. And it is the good behavior of the $\leq_k$ ordering that saves things.

Proposition 4.4 *The truth revision operator Φ is monotonic with respect to the $\leq_k$ ordering. That is, if $v_1 \leq_k v_2$ then $\Phi(v_1) \leq_k \Phi(v_2)$.*

Proof Suppose $v_1 \leq_k v_2$. By the previous Proposition, $v_1(X) \leq_k v_2(X)$ for every sentence X, so by definition $\Phi(v_1)(T(\ulcorner X \urcorner)) \leq_k \Phi(v_2)(T(\ulcorner X \urcorner))$. It follows that $\Phi(v_1) \leq_k \Phi(v_2)$. ∎

There is a well-known theorem of Knaster and Tarski [16] that says a monotonic function on a complete lattice has a smallest, and a biggest fixed point. (More generally, the collection of fixed points will, itself, be a complete lattice.) Since Φ is monotonic under $\leq_k$, and the space of valuations is a complete lattice under this ordering, Φ has a smallest, and a biggest, fixed point. In particular, fixed points exist. The smallest fixed point of Φ is a natural candidate for a truth assignment since it contains the minimum information to which we are entitled. Kripke also singled out another fixed point of interest, the largest intrinsic one. We will say more about this in Section 10.

In point of fact, Kripke's argument did not look quite like this. Instead of a four-valued logic he used a three-valued one; consequently he did not have a complete lattice to work with, so the Knaster-Tarski theorem was not applicable. Still, the structure he used was that of a complete *semi-lattice*, and a Knaster-Tarski like theorem still applies (see [5]). His valuations were not given directly as mappings to a many-valued logic, but rather in terms of extension/anti-extension pairs. It is not hard to see that this is an equivalent alternative. Thus what we have presented is essentially Kripke's version, despite the dissimilar appearance.

5 GLF-Stable fixed points

As we said earlier, whenever possible we want to give preference to falsehood over truth. But this implies we should distinguish those occurrences of the truth predicate in which it essentially represents truth from those in which it essentially represents falsehood. That is, we should distinguish between positive and negative occurrences, between those inside an even number of negation signs and those inside an odd number. Actually, to keep notation and terminology simple, we will assume formulas have been 'normalized.'

Normal Form Assumption From now on we assume all formulas are in negation normal form: all occurrences of the negation symbol are at the atomic level.

Since the DeMorgan laws hold for the $\leq_t$ connectives, every formula is equivalent to one in negation normal form, so the Assumption is no restriction.

With formulas in negation normal form, we can think of occurrences of $\neg T(x)$ as if they were occurrences of a new atom, a falsehood atom, now disconnected from $T(x)$. We introduce the notion of a *pseudo-valuation*, which is a mapping from sentences of the forms $T(t)$ and $\neg T(t)$ to $\mathcal{FOUR}$. Pseudo-valuations can be concocted from valuations in a convenient way.

Definition 5.1 Let v_1 and v_2 be valuations. We define a pseudo-valuation $v_1 \triangle v_2$ as follows.

$$\begin{array}{rcl}(v_1 \triangle v_2)(T(t)) & = & v_1(T(t)) \\ (v_1 \triangle v_2)(\neg T(t)) & = & \neg v_2(T(t))\end{array}$$

Pseudo-valuations are extended to arbitrary sentences following the usual inductive rules (except for negation, which no longer comes into things).

The idea is, in $v_1 \triangle v_2$, v_1 supplies positive information, about 'truth,' while v_2 supplies negative information, about 'falsehood.' Now we generalize the truth revision operator in a straightforward way, to use separate inputs for positive and for negative occurrences of T.

Definition 5.2 Let v_1 and v_2 be valuations. Then $\Psi(v_1, v_2)$ is the valuation such that, for each sentence X of L,

$$\Psi(v_1, v_2)(T(\ulcorner X \urcorner)) = (v_1 \triangle v_2)(X).$$

The mapping Φ was monotonic under $\leq_k$ but not under $\leq_t$. In fact, with respect to $\leq_t$ little of use could be said. The mapping Ψ is considerably better behaved. We leave it to you to check the following.

Proposition 5.3

1. *Ψ is monotonic in both inputs, under $\leq_k$. That is, if $v_1 \leq_k v_2$ and $w_1 \leq_k w_2$ then $\Psi(v_1, w_1) \leq_k \Psi(v_2, w_2)$.*

2. *Ψ is monotonic in its first input, under $\leq_t$. That is, if $v_1 \leq_t v_2$ then $\Psi(v_1, w) \leq_t \Psi(v_2, w)$.*

3. *Ψ is anti-monotonic in its second input, under $\leq_t$. That is, if $w_1 \leq_t w_2$ then $\Psi(v, w_1) \geq_t \Psi(v, w_2)$.*

The space of valuations is a complete lattice under $\leq_t$ as well as under $\leq_k$, though no use could be made of this when dealing with Φ. But since Ψ is monotonic under $\leq_t$ in its first input, if we hold the second input fixed, as a parameter, and consider it only as a function of its first input, we have a monotonic function on a complete lattice, and the Knaster-Tarski Theorem applies. It is here that the notion of giving preference to falsehood is explicitly introduced.

Definition 5.4 The *derived operator* of Ψ is the single input mapping Ψ' given by: $\Psi'(v)$ is the smallest fixed point, in the $\leq_t$ ordering, of the function $(\lambda x)\Psi(x, v)$.

The mapping Ψ' can be thought of as a new candidate for a truth revision operator. The intuition is a little more complex than it was with Φ, however. Suppose we make a 'guess' at falsehood — that is, we choose a valuation v telling us how occurrences of $\neg T(t)$ behave. Subject to this guess at falsehood behavior, the least fixed point of $(\lambda x)\Psi(x, v)$ will reflect how truth — occurrences of $T(t)$ — should be valued, provided we use *false* as the value whenever possible. If the result is the guess we began with, then v was a good guess; it is a valuation we can not reason ourselves away from, given that *false* is our 'default' truth value. This is formalized in the following.

Definition 5.5 A *GLF-stable* valuation is a fixed point of Ψ'.

Now we face technical problems similar to those of Kripke: for instance, are there any GLF-stable valuations? We will show there are and the family of them has a nice mathematical structure worth further investigation. Moreover, we will show that every GLF-stable valuation is one of Kripke's fixed points, so we are singling out a subclass of the valuations Kripke considered, for special attention.

Assuming there are GLF-stable valuations, how do various problematic sentences behave in them? Suppose X is a liar sentence, $X \leftrightarrow \neg T(\ulcorner X \urcorner)$. We will produce a GLF-stable valuation in which X is $\perp$ and another in which it is $\top$. There are none in which it is *true* or *false*, since these values are impossible in Kripke's fixed points. Thus the liar sentence behaves essentially as it does in Kripke's theory.

Now suppose X is a truth-teller, $X \leftrightarrow T(\ulcorner X \urcorner)$. We will see later on that X evaluates to *false* in every GLF-stable valuation. Since it can take on any of the four truth values in various of Kripke's fixed points, it follows that GLF-stable valuations are a proper subset of Kripke's fixed points.

For a slightly more complicated example, suppose X and Y each say the other is lying. That is, $X \leftrightarrow \neg T(\ulcorner Y \urcorner)$ and $Y \leftrightarrow \neg T(\ulcorner X \urcorner)$. In one of Kripke's fixed points X is *true* and Y is *false*; in another X is *false* and Y is *true*; in still others both X and Y are $\perp$, or both X and Y are $\top$. GLF-stable valuations will assign a sentence the value *false* whenever possible; the problem now is the symmetry between X and Y — which of the two should be assigned *false*? We will see that in all GLF-stable valuations either X and Y are $\top$, or both are $\perp$. An assignment of *false* to one and *true* to the other is not possible in a GLF-stable valuation.

As a final example, consider a sentence that asserts either it or its negation is true. That is, construct a sentence X so that $X \leftrightarrow T(\ulcorner X \urcorner) \vee T(\ulcorner \neg X \urcorner)$. This will evaluate to $\perp$ in every GLF-valuation. By way of contrast, in Kripke's fixed points it can be either *true* or $\perp$, though never *false*. Incidentally, in the "truth" ordering, $\perp <_t$ *true*, so our general preference for falsehood has prevailed in the sense that we minimized values in the $\leq_t$ ordering as far as possible.

6 Background on the Knaster-Tarski theorem

We have used the Knaster-Tarski theorem several times, to ensure that various fixed points exist. It has two essentially different proofs which we will need, so we take a moment to describe them. Also we introduce an *anti-monotonic* analog that apparently first appeared in a paper of Stephen Yablo's on the theory of truth [19], and which will play a curious role here.

Suppose M is a complete lattice, and f is a monotonic function on it. In particular we are assuming that in M there is a partial ordering, $\leq$, and every subset of M has a least upper bound and a greatest lower bound with respect to this ordering. If $S \subseteq M$, the least upper bound of S is customarily denoted $\bigvee S$, and the greatest lower bound is denoted $\bigwedge S$. Further, we are assuming that if $x \leq y$ then $f(x) \leq f(y)$. The Knaster-Tarski theorem says that under these circumstances f will have a smallest (and by duality a largest) fixed point. There are two quite different ways of characterizing this fixed point: from below, and from above. We sketch the techniques.

In the 'from below' approach, one draws near to the smallest fixed point of f by a sequence of approximations, each of which is below it. The sequence of approximations may need a transfinite number of steps, so it is indexed by ordinal numbers. More specifically, a sequence, $f_0, f_1, \ldots, f_\alpha, \ldots$ is defined as follows. f_0 is the smallest member of M (which exists in every complete lattice). For a successor ordinal $\alpha+1$ we set $f_{\alpha+1} = f(f_\alpha)$. And for a limit ordinal λ we set $f_\lambda = \bigvee\{f_\alpha \mid \alpha < \lambda\}$. One can show this is an increasing sequence, that is, if $\alpha < \beta$ then $f_\alpha \leq f_\beta$. If the inequality $f_\alpha \leq f_\beta$ were always strict, always $<$, we would have a $1-1$ mapping from the class of ordinals into M. Since we are assuming that M is a set such a thing is impossible, so there must be an ordinal α such that $f_\alpha = f_{\alpha+1} = f(f_\alpha)$. Thus f has a fixed point.

If it can be shown by transfinite induction that every f_α has a certain property, that property will be shared by the fixed point just produced, since it is one of the f_α. As one example, it can be shown that if x is some fixed point of f, then for each α, f_α will be $\leq x$. Consequently the fixed point constructed by this approximation technique will be below any fixed point, so it is the *least* fixed point.

In Kripke's paper [13] this is the approach he describes, approximating to the least fixed point through a transfinite sequence of extension/anti-extension pairs.

There is quite another way of producing the least fixed point, yielding a different method of proving things about it. Let $S = \{x \in M \mid f(x) \leq x\}$. S is not empty since it contains the biggest member of M (which always exists in a complete lattice). Set $s = \bigwedge S$ — s turns out to be the smallest fixed point of S by the following simple argument. First, if $x \in S$ then $s \leq x$ so by monotonicity $f(s) \leq f(x) \leq x$ (the latter since $x \in S$). Since x was an arbitrary member of S, it follows that $f(s) \leq s$, so $s \in S$. Next S is closed under f since, if $x \in S$, $f(x) \leq x$, so by monotonicity $f(f(x)) \leq f(x)$, and this says $f(x) \in S$. Consequently $f(s) \in S$, so $s \leq f(s)$. Combining this with an earlier inequality, $f(s) = s$. Finally, if x is any fixed point of f, $x \in S$, so $s \leq x$, and hence s is the *least* fixed point.

Notice that this method of producing the least fixed point also gives a way of proving things about it. Suppose z is any point in M such that $f(z) \leq z$. Then $z \in S$, so $s \leq z$. This gives an easy way of showing various things are upper bounds for the least fixed point of f — it is something we will use often.

Finally we turn to Yablo's curious modification of the Knaster-Tarski theorem.

Definition 6.1 Let M be a complete lattice and let μ and ν be in M. We say these are *oscillation points* for a function f if $f(\mu) = \nu$ and $f(\nu) = \mu$. We say these are *extreme* oscillation points if: 1) $\mu \leq \nu$ and, 2) if x and y are any pair of oscillation points for f, then x and y lie between μ and ν.

Now Yablo's result is easily stated.

Proposition 6.2 *Let f be anti-monotonic on the complete lattice M. Then f has a pair of extreme oscillation points.*

If f has a pair of extreme oscillation points they must be unique, since if there were another extreme oscillation pair, each pair must be between the other. Yablo established his result using a proof much like the approximation sequence argument for the Knaster-Tarski result, under approximating to one of the points, and over approximating to the other. There is a different, simpler, argument whose details we leave to you. If f is anti-monotonic, f^2 is monotonic. Then by the Knaster-Tarski theorem f^2 has a smallest fixed point, and by dualizing the argument, it also has a biggest fixed point. These can be shown to be a pair of extreme oscillation points for the function f itself.

7 GLF-Stable valuations exist

The mapping Ψ', despite its somewhat elaborate characterization, is quite well-behaved. The proof that it is so is essentially algebraic, once some basic facts about monotonicity have been established. This is where we begin.

Theorem 7.1 *The function Ψ' is monotonic in the $\leq_k$ ordering, but is anti-monotonic in the $\leq_t$ ordering.*

Proof In order to show Ψ' is anti-monotonic in the $\leq_t$ ordering we can use one of the arguments for the Knaster-Tarski theorem in quite a direct way. Recall that if a function f is monotonic on a complete lattice, and if $f(x) \leq x$, then the least fixed point of f will be $\leq x$.

Now suppose $v_1 \leq_t v_2$; we show $\Psi'(v_2) \leq_t \Psi'(v_1)$. Since Ψ is anti-monotonic in its second argument under $\leq_t$,

$$\Psi(\Psi'(v_1), v_2) \leq_t \Psi(\Psi'(v_1), v_1).$$

Since $\Psi'(v_1)$ is a fixed point of $(\lambda x)\Psi(x, v_1)$ this yields:

$$\Psi(\Psi'(v_1), v_2) \leq_t \Psi'(v_1).$$

Since $\Psi'(v_2)$ is the least fixed point of $(\lambda x)\Psi(x, v_2)$ under $\leq_t$, it follows that

$$\Psi'(v_2) \leq_t \Psi'(v_1).$$

In showing monotonicity under $\leq_k$ a similar argument is not as straightforward, since Ψ' is defined in terms of the least fixed point operation of $\leq_t$, and we are interested in behavior relative to $\leq_k$, so the two orderings are rather badly mixed up. Konstantinos Georgatos recently showed that one can still establish monotonicity by a slightly more complicated version of this proof, but for the sake of variety we base our approach on the other proof of the Knaster-Tarski theorem, the one that approximates to a least fixed point from below.

Suppose $v_1 \leq_k v_2$; we will show $\Psi'(v_1) \leq_k \Psi'(v_2)$. Define two sequences of valuations, a_α and b_α, as follows. $a_0 = b_0$ is the valuation that assigns *false* to every closed instance of $T(x)$ — thus we have the least valuation in the $\leq_t$ ordering. Set $a_{\alpha+1} = \Psi(a_\alpha, v_1)$ and $b_{\alpha+1} = \Psi(b_\alpha, v_2)$. Finally, for a limit ordinal λ, $a_\lambda = \bigvee_{\alpha<\lambda} a_\alpha$ and $b_\lambda = \bigvee_{\alpha<\lambda} b_\alpha$. Both sequences are increasing in the $\leq_t$

ordering since Ψ is monotone in its first argument. The a_α sequence approximates to $\Psi'(v_1)$ while the b_α sequence approximates to $\Psi'(v_2)$. So it is enough to show that for each α, $a_\alpha \leq_k b_\alpha$.

If $\alpha = 0$ the result is trivial. Now suppose $a_\alpha \leq_k b_\alpha$. Then $a_{\alpha+1} = \Psi(a_\alpha, v_1) \leq_k \Psi(b_\alpha, v_2) = b_{\alpha+1}$, using the monotonicity of Ψ in both arguments, under $\leq_k$. Finally, if $a_\alpha \leq_k b_\alpha$ for each $\alpha < \lambda$ it follows that $\bigvee_{\alpha<\lambda} a_\alpha \leq_k \bigvee_{\alpha<\lambda} b_\alpha$ using the interlacing conditions. This yields that $a_\lambda \leq_k b_\lambda$ and completes the transfinite induction proof that $a_\alpha \leq_k b_\alpha$ for all α, establishing the rest of the theorem. ■

Now we have enough machinery to show there are GLF-stable valuations. Very simply, since Ψ' is monotonic in the $\leq_k$ ordering, an application of the Knaster-Tarski theorem yields that it has a smallest and a biggest fixed point with respect to this ordering. We will denote the smallest fixed point under $\leq_k$ by s_k, and the biggest by S_k.

Next we want to get some feeling for the behavior of GLF-stable valuations. Suppose we construct a truth-teller in the usual way, which we briefly sketch. The formula $T(d^1(x,x))$ is $\phi^1_n(x_1)$ for some n (we are using our effective enumeration of all formulas with only x_1 free). Let $X = \phi^1_n(\overline{n}) = T(d^1(\overline{n},\overline{n}))$. The function symbol d^1 is always interpreted by D^1, and $D^1(n,n) = \ulcorner\phi^1_n(\overline{n})\urcorner$, so X is a truth-teller, since $X = T(d^1(\overline{n},\overline{n})) \leftrightarrow T(\ulcorner\phi^1_n(\overline{n})\urcorner) = T(\ulcorner X\urcorner)$.

Now let v be any valuation, we calculate the behavior of $\Psi'(v)$ on $X = T(d^1(\overline{n},\overline{n}))$ using the ordinal approximation technique. $\Psi'(v)$ is the least fixed point, with respect to $\leq_t$, of $(\lambda x)\Psi(x,v) = f(x)$ and we approximate to this via f_0, f_1, ..., f_α, ...

f_0 is the smallest valuation in the $\leq_t$ ordering — it assigns *false* to every instance of T, in particular to $T(\ulcorner X\urcorner)$, and hence also to X since $X \leftrightarrow T(\ulcorner X\urcorner)$.

Suppose $f_\alpha(X) = f_\alpha(T(\ulcorner X\urcorner)) = \mathit{false}$ Then

$$\begin{array}{rcl} f_{\alpha+1}(T(\ulcorner X\urcorner)) & = & (f(f_\alpha))(T(\ulcorner X\urcorner)) \\ & = & \Psi(f_\alpha, v)(T(\ulcorner X\urcorner)) \\ & = & (f_\alpha \triangle v)(X) \\ & = & (f_\alpha \triangle v)(T(d^1(\overline{n},\overline{n}))) \\ & = & f_\alpha(T(d^1(\overline{n},\overline{n}))) \\ & = & f_\alpha(X) \\ & = & \mathit{false} \end{array}$$

Hence also $f_{\alpha+1}(X) = \mathit{false}$ since $X \leftrightarrow T(\ulcorner X\urcorner)$.

Finally, if $f_\alpha(T(\ulcorner X\urcorner)) = \mathit{false}$ for every $\alpha < \lambda$, it follows immediately that $f_\lambda(T(\ulcorner X\urcorner)) = \mathit{false}$.

All this establishes that the limit of the f_α sequence assigns *false* to $T(\ulcorner X\urcorner)$, and hence to X. Thus $\Psi'(v)(X) = \mathit{false}$. Now, if v is a GLF-stable valuation, $\Psi'(v) = v$, hence $v(X) = \mathit{false}$. That is, the truth-teller is *false* in every GLF-stable valuation.

In a similar way we can construct mutual truth-tellers, $X \leftrightarrow T(\ulcorner Y\urcorner)$ and $Y \leftrightarrow T(\ulcorner X\urcorner)$. These too turn out to be *false* in every GLF-valuation.

Liar sentences are a little more complicated in behavior. If $\neg T(d^1(x,x))$ is $\phi^1_n(x)$, then $Z = \phi^1_n(\overline{n})$ will be a liar sentence: $Z \leftrightarrow \neg T(\ulcorner Z\urcorner)$. It can be shown by an argument similar to the one above, that in the smallest fixed point of Ψ', Z evaluates to $\perp$. That is, this liar sentence is $\perp$ in the $\leq_k$ smallest GLF-stable valuation. Dualizing the argument shows that in the $\leq_k$ largest GLF-stable valuation it evaluates to $\top$. Since (as we will show later) every GLF-valuation is a fixed point of Kripke's operator Φ, it follows that no GLF-stable valuation can make the liar sentence either *true* or *false*.

The situation is similar with sentences X and Y constructed so that $X \leftrightarrow \neg T(\ulcorner Y\urcorner)$ and $Y \leftrightarrow \neg T(\ulcorner X\urcorner)$. That is, they are both $\perp$ or both $\top$ in every GLF-stable valuation.

Finally, we can construct two sentences, X and Y, so that

$$\begin{array}{rcl} X & \leftrightarrow & T(\ulcorner X \urcorner) \vee T(\ulcorner Y \urcorner) \\ Y & \leftrightarrow & \neg T(\ulcorner X \urcorner) \wedge \neg T(\ulcorner Y \urcorner) \end{array}$$

Then $Y \leftrightarrow \neg X$, so in effect $X \leftrightarrow T(\ulcorner X \urcorner) \vee T(\ulcorner \neg X \urcorner)$. We leave it to you to verify that there are no GLF-stable valuations in which X is *true*, but there are some in which X is $\perp$, and others in which it is $\top$.

8 Structural results

We have said several times now that the GLF-stable valuations are among Kripke's fixed points. The verification is rather simple.

Theorem 8.1 *Every GLF-stable valuation is a Kripke fixed point. That is, every fixed point of Ψ' is a fixed point of Φ.*

Proof It follows directly from the definitions of Φ and Ψ that $\Phi(x) = \Psi(x, x)$. Also $\Psi'(t)$ is a fixed point of $(\lambda x)\Psi(x, t)$ so $\Psi(\Psi'(t), t) = \Psi'(t)$. Now suppose s is a fixed point of Ψ'. Then:

$$\Phi(s) = \Psi(s, s) = \Psi(\Psi'(s), s) = \Psi'(s) = s.$$

■

Next we will make use of Yablo's anti-monotonicity result, Proposition 6.2. Since Ψ' is anti-monotonic in the $\leq_t$ ordering, there is a pair of extreme oscillation points, let us call them s_t and S_t. These are not GLF-stable valuations. For instance, s_t makes the liar sentence *false* while S_t makes it *true*, and we already observed the liar sentence must come out either $\perp$ or $\top$ in every GLF-stable valuation. But there is a rather remarkable way of obtaining GLF-stable valuations from s_t and S_t. Recall, the operations $\wedge$, $\vee$, $\otimes$ and $\oplus$ extended from $\mathcal{FOUR}$ to the space of valuations using pointwise extensions: $(v \otimes w)(T(t)) = v(T(t)) \otimes w(T(t))$, and so on. We will show both $s_t \otimes S_t$ and $s_t \oplus S_t$ are GLF-stable valuations. We first need a simple lemma, then we can prove the principle result.

Lemma 8.2 *Let v_1, v_2, v_3 be valuations. Then:*

1. *if $v_1 \leq_t v_2 \leq_t v_3$ then $v_1 \otimes v_3 \leq_k v_2$;*
2. *if $v_1 \leq_t v_2 \leq_t v_3$ then $v_2 \leq_k v_1 \oplus v_3$;*
3. *if $v_1 \leq_k v_2 \leq_k v_3$ then $v_1 \wedge v_3 \leq_t v_2$;*
4. *if $v_1 \leq_k v_2 \leq_k v_3$ then $v_2 \leq_t v_1 \vee v_3$.*

Proof We only show item 1; the others are similar. Suppose $v_1 \leq_t v_2 \leq_t v_3$. Then for an atomic sentence A of the form $T(t)$, $v_1(A) \leq_t v_2(A) \leq_t v_3(A)$. We must show $v_1(A) \otimes v_3(A) \leq_k v_2(A)$. Now, if any of the two $\leq_t$ are actually $=$, then it is immediate that $v_1(A) \otimes v_3(A) \leq_k v_2(A)$. If all the inequalities are strict it must be that $v_1(A) = $ *false*, $v_3(A) = $ *true*, and $v_2(A)$ is either $\perp$ or $\top$; the result is immediate in either case. ■

Theorem 8.3 *The valuations $s_t \otimes S_t$ and $s_t \oplus S_t$ are both GLF-stable.*

Proof We show $s_t \otimes S_t$ is a fixed point of Ψ'; the other part has a similar proof.

We know Ψ' is monotonic under $\leq_k$. Also $\Psi'(s_t) = S_t$, and the other way around as well, since these are oscillation points for Ψ'. Finally $s_t \otimes S_t$ is below both s_t and S_t in the $\leq_k$ ordering. Hence:

$$\Psi'(s_t \otimes S_t) \leq_k \Psi'(s_t) = S_t$$

$$\Psi'(s_t \otimes S_t) \leq_k \Psi'(S_t) = s_t$$

It follows that:

$$\Psi'(s_t \otimes S_t) \leq_k s_t \otimes S_t.$$

Next, $s_t \leq_t S_t$ by Proposition 6.2. Then, using the interlacing conditions:

$$s_t = s_t \otimes s_t \leq_t s_t \otimes S_t \leq_t S_t \otimes S_t = S_t$$

So, using anti-monotonicity,

$$\Psi'(S_t) \leq_t \Psi'(s_t \otimes S_t) \leq_t \Psi'(s_t)$$

from which we obtain

$$s_t \leq_t \Psi'(s_t \otimes S_t) \leq_t S_t.$$

From this, using Lemma 8.2,

$$s_t \otimes S_t \leq_k \Psi'(s_t \otimes S_t).$$

This, together with the earlier inequality, establishes the fixpoint result. ■

This theorem can be considerably improved — we can say which GLF-stable valuations $s_t \otimes S_t$ and $s_t \oplus S_t$ are. Recall, s_k and S_k are the smallest and biggest fixed points of Ψ' in the $\leq_k$ ordering.

Theorem 8.4

1. $s_k = s_t \otimes S_t$
2. $S_k = s_t \oplus S_t$.

Proof By definition, s_k is the least fixed point of Ψ' under $\leq_k$. By the previous Theorem, $s_t \otimes S_t$ is a fixed point. Hence $s_k \leq_k s_t \otimes S_t$.

Also by definition, s_t and S_t are extreme oscillation points of Ψ' under $\leq_t$. Since s_k is a fixed point, s_k, s_k is an oscillation pair, so it must lie between the extreme pair. That is, $s_t \leq_t s_k \leq_t S_t$. Then it follows from Lemma 8.2 that $s_t \otimes S_t \leq_k s_k$.

The other item has a similar proof. ■

We have seen that the least and greatest GLF-stable valuations, under $\leq_k$, can be expressed easily, using the extreme oscillation pair under $\leq_t$. Now we show the connection goes the other way as well. First we need a simple lemma concerning fixpoints.

Lemma 8.5 *If f is a monotonic function on a complete lattice, then f and f^2 have the same least and greatest fixed points.*

Proof Let a be the least fixed point of f, and let b be the least fixed point of f^2. Since every fixed point of f is also a fixed point of f^2, $f^2(a) = a$. Since b is the least fixed point of f^2, $b \leq a$.

In the other direction, if x is a fixed point of f^2, so is $f(x)$, because $f^2(f(x)) = f(f^2(x)) = f(x)$. But then $f(b)$ must be a fixed point of f^2, and since b is the least fixed point of f^2, $b \leq f(b)$. Using monotonicity, $f(b) \leq f^2(b)$, so $f(b) \leq b$, and b is a fixed point of f. Since a is the least fixed point of f, $a \leq b$. ■

Now we show a companion to Theorem 8.4.

Theorem 8.6

1. $s_t = s_k \wedge S_k$
2. $S_t = s_k \vee S_k$.

Proof The function Ψ' is monotonic under $\leq_k$, with s_k and S_k as least and greatest fixed points. Then $(\Psi')^2$ is also monotonic under $\leq_k$ and, by the Lemma, it has the same least and greatest fixed points. Now, $(\Psi')^2$ is also monotonic under $\leq_t$, with s_t and S_t as least and greatest fixed points. By definition,

$$s_k \wedge S_k \leq_t s_k$$

so

$$(\Psi')^2(s_k \wedge S_k) \leq_t (\Psi')^2(s_k) = s_k$$

and similarly

$$(\Psi')^2(s_k \wedge S_k) \leq_t S_k$$

so

$$(\Psi')^2(s_k \wedge S_k) \leq_t (s_k \wedge S_k).$$

Now, since s_t is the least fixed point of $(\Psi')^2$ under $\leq_t$,

$$s_t \leq_t s_k \wedge S_k.$$

But also, since s_t is a fixed point of $(\Psi')^2$, and s_k and S_k are its least and greatest fixed points under $\leq_k$,

$$s_k \leq_k s_t \leq_k S_k$$

and so by Lemma 8.2

$$s_k \wedge S_k \leq_t s_t.$$

The other part of the Theorem has a similar proof. ■

We noted earlier that s_t values the liar sentence as *false*, while S_t values it *true*. It is now simple to verify this using the Theorem above. If X is the liar sentence, we know from earlier that $s_k(X) = \bot$ and $S_k(X) = \top$. But $\bot \wedge \top =$ *false* and $\bot \vee \top =$ *true*.

It follows from Theorems 8.4 and 8.6 that the collection of GLF-stable valuations can be bounded rather neatly. The set-up is displayed in Figure 2. Briefly, all GLF-stable valuations lie between s_k and S_k under the $\leq_k$ ordering, with these particular valuations included. Also they all lie between s_t and S_t under the $\leq_t$ ordering, with these valuations not included. And each set of these valuations can be calculated from the other, using the Theorems above.

9 Why $\mathcal{FOUR}$?

Why was a four-valued logic used in this work? There are several reasons which will be discussed here, but the fact of the matter is that the use of four truth values is not essential. We could have used more — or less.

$\mathcal{FOUR}$ is the simplest example of a *bilattice*, a notion due to Matt Ginsberg [10]. Bilattices constitute a special class of many-valued logics with very nice properties. We will not go into a

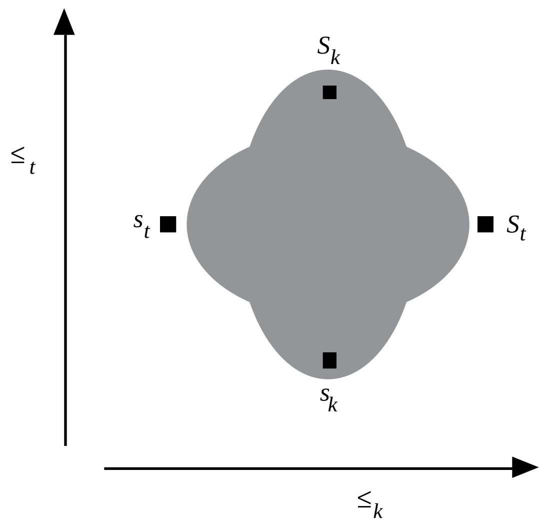

Figure 2: GLF-stable valuations

discussion of them here — [6] gives a general idea. Suffice it to say that bilattices, like $\mathcal{FOUR}$, have two orderings, $\leq_k$ and $\leq_t$, which are inter-connected in various ways. They arise naturally. The collection of valuations as used in this paper has $\leq_k$ and $\leq_t$ orderings and many of the properties of $\mathcal{FOUR}$ carry over to it; in fact, it is a bilattice. Bilattices come up if one is discussing imperfect knowledge of several people; reasoning in which one assigns probabilities for and against sentences; and truth across several possible worlds.

Every fundamental result established in this paper continues to apply if $\mathcal{FOUR}$ is replaced by an arbitrary bilattice in which the operations satisfy distributive laws (this includes all the bilattices mentioned in the previous paragraph). Not only that, but virtually all the proofs continue to apply with no changes. The proof of Lemma 8.2 is a exception since we used the particular values of $\mathcal{FOUR}$ explicitly, but the Lemma holds generally with a more complicated proof. Thus the results presented here are really much broader than they would seem at first glance.

Probably not having enough truth values is not the objection to $\mathcal{FOUR}$ that would occur to most people. More commonly an objection is made to the use of $\top$ as one of the truth values. Kripke used three truth values — why bring in four? Well in fact, we didn't need to. The general idea of GLF-stable valuations could have been introduced in the framework of Kleene's strong (or weak) three-valued logic instead. In logic programming, where the notion originated, this is the way it is most commonly done. We would, of course, lose some results. Theorem 8.6 could not be stated if we did not have S_k, and it would not be available if a three-valued logic were used. Similarly part 2 of Theorem 8.4 could not be stated. Still, the essential ideas of a GLF-stable valuation can be developed in a three-valued setting.

The use of three instead of four truth values does make the mathematics a little more complicated. Without $\top$ the space of truth values is not a complete lattice under $\leq_k$, so the Knaster-Tarski theorem can not be used to establish the existence of a GLF-stable valuation. Fortunately there are generalizations that apply instead. The structure under $\leq_t$ is still that of a complete lattice when there are three truth values, so Definition 5.4 continues to make sense. Nothing basic disappears.

Finally, Kripke considered the use of supervaluations as well as Kleene's three valued logics. Presumably this could be applied here as well. It remains to be investigated.

10 Intrinsic fixed points

From the class of his fixed points, Kripke singled out a subclass for special attention, the *intrinsic* ones. These were the ones compatible with every fixed point. We must modify the definition somewhat, because we have four truth values instead of Kripke's three, but our version is equivalent.

Definition 10.1 A valuation v is *consistent* if $v(T(t)) \neq \top$ for every closed atomic sentence $T(t)$.

In Kripke's version all valuations must be consistent since there is no $\top$ truth value. So the issue of consistency for us corresponds to the issue of existence in Kripke's approach.

Definition 10.2 A fixed point v of Φ is *intrinsic* if $v \oplus w$ is consistent, for every consistent fixed point w.

An intrinsic fixed point must, itself, be consistent. The argument is elementary. Suppose v is intrinsic. Let v_0 be the smallest fixed point with respect to $\leq_k$; v_0 will be consistent. Since $v_0 \leq_k v$, $v \oplus v_0 = v$, and this must be consistent according to the definition. This same little argument also shows that v_0, the smallest fixed point, must itself be intrinsic, so intrinsic fixed points exist.

Let $\mathcal{I}$ be the collection of all intrinsic fixed points of Φ. $\mathcal{I}$ is *directed*, that is if $v_1, v_2 \in \mathcal{I}$, there is some $v_3 \in \mathcal{I}$ with $v_1 \leq_k v_3$ and $v_2 \leq_k v_3$. We sketch the argument for this. Since v_1 is intrinsic and v_2 is a consistent fixed point, $v_1 \oplus v_2$ is consistent. Now, $v_1 \leq_k v_1 \oplus v_2$, so by monotonicity, $v_1 = \Phi(v_1) \leq_k \Phi(v_1 \oplus v_2)$. Similarly $v_2 \leq_k \Phi(v_1 \oplus v_2)$. Consequently $v_1 \oplus v_2 \leq_k \Phi(v_1 \oplus v_2)$. The argument that Φ has a consistent least fixed point under $\leq_k$ generalizes directly to show that if $v \leq_k \Phi(v)$, where v is consistent, there is a consistent fixed point of Φ above v that is least among all the fixed points above Φ. It follows that there is a fixed point, v_3, that is least among those above $v_1 \oplus v_2$, and v_3 is consistent. A proof that v_3 exists can be produced that follows the 'approximate by a transfinite sequence' approach of the Knaster-Tarski theorem. One can show that every member of this sequence will be compatible with every consistent fixed point of Φ, and consequently the limit, v_3 will have this property. Then v_3 must be intrinsic.

Since the space of valuations is a complete lattice under $\leq_k$, $\bigvee \mathcal{I}$ exists. Once it has been shown that $\mathcal{I}$ is directed, it is not difficult to show $\bigvee \mathcal{I}$ is, itself, intrinsic. Consequently a biggest intrinsic fixed point of Φ exists.

Let X be a truth teller. In some of Kripke's fixed points X is *true*, in some, *false*. It follows that if v is an *intrinsic* fixed point, X can be neither *true* nor *false* in v. Consistency rules out X being $\top$. Consequently X must be $\bot$ in every intrinsic fixed point. But we have already seen that a truth teller is *false* in every GLF-stable valuation. *No GLF-stable valuation is intrinsic.*

This does not mean the notion of intrinsic plays no role in the present development. Rather, it must be relativized suitably.

Definition 10.3 A GLF-stable valuation v is *GLF-intrinsic* if $v \oplus w$ is consistent, for every consistent GLF-stable valuation w.

Arguments like those above show that the smallest GLF-stable valuation is GLF-intrinsic, and a largest GLF-intrinsic valuation exists. Beyond this, very little is known about the family of GLF-intrinsic valuations. We leave this as an open problem: what are the properties of GLF-intrinsic valuations, and how does the notion relate to Kripke's version of intrinsic.

References

[1] N. D. Belnap, Jr. A useful four-valued logic. In J. M. Dunn and G. Epstein, editors, *Modern Uses of Multiple-Valued Logic*. D. Reidel, 1977.

[2] J. M. Dunn. Intuitive semantics for first-degree entailments and coupled trees. *Philosophical Studies*, 29:149–168, 1976.

[3] K. Fine. The justification of negation as failure. In J. E. Fenstad, I. T. Frolov, and R. Hilpinen, editors, *Logic, Methodology and Philosophy of Science VIII*, pages 263–301, Amsterdam, 1989. North-Holland.

[4] M. C. Fitting. A Kripke/Kleene semantics for logic programs. *Journal of Logic Programming*, 2:295–312, 1985.

[5] M. C. Fitting. Notes on the mathematical aspects of Kripke's theory of truth. *Notre Dame Journal of Formal Logic*, 27:75–88, 1986.

[6] M. C. Fitting. Bilattices and the theory of truth. *Journal of Philosophical Logic*, 18:225–256, 1989.

[7] M. C. Fitting. Kleene's logic, generalized. *Journal of Logic and Computation*, 1:797–810, 1992.

[8] M. C. Fitting. The family of stable models. *Journal of Logic Programming*, 17:197–225, 1993.

[9] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In R. Kowalski and K. Bowen, editors, *Proc. of the Fifth Logic Programming Symposium*, pages 1070–1080, Cambridge, MA, 1988. MIT Press.

[10] M. L. Ginsberg. Multivalued logics: a uniform approach to reasoning in artificial intelligence. *Computational Intelligence*, 4:265–316, 1988.

[11] A. K. Gupta. Truth and paradox. *Journal of Philosophical Logic*, 11:1–60, 1982.

[12] H. G. Herzberger. Naive semantics and the liar paradox. *Journal of Philosophy*, 79:479–497, 1982.

[13] S. Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72:690–716, 1975. Reprinted in *New Essays on Truth and the Liar Paradox*, R. L. Martin, ed., Oxford (1983).

[14] Z. Manna and A. Shamir. The optimal approach to recursive programs. *Comm. ACM*, 20:824–831, 1977.

[15] R. L. Martin and P. W. Woodruff. On representing "true-in-l" in l. *Philosophia*, 5:217–221, 1975. Reprinted in *Recent Essays on Truth and the Liar paradox*, R. L. Martin editor, Oxford, 1984.

[16] A. Tarski. A lattice-theoretical theorem and its applications. *Pacific Journal of Mathematics*, 5:285–309, 1955.

[17] A. Visser. Four valued semantics and the liar. *Journal of Philosophical Logic*, 13:181–212, 1984.

[18] P. W. Woodruff. Paradox, truth and logic, part I. *Journal of Philosophical Logic*, 13:213–232, 1984.

[19] S. Yablo. Truth and reflection. *Journal of Philosophical Logic*, 14:297–349, 1985.