To appear in THE JOURNAL OF SYMBOLIC LOGIC

Review of
TRUTH AND MODALITY FOR KNOWLEDGE REPRESENTATION *
by Raymond Turner

Simone Martini
Udine, Italy

August 1995

Knowledge representation (the development and use of expressive formal tools and techniques to represent human knowledge on specific subjects) is an important topic in (symbolic) Artificial Intelligence, and the specialized literature is full of approaches and proposals, some of them showing little care for their logical status.

Logics of truth, on the other hand, where a suitably axiomatized truth predicate may be applied to (codes of) formulas, have been extensively studied by logicians in the last twenty years, starting with the seminal Kripke's *An outline of a theory of truth* (L 1068). Many elegant and powerful formal systems have been proposed, and much mathematics has been developed. The field, however, still misses an authoritative comprehensive reference, although Sheard's roadmap (*A guide to truth predicates in the modern era*, this JOURNAL, vol. 59 (1994), pp. 1032–1054) and Visser's Handbook paper (*Semantics and the Liar paradox*, in *Handbook of Philosophical Logic*, vol. IV, Gabbay and Guenthner Eds, Reidel, 1989, pp. 617–706) are valuable introductions.

The book by Turner presents itself as an introduction to several of these logics of truth, with the aim to further extend them with modalities and to present the resulting systems as a suitable tool for knowledge representation. However, this knowledge representation perspective (and the title of the book) should not mislead the reader. This is a logic book and very little is said (in the first chapter and the conclusion) on how and why this formal systems should be used as a knowledge representation tool. Before saying how successful this introduction is, we review in detail the content of this rather short book (117 pages of text), which may be conceptually divided in an introduction and three parts.

The introduction (Chapter 1, 12 pages) gives a succinct overview of the problems a logic of knowledge has to face. The notions of proposition and

*Pitman, London 1990, v+122 pp., and The MIT Press, Cambridge, Mass., 1991, vii+126 pp.

intensionality are briefly discussed, with the goal to motivate the difference between formal sentences and propositions, and then, after showing a version of the Liar paradox, the need for a logic of truth where not all sentences are propositions.

The first part (chapters 2 to 5, 62 pages) is devoted to the logics of truth. Chapter 2 introduces the formal (standard) background: first, the lambda-calculus, with syntax—terms, substitution (but a side-condition is missing in the definition), conversion, fixed-point combinator—and semantics (based on complete partially ordered sets, with the construction of the D_∞ model). Second, the first order predicate calculus is introduced, the terms being those of lambda-calculus; the signature always contains the predicate symbols T and F, whose intended interpretation will be the truth predicates. Finally, we are given a way to code terms and formulas with lambda-terms (but not a completely satisfactory one, see the following) and we deduce the inconsistency of the calculus with the Tarski biconditional schema, $A \leftrightarrow T(\hat{\ }A)$, where $\hat{\ }A$ is the code of the wff A (in the following we will write $T(A)$ for $T(\hat{\ }A)$). Chapter 3 introduces the logic of truth **KFG**, based on Kleene's (strong) three-valued logic. The axioms for T and F, describing Kleene's three-valued tables, are listed and showed consistent, via the usual fixed-point construction over the (monotone) notion of Kleene revision of an interpretation (informally, in the revised interpretation $F(A)$ is true when A is false in the original interpretation). It is then discussed a reformulation of **KFG** along Feferman's work and, finally, it is shown how **KFG** gives rise to a natural theory of truth and propositions (A is a proposition iff $T(A) \vee F(A)$). Chapter 4 studies the different (non-monotonic) notion of Tarski revision (informally, in the revised interpretation $F(A)$ is true when $\neg A$ is true in the original interpretation). The theory of stabilization ordinals and safeness is briefly presented and Herzberger's theorem on the existence of a stabilization ordinal is stated without proof. Several formal systems motivated by this semantics are now introduced and studied, taking as guideline for their axioms the normal modal logics D,T, S4 and S5 (with the Barcan formula). The logic corresponding to D proves stably true theorems, while the others prove only safe theorems. Restrictions to the axioms are then presented in order to get stability (e.g., instead of the T-axiom, $T(A) \rightarrow A$, the axiom $T(T(A) \rightarrow A) \rightarrow P(A)$ is used, where $P(A)$ is $T(A) \vee F(A)$). Chapter 5 is a quick introduction to the theory of Frege Structure, à la Scott-Aczel, as a theory of truth and propositions and to its relations with the theory developed in Chapter 3.

The second part of the book (Chapters 6 and 7, 24 pages) deals with standard modal logics and their extensions with a truth predicate. Chapter 6 surveys the possible world semantics and the normal logics K, D, T, S4, S5. Chapter 7 takes the three approaches to the truth predicate studied in the first part and shows how each of them extends rather smoothly to the modal case. For the theory of stable truth there are many possibilities and the book lists many of them.

The last part (Chapter 8, 11 pages; Chapter 9 merely concludes) deals with predicative modality. Instead of defining the modality as a unary connective, a predicate $\triangle$ (over the code of a formula) is introduced. After presenting Montague's classical result on the inconsistency of such an approach already for a very weak modal logic (axioms T, K, $\triangle(\triangle A \rightarrow A)$, restricted necessitation), the author exploits the results of Chapter 7 to define consistent systems. Consistency is established by a relative interpretation into the logics of Chapter 7, whose main clause is $\triangle x \equiv_{\text{def}} \Box T(x)$. The space of possibilities is now rather large (which axioms for the truth predicate; which axioms for the modality; full or restricted necessitation) and for all of them consistency is stated.

From this summary and the length of the book, it is clear that many arguments are necessarily just touched upon. Consistently, the author repeats many times that the development (and not only the details) of many topics is left to the reader (e.g. pag. 57, 60, 63, 72, 98, 114 and the conclusions).

The book is written for a reader with few prerequisites: the notions of a standard first course in logic suffice (and are recalled in Chapter 2). Since it is concerned mainly with the consistency of the various logics, there are no difficult theorems to prove (except Herzberger's one). For all the stated theorems, the given sketch of the proof should suffice for most readers.

The exposition is simple and never goes in deep technicalities. In several points, however, this simplicity is misleading.

A first problematic issue concerns the coding of wffs with lambda-terms. On page 24 an *independence property* is claimed: if $\hat{}A$ and $\hat{}B$ are equated in a model, then "A and B will be the same wff". Even admitting that "same wff" should be understood modulo β-conversion, this in general will be false if no restriction is issued on the lambda-terms present in A and B. The model D_∞, for instance, equates all the unsolvable lambda-terms, which are not equated by β-conversion. On the same page, in the proof of Lemma 2.16, it is assumed that $(\hat{}A)[t/x] = \hat{}(A[t/x])$, which is false, for the proposed coding, if x occurs bound in A. This substitution property is used many times throughout the book, e.g. page 25 in the derivation of Russell paradox; and page 40, in the table; every time it is used, it should be assumed that x is not bound, which is obviously always possible. Or, better, a different coding has to be adopted, for which $FV(A) = FV(\hat{}A)$. Similar problems will arise on page 68 (Def. 5.0), where the clauses for the coding of Ξ and Θ generate free variables ($x \in FV(\Xi(\lambda x.t))$).

The proofs of Theorems 4.5 (page 49) and 4.9 (page 52) are inconclusive for the Barcan formula, in the case of a limit ordinal. Theorem 4.9 should also be contrasted with the result of McGee (**Journal of Philosophical Logic**, vol. 14 (1985), pp. 399–410) that the considered system is ω-inconsistent when the ground structure contains the natural numbers.

On page 59, in the proof of Theorem 4.19, axiom R is used. However, R is not an axiom of the formal system considered, and we only know that it is safe (and hence a sound reasoning principle at any stabilization ordinal). Thus the proof only establishes that E cannot be true at any stabilization ordinal. But it does not rule out the possibility that other, completely different, models exist.

Also, there are several misprints, some of them in crucial positions (e.g. the side conditions of the table on page 40).

Another point in which exposition could have been improved is the choice of the formal systems to present. I have to admit that—and not being an expert of the field I think I may be considered an average reader—I have been very "impressed by the number of possibilities that are available" (page 114) and sometimes I felt lost in "the labyrinth of the logics" (page 115) and of the notations for their names (especially on pages 98 and 111). Fewer systems in more details, and some exercises (no exercise is provided at all) could have given the reader more opportunities to assimilate the material (see, for instance, page 72, with one line on "elementary in the sense of Feferman" and half a page on Martin-Löf Type Theory and dependent types, which leave the reader—being him novice or expert—completely unsatisfied).

These things said and errors corrected, the book appears as a first introduction to the topics and problems of the logics of truth and modality, probably suited for students of applied logic (and under this respect the title appears better chosen), for which the ready availability of a system is more important than the problems that system raises. It is certainly not sufficient for logicians, for which it is too quick and not problematic enough, unless used for a very first approach, to be continuously supplemented with the original sources.

As for the thesis the author maintains, that the theory of truth and predicative modality is of potential use in knowledge representation, I do not have strong arguments against it. But the fact that it "reflects the uniformity of natural language in regard to its representation of" truth and modality (page 103), seems also too weak an argument in favor.

MINOR COMMENTS AND MISPRINTS

page 6, formulas (7') and (8'): on page 4 B_A is used instead of bel.
page 15, clause (b2): "z not free in s" should be "z not free in st'".
page 19, line -5: $\Psi_n(\Phi_n(f')) \subseteq f'$ should be $\Phi_n(\Psi_n(f')) \subseteq f'$.
page 24, proof of 2.16: $\approx$ ("syntactic identity of terms", page 14) would have been better than $=$ in the definition of f, t and B. In general, $\approx$ is used only erratically in the book (see notes on page 37 and page 65).
page 25, derivation of Russell paradox: in the first line $\lambda x.B$ should be $\lambda x.\hat{}B$; in the second line, $\neg(t \in t)$ should be $\hat{}(\neg(t \in t))$.
page 25, axiom **DIS**: among the things the reader may be puzzled by, is why the axioms for truth are sometimes stated as schemata (that is instantiated with codes for wffs, e.g. **FT**) sometimes as predicates on free variables (e.g. **DIS** here, but not on page 29).
page 33, Def. 3.0: the independence property is needed again. The right notion seems to be: $\mathcal{I}(\hat{}A) = \mathcal{I}(\hat{}B)$ iff A and B have the same structure as wffs and their corresponding terms are equated in the model.
page 35, Thm. 3.3: another example of bad use of equality symbols (both $=$ and $\equiv$ are used for set equality; $\equiv$ never introduced first).
page 34, line 6: "and disjunction" should be "and conjunction".
page 36, Def. 3.5: subscript g is omitted in $\mathcal{M} \vdash_g A$ and $\mathcal{M} \dashv_g A$.

page 37, Def. 3.6: a particularly nasty example of the non-use of $\approx$; see especially clause (i).
page 40, line -3: $A[c]$ should be $T(A[c])$.
page 40, line -1: "$T(A[c])$ depends" should be "$T(C)$ depends, except at most $T(A[c])$".
page 42, failure of necessitation for **KFG**: Since the book is directed to novices, it should have helped a remark on the consequences of this failure, in particular the failure of substitution inside the truth predicate.
page 63, truth teller: another point where more care in the use of codes, $\approx$, and $=$, would have helped the reader.
page 63, the displayed equation in the proof of Thm. 4.26 should be $T(a = b) \rightarrow T(T(a) \leftrightarrow T(b))$.
page 65, line -5: $\Leftrightarrow$ is in the language, but in the following page is treated as an abbreviation; $\approx$ becomes now the combinator for intensional equality.
page 66, line 3: the same symbol, $\vee$, is used for the combinator and the connective of disjunction (repeated many times, in all the places where disjunction is used in Chapters 5 (pages 65-70) and 7 (pages 98-99)).
page 66, displayed equation (1): T should be True.
page 104, first displayed equation:) missing.
page 106, axiom **imp**: the name **ip** was used before (same in line 4 of Section 8.3 on page 107).
page 108, clauses (i)–(x): the clause for atomic equation is missing.
page 109, line -11: first "ip" should be "IP".