Truth and the Liar
in De Morgan-Valued Models

HANNES LEITGEB

Abstract   The aim of this paper is to give a certain *algebraic* account of truth: we want to define what we mean by *De Morgan-valued truth models* and show their *existence* even in the case of *semantical closure*: that is, languages may contain their own truth predicate if they are interpreted by De Morgan-valued models. Before we can prove this result, we have to repeat some basic facts concerning De Morgan-valued models in general, and we will introduce a notion of truth both on the object- and on the metalanguage level appropriate for such models. The definitions and the existence theorem are extensions of Kripke's, Woodruff's, and Visser's concepts and results concerning three- and four-valued truth models.

1 Preliminaries

1.1 De Morgan lattices   De Morgan lattices are lattices with an associated unary complement function having some intuitively attractive properties; they have been introduced in the late 1950s by various authors independently (Bialynicki-Birula and Rasiowa [4], Kalman [9], and Monteiro [15]) and have been used to give a semantics for Relevance Logic (Dunn [5]; in this context they are also called *intensional* lattices).

Definition 1.1   We call $\mathbb{M} = \langle \mathbf{M}, \leqq_{\mathbb{M}}, \wedge_{\mathbb{M}}, \vee_{\mathbb{M}}, \neg_{\mathbb{M}} \rangle$ a *De Morgan lattice* if

1. $\langle \mathbf{M}, \leqq_{\mathbb{M}}, \wedge_{\mathbb{M}}, \vee_{\mathbb{M}} \rangle$ is a distributive lattice ($\leqq_{\mathbb{M}}$ is the partial order of $\mathbb{M}$, $\wedge_{\mathbb{M}}$ and $\vee_{\mathbb{M}}$ are the corresponding binary infimum and supremum functions), and
2. $\neg_{\mathbb{M}} : \mathbf{M} \longrightarrow \mathbf{M}$ such that

   (a) for all $u \in \mathbf{M}$: $\neg_{\mathbb{M}} \neg_{\mathbb{M}} u = u$,

   (b) for all u,

   $$v \in \mathbf{M}: \neg_{\mathbb{M}}(u \wedge_{\mathbb{M}} v) = \neg_{\mathbb{M}} u \vee_{\mathbb{M}} \neg_{\mathbb{M}} v, \ \neg_{\mathbb{M}}(u \vee_{\mathbb{M}} v) = \neg_{\mathbb{M}} u \wedge_{\mathbb{M}} \neg_{\mathbb{M}} v.$$

That means, De Morgan lattices are distributive lattices with a complement function that is a *dual automorphism of period two*, that is, it satisfies the *law of double complement* and the *De Morgan rules*; it is easy to see that in a De Morgan lattice also the *law of contraposition* holds: for all u, $v \in \mathbf{M}$ $u \leqq_{\mathbb{M}} v$ if and only if $\neg_{\mathbb{M}} v \leqq_{\mathbb{M}} \neg_{\mathbb{M}} u$.

At first glance, De Morgan lattices are quite similar to Boolean algebras: indeed, every Boolean algebra is also a De Morgan lattice, and the variety of Boolean algebras is a subvariety of the variety of De Morgan lattices; but in contrast to Boolean algebras $u \wedge_{\mathbb{M}} \neg_{\mathbb{M}} u$ is *not* necessarily the bottom in a De Morgan lattice—actually a De Morgan lattice is not necessarily bounded at all—and $u \vee_{\mathbb{M}} \neg_{\mathbb{M}} u$ is not necessarily its top.

The following examples show that De Morgan lattices may indeed differ a lot from Boolean algebras:
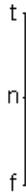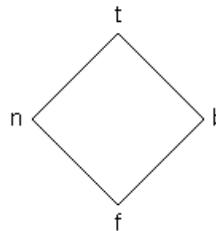
Figure 1: **Three**   Figure 2: **Four**

As you can read off from the way the complement function is defined in each of the two examples, it is possible that a De Morgan lattice has members which are *identical to their own De Morgan complement*. Let us call the De Morgan lattice in Figure 1 '**Three**' and the one in Figure 2 '**Four**'. According to a theorem in [9] every DeMorgan lattice is isomorphic to a sublattice of a product lattice of **Four**: thus this lattice plays the same role for the variety of De Morgan lattices as the two-member Boolean algebra for the variety of Boolean algebras.

Another interesting example of a De Morgan lattice—this time an *infinite* one—is the compact interval [0, 1] with real number order and the complement function $\neg_{\mathbb{M}} u = 1 - u$. If a De Morgan lattice is *bounded* we speak of a *De Morgan algebra*: in this case we call the smallest element **0** and the largest element **1**. Note that always $\neg_{\mathbb{M}} \mathbf{0} = \mathbf{1}$ and $\neg_{\mathbb{M}} \mathbf{1} = \mathbf{0}$.

1.2 De Morgan-valued models   Next, we suggest to take the members of a De Morgan lattice as candidates for values which are assigned to sentences. Of course, one immediately thinks of such values as truth values, that is, values representing the truth status of a formula. In Section 1.3 we will deal with such interpretations. But, at the moment, let us consider a *De Morgan-valued model* just as a way to connect the structure of a first-order language to the structure of a De Morgan lattice: by the structure of a language we mean its *syntax*, that is, that formulas are built from atomic formulas by negation, conjunction, disjunction, universal, and existential quantification; by the structure of a De Morgan lattice we mean its partial order and its complement function.

Let $\mathcal{L}$ be a first-order language corresponding to an alphabet $\mathcal{A}_S$ consisting of variables, the usual logical connectives and quantifiers, parentheses, and a symbol

set $\mathcal{S}$ of n-ary predicates P_k^n (for $n = 1, 2, \ldots$) and individual constants c_i. Let $\mathbb{M} = \langle \mathbf{M}, \leqq_{\mathbb{M}}, \wedge_{\mathbb{M}}, \vee_{\mathbb{M}}, \neg_{\mathbb{M}} \rangle$ be a De Morgan lattice.

Definition 1.2   An $\mathbb{M}$-valued $\mathcal{S}$-model (a *De Morgan-valued model*) is a pair $\mathfrak{M} = \langle D, \mathfrak{I} \rangle$ such that

1. D is a nonempty set (the *domain*).
2. $\mathfrak{I}$ is a function (the *interpretation function*) such that

   (a) for every n-ary predicate $P \in \mathcal{S}$: $\mathfrak{I}(P) : D^n \longrightarrow \mathbf{M}$,
   (b) for every constant $c \in \mathcal{S}$: $\mathfrak{I}(c) \in D$.

Later, we will be interested in extending $\mathbb{M}$-valued $\mathcal{S}$-models $\mathfrak{M} = \langle D, \mathfrak{I} \rangle$ to other De Morgan-valued models. In this context we will sometimes use the more complicated but somewhat more "transparent" notation,

$$\mathfrak{M} = \langle \mathcal{S}, \mathcal{L}, \mathbb{M}, D, \mathfrak{I}, \rangle$$

for models.

Variable assignments s are defined the same way as for classical models. If s is an assignment, $s\frac{d}{x}$ is the assignment identical to s, except that it maps x to d.

The sets of $\mathcal{S}$-terms and $\mathcal{S}$-formulas are defined inductively in the usual way. Accordingly, *evaluations* $Val_{\mathfrak{M},s}$ are also defined inductively: of course, the values of negations, conjunctions, and disjunctions are drawn back to complements, infima, and suprema; the quantification cases are considered as (possibly) infinite conjunctions and disjunctions.

Definition 1.3   If φ is an $\mathcal{S}$-formula and x is a variable, then

1. $Val_{\mathfrak{M},s}(\forall x \varphi) = \inf_{\mathbb{M}}\{Val_{\mathfrak{M},s\frac{d}{x}}(\varphi) | d \in D\}$;
2. $Val_{\mathfrak{M},s}(\exists x \varphi) = \sup_{\mathbb{M}}\{Val_{\mathfrak{M},s\frac{d}{x}}(\varphi) | d \in D\}$.

When we say that the values of quantified formulas are given by the infima or suprema of certain sets of values, we assume that these sets of values actually *do* have an infimum or supremum in $\mathbb{M}$! Of course, if we assume that $\mathbb{M}$ is finite or, more generally, *complete*, then we will not get into trouble at this point; but such a restriction is by no means necessary, since we only need that *certain* infinite infima and suprema exist: those, which are given by quantification, that is, definable in $\mathcal{L}$.

Important examples of De Morgan-valued models are, of course, those where $\mathbb{M}$ is the smallest nontrivial Boolean algebra, that is, the algebra of classical truth values. In this case *classical models* are back on stage again. If $\mathbb{M}$ is any Boolean algebra, we are dealing with *Boolean-valued models*, which have, for example, been used to prove independence results in set theory (see, e.g., Rosser [17]). If $\mathbb{M}$ is identical to **Three**, the corresponding $\mathbb{M}$-valued models are three-valued models the semantics of which is given by the *Strong Kleene scheme*; and if $\mathbb{M}$ is identical to **Four**, we are considering Belnap's *'useful four-valued logic'* [3]. As a final example, the compact unit interval of the reals together with the complement function from above leads to the infinitely-valued system $S_{\aleph}^{\supseteq}$ (we use the notation of Rescher [16], p. 344), which is also known from fuzzy logic. The notion of a De Morgan-valued model is the result of abstracting from all these examples.

A further natural constraint on De Morgan algebras which are to be used for De Morgan-valued models, is to demand some strengthening of finite distributivity to a version of *infinite* distributivity. But this may turn out to be a rather strong restriction: if, for example, $\mathbb{M}$ is Boolean and *completely* distributive, then $\mathbb{M}$ may be shown to be atomic (see Koppelberg [10], p. 215).

1.3 De Morgan-values considered as intensions   In this section we will present some ideas on how to interpret De Morgan-valued models properly. If we reconsider our examples from above, we see that—with the exception of general Boolean-valued models and $S_{\aleph}^{\supseteq}$—all of them are finitely valued, and each value is usually interpreted as a *truth* value: *t* and *f* as *true* and *false*, *n* as *neither true nor false*, *b* as *both true and false*.

Now we are heading for an interpretation of an *arbitrary* De Morgan-valued model: consequently, we cannot be sure that its corresponding De Morgan lattice is *finite*; therefore it will no longer do to attach finitely many labels to the members of the lattice, which denote some truth status. It is also very questionable, whether an *infinite* De Morgan lattice allows *any* truth status interpretation at all: the only interpretation of that kind I could think of is a *probability* or *possibility* interpretation; but such an interpretation does not seem to make sense for an arbitrary De Morgan lattice (though it does make sense for the compact real unit interval). Instead, in the context of a De Morgan-valued model, I suggest to interpret the members of a De Morgan lattice as *intensional* entities of a certain kind: as *propositions*. Let φ be a sentence in $\mathcal{L}$. Let $u \in \mathbb{M}$. We say that φ *expresses* (the proposition) u, if $Val_{\mathfrak{M}}(\varphi) = u$.

Propositions are sentence meanings. In the way I present them here, they are not (necessarily) objects which have a structure in themselves—contrary to, for example, Barwise [1], where propositions are considered as complex set theoretic objects; they are just members of a lattice. On the other hand, because they are members of a lattice, they share the De Morgan lattice structure: let $u, v \in \mathbb{M}$; we say that *u contains v*, if $u \leqq_{\mathbb{M}} v$.

In possible world semantics, propositions are analyzed as sets of possible worlds, φ expresses the very set of worlds where it is true and the relation of containment is defined by the subset relation. Something similar to a possible world semantics may also be given for De Morgan-valued models. Consider the De Morgan lattice $\mathbb{M} = \langle \mathbf{M}, \leqq, \wedge, \vee, \neg \rangle$ with

1. $\mathbf{M} := \mathfrak{X} \times \mathfrak{X}$;
2. $\mathfrak{X}$ is a ring of subsets of a fixed set X;
3. $\langle U_1, V_1 \rangle \leqq \langle U_2, V_2 \rangle$ if $U_1 \subseteq U_2$, $V_1 \supseteq V_2$;
4. $\langle U_1, V_1 \rangle \wedge \langle U_2, V_2 \rangle = \langle U_1 \cap U_2, V_1 \cup V_2 \rangle$;
5. $\langle U_1, V_1 \rangle \vee \langle U_2, V_2 \rangle = \langle U_1 \cup U_2, V_1 \cap V_2 \rangle$;
6. $\neg \langle U, V \rangle = \langle V, U \rangle$.

Dunn [5] calls De Morgan lattices of this kind *fields of polarities*. He has shown that every De Morgan lattice is isomorphic to such a field. Dunn's theorem shows the "degree of freedom" concerning complement functions one gains in De Morgan lattices compared to Boolean algebras: in the latter for any U we have *one and only one* V such that $\langle U, V \rangle \in \mathbf{M} = \mathbf{B}$; in the former there may be many such sets V. More

important, by Dunn's theorem we know that there is also a possible worlds semantics for De Morgan-valued models: it is just that a sentence now expresses a *pair* of world sets; the first one is the set of worlds where φ is definitely true; the second one is the set of worlds where φ is definitely false. The difference to classical possible worlds is constituted by the fact that these two sets are neither necessarily disjoint nor do they necessarily exhaust the set of *all* possible worlds.

Let us return to the language level. If φ expresses u, ψ expresses v, and u contains v, we say that φ *semantically implies* ψ; that is, by the concept of containment on the propositional level we also get a concept of semantical implication for sentences. If two sentences express the same proposition, we call them *semantically equivalent* or *synonymous*.

As we can see, De Morgan-valued models define some binary *semantical implication relation* on a language and they do this in a neat way. The resulting implication relation has some of the properties one usually wants to have: for example, φ semantically implies $\varphi \vee \psi$; if φ semantically implies ρ and ψ semantically implies ρ, then also $\varphi \vee \psi$ semantically implies ρ; $\varphi\frac{t}{x}$ semantically implies $\exists x \varphi$; and so on. Since De Morgan lattices are distributive we also have that $\varphi \wedge (\psi \vee \rho)$ is semantically equivalent to $(\varphi \wedge \psi) \vee (\varphi \wedge \rho)$. Moreover, φ is semantically equivalent to $\neg\neg\varphi$, $\neg(\varphi \wedge \psi)$ is semantically equivalent to $\neg\varphi \vee \neg\psi$, and $\neg\forall x \varphi$ is semantically equivalent to $\exists x \neg\varphi$. I have said that we usually want to have such properties when we think of an implication relation. But do we really? This is, of course, a matter of taste; for relevance logicians the properties above seem to be on entailment's shopping list; for the paraconsistent logician the possible absence of $(\varphi \wedge \neg\varphi)$ semantically implying ψ for arbitrary ψ is what she likes, since she does not want contradiction to spread over the whole language; on the other hand, it is just this facet that is rather untenable for the classical mathematician, who definitely wants to use his reductio ad absurdum: what he could do, at least, is to consider a certain subclass of De Morgan-valued models, namely, Boolean-valued models. But for intuitionists, the law of double negation makes De Morgan-valued models completely inadequate! In a nutshell, the class of De Morgan-valued models is large enough to contain *some* interesting semantical implication relations on languages. I am sure that there are many other ways of understanding De Morgan-valued models (some are discussed in Dunn [6]).

Note that, when we say that introducing De Morgan-valued models is a way to define a semantical implication relation on a language, it is important to keep in mind that such models do actually much more: they are *semantical* objects, that is, they also relate language to a domain of entities! This is the main difference between De Morgan-valued models and axiom systems for De Morgan-like implication; the notion of a universe does not play a role for the latter, but it does for models.

1.4 Designation in De Morgan-valued models   Nonclassical De Morgan-valued models give us a much more fine-grained implicational structure than classical models do; indeed, from the viewpoint of semantical implication, classical models are quite unacceptable, as one may easily see from the so-called paradoxes of material implication.

However, contrary to classical models, we can no longer say which sentences are evaluated as *true* and which sentences as *false* in a De Morgan-valued model. Worse,

we have no semantical analogue to *theories*: we cannot say that some sentence is a member of a theory while another is not.

In many-valued logic one usually solves this problem by taking a subset of the set of truth values as the set of so-called *designated* truth values: if some sentence has a designated truth value, we say that it is evaluated as true, else as false. In classical possible world semantics something similar is done when a world is designated as "the" real world. The natural way to define a set of designated values in a De Morgan-valued model is given by the concept of a *truth filter*: let $\mathfrak{M} = \langle \mathcal{S}, \mathcal{L}, \mathbb{M}, D, \mathfrak{I} \rangle$ be an $\mathbb{M}$-valued $\mathcal{S}$-model.

Definition 1.4   Let TF be a proper filter in $\mathbb{M}$ (a *truth filter*). Let φ be a sentence in $\mathcal{L}$. We say that

1. φ is *evaluated as true*, if $Val_{\mathfrak{M}}(\varphi) \in TF$,
2. φ is *evaluated as false*, else.

The notion of a *truth filter* is used by many authors; probably among the earliest by Belnap [2]: there truth filters had to be *ultra*filters in De Morgan lattices. On the side of languages, truth filters have characteristics quite similar to *theories*: for example, if φ is evaluated as true, and φ semantically implies ψ, then ψ is also evaluated as true. A word has to be said about our definition of 'evaluated as false': another possibility of defining falsehood would have been to say that a sentence $\varphi \in \mathcal{L}$ is *evaluated as false* if $Val_{\mathfrak{M}}(\neg\varphi) \in TF$. We leave the question to the reader, which definition is the more appropriate (if any). From now on, if we speak of truth filters, we will restrict ourselves to the definition given above. Thus no formula is *both* evaluated as true *and* as false.

Since the filter property is still rather weak, we should think about ways to strengthen it. In the literature we find three conditions that could be added (for Boolean-valued models these three conditions coincide):

1. a truth filter should be *prime*, that is, $u \vee_{\mathbb{M}} v \in TF$ if and only if $u \in TF$ or $v \in TF$;
2. a truth filter should be *maximal*, that is, it is not properly contained in any other proper filter;
3. a truth filter should be an *ultrafilter*, that is, for all $u \in \mathbb{M}$ it contains either u or $\neg u$, but not both.

Generally, it cannot be said which way of strengthening is "the" right one, since it is not so clear what counts as "right". Since De Morgan lattices are distributive, every maximal filter is prime, and since the De Morgan complement obeys the De Morgan rules, also every ultrafilter is prime. Therefore, we shall at least *add the condition of primeness to the definition of a truth filter*. Note that De Morgan lattices do not necessarily have maximal proper filters, whereas De Morgan algebras do. De Morgan algebras do not necessarily have ultrafilters. But both De Morgan lattices and De Morgan algebras always have prime filters.

If TF is a prime filter, conjunction and disjunction behave truth functionally as they should. This is not necessarily so in the case of negation.

Concerning quantification it is natural to consider prime truth filters having the following properties:

1. $\forall x\varphi$ is evaluated as true if and only if for all assignments s, $Val_{\mathfrak{M},s}(\varphi) \in TF$.
2. $\exists x\varphi$ is evaluated as true if and only if there is an assignment s such that $Val_{\mathfrak{M},s}(\varphi) \in TF$.

Of course, it would be enough to postulate just one of the two statements. If a truth filter has this property—which a prime filter does not necessarily have—also quantified sentences will behave truth functionally as expected.

2 De Morgan-valued truth models and semantical paradoxes   Now we want to give an application of our concept of De Morgan-valued models: we will show that, if languages are understood as interpreted by such models, they may contain their own truth predicates.

2.1 De Morgan-valued truth models   In his classical papers [18] and [19], Tarski presents, essentially, the following account of truth. Let $\mathcal{L}_1$ be a language (= a set of sentences) corresponding to a symbol set $\mathcal{S}_1$. Let $\mathcal{L}_2$ be a language corresponding to a symbol set $\mathcal{S}_2$ such that $\mathcal{S}_2$ contains a unary predicate *True* (the *truth predicate*), there is a partial surjection *ps* from the terms of $\mathcal{S}_2$ onto the sentences of $\mathcal{L}_1$, and $\mathcal{S}_1 \subseteq \mathcal{S}_2$, that is, $\mathcal{L}_1$ is a sublanguage of $\mathcal{L}_2$. $\mathcal{L}_1$ is the *object language*, $\mathcal{L}_2$ the *metalanguage*.

Now we can define what a truth theory is: let $\mathcal{TH}$ be a first-order theory.

Definition 2.1 (Convention T) $\mathcal{TH}$ is a (classical) *truth theory* for $\mathcal{L}_1$ in $\mathcal{L}_2$ if

$$\mathcal{TH} \vdash True(t) \longleftrightarrow \varphi$$

for all t, φ such that t is a term in $\mathcal{S}_2$ such that t is mapped to the sentence $\varphi \in \mathcal{L}_1$ by *ps* (thus, Convention T is actually relativized to the stipulated *ps*).

Thus, according to Tarski, $True(t)$ should (materially) imply the sentence φ, and vice versa.

Now we want to use Definition 2.1 as a guide which is to lead us to an intuitively plausible notion of a *De Morgan-valued truth model*, that is, a De Morgan-valued model which interprets the truth predicate in a plausible way. First of all, we suggest to replace the mapping *ps* from above by the interpretation mapping $\mathfrak{I}$ of a De Morgan-valued model such that $ps(t) = \varphi$ if and only if $\mathfrak{I}(t) = \varphi$. Secondly, we have to "put" the T-biconditionals of Definition 2.1 into the context of De Morgan-valued models. This can be done in various ways. One obvious way is to demand that $True(t)$ and φ should be semantically equivalent in every De Morgan-valued truth model. Call this the *implication reading* of Convention T.

Another possible reading of Convention T in this context is the *truth reading*: $True(t)$ and φ should either both be *evaluated as true* or both of them as *false* in a De Morgan-valued truth model. In classical models, the implication and the truth reading coincide; in De Morgan-valued models the latter reading corresponds to saying that the values of $True(t)$ and φ should either both be members of the given truth filter or not (recall Section 1.4): this is something completely different compared to the first reading! But, of course, if $True(t)$ and φ semantically imply each other, they are also either both members of the truth filter or not: that is, the implication reading *implies* the truth reading, but not necessarily vice versa! This is another reason why

we will basically be interested in the semantical implication reading of Convention T.

A third possible reading of Convention T in the light of general De Morgan-valued models is the *biconditional reading*: the sentence $True(t) \longleftrightarrow \varphi$ should be evaluated as true, that is, its value should be a member of the truth filter. But what is $True(t) \longleftrightarrow \varphi$? Now and in the following, we will understand $\varphi \longleftrightarrow \psi$ as short for $(\varphi \longrightarrow \psi) \wedge (\psi \longrightarrow \varphi)$ where $\varphi \longrightarrow \psi$ is a metalinguistic abbreviation of $\neg\varphi \vee \psi$. The problem with this reading is that, if $True(t) \longleftrightarrow \varphi$ is evaluated as true, this does not necessarily entail that $True(t)$ and φ semantically imply each other; corresponding counterexamples are easily constructed. Put differently, the connective $\longrightarrow$ as defined above does *not* in general represent the metalinguistic '$\leqq$' in the object language, as it does in the classical case. Of course, one could introduce such a connective in the object language, and we could correspondingly extend De Morgan algebras by a component used for the interpretation of this new connective (e.g., we might consider Heyting algebras with De Morgan complement). But note that this could have a negative effect on the general existence of truth models (e.g., in the Heyting algebra approach we might suffer from the Curry-Löb paradox). Therefore, we neglect the biconditional reading of Convention T in our considerations.

Since Tarski's efforts we know that if $\mathcal{L}_1 = \mathcal{L}_2 = \mathcal{L}_{True}$, that is, in the extreme case where a language contains *its own* truth predicate, there is generally *no classical truth model* for $\mathcal{L}_{True}$ in $\mathcal{L}_{True}$. The reason for this is that $\mathcal{L}_{True}$ may contain *self-referent* sentences. Not all of them are necessarily vicious, but some definitely are: the classical example is the so-called *Liar* sentence, which says about itself that it is not true. Intuitively, the Liar is true if and only if it is not. But, of course, this is impossible in a *classical* truth model. Languages that contain their own truth predicate or that are rather supposed to contain it, have been called *semantically closed* by Tarski.

Starting with Martin [13] and Kripke [11] in the mid seventies, we perceive an avalanche of publications which show that semantically closed languages indeed exist, if only the underlying classical semantics is altered. Among these new approaches we also find existence results for truth models, the truth values of which are members of a lattice, though these models are no longer Boolean-valued: examples are Kripke's Strong Kleene-valued truth models and Woodruff's [22] and Visser's [20] Belnap-valued truth models. As we have seen before, both of these models are actually *De Morgan-valued*; and indeed, in Kripke's and Woodruff's/Visser's accounts Convention T is understood in the implication reading. The concepts and results we present below, and which have been stimulated by these famous predecessors, are extensions of these accounts. Particularly, Kalman's theorem cited on p. 497 sheds some light on Woodruff's/Visser's existence result for truth models with a semantics based on **Four**: it indicates that an analogous existence result should be derivable for *all* De Morgan-valued models; however, we will prove existence later *without* making use of Woodruff's/Visser's or Kalman's theorems.

Let us now introduce our notion of a *De Morgan-valued truth model*: let $\mathcal{L}_1$ and $\mathcal{L}_2$ be as above.

Definition 2.2   We call an $\mathbb{M}^*$-valued $\mathcal{S}_2$-model $\mathfrak{M}^* = \langle \mathcal{S}_2, \mathcal{L}_2, \mathbb{M}^*, D^*, \mathfrak{I}^* \rangle$ a (De Morgan-valued) *truth model* for $\mathcal{L}_1$ in $\mathcal{L}_2$ if

1. $D^* \supseteq \mathcal{L}_1$;
2. for all $\varphi \in \mathcal{L}_1$ there is a constant t in $\mathcal{S}_2$, such that $\mathfrak{I}^*(t) = \varphi$;
3. for all $\varphi \in \mathcal{L}_1$ and all constants t with $\mathfrak{I}^*(t) = \varphi$,

$$Val_{\mathfrak{M}^*}(True(t)) = Val_{\mathfrak{M}^*}(\varphi).$$

The point of this definition is to postulate that $True(t)$ should express the same proposition as φ if $\mathfrak{I}^*(t) = \varphi$. Note that we use the notation '*' because we want to extend models $\mathfrak{M}$ to truth models $\mathfrak{M}^*$ later. The object language is a subset of the domain of the truth model, since we want to talk about properties of *sentences* in the truth model. An alternative would have been to use some objects in the domain, for example, natural numbers, as *codes* of sentences; but we refrain from this strategy here for the sake of simplicity.

Since $\mathcal{L}_1 = \mathcal{L}_2$ seems to be the problematic case concerning the existence of truth models, we will concentrate on such truth models. Let $\mathcal{L}_{True} = \mathcal{L}_1 = \mathcal{L}_2$. The question is: Are there truth models for $\mathcal{L}_{True}$ in $\mathcal{L}_{True}$? Surely, we do not only want to give instances of such truth models but we want to show a general existence result; to do so we have to introduce yet another notion: De Morgan-valued truth extensions.

2.2 De Morgan-valued truth extensions   We are going to present a "natural" class of truth models: the idea is to start with a *ground model*, that is, a model of a language $\mathcal{L}$ *without* truth predicate, and just as we extend $\mathcal{L}$ *syntactically* to a language $\mathcal{L}_{True}$ *with* truth predicate, we extend the ground model *semantically* to a truth model. Consequently, this truth model is called an *extension* of the ground model:

Definition 2.3   By a *truth extension* of an $\mathbb{M}$-valued $\mathcal{S}$-model $\mathfrak{M} = \langle \mathcal{S}, \mathcal{L}, \mathbb{M}, D, \mathfrak{I} \rangle$ we mean an $\mathbb{M}^*$-valued $\mathcal{S}_{True}$-truth model $\mathfrak{M}^* = \langle \mathcal{S}_{True}, \mathcal{L}_{True}, \mathbb{M}^*, D^*, \mathfrak{I}^* \rangle$ such that

1. $\mathcal{S} \subseteq \mathcal{S}_{True}$, $\mathcal{L}$ is a sublanguage of $\mathcal{L}_{True}$;
2. $\mathbb{M}$ is a sublattice of $\mathbb{M}^*$;
3. $D \subseteq D^*$;
4. for all constants c in $\mathcal{S}$, $\mathfrak{I}^*(c) = \mathfrak{I}(c)$,
   for all n-ary predicates P in $\mathcal{S}$, $\mathfrak{I}^*(P) \restriction_{D^n} \equiv \mathfrak{I}(P)$;
5. for all $\varphi \in \mathcal{L}$: $Val_{\mathfrak{M}}(\varphi) = Val_{\mathfrak{M}^*}(\varphi)$.

We call $\mathfrak{M}$ a *ground model* (for the ground language $\mathcal{L}$).

Remark 2.4   The transition from a ground model to a truth model by truth extension is the generalization of a common procedure in the literature; it should not be mistaken for some transition from object- to metalanguage: we want to emphasize again that $\mathcal{L}_{True}$ functions as object- and metalanguage at the same time. The most important point regarding this transition is that all semantical implications in the ground model are *preserved* in the truth extension and *no semantical implications between sentences in the ground language are "added"* in the transition.

Remark 2.5   If a truth filter TF is associated with the ground model $\mathfrak{M}$ it is plausible to add a truth filter TF^*—if possible—to the truth model $\mathfrak{M}^*$ which satisfies the following condition: $TF^* \cap \mathbb{M} = TF$. This means $\varphi \in \mathcal{L}$ is evaluated as true/false in the ground model $\mathfrak{M}$ if and only if it is evaluated as true/false in the truth extension $\mathfrak{M}^*$.

2.3 The existence of De Morgan-valued truth extensions   The basic question is now: Given some ground model, is there a truth extension? First we will describe what the ground models we want to extend should look like; afterward we will show that they *can* be extended to truth models.

2.3.1 Our ground models $\mathfrak{M}$ One peculiar property of the ground models which we will use is that they are already "prepared" to be extended to truth models, in the sense that they already contain $\mathcal{L}_{True}$ in their domain. But, of course, this is no *real* restriction: otherwise we would embed the ground model into one with this property. To make things more readable, we will furthermore assume that $\mathcal{L}_{True}$ contains quotation marks, that is, if φ is a sentence in $\mathcal{L}_{True}$ then $\ulcorner\varphi\urcorner$ is a term.

Let $\mathcal{S} = \mathcal{S}_{True}\backslash\{True\}$. Moreover, let $C\mathcal{T} = C\mathcal{T}_{\mathcal{S}_{True}}$ be the set of constant terms, $\mathcal{T} = \mathcal{T}_{\mathcal{S}_{True}}$ the set of terms, $\mathcal{F} = \mathcal{F}_{\mathcal{S}_{True}}$ the set of formulas. Finally, let $\mathcal{L} = \{\varphi \in \mathcal{L}_{True}|$ if $True$ occurs in φ, it occurs under quotation marks$\}$.

This is what our ground models look like:

1. Let $\mathfrak{M} = \langle \mathcal{S}, \mathcal{L}, \mathbb{M}, D, \mathfrak{I}\rangle$ with

    (a) $D \supseteq \mathcal{L}_{True}$,
    (b) for all $\varphi \in \mathcal{L}_{True}$: $\mathfrak{I}(\ulcorner\varphi\urcorner) = \varphi$.

2. Moreover, we assume that for every $u \in \mathbb{M}$ there is a $\varphi \in \mathcal{L}$ such that $Val_{\mathfrak{M}}(\varphi) = u$. This property says that $\mathbb{M}$ is already "covered" by the values of sentences. This is fulfilled if, for example, $\mathbb{M}$ is the Boolean algebra with two members; moreover, if each member in the domain has an individual constant as its name, one can simply take $\{Val_{\mathfrak{M}}(\varphi)|\varphi \in \mathcal{L}\}$ as "new" lattice $\mathbb{M}$ which has our desired property.

3. *Prime Filter Property*   For all formulas $\exists x\varphi$ in $\mathcal{F}$ and all prime filters PF of $\mathbb{M}$ we assume that

$$Val_{\mathfrak{M}}(\exists x\varphi) \in PF$$

    if and only if there is an assignment s such that $Val_{\mathfrak{M},s}(\varphi) \in PF$. This is kind of a technical property; the restriction to ground models of this kind is severe, but it still includes enough interesting cases.

If $\mathbb{M}$ has the property that for all $X \subseteq \mathbf{M}$, for all prime filters PF of $\mathbb{M}$ we have that $sup(X) \in PF$ if and only if some member of X is contained in PF, then a model based on $\mathbb{M}$ has the prime filter property a fortiori (in such a case, we might say that $\mathbb{M}$ has the *full prime filter property*; this was suggested to us by an anonymous referee).

Example 2.6   If $\mathbb{M}$ satisfies the ascending chain condition (and thus, by De Morgan properties, the descending chain condition), in particular, if $\mathbb{M}$ is finite, then $\mathbb{M}$ has the full prime filter property. For example, the De Morgan algebra with eight members, which is referred to as $\mathbb{M}_0$ in Dunn ([5], p. 198), plays an important role for the algebraic semantics of Relevance Logic; since $\mathbb{M}_0$ is finite, it satisfies, of course, the full prime filter property.

Example 2.7   Let $\mathbb{M}$ be a subalgebra of the compact unit interval such that $\mathbf{M} = \{\frac{1}{2^n}|n \in \mathbb{N}_0\} \cup \{1 - \frac{1}{2^n}|n \in \mathbb{N}_0\}$. $\mathbb{M}$ is a complete De Morgan algebra. Now let $\mathfrak{M}$

be a ground model such that for all predicates P, $X_P = \{u \in \mathbf{M} | \exists \langle d_1, \dots, d_n \rangle \in D^n$ such that $\mathfrak{I}(P)(d_1, \dots, d_n) = u\}$ has the property that (i) if there is a sequence of members of X_P which approach 1, then $1 \in X_P$ and (ii) if there is a sequence of members of X_P which approach 0, then $0 \in X_P$. It is easy to see that in this case $\mathfrak{M}$ has the prime filter property.

Example 2.8   If $\mathbb{M}$ is Boolean, it may be embedded into a Boolean algebra which has the full prime filter property (this may be done by a kind of *Henkin* extension; see Lemma 56 in Leitgeb [12]).

On the other hand, for example, the compact unit interval $[0, 1]$ does *not* have the full prime filter property.

Remark 2.9   Note that in the ground model we may also have any (interpreted) *syntactical* theory of any strength and expressiveness you want: there may be predicates the interpretation of which is 'ρ is the conjunction of φ and ψ' or 'φ is what you get when you substitute t for x in ψ' and so on.

If our ground model satisfies these three conditions, it may be shown to have a truth extension.

2.3.2 The existence theorem

Theorem 2.10   *For any $\mathbb{M}$-valued $\mathcal{S}$-model $\mathfrak{M} = \langle \mathcal{S}, \mathcal{L}, \mathbb{M}, D, \mathfrak{I} \rangle$ satisfying properties 2.3.1 (1)–(3) from above there is a truth extension*

$$\mathfrak{M}^* = \langle \mathcal{S}_{True}, \mathcal{L}_{True}, \mathbb{M}^*, D, \mathfrak{I}^* \rangle.$$

Remark 2.11   Actually, we will show more: there is always a truth extension with the properties

1. $\mathbb{M}^*$ is a *complete* De Morgan algebra,
2. if $\mathbb{M}$ is a De Morgan *algebra*, bottom and top of $\mathbb{M}$ and $\mathbb{M}^*$ coincide.

Proof of Theorem 2.10:   The plan of the proof is the following. We extend $\mathfrak{M}$ to its diagram expansion. Then we will basically embed the resulting model into one which is defined over a *complete* field of polarities, of which $\mathbb{M}$ is a sublattice. Note that this will need some kind of "cosmetic" procedure: if we had defined truth extensions differently, it would have been enough, if $\mathbb{M}$ were just *isomorphic* to a sublattice of $\mathbb{M}^*$. In this latter model we will study the possible interpretations of the truth predicate; we choose one which is a *fixed interpretation* under some appropriately defined "jump" operation: such an interpretation will be an adequate interpretation of the truth predicate. Its existence is ensured by the *Knaster-Tarski Fixed Point Theorem*. The theorem may be applied, although, for example, negation does *not* give rise to monotonic jump operators per se: but first, odd number occurences of the negation sign may be eliminated over De Morgan-valued models, and second, an atomic sentence and its negation are not so "tightly" related as in Boolean-valued models, indeed, they may be considered independently; this additional degree of freedom is enough to prove existence of a truth extension. Put in a different way, truth may be defined by a *positive inductive definition* in De Morgan-valued models; for more on such inductive definitions see, for example, McGee [14], Chapter 5.

1. First, we extend $\mathfrak{M} = \langle S, L, \mathbb{M}, D, \mathfrak{I} \rangle$ to the diagram expansion $\mathfrak{M}_1 = \langle S_1, L_1, \mathbb{M}, D, \mathfrak{I}_1 \rangle$; that is, add individual constants for each member of D and interpret them in the obvious way such that each member of the domain is denoted by a constant; what we get is $S_1 \supseteq S$, $L_1 \supseteq L$, and $\mathfrak{I}_1$ is an extension of $\mathfrak{I}$. It follows that quantification may now be understood substitutionally, since every member of the domain has a name. $\mathfrak{M}_1$ is still a model since no new definable sets of values in $\mathbf{M}$ are introduced by adding the constants. It follows that for all $\varphi \in L$, $Val_{\mathfrak{M}}(\varphi) = Val_{\mathfrak{M}_1}(\varphi)$.

2. Let $u \in \mathbf{M}$. Then we define

$$PF(u) := \{X \subseteq \mathbf{M} \ \left| \ \begin{array}{l} \text{(i) } u \in X, \\ \text{(ii) } X \text{ is a proper prime filter} \end{array} \right\}.$$

3. From Stone's Representation Theorem for distributive lattices we know that PF considered as a mapping

$$PF: \begin{array}{l} \mathbf{M} \longrightarrow \{PF(u) | u \in \mathbf{M}\} \\ u \longmapsto PF(u) \end{array}$$

is an isomorphism from $\mathbb{M}$ onto the lattice $\langle \{PF(u) | u \in \mathbf{M}\}, \subseteq, \cap, \cup \rangle$. Note that if $\mathbb{M}$ is a De Morgan algebra, its bottom is mapped to $\varnothing$ by PF and its top to the set of all proper prime filters of $\mathbb{M}$.

4. Since PF is an isomorphism and quantification in $\mathfrak{M}_1$ is substitutional, we have (let $CT_1 = CT_{S_1}$)

$$\begin{aligned} PF(Val_{\mathfrak{M}_1}(\exists x \varphi)) &= PF(\sup\{Val_{\mathfrak{M}_1}(\varphi\tfrac{t}{x}) | t \in CT_1\}) \\ &= \sup\{PF(Val_{\mathfrak{M}_1}(\varphi\tfrac{t}{x})) | t \in CT_1\} \text{ (because the prime} \\ &\qquad\qquad\qquad\qquad\qquad \text{filters in } \mathbb{M} \text{ have the prime filter property)} \\ &= \bigcup_{t \in CT_1} PF(Val_{\mathfrak{M}_1}(\varphi\tfrac{t}{x})). \end{aligned}$$

Similarly, $PF(Val_{\mathfrak{M}_1}(\forall x \varphi)) = \bigcap_{t \in CT_1} PF(Val_{\mathfrak{M}_1}(\varphi\tfrac{t}{x}))$, since for any prime filter F, for all definable sets $X \subseteq F$, it holds that $\inf X \in F$. This follows from the prime filter property by applying it to the set-complement $(\neg F)^C$ of the set $\neg F$ of complemented values of F $((\neg F)^C$ is also a prime filter).

5. In the following we will denote the set $\{X \subseteq \mathbf{M} | X \text{ is a proper prime filter}\}$ by 'Pr$(\mathbf{M})$'. Let $\mathbb{M}_2 = \langle \mathbf{M}_2, \leq_2, \wedge_2, \vee_2, \neg_2 \rangle$ be the field of polarities where

$$\mathbf{M}_2 := \wp(\text{Pr}(\mathbf{M})) \times \wp(\text{Pr}(\mathbf{M})).$$

Obviously, $\mathbb{M}_2$ is complete and for $A \subseteq \mathbf{M}_2$,

$$\inf A = \left\langle \bigcap_{\langle U, V \rangle \in A} U, \bigcup_{\langle U, V \rangle \in A} V \right\rangle, \sup A = \left\langle \bigcup_{\langle U, V \rangle \in A} U, \bigcap_{\langle U, V \rangle \in A} V \right\rangle.$$

6. Let $\mathbb{L} := \langle \wp(\text{Pr}(\mathbf{M})), \subseteq, \cap, \cup \rangle$. $\mathbb{L}$ is a complete lattice. We will need that $\bigcap$ and $\bigcup$ are monotone on $\mathcal{L}$, that is, if $\{U_j | j \in J\}$, $\{V_j | j \in J\}$ are families of subsets of Pr$(\mathbf{M})$ with $U_j \subseteq V_j$, where J is an arbitrary set of indices, it follows that $\bigcap_{j \in J} U_j \subseteq \bigcap_{j \in J} V_j$ and $\bigcup_{j \in J} U_j \subseteq \bigcup_{j \in J} V_j$.

7. Let $G = \{g | g : \mathcal{L}_{True} \longrightarrow \wp(\text{Pr}(\mathbf{M}))\}$. Let $g_1 \leqq_G g_2$ if for all $\varphi \in \mathcal{L}_{True}$: $g_1(\varphi) \subseteq g_2(\varphi)$. G with $\leqq_G$ is a complete lattice because $\mathbb{L}$ is.

8. Let $\mathcal{L}' = \mathcal{L}'_{\mathcal{S}'}$, where $\mathcal{S}' := \mathcal{S}_1 \cup \{True\}$.

9. Moreover, we fix some member $\mathbf{A}_0$ of $\wp(\text{Pr}(\mathbf{M}))$.

10. Now we define inductively a family of functions $h_g : \mathcal{L}' \longrightarrow \wp(\text{Pr}(\mathbf{M}))$, each relative to some $g \in G$.

    (a) For atomic φ, with $\varphi \in \mathcal{L}_1$,
    $$h_g(\varphi) := PF(Val_{\mathfrak{M}_1}(\varphi)).$$

    (b) For atomic φ, with $\varphi = True(t)$,
    $$h_g(\varphi) := \begin{cases} g(\psi) & \text{if } \mathfrak{I}_1(t) = \psi \text{ for } \psi \in \mathcal{L}_{True}, \\ \mathbf{A}_0 & \text{else.} \end{cases}$$

    (c) For $\neg\varphi$, with φ atomic and $\varphi \in \mathcal{L}_1$,
    $$h_g(\neg\varphi) := PF(Val_{\mathfrak{M}_1}(\neg\varphi)).$$

    (d) For $\neg\varphi$, with φ atomic and $\varphi = True(t)$,
    $$h_g(\neg\varphi) := \begin{cases} g(\neg\psi) & \text{if } \mathfrak{I}_1(t) = \psi \text{ for } \psi \in \mathcal{L}_{True}, \\ \mathbf{A}_0 & \text{else.} \end{cases}$$

    (e) $h_g(\varphi \wedge \psi) := h_g(\varphi) \cap h_g(\psi)$, $h_g(\varphi \vee \psi) := h_g(\varphi) \cup h_g(\psi)$.

    (f) $h_g(\neg(\varphi \wedge \psi)) := h_g(\neg\varphi \vee \neg\psi)$, $h_g(\neg(\varphi \vee \psi)) := h_g(\neg\varphi \wedge \neg\psi)$.

    (g) $h_g(\forall x \varphi) := \bigcap_{t \in CT_1} h_g(\varphi_x^t)$, $h_g(\exists x \varphi) := \bigcup_{t \in CT_1} h_g(\varphi_x^t)$.

    (h) $h_g(\neg\forall x \varphi) := h_g(\exists x \neg\varphi)$, $h_g(\neg\exists x \varphi) := h_g(\forall x \neg\varphi)$.

    (i) $h_g(\neg\neg\varphi) := h_g(\varphi)$.

    Since the value of $h_g(\varphi)$ is either directly set or drawn back to some value(s) $h_g(\psi)$ such that ψ has smaller rank than φ, the definition of h_g is sound.

11. Next we define a family of $\mathbb{M}_2$-valued models $\mathfrak{M}_g$ of $\mathcal{L}'$. Let $D_g := D$. Let $\mathfrak{I}_g$ be an interpretation function such that

    (a) $\mathfrak{I}_g(t) := \mathfrak{I}_1(t)$;

    (b) $[\mathfrak{I}_g(P)](d_0, \ldots, d_{n-1}) :=$
    $$\langle PF([\mathfrak{I}_1(P)](d_0, \ldots, d_{n-1})), PF(\neg[\mathfrak{I}_1(P)](d_0, \ldots, d_{n-1})) \rangle;$$

    (c) $[\mathfrak{I}_g(True)](d) := \begin{cases} \langle g(d), g(\neg d) \rangle & \text{if } d = \varphi \text{ for } \varphi \in \mathcal{L}_{True}, \\ \langle \mathbf{A}_0, \mathbf{A}_0 \rangle & \text{else.} \end{cases}$

    Thus, $\mathfrak{M}_g = \langle \mathcal{S}', \mathcal{L}', \mathbb{M}_2, D, \mathfrak{I}_g \rangle$. Each $\mathfrak{M}_g$ is defined relative to some $g \in G$. All infima and suprema arising from $Val_{\mathfrak{M}_g}$ exist in $\mathbb{M}_2$ since $\mathbb{M}_2$ is complete.

12. It follows from the definition of $\mathfrak{M}_g$ that for all $\varphi \in \mathcal{L}'$,

$$Val_{\mathfrak{M}_g}(\varphi) = \langle h_g(\varphi), h_g(\neg\varphi) \rangle.$$

    The proof is by standard induction. We use the definitions of h_g and $\mathbb{M}_2$, and that $\mathfrak{I}_g$ and $\mathfrak{I}_1$ do not differ on terms.

13. Furthermore, it follows (again by standard induction) from the definition of $\mathfrak{M}_g$ that for all $\varphi \in \mathcal{L}_1$,

$$Val_{\mathfrak{M}_g}(\varphi) = \langle PF(Val_{\mathfrak{M}_1}(\varphi)), PF(Val_{\mathfrak{M}_1}(\neg\varphi)) \rangle.$$

14. Thus, for all $\varphi, \psi \in \mathcal{L}_1$,

$$Val_{\mathfrak{M}_1}(\varphi) \leqq Val_{\mathfrak{M}_1}(\psi) \text{ iff } Val_{\mathfrak{M}_g}(\varphi) \leqq_2 Val_{\mathfrak{M}_g}(\psi).$$

For the proof, use (13) contraposition, and the fact that PF is an isomorphism. This means that the sublattice $\{Val_{\mathfrak{M}_g}(\varphi)|\varphi \in \mathcal{L}_1\}$ of $\mathbb{M}_2$ is isomorphic to $\mathbb{M}$ via the isomorphism $f(Val_{\mathfrak{M}_g}(\varphi)) = Val_{\mathfrak{M}_1}(\varphi)$. Note that f is onto because $\mathbb{M}$ is "covered" by the values $Val_{\mathfrak{M}_1}(\varphi) = Val_{\mathfrak{M}}(\varphi)$ of sentences $\varphi \in \mathcal{L} \subseteq \mathcal{L}_1$.

15. Let $\mathbb{M}_2'$ be the lattice you get when you replace $\{Val_{\mathfrak{M}_g}(\varphi)|\varphi \in \mathcal{L}_1\}$ in $\mathbb{M}_2$ by $\mathbb{M}$ via an isomorphism $f' : \mathbb{M}_2 \longrightarrow \mathbb{M}_2'$ with $f' \restriction_{\{Val_{\mathfrak{M}_g}(\varphi)|\varphi \in \mathcal{L}_1\}} \equiv f$. $\mathbb{M}$ is a sublattice of $\mathbb{M}_2'$; by (13) the bottom and the top of $\mathbb{M}$ and $\mathbb{M}_2'$ coincide.

16. Now, we "redefine" our model $\mathfrak{M}_g$ in a way that its associated lattice is $\mathbb{M}_2' \supseteq \mathbb{M}$: let $D_g' := D$; let $\mathfrak{I}_g'(t) := \mathfrak{I}_1(t)$, $[\mathfrak{I}_g'(P)](d_0, \ldots, d_{n-1}) := f'([\mathfrak{I}_g(P)](d_0, \ldots, d_{n-1}))$, $[\mathfrak{I}_g'(True)](d) := f'([\mathfrak{I}_g(True)](d))$. $\mathfrak{M}_g' = \langle S', \mathcal{L}', \mathbb{M}_2', D, \mathfrak{I}_g' \rangle$ is a model. Each $\mathfrak{M}_g'$ is defined relative to some $g \in G$.

17. It follows immediately that for all $\varphi \in \mathcal{L}'$,

$$Val_{\mathfrak{M}_g'}(\varphi) = f'(Val_{\mathfrak{M}_g}(\varphi)).$$

18. This implies for all $\varphi \in \mathcal{L}$, $Val_{\mathfrak{M}_g'}(\varphi) = Val_{\mathfrak{M}}(\varphi)$, since for all such φ we have

$$Val_{\mathfrak{M}_g'}(\varphi) = f'(Val_{\mathfrak{M}_g}(\varphi)) = f(Val_{\mathfrak{M}_g}(\varphi)) = Val_{\mathfrak{M}_1}(\varphi) = Val_{\mathfrak{M}}(\varphi).$$

We have used (17), (15), and (1).

19. Let $g_1, g_2 \in G$ with $g_1 \leqq_G g_2$. This implies that for all $\varphi \in \mathcal{L}'$, $h_{g_1}(\varphi) \subseteq h_{g_2}(\varphi)$.

    Proof: This is easily shown by induction.

    (a) For atomic φ, with $\varphi \in \mathcal{L}_1$, $h_{g_1}(\varphi) = PF(Val_{\mathfrak{M}_1}(\varphi)) = h_{g_2}(\varphi)$.

    (b) For atomic φ, with $\varphi = True(t)$,

    $$h_g(\varphi) = \begin{cases} g(\psi) & \text{if } \mathfrak{I}_1(t) = \psi \text{ for } \psi \in \mathcal{L}_{True}, \\ \mathbf{A}_0 & \text{else,} \end{cases}$$

    but since $g_1(\psi) \subseteq g_2(\psi)$ by assumption and $\mathbf{A}_0 \subseteq \mathbf{A}_0$, $h_{g_1}(\varphi) \subseteq h_{g_2}(\varphi)$.

    (c) For $\neg\varphi$, with φ atomic and $\varphi \in \mathcal{L}_1$,

    $$h_{g_1}(\neg\varphi) = PF(Val_{\mathfrak{M}_1}(\neg\varphi)) = h_{g_2}(\neg\varphi).$$

    (d) For $\neg\varphi$, with φ atomic and $\varphi = True(t)$,

    $$h_g(\neg\varphi) = \begin{cases} g(\neg\psi) & \text{if } \mathfrak{I}_1(t) = \psi \text{ for } \psi \in \mathcal{L}_{True}, \\ \mathbf{A}_0 & \text{else,} \end{cases}$$

    but since $g_1(\neg\psi) \subseteq g_2(\neg\psi)$ by assumption and $\mathbf{A}_0 \subseteq \mathbf{A}_0$,

    $$h_{g_1}(\neg\varphi) \subseteq h_{g_2}(\neg\varphi).$$

    (e) For all further cases use what we said in (6).   □

20. Now let $J : \qquad G \longrightarrow G$

$$[J(g)](\varphi) := h_g(\varphi)$$

for all $\varphi \in \mathcal{L}_{True}$. In the literature such a function J is often called a "Jump".

21. J is monotone on G since, if $g_1, g_2 \in G$ with $g_1 \leqq_G g_2$, according to (19) it holds that for all $\varphi \in \mathcal{L}'$, $h_{g_1}(\varphi) \subseteq h_{g_2}(\varphi)$. Therefore, also for all $\varphi \in \mathcal{L}_{True}$, $[J(g_1)](\varphi) = h_{g_1}(\varphi) \subseteq h_{g_2}(\varphi) = [J(g_2)](\varphi)$. That is, $J(g_1) \leqq_G J(g_2)$.

22. Since G is complete and J is monotone on G, it follows from the Knaster-Tarski Fixed Point Theorem (see, e.g., Hermes [8]) that there is a function g^* such that $J(g^*) = g^*$.

23. This implies

(a) For all $\varphi \in \mathcal{L}_{True}$, $[J(g^*)](\varphi) := h_{g^*}(\varphi)$, according to the definition of J.

(b) For all $\varphi \in \mathcal{L}_{True}$, $[J(g^*)](\varphi) := g^*(\varphi)$, since g^* is a fixed function under J.

(c) Therefore, for all $\varphi \in \mathcal{L}_{True}$, $h_{g^*}(\varphi) := g^*(\varphi)$.

(d) Put together, we have for $\mathfrak{I}_{g^*}(t) = \mathfrak{I}_1(t) = \psi$ with $\psi \in \mathcal{L}_{True}$,

$$
\begin{aligned}
Val_{\mathfrak{M}'_{g^*}}(True(t)) &= [\mathfrak{I}'_{g^*}(True)](\mathfrak{I}'_{g^*}(t)), \text{ according to definition,} \\
&= f'([\mathfrak{I}_{g^*}(True)](\mathfrak{I}_{g^*}(t))) \\
&= f'(\langle g^*(\mathfrak{I}_1(t)), g^*(\neg \mathfrak{I}_1(t))\rangle), \\
&\qquad \text{according to what we have just seen,} \\
&= f'(\langle h_{g^*}(\psi), h_{g^*}(\neg\psi)\rangle), \text{ since (12) holds,} \\
&= f'(Val_{\mathfrak{M}_{g^*}}(\psi)), \text{ because of (17),} \\
&= Val_{\mathfrak{M}'_{g^*}}(\psi).
\end{aligned}
$$

24. If we now only look at $\mathcal{S}_{True}$ instead of $\mathcal{S}'$ and, respectively, at $\mathcal{L}_{True}$ instead of $\mathcal{L}'$, we have that there is a truth extension $\mathfrak{M}^* = \langle \mathcal{S}_{True}, \mathcal{L}_{True}, \mathbb{M}^*, D, \mathfrak{I}^* \rangle$, where

(a) $\mathbb{M}^* := \mathbb{M}'_2$,

(b) $\mathfrak{I}^* := \mathfrak{I}'_{g^*} \restriction_{\mathcal{S}_{True}}$.

Proof: First we have two remarks.

Remark 2.12   For all $\varphi \in \mathcal{L}_{True}$, $Val_{\mathfrak{M}'_{g^*}}(\varphi) = Val_{\mathfrak{M}^*}(\varphi)$. This is true, because $Val_{\mathfrak{M}'_{g^*},s}(\varphi) = Val_{\mathfrak{M}^*,s}(\varphi)$ for all atomic formulas $\varphi \in \mathcal{F}_{True}$ and all assignments s in $\mathfrak{M}'_{g^*}$ and $\mathfrak{M}^*$ ($\mathfrak{M}'_{g^*}$ and $\mathfrak{M}^*$ have the same domain), and by induction, for all formulas $\varphi \in \mathcal{F}_{True}$ and all assignments. For example, in the case of $\varphi = \forall x\psi \in \mathcal{F}_{True}$,

$$
\begin{aligned}
Val_{\mathfrak{M}'_{g^*},s}(\forall x\psi) &= \inf_{\mathbb{M}'_2}\{Val_{\mathfrak{M}'_{g^*},s\frac{d}{x}}(\psi)|d \in D\}, \text{ by inductive hypothesis,} \\
&= \inf_{\mathbb{M}'_2}\{Val_{\mathfrak{M}^*,s\frac{d}{x}}(\psi)|d \in\} \\
&= \inf_{\mathbb{M}^*}\{Val_{\mathfrak{M}^*,s\frac{d}{x}}(\psi)|d \in D^*\} \\
&= Val_{\mathfrak{M}^*,s}(\forall x\psi).
\end{aligned}
$$

Remark 2.13   For all $\varphi \in L$, $Val_{\mathfrak{M}}(\varphi) = Val_{\mathfrak{M}^*}(\varphi)$. The reason for this is that for all $\varphi \in L$, $Val_{\mathfrak{M}}(\varphi) = Val_{\mathfrak{M}'_{g*}}(\varphi)$ because of (18), and $Val_{\mathfrak{M}^*}(\varphi) = Val_{\mathfrak{M}'_{g*}}(\varphi)$, as we have seen in Remark 2.12.

Now we know that $\mathfrak{M}^*$ is a truth model, since for all $\varphi \in L_{True}$ and all constants t with $\mathfrak{I}^*(t) = \varphi$

$$Val_{\mathfrak{M}^*}(True(t)) = Val_{\mathfrak{M}'_{g*}}(True(t)) = Val_{\mathfrak{M}'_{g*}}(\varphi) = Val_{\mathfrak{M}^*}(\varphi). \qquad \square$$

We have used Remark 2.12 again and proof item 23. Finally, this implies that $\mathfrak{M}^*$ is a truth extension of $\mathfrak{M}$. This completes the proof of Theorem 2.10. $\square$

Corollary 2.14   *For any Boolean-valued model there is a truth extension.*

For the proof of Corollary 2.14, extend the given Boolean-valued model to one with the prime filter property (recall Example 2.8) and use Theorem 2.10. Corollary 2.14 is thus an instance of a possible application of Theorem 2.10 and it also eases the burden a little bit constituted by the fact that we have to assume ground models which satisfy the prime filter property.

Remark 2.15   If $\mathfrak{M}$ is a ground model satisfying (1) – (2) from Section 2.3.1, such that $\mathbb{M}$ is complete and atomic, then we can prove the existence of a truth extension of $\mathfrak{M}$ without making use of the Stone mapping PF, but rather by using the mapping AT such that $AT(u) := \{v \in \mathbf{M} | v \leq_{\mathbb{M}} u, v \text{ is an atom}\}$ (see Koppelberg [10], pp. 29–30); in this case we take $\wp(At(\mathbf{M})) \times \wp(At(\mathbf{M}))$ to be our field of polarities, where $At(\mathbf{M})$ is the set of atoms in $\mathbf{M}$.

2.3.3 Designation in De Morgan-valued truth extensions

As explained at the end of Section 1.4, a truth filter TF^* may be added to the truth extension in order to get some notions of *evaluated as true* and *evaluated as false*. We will show that, if $\mathbb{M}$ is a De Morgan *algebra*, any maximal truth filter TF (given with the ground model $\mathfrak{M}$) may be extended to a maximal truth filter TF^* such that $TF^* \cap \mathbb{M} = TF$. Assume that $\mathbb{M}$ is a De Morgan algebra and that TF is maximal in $\mathbb{M}$. Take the filter in $\mathbb{M}^*$ which is generated by TF; obviously this filter is proper, because otherwise TF would not have been proper in $\mathbb{M}$. Furthermore, the intersection of the generated filter with $\mathbb{M}$ is a filter of $\mathbb{M}$, which cannot be larger than TF; that is, it is identical with TF.

Since $\mathbb{M}^*$ has a bottom, we can even extend the generated filter to a maximal truth filter, which is prime, because $\mathbb{M}^*$ is distributive. The intersection of this maximal filter with $\mathbb{M}$ is a proper filter of $\mathbb{M}$, because otherwise it would include the bottom of $\mathbb{M}$, which is by Theorem 2.10 also the bottom of $\mathbb{M}^*$, and then the maximal filter would not be proper in $\mathbb{M}^*$. It cannot be larger than TF since the latter is maximal by assumption. Thus, we know that TF^* is some maximal prime truth filter extending TF.

Do we also know whether TF^* will preserve all infima and suprema given by quantification in L_{True}? Generally: no. However, since TF is prime it preserves all infima and suprema arising from quantification in L because of the Prime Filter Property; the same must hold for TF^* regarding quantification in L. If we additionally

assume that $\mathbb{M}$ is *itself* a field of polarities of the kind used in our proof we see from the proof of Theorem 2.10 that we also can assume $\mathbb{M}^* = \mathbb{M}$ and then $TF^* = TF$ preserves all infima and suprema given by quantification in $\mathcal{L}_{True}$.

Since $Val_{\mathfrak{M}^*}(True(\ulcorner\varphi\urcorner)) = Val_{\mathfrak{M}^*}(\varphi)$ for all $\varphi \in \mathcal{L}_{True}$ it follows that

1. For all $\varphi \in \mathcal{L}_{True}$, φ is evaluated as true if and only if $True(\ulcorner\varphi\urcorner)$ is evaluated as true.
2. For all $\varphi \in \mathcal{L}_{True}$, φ is evaluated as false if and only if $True(\ulcorner\varphi\urcorner)$ is evaluated as false.

That means that if we look at the interpretation of the truth predicate we may deduce whether some formula φ is evaluated as true: this is the case if and only if $True(\ulcorner\varphi\urcorner)$ is evaluated as true. Therefore we can conclude that our De Morgan-valued truth extension $\mathfrak{M}^*$ satisfies Tarski's Convention T both in the implication *and* in the truth reading.

2.3.4 Example   Now, what does this mean for the Liar sentence? Let $\mathfrak{M} = \langle \mathcal{S}, \mathcal{L}, \mathbb{M}, D, \mathfrak{I}\rangle$ such that $\mathfrak{I}(c_1) = \neg True(c_1)$. Then we know a truth model $\mathfrak{M}^* = \langle \mathcal{S}_{True}, \mathcal{L}_{True}, \mathbb{M}^*, D, \mathfrak{I}^*\rangle$ exists such that for all $\varphi \in \mathcal{L}_{True}$ and all constants t with $\mathfrak{I}(t) = \varphi$

$$Val_{\mathfrak{M}^*}(True(t)) = Val_{\mathfrak{M}^*}(\varphi).$$

In the case of the Liar this implies

$$Val_{\mathfrak{M}^*}(True(c_1)) = Val_{\mathfrak{M}^*}(\neg True(c_1))\ (\ = \neg_{\mathbb{M}^*} Val_{\mathfrak{M}^*}(True(c_1))).$$

Thus, the Liar and its negation are semantically equivalent in De Morgan-valued truth extensions since they express the same proposition. Note that whether the Liar is evaluated as true or false depends on which truth filter TF^* is chosen. But, in any case, the Liar is evaluated as true/false if and only if its negation is evaluated as true/false. This corresponds to the intuition that the Liar should be true if and only if the same holds for its negation.

3 Summary   We have shown that a language, if understood as given by a De Morgan-valued model, may indeed contain its own truth predicate. On the technical side, the most relevant open question is whether the Prime Filter Property from Section 2.3.1 can be weakened, and thus whether a more general result could be obtained.

A different question is how to interpret our result: it is clear that our proof only works because negation is a much weaker notion in De Morgan-valued models than in classical models, or, more generally, in Boolean-valued models. As a consequence, the metalinguistic semantical implication relation between sentences of the object language is no longer represented in the latter by some composition of logical connectives, contrary to classical and, in some sense, also Boolean-valued models. From this point of view, we have only *shifted* the problems connected to semantically closed languages: from the unrepresentability of truth to the unrepresentability of semantical implication (a similar point has already been made by Gupta [7], pp. 91–104). It needs a separate paper to investigate this in more detail.

As you may have noted, all our efforts have a certain algebraic flavor: indeed, you may see this paper as a kind of advertisement to deal with problems of truth and

self-reference in an algebraic mood. Of course, this conception is not new at all, since, for example, Visser's contributions to this area are basically algebraic in style (see Visser [21]). The nomenclature and concepts developed in this paper up to Section 2.3 are general enough to be applicable to many kinds of lattice-valued models, where a complement function is associated with the lattice, and not just to De Morgan-valued models. In particular, our notions of a truth model and a truth extension do not contain any trace of De Morgan structure at all. By the means of our background theory it is therefore also possible to ask for which *other* algebraic structures existence or non-existence of truth extensions may be proved, and which (necessary and/or sufficient) criteria can be found for such results.

REFERENCES

[1] Barwise, J., and J. Etchemendy, *The Liar*, Oxford University Press, Oxford, 1987. Zbl 0678.03001 MR 88k:03009   1.3

[2] Belnap, N. D. Jr., and J. H. Spencer, "Intensionally complemented distributive lattices," *Portugaliae Mathematica*, vol. 25 (1966), pp. 99–104. Zbl 0158.27103 MR 35:1519   1.4

[3] Belnap, N. D. Jr., "A useful four-valued logic," pp. 8–37 in *Modern Uses of Multiple-Valued Logic*, edited by J. M. Dunn and G. Epstein, D. Reidel, Dordrecht, 1977. Zbl 0424.03012 MR 58 #5021   1.2

[4] Bialynicki-Birula, A., and H. Rasiowa, "On the representation of quasi-boolean algebras," *Bulletin Academie Polen. Sci. Cl.*, vol. 3 (1957), pp. 259–61. Zbl 0082.01403 MR 19,380b   1.1

[5] Dunn, J. M., "Intensional algebras," pp. 180–206 in *Entailment*, edited by A. R. Anderson and N. D. Belnap Jr., Princeton University Press, Princeton, 1975.   1.1, 1.3, 2.6

[6] Dunn, J. M., "Relevance logic and entailment," pp. 117–224 in *Handbook of Philosophical Logic III*, edited by D. Gabbay and F. Guenthner, D. Reidel, Dordrecht, 1986. Zbl 0875.03051   1.3

[7] Gupta, A., and N. Belnap, *The Revision Theory of Truth*, The MIT Press, Cambridge, 1993. Zbl 0858.03010 MR 95f:03003   3

[8] Hermes, H., *Einführung in die Verbandstheorie*, Springer-Verlag, Berlin, 1967. Zbl 0153.33203 MR 36 #3686   22

[9] Kalman, J. A., "Lattices with involution," *Transactions of the American Mathematical Society*, vol. 87 (1958), 485–91. Zbl 0228.06003 MR 20 #1641   1.1, 1.1

[10] Koppelberg, S., *Handbook of Boolean Algebras*, vol. 1, North-Holland, Amsterdam, 1989. Zbl 0671.06001 MR 90k:06002   1.2, 2.15

[11] Kripke, S. A., "Outline of a theory of truth," *The Journal of Philosophy*, vol. 72 (1975), pp. 690–716. Zbl 0952.03513   2.1

[12] Leitgeb, H., *Truth as Translation*, Ph.D. thesis, University of Salzburg, Salzburg, 1998.   2.8

[13] Martin, R. L., and P. W. Woodruff, "On representing 'True-in-*L*' in *L*," *Philosophia*, vol. 5 (1975), pp. 213–17. Zbl 0386.03001 2.1

[14] McGee, V., *Truth, Vagueness and Paradox*, Hackett Publishing Company, Indianapolis, 1991. Zbl 0734.03001 MR 92k:03004 2.3.2

[15] Monteiro, A., "Matrices de Morgan caractèristiques pour le calcul propositionnel classique," *An. Acad. Brasil, Ci.*, vol. 32 (1960), pp. 1–7. Zbl 0094.00605 MR 28:18 1.1

[16] Rescher, N., *Many-valued Logic,* McGraw-Hill, New York, 1969. Zbl 0248.02023 1.2

[17] Rosser, J. B., *Simplified Independence Proofs*, Academic Press, New York, 1969. Zbl 0209.30502 MR 40 #2536 1.2

[18] Tarski, A., "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, vol. 1 (1935), pp. 261–405. Zbl 0013.28903 2.1

[19] Tarski, A., "The semantic conception of truth and the foundations of semantics," *Philosophy and Phenomenological Research*, vol. 4 (1944), pp. 341–76. Zbl 0061.00807 2.1

[20] Visser, A., "Four valued semantics and the Liar," *Journal of Philosophical Logic*, vol. 13 (1984), pp. 695–708. Zbl 0546.03007 MR 86f:03042 2.1

[21] Visser, A., "Semantics and the Liar Paradox," pp. 617–706, in *Handbook of Philosophical Logic IV*, edited by D. Gabbay and F. Guenthner, D. Reidel, Dordrecht, 1989. Zbl 0875.03030 3

[22] Woodruff, P. W., "Paradox, truth and logic. Part 1: paradox and truth," *Journal of Philosophical Logic*, vol. 13 (1984), pp. 213–32. Zbl 0546.03006 MR 86d:03005d 2.1

Department of Philosophy
University of Salzburg
Franziskanergasse 1
A-5020 Salzburg
AUSTRIA
email: hannes.leitgeb@mh.sbg.ac.at