# TRUTH AND MODALITY FOR KNOWLEDGE REPRESENTATION

RAYMOND TURNER

# Truth and Modality for Knowledge Representation

Raymond Turner

The MIT Press
Cambridge, Massachusetts

First MIT Press edition 1991

Printed and bound in the United States of America.

Library of Congress Cataloging in Publication Data

Turner, Raymond, 1947–
Truth and modality for knowledge representation / Raymond Turner.
p. cm. — (Artificial intelligence)
Includes bibliographical references (p. ) and index.
ISBN 0-262-20080-5
1. Knowledge representation (Information theory) 2. Artificial intelligence. 3. Truth.
4. Modality (Theory of knowledge) I. Title. II. Series: Artificial intelligence
(Cambridge, Mass.)
Q387.T87 1991
003'.54—dc20
90-19128
CIP

# Truth and Modality for Knowledge Representation

# Series Foreword

Artificial intelligence is the study of intelligence using the ideas and methods of computation. Unfortunately, a definition of intelligence seems impossible at the moment because intelligence appears to be an amalgam of so many information-processing and information-representation abilities. Of course psychology, philosophy, linguistics, and related disciplines offer various perspectives and methodologies for studying intelligence. For the most part, however, the theories proposed in these fields are too incomplete and too vaguely stated to be realized in computational terms. Something more is needed, even though valuable ideas, relationships, and constraints can be gleaned from traditional studies of what are, after all, impressive existence proofs that intelligence is in fact possible. Artificial intelligence offers a new perspective and a new methodology. Its central goal is to make computers intelligent, both to make them more useful and to understand the principles that make intelligence possible. That intelligent computers will be extremely useful is obvious. The more profound point is that artificial intelligence aims to understand intelligence using the ideas and methods of computation, thus offering a radically new and different basis for theory formation. Most of the people doing work in artificial intelligence believe that these theories will apply to any intelligent information processor, whether biological or solid state.

There are side effects that deserve attention, too. Any program that will successfully model even a small part of intelligence will be inherently massive and complex. Consequently, artificial intelligence continually confronts the limits of computer-science technology. The problems encountered have been hard enough and interesting enough to seduce artificial intelligence people into working on them with enthusiasm. It is natural, then, that there has been a steady flow of ideas from artificial intelligence to computer science, and the flow shows no sign of abating.

The purpose of The MIT Press Series in Artificial Intelligence is to provide people in many areas, both professionals and students, with timely, detailed in formation about what is happening on the frontiers in research centers all over the world.

Patrick Henry Winston  
J. Michael Brady  
Daniel Bobrow

# Preface

The need for more expressive systems of knowledge representation is not controversial although it is still debatable whether or not such systems have to be based on formal logic. In this book we shall take it is as read that the formal approach is a worthy one. Our objective is to explore the development of formal languages and appropriate logics for that aspect of knowledge representation concerned with reasoning about truth and modality. A great deal of the current literature in Artificial Intelligence is devoted to the development of formalisms which facilitate the expression of modal concepts. Much of this work, however, is based upon the theories of modality and truth which were developed in the period 1960–1980. In the last ten years there has been a great deal of activity within the logical community centered upon the development of logics of truth and modality. Our objective is to bring this material to the attention of AI researchers by putting it in a context where it might be directly applicable to AI knowledge representation.

## Acknowledgements

I must thank Sam Steel and Allan Ramsay for encouraging me to write this book. Nicholas Davies is to be thanked for proof reading the final draft. Roxanna Berretta also proof read the final script. Finally, I cannot find words to convey my thanks to Marisa Bostock and Ann Cook who typed the manuscript.

# Contents

# 1 Reasoning Agents

A great deal of research in Artificial Intelligence is concerned with the introduction and exposition of languages of *knowledge representation*. Many of these new languages are variations or extensions of languages which have their origins in the literature of formal logic. The reasons for this are obvious. Formal logic has been concerned throughout its history with the representation of informal argument, its primary goal being the representation of informal arguments in a language where the *form* of the argument and its component sentences are made explicit and unambiguous.

The majority of this formal work in knowledge representation has been carried out within the language and proof theory of the first-order Predicate Calculus. Moreover, many *informal* systems of knowledge representation have been recast within the Predicate Calculus with the implicit belief that such recasting provides a formal semantics for these formalisms and thereby assigns them a degree of clarity and respectability. This may well be true but what is not clear is that the Predicate Calculus is rich enough in expressive power for the full range of applications which knowledge representation seems to subsume. There are many areas where the *natural* representation of informal arguments appears to demand more facilities than those available within the Calculus. The representation of arguments involving time and modality are two topics which immediately spring to mind. Indeed, within the AI community itself (Moore[1980], [1984], Konolige[1986], McDermott[1982], Allen [1984]) such extensions to the Predicate Calculus have been explored with much benefit.

This book is largely concerned with the representation of certain modal notions where *modality* will be given a fairly liberal interpretation. However, before we embark on any lengthy discussion of the nature of modality it seems prudent to examine the general goals of knowledge representation. What exactly is it that we need to represent and reason about?

## 1.1 Modality and Knowledge Representation

Reasoning agents must be capable of representing and reasoning about what they and other agents *believe, know* and *hold true*. This ability seems to be common ground for many different areas of AI. For example, any natural language program which has to form part of a dialogue system must be capable of representing and reasoning about the knowledge and belief of the players or agents in the system. Moreover, any reasonable system will have to be able to distinguish between these various modalities: to believe a proposition is not the same as knowing it since it is clearly possible for an agent to believe false propositions. A certain tradition in epistemology interprets knowledge as *justified true belief*, and under this interpretation knowing a proposition implies it is both believed and true. On the other hand, it may be both true and believed by some agent but this is not sufficient grounds for claiming that the agent knows the proposition since it may be believed for completely spurious reasons. Consequently, a theory of knowledge representation must be capable of representing and distinguishing between assertions of the form:

(1) Agent A believes that p
(2) Agent A knows that p
(3) Agent A believes some proposition which is false

The upshot of these simple observations seems to be that any adequate theory of knowledge representation must in part be a *theory* of truth and modality. The central concern of such a theory must therefore be with the development of a language in which such assertions as (1), (2) and (3) can be represented together with the formulation of appropriate logics for such modal notions. Much of the present study is concerned with the development of such languages and logics. Not that we shall provide anything like a definitive answer. We shall not enter into debate about the logics which are most appropriate for any particular modality; this is already a well documented, even if controversial, area. Our goal is much less ambitious: we aim to draw certain boundaries around the possible form of such theories. The need for this stems from the particular nature of the theories we shall advocate. We firmly believe that the most natural and computationally tractable theory is a first-order one. This imposes limitations on the logics of modality and truth.

To some extent this goal might be taken as identical to that of natural language semantics, or rather that part of semantics concerned with *truth* and *modality*. One important aspect in which our task is different concerns the concentration in semantics on the systematic relation between syntactic structure and semantic representation. We are at liberty to concentrate on the language of semantic representation and only appeal to natural language as a guide since we are not primarily concerned with the syntax–semantics connection. Rather we are concerned with the language and nature of semantic representation. This is important since it enables us largely to divorce our study from considerations which relate directly to natural language syntax and how the syntactic or grammatical form of a sentence contributes to its logical representation. However, this separation is a delicate one and has to be exercised with care. In the end, our languages of knowledge representation must be as naturally expressive as natural language itself and the insights of natural language semantics cannot be ignored.

## 1.2 Propositions and Modality

One of the first questions which arises in such an endeavour concerns the arguments of these modal connectives: what is it that is believed, known or taken to be true? This is, of course, a non-trivial philosophical question and one that perhaps has no definitive answer. Nevertheless it is not one that can be avoided in any serious study of the issues at hand: the answer given to this question either explicitly or implicitly dictates the form of the theory that will be championed. Traditionally, *propositions* are taken to be the objects of belief, knowledge and truth. Of course, this is just a name to hang the problem on since now we are forced to face the question *what is a proposition*? Fortunately, the literature is full of possible answers and the formal theory of propositions largely determines the language and content of the resulting theory of modality. In what follows we shall explore various proposals from the AI and philosphical literature and see how they meet the needs of knowledge representation.

## 1.3 Possible Worlds

In one of the major paradigms, the notion of possible worlds plays the leading role. Propositions are taken to be sets of possible worlds and properties are understood as functions from individuals to propositions. The modal and doxastic operators are then treated as functions from propositions to propositions. In general, modal operators are analysed as functions which send a proposition P to that proposition which consists of all those worlds which are accessible from elements of P. Different choices of the relation of accessibility lead to different modal and doxastic notions (Hintikka [1962], Kripke [1963]).

This is the *classical* theory and forms the underlying semantic theory of first-order modal logic. The language of the latter is derived from the language of first-order logic by the addition of new sentential operators (the modal connectives) whose logic is given by one of the standard systems of modal logic. The different logics correspond to different properties of the relation of accessibility. In this approach, the modal connectives operate on sentences of the language where the latter are taken to denote propositions (i.e. sets of worlds).

The elegance of this simple theory together with the fact that the different modal and doxastic notions can be distinguished by varying the constraints on the accessibility relation are largely responsible for its abiding influence on both logicians and computer scientists. As it stands, however, it will not serve our purposes.

## 1.4 Higher-Order Modal Logic

This first-order approach will not be expressive enough for the goals of knowledge representation. For one thing there is a prima facie need to quantify over propositions and properties. Consider the sentences:

(4) Agent A believes a false proposition
(5) Every proposition agent A believes is false
(6) Agent A can perform every task agent B can

The expression of (4), (5) and (6) demands more than first-order modal logic since they involve quantification over propositions and properties and this is not available in first-order modal logic. Other reasons might be marshalled to justify such quantificational facilities. For example, in Ramsay[1988] the need for quantification over properties is illustrated by the desire to formulate *frame* axioms in a natural and succinct fashion. Here one has to stipulate those aspects of a situation which remain unchanged under a specified revision, and the natural and elegant way of achieving this involves quantification over properties.

The introduction of such quantification moves us into the domain of higher-order intensional logic, the most highly developed version of which is due to Montague[1973]. In this logic we are able to quantify over propositions and properties and much else besides. The different kinds of objects are delineated by the formal notion of *type*: types are the basic ones [individuals ($e$) and truth values ($t$)], or function types (the type of functions from one type to a second), or the type of functions from the type of possible worlds (or some more complex index) to any existing type. Propositions are then analysed as functions from worlds to truth values (or equivalently sets of worlds) and properties as functions from individuals to propositions. The variables of the language are decorated with these types so, for example, the content of (4), (5), (6) can be formally expressed as

(4′) $\exists x_{\langle w,t\rangle}(\mathrm{B_A}(x) \& \sim(x))$
(5′) $\forall x_{\langle w,t\rangle}(\mathrm{B_A}(x) \rightarrow \sim(x))$
(6′) $\forall x_{\langle i,\langle w,t\rangle\rangle}(\mathrm{Perform}(\mathrm{B}, x) \rightarrow \mathrm{Perform}(\mathrm{A}, x))$

where the variables $x_{\langle w,t\rangle}$ range over propositions and the $x_{\langle i,\langle w,t\rangle\rangle}$ over properties. In the higher-order intensional logics, quantification over individuals, propositions and properties is built into the language of the theory. Indeed, quantification over even higher type objects is permitted. This approach thus appears to be expressive enough for our needs but suffers from a possible conceptual drawback.

## 1.5 *Fine-Grained* Propositions

The major objection to the whole possible world approach, at least in regard to its application to the doxastic modalities such as belief and knowledge, concerns the nature of propositions within this regime. It is argued that propositions, as sets of possible worlds, are too *coarse-grained* to serve as arguments to the doxastic operators of knowledge and belief: if two sentences denote exactly the same set of possible worlds, then an agent who believes one is committed to believing the other. While this may be acceptable for certain notions of belief (e.g. rational belief) it does not seem persuasive for all notions. Mathematical belief seems to be a case in point. Since mathematical assertions are naturally taken to be necessarily true or necessarily false, under this account of belief an agent who believes one true assertion of mathematics (e.g. $2 + 2 = 4$) is thereby committed to believing all true assertions of mathematics. This seems not to supply an adequate account of mathematical belief. This criticism of the possible world approach to the attitudes of belief and knowledge naturally leads to the demand for a more *fine-grained* notion of proposition: one that will not commit an agent to believing all the logical consequences of his or her basic beliefs.

## 1.6 Propositions as Primitive

One proposal for overcoming this problem involves kicking away the possible world ladder which supports the notion of proposition. Propositions are then not unpacked in terms of possible worlds or any other supposedly more fundamental notion but taken as primitive. Subsequently, we are not forced to take the equality of propositions to be given by the extensional equivalence between sets of worlds. Rather, we are free to invest the notion of proposition with the properties we see fit. Thomason [1980] develops such an approach for higher-order intensional logic and thus combines the expressive power of Montague's intensional logic with a more *fine-grained* notion of proposition. In Thomason's simple type theory there are three basic types: $e$, $t$, and $p$, where $e$ is the type of individuals, $t$ the type of truth-values, and $p$ the type of propositions. Higher-order types are constructed from these in the standard way. In addition, Thomason introduces a simple truth predicate to express the fact that a proposition is true. The Lambda Calculus (typed) is built into the system in the way familiar from Montague. Thomason's logic is quite complex and perhaps hard to work with in terms of the application at hand. Nevertheless, this is a step in the right direction.

## 1.7 Higher-Order Theories and Computational Tractability

Both the approaches of Montague and Thomason are versions of higher-order intensional logic. Indeed, it is exactly this quality which enables the expression of the logical content of sentences such as

(7) A believes everything B believes
(8) Something A believes is true

These can be expressed in the language of Montague's higher-order intensional logic (or with suitable modifications in Thomason's) by

(7′) $\forall x_{\langle w,t\rangle}(\mathrm{bel}(B, x) \rightarrow \mathrm{bel}(A, x))$
(8′) $\exists x_{\langle w,t\rangle}(\mathrm{bel}(A, x) \,\&\, x)$

Unfortunately, this expressive power comes with a price tag. While it is true that first-order logic is only semi-decidable, higher-order logic is much worse. In first-order logic we can construct theorem provers which return a proof for any valid well-formed formulae even if they may fail to terminate when applied to invalid ones; in higher-order logic the theorem provers may fail to terminate even on the valid formulae. This is, of course, not a sufficient reason for discarding the higher-order approach and certainly not in the absence of a better alternative. But it is at least a reason for being somewhat circumspect about rushing headlong into higher-order logic.

## 1.8 Propositions as Sentences

A second and completely different proposal for the analysis of intensionality emanates largely from the AI community itself and seeks to view propositions as sentences in some language of semantic representation. Konolige [1986] and Moore [1980] implicitly seem to advocate such a view. This certainly addresses the fact that propositions need to be *fine-grained*, but there are obvious philosophical objections to such an approach. What is believed, known, deemed to be possible, etc. are not sentences in some language: sentences are just marks or symbols and the objects of knowledge or belief have semantic *content*; one cannot be said to believe a sentence but rather one stands in relation to its semantic content. Whatever the merits of this opinion, there are more devastating problems for this *syntactic* analysis of propositions.

Let $L$ be the language in which such propositions are to be expressed. Then, writing $x \in L$ to indicate that $x$ is a sentence of the language $L$, we can express:

(9) John believes something false

as

(9′) $\exists x(x \in L \ \& \ \text{bel}(\text{John}, x) \ \& \ \text{false}(x))$

This looks promising in that we have expressed the fact that John believes something false or believes a proposition which is false where propositions are identified with the sentence of $L$. But now we face a problem with regard to the language in which (9′) is expressed. This is not $L$ and cannot be without violating predicativity. Suppose there are two believers, John and Peter, and Peter wishes to express the assertion that John believes something false. Then Peter's language must be expressive enough to express (9′) and consequently must have access to the *propositions* of John's language in that its variables of quantification must range over the sentences of $L$. One way of achieving this is to employ the Tarskian Object/Metalanguage distinction.

The main problem with such an approach concerns its expressive power. Consider the sentences:

(10) A believes that everything that B believes is true
(11) B believes that everything that A believes is true

Let $O_{\mathrm{A}}$ $(= M_{O_{\mathrm{B}}})$ be the object language of A which also serves as the metalanguage for the object language $O_{\mathrm{B}}$ of B. Then we can attempt to express (10) as

(10′) $\text{bel}(\mathrm{A}, \forall x \in O_{\mathrm{B}}(\text{bel}(\mathrm{B}, x) \rightarrow \text{true}(x)))$

where $x \in O_{\mathrm{B}}$ has its obvious interpretation and facilitates quantification over the wff of B's language. In order to express (11), however, we require a language which has the facility to quantify over the the wff/beliefs of A. This cannot be achieved in $O_{\mathrm{B}}$ or $O_{\mathrm{A}}$ since we now wish to quantify over A's beliefs. We must resort to a metalanguage $M_{O_{\mathrm{A}}}$. We can then attempt a statement of (11):

(11′) $\text{bel}(\mathrm{B}, \forall x \in O_{\mathrm{A}}(\text{bel}(\mathrm{A}, x) \rightarrow \text{true}(x)))$

But now observe that (10′) does not include the belief of B expressed by (11′) since in $O_{\mathrm{A}}$ we can only quantify over those beliefs expressible in $O_{\mathrm{B}}$. The belief expressed by (11′) is only expressible in a further metalanguage $M_{O_{\mathrm{A}}}$. We can, of course, replace (10′) by (10″):

(10″) $\text{bel}(\mathrm{A}, \forall x \in M_{O_{\mathrm{A}}}(\text{bel}(\mathrm{B}, x) \rightarrow \text{true}(x)))$

but then (11′) does not express what we think it does. Indeed, no matter how far we climb up the object/metalanguage hierarchy we will not be able to capture the intuitive content of (10) and (11); some of the beliefs of A or B will always be left out. This whole approach runs into problems of expressive power: we are unable to express mutual belief and the reason stems from the hierarchical impositions placed on the organization of the languages of representation. This should be seen in contrast to the higher-order approach where the proposition variables range over the domain of propositions and include the denotations of the sentences expressed by (10) and (11). This is formally sanctioned by the comprehension schema of higher-order intensional logic which in particular guarantees that for every sentence of the formal language there is a proposition which furnishes its denotation.

This object/metalanguage analysis parallels the Tarski[1937] account of truth. In this theory there is a hierarchy of object/metalanguages and the metalanguage at each stage contains the truth predicate of the previous language. Tarski's theory of truth suffers similar drawbacks of expressive power. Moreover, there are obvious intuitive objections to this object/metalanguage approach. Natural language has no markings corresponding to these levels. According to this account every sentence must live somewhere in the hierarchy but given an arbitrary natural language sentence we seem unable to place it in such a hierarchy. Indeed, such cumbersome information is totally irrelevant to everyday communication. There are no good reasons to think that theories of knowledge representation, based upon such an explicit marking of levels, would be any less cumbersome and inefficient.

## 1.9 Syntactic Modality

One way out of this impasse is to remove the object/metalanguage distinction and identify the languages. The modal and truth operators now take sentences of the language (or their quoted relations) as arguments. Perlis[1988] has advocated such an approach to knowledge representation. This approach would certainly increase the expressive power of the language. Moreover, we would have a simple first-order system in which to express such modal notions since under this regime the modal operators are naturally analysed as simple first-order predicates. This appears to be in keeping with natural language itself. To see this consider the following sentences:

(12) John believes that Peter sings
(13) Mary thinks that Harry is a man
(14) Peter knows that he will win
(15) It is true that Peter believes that John sings