DISCOURSE AND INFERENCE

Jerry R. Hobbs

USC Information Sciences Institute
Marina del Rey, California

November 3, 2003

Chapter 2

The Logical Notation: Ontological Promiscuity

2.1 Desiderata for a Logical Notation

The role of a logical notation in a theory of discourse interpretation is for representing the explicit information in English texts, for expressing the knowledge required for understanding texts (which in ordinary life we express in English), and for manipulation by the interpretation process. These uses lead to two principal criteria for a logical notation.

Criterion I: The notation should be as close to English as possible. This makes it easier to specify the rules for translation between English and the formal language, and also makes it easier to encode in logical notation facts we normally think of in English. The ideal choice by this criterion is English itself, but it fails monumentally on the second criterion.

Criterion II: The notation should be syntactically simple. Since inference is defined in terms of manipulations performed on expressions in the logical notation, the simpler that notation, the easier it will be to define the inference process.

A logical notation is proposed here which is first-order and nonintensional and for which semantic translation can be naively compositional. The key move is to expand what kinds of entities one allows in one's ontology, rather than complicating the logical notation, the logical form of sentences, or the semantic translation process.

In particular, certain decisions are embodied in the notation.

1. In the interests of simple syntax, the notation is a variety of first-order

predicate calculus. Much of the complexity of English syntax, e.g. the division of predicates into nouns, adjectives, verbs, adverbs, and prepositions, reflects a conceptual scheme that is better captured in the axioms than in the syntax of our formal language. There has been an attempt to make the notation as "flat" as possible. Modal operators are avoided, and there are no functionals, and there is no quantification over predicates. Functions are sometimes used for expository purposes, but play no role in the proper logical notation. Standard logical quantifiers are used sparingly for specific notational purposes; they are not used to represent natural language quantifiers, however. All of this can be summarized in the motto:

> All morphemes are created equal.

Morphemes introduce predications, and that's all.

2. In the interests of uniformity in the syntax of the logical notation, all semantic content is in the predicates and none is in the constants. All knowledge is knowledge of predications. Constants are only handles. Intuitively, the reason for this is that in natural language we cannot communicate entities directly. We can only communicate properties and hope that the listener can determine the entity we are attempting to refer to. Formally, this decision means that predicates and constants play quite distinct roles in the inferential process. This approach is sometimes carried to extremes for the purpose of shedding light on several thorny issues. For example, proper names and sequences of numbers, when they are the focus of concern, are treated as predicates, and *de re* belief is also treated via predicates. Constants are occasionally given suggestive names—J for John—but this is strictly for the convenience of the reader.

3. In the interests of closeness to English, for most of the predicates in the formal language we simply use the corresponding morphemes in English. An extreme example of this is seen in the representation of the propositional content of the pronoun "he": rather than raising the register to express it as $human(X) \wedge masculine(X)$, the predication is simply $he(X)$. Since predicate names are arbitrary, we may as well use the ones English already supplies. We need, in addition, some predicates beyond just the morphemes of English. For example, where several senses of a word need to be distinguished, the morphemes are subscripted, e.g. $bank_1$, $bank_2$.

4. In the interests of closeness to English, the initial logical form consists of about one predication and one constant for every morpheme in the sentence, as well as predications corresponding to certain structural combinations, such as compound nominals and adverbial measure phrases. No

obligatory lexical decompositions are done in semantic translation, although such decompositions may in effect be performed as inferences during subsequent processing.

The development of such a logical notation is usually taken to be a very hard problem. I believe this is because researchers have imposed upon themselves several additional constraints—to adhere to stringent ontological scruples, to explain a number of mysterious syntactic facts as a by-product of the notation, and to encode efficient deduction techniques in the notation. Most representational difficulties go away if one rejects these constraints, and there are good reasons in this enterprise for rejecting each of these constraints.

Ontological Scruples: Researchers in philosophy and linguistics have typically restricted themselves to very few (although a strange assortment of) kinds of entities – physical objects, numbers, sets, times, possible worlds, propositions, events, and situations – and all of these but the first have been controversial. Quine has been the greatest exponent of ontological chastity. His argument is that in any scientific theory, "we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of our experience can be fitted and arranged." (Quine, 1953, p. 16.) But he goes on to say that "simplicity ... is not a clear and unambiguous idea; and it is quite capable of presenting a double or multiple standard." (*Ibid.*, p. 17.) Minimizing kinds of entities is not the only way to achieve simplicity in a theory. The aim in this enterprise is to achieve simplicity by minimizing the complexity of the inference process. It turns out this can be achieved by multiplying kinds of entities, by allowing as an entity everything that can be referred to by a noun phrase.

Syntactic Explanation: The argument here is easy. It would be pleasant if an explanation of, say, the syntactic behavior of count nouns and mass nouns fell out of our underlying ontological structure at no extra cost, but if the extra cost is great complication in the definition of the inference process, it would be quite unpleasant. In constructing a theory of discourse interpretation, it doesn't make sense for us to tie our hands by requiring syntactic explanations as well. The problem of discourse is at least an order of magnitude harder than the problem of syntax, and syntax shouldn't be in the driver's seat.

Efficient Deduction: There is a long tradition in artificial intelligence of building control information into the notation, and indeed much early work in knowledge representation wass driven by this consideration. Semantic networks, description logics, and other notational systems built around hierarchies (Quillian, 1968; Simmons, 1973; Hendrix, 1975; ??, 19??)) implicitly assign a low cost to certain types of syllogistic reasoning. The KL-ONE representation language (Schmolze and Brachman, 1982) has a variety of notational devices, each with an associated efficient deduction procedure. Hayes (1979) argued that frame representations (Minsky, 1975, Bobrow and Winograd, 1977) should be viewed as sets of predicate calculus axioms together with a control component for drawing certain kinds of inferences quickly. In quite a different vein, Moore (1980) used a possible worlds notation to model knowledge and action in part to avoid inefficiencies in theorem-proving.

By contrast, I have not attempted to build efficiencies into the notation, and the inference process as well is specified without regard to data structures or efficiency considerations. From a psychological point of view, this allows us to abstract away from the details of implementation on a particular computational device, increasing the generality of the theory. From a technological point of view, it reflects a decision to first determine empirically the most common and/or problematic classes of inferences required for discourse interpretation and only then to seek algorithms for optimizing them.

This book employs a flat logical notation with an ontologically promiscuous semantics.

One's first naive guess as to how to represent a simple sentence like

A boy builds a boat.

is as follows:

$$(\exists x, y) build(x, y) \wedge boy(x) \wedge boat(y)$$

This simple approach seems to break down when we encounter the more difficult phenomena of natural language, like tense, intensional contexts, and adverbials, as in the sentence

Maybe the boy wanted to build a boat quickly.

These phenomena have led students of language to introduce significant complications in their logical notations for representing sentences (Schubert & Chung, 19??). For example, combining a number of proposed notations, we might represent the content of this sentence as follows:

$$(\iota x : BOY)[\diamond PAST[WANT(x, \lambda z[(\exists\, y : BOAT) Quick(build)(z, y)]]]$$

The definite article is encoded in the iota operator ι. Possibility is represented as the modal operator $\diamond$. The past tense is represented with a temporal operator $PAST$ on expressions. The facts that x is a boy and y is a boat are encoded in type constraints associated with quantifiers. Wanting is an operator $WANT$ that takes an individual and a lambda expression as its arguments. The adverb "quickly" is represented as a operator mapping predicates into predicates. The verb "build" is represented as an ordinary predicate taking individuals as its arguments.

But consider the following set of possible next sentences that refer anaphorically to material in this sentence. The piece of the complicated logical form that is being referred to is indicated in parentheses after the example.

> *He* is impatient. (x)
> *It* should be seaworthy. (y)
> But *it* always takes a long time. ($build$)
> *This desire* is impractical. ($WANT$)
> *The possibility* distresses me. ($\diamond$)
> Maybe he wanted to *then*, but he no longer does. ($PAST$)

Now imagine a theory of discourse interpretation that must say how coreference of anaphoric expressions is resolved. There would have to be special cases for each of these examples, and for every other example that exhibited a phenomenon that had given rise to a specialized notation.

Specialized notations are very useful for specialized inquiries. When we are studying only time, it may be useful to have a specialized temporal logic. But it becomes a hinderance when we expand our scope to all of the content of discourse.

The approach used here will be to maintain the syntactic simplicity of the logical notation and expand the theory of the world implicit in the semantics to accommodate this simplicity. The representation of the above sentence,[1] as is justified below, is

$$\begin{aligned}&(\exists\, e_1, e_2, e_3, e_4, x, y) Rexist(e_0) \wedge Past'(e_0, e_1) \wedge possible'(e_1, e_2)\\ &\quad \wedge want'(e_2, x, e_3) \wedge quick'(e_3, e_4) \wedge build'(e_4, x, y) \wedge boy(x)\\ &\quad \wedge boat(y)\end{aligned}$$

That is, the eventuality e_0 really exists, where e_0 is e_1's occurrence in the past, where e_1 is the possibility of e_2's holding, where e_2 is x's wanting e_3, which is the quickness of e_4, which is x's building of y, where x is a boy and

[1] Or at least an approximation to it.

y is a boat. (The treatment of the determiners is discussed in Chapter 4.) Then the referents of the italicized anaphoric expressions above are x, y, e_4, e_2, e_1, and e_0, respectively, all individual constants.

In brief, the logical form of natural language sentences will be a conjunction of atomic predications in which all variables are existentially quantified with the widest possible scope. Predicates will be identical or nearly identical to natural language morphemes. There will be no functions, functionals, nested quantifiers, disjunctions, negations, or modal or intensional operators in logical forms.

This notation can also be seen as one solution to a problem that has driven much research in natural language semantics in recent years: How do quantifiers scope across sentences? Kohlhase (19??) has identified three possible approaches, each spawning a program of research:

1. Replace existential quantifiers by something else: Discourse representation theory (Kamp, 19??).
2. Extend the scope of quantifiers dynamically: Dynamic predicate logic (Groendijk & Stendhal, 19??).
3. Composition of sentences by conjunction inside the scope of existential quantifiers: Ontological promiscuity.

This chapter is an exposition of the third approach.

2.2 Ontological Promiscuity

2.2.1 Motivation

Davidson (1967) proposed a treatment of action sentences in which events are treated as individuals. This facilitated the representation of sentences with time and place adverbials. Thus we can view the sentences

John ran on Monday.

John ran in San Francisco.

as asserting the existence of a running event by John and asserting a relation between the event and Monday or San Francisco. We can similarly view the sentence

John ran slowly.

as expressing an attribute about a running event. Treating events as individuals is also useful because they can be arguments of statements about causes:

> Because John ran, he arrived sooner than anyone else.
>
> Because he wanted to get there first, John ran.

They can be the objects of propositional attitudes:

> Bill was surprised that John ran.

Finally, this approach accomodates the facts that events can be nominalized and can be referred to pronominally:

> John's running tired him out.
>
> John ran, and Bill saw it.

But virtually every predication that can be made in natural language can be specified as to time and place, be modified adverbially, function as a cause or effect of something else, be the object of a propositional attitude, be nominalized, and be referred to by a pronoun. It is therefore convenient to extend Davidson's approach to all predications. That is, corresponding to any predication that can be made in natural language, we will say there is an event, or state, or condition, or situation, or "eventuality", or whatever, in the world that it refers to. This approach might be called "ontological promiscuity". One abandons all ontological scruples.

Thus we would like to have in our logical notation the possibility of an extra argument in each predication referring to the "condition" that exists when that predication is true. However, especially for expository convenience, we would like to retain the option of not specifying that extra argument when it is not needed and would only get in our way. Hence, the logical notation will provide two sets of predicates that are systematically related, by introducing what might be called a "nominalization" operator $'$. Corresponding to every n-ary predicate p there will be an $n+1$-ary predicate p' whose first argument can be thought of as the condition that holds when p is true of the subsequent arguments. Thus, if $run(J)$ means that John runs, $run'(E, J)$ means that E is a running event by John, or John's running. If $slippery(F)$ means that floor F is slippery, then $slippery'(E, F)$ means that E is the condition of F's being slippery, or F's slipperiness. The effect of this notational maneuver is to provide handles by which various predications can be grasped by higher predications. A similar approach has been used in many AI systems.

It should be pointed out that, although I refer to the prime as a "nominalization operator", I do so only informally. This is not an extension to first-order logic. Rather it is parallel set of ordinary predicates, with similar names for mnemonic purposes.

In discourse one not only makes predications about such ephemera as events, states and conditions. One also refers to entities that do not actually exist. Our notation must thus have a way of referring to such entities. We therefore take our model to be a Platonic universe which contains everything that can be spoken of—objects, events, states, conditions—whether they exist in the real world or not. It then may or may not be a property of such entities that they exist in the real world. In the sentence

(2.1) John worships Zeus.

the worshipping event and John, but not Zeus, exist in the real world, but all three exist in the (overpopulated) Platonic universe. Similarly, in

John wants to fly.

John's flying exists in the Platonic universe but not in the real world.[2,3,4]

2.2.2 Notation

The logical notation is just first-order predicate calculus. Among the notational conventions used throughout the book are that constants are upper-case letters or words in upper-case letters, possibly subscripted; variables

[2] I have my tongue in my cheek, of course, when I refer to a "Platonic universe". I do not (I think) advocate Platonism. The Platonic universe should be viewed as a socially constituted, or conventional, construction, which is nevertheless highly constrained by the way the (not directly accessible) material world is. The degree of constraint is variable. We are more constrained by the material world to believe in trees and chairs, less so to believe in patriotism or ghosts.

[3] The reader might choose to think of the Platonic universe as the universe of possible individuals, although I do not want to exclude *logically* impossible individuals, such as the condition John believes to exist when he believes $6 + 7 = 15$.

[4] Rappaport has urged that I call this a "Meinongian" universe rather than a "Platonic" universe since it contains nonexistent individuals (Rappaport, 19??, Meinong, 18??). In any case, neither term plays a serious terminological role in this treatment.

are lower-case letters, often subscripted; predicates closely related to English morphemes are in lower-case letters, and other predicates have their first letter capitalized. To reduce the use of parentheses and brackets, we assume $\neg$ has the highest precedence, $\wedge$ and $\vee$ have higher precedence than $\supset$ and $\equiv$, and all of these operators have precedence over quantifiers. Instead of writing

$$(\forall\, x)[[p(x) \wedge [\neg q(x)]] \supset (\exists\, y)[r(x, y) \vee s(x, y)]]$$

we may write

$$(\forall\, x)p(x) \wedge \neg q(x) \supset (\exists\, y)r(x, y) \vee s(x, y)$$

Quantification in this notation is always over entities in the Platonic universe. Existence in the real world is expressed by a separate predication, with the predicate *Rexist* described below. Further abbreviations are introduced in Section 2.5.1. Particularly complex logical expressions may be written out fully bracketed for clarity under inspection.

The translation of a large and representative subset of English into the logical notation is specified precisely in Chapter 4. For now, the following examples should give the reader an intuitive feel for the relation between the notation and English sentences.

$kiss(JOHN, MARY)$: John kisses Mary.

$kiss'(E_1, JOHN, MARY)$: E_1 is the (existent or nonexistent) event of John's kissing Mary.

$in(BALL_1, BOX_1)$: $BALL_1$ is in BOX_1.

$in'(E_2, BALL_1, BOX_1)$: E_2 is the condition of $BALL_1$'s being in BOX_1.

$believe(BILL, E_2)$: Bill believes condition E_2 to exist, i.e., that $BALL_1$ is in BOX_1.

$almost(E_2)$: Condition E_2 almost exists, i.e., the ball is almost in the box.

$not(E_2)$: Condition E_2 does not exist, i.e., the ball is not in the box.

$not'(E_3, E_2)$: E_3 is the condition of E_2's not existing.

$man'(E_4, JOHN)$: E_4 is the condition of John's being a man.

$white'(E_5, BLOCK_1)$: E_5 is the condition of $BLOCK_1$'s being white.

$slow'(E_7, E_6) \wedge run'(E_6, JOHN)$: E_7 is the condition that the event E_6 of John's running happens in a slow manner.

$and(E_8, E_9) \wedge on'(E_8, BLOCK_1, TABLE) \wedge on'(E_9, BLOCK_2, BLOCK_1))$: Conditions E_8 and E_9 both exist, where E_8 is the condition of $BLOCK_1$'s being on $TABLE$ and E_9 is the condition of $BLOCK_2$'s being on $BLOCK_1$; more naturally, $BLOCK_1$ is on $TABLE$ and $BLOCK_2$ is on $BLOCK_1$.

$and'(E_{10}, E_8, E_9)$: E_{10} is the condition of both E_8 and E_9 existing.

$cause(JOHN, E_{13}) \wedge change'(E_{13}, E_{11}, E_{12}) \wedge alive'(E_{11}, BILL) \wedge not'(E_{12}, E_{11})$: John causes a change event E_{13} from a condition E_{11} of Bill's being alive to a condition E_{12} of condition E_{11} not existing; that is, John causes Bill to be not alive.

Existence and truth in the actual universe are treated as predications about individuals in the Platonic universe. For this purpose, we use a predicate *Rexist*. The formula $Rexist(JOHN)$ says that the individual in the Platonic universe denoted by $JOHN$ exists in the actual universe.[5] The formula

$$Rexist(E6) \wedge run'(E6, JOHN) \tag{2.2}$$

says that the condition E_6 of John's running exists in the actual universe, or more simply that "John runs" is true, or still more simply, that John runs. A shorter way to write it is $run(JOHN)$.

Although for a simple sentence like "John runs", a logical form like (2.2) seems a bit overblown, when we come to real sentences in English discourse with their variety of tenses, modalities and adverbial modifiers, the more elaborated logical form is necessary. In fact, when we deal with time in Section 2.3.3, we find we must think carefully about just what the expression $run(JOHN)$ means, and whether we would ever want it to be the logical form of a sentence.

[5]McCarthy (1977) employs a similar technique.

Adopting the notation of (2.2) has the effect of splitting a sentence into its propositional content—$run'(E6, JOHN)$—and its assertional claim—$Rexist(E6)$. This frequently turns out to be useful, as the latter is often in doubt until substantial inferential work has been done. An entire sentence may be embedded within an indirect proof or other extended counterfactual.

2.2.3 Eventualities

We are now in a position to state formally the systematic relation between the unprimed and primed predicates as an axiom schema. For every n-ary predicate p,

$$(\forall\, x_1, \ldots, x_n) p(x_1, \ldots, x_n) \;\supset\; (\exists\, e) Rexist(e) \;\wedge\; p'(e, x_1, \ldots, x_n)$$

That is, if p is true of $x_1, \ldots, x_n$, then there is a condition e of p's being true of $x_1, \ldots, x_n$, and e really exists. Conversely,

$$(\forall\, e, x_1, \ldots, x_n) Rexist(e) \;\wedge\; p'(e, x_1, \ldots, x_n) \;\supset\; p(x_1, \ldots, x_n)$$

That is, if e is the condition of p's being true of $x_1, \ldots, x_n$, and e really exists, then p is true of $x_1, \ldots, x_n$. For future reference, we compress these axiom schemas into one formula:

(A1) $\quad (\forall\, x_1, \ldots, x_n) p(x_1, \ldots, x_n) \;\equiv\; (\exists\, e) Rexist(e) \;\wedge\; p'(e, x_1, \ldots, x_n)$

This may be seen as a weak sort of comprehension axiom. A comprehension axiom is one that relates descriptions to existing entities. Thus, there would be a functional F such that, given a description of a situation $p(x)$, it would produce the corresponding really existing eventuality $F(p(x))$ such that $p'(F(p(x)), x)$ holds. The above axiom is weaker in that the existence of the eventuality is guaranteed, but there is not necessarily any *constructive* way of getting from a predication to a *unique* eventuality.

Axiom schema guarantees the existence in the Platonic universe of any eventuality that also exists in the real world. It will also be useful to have an eventuality in the Platonic universe, though not necessarily in the real world, for any possible predication. The following axiom schema does this:

(A2) $\quad (\forall\, x_1, \ldots, x_n)(\exists\, e) p'(e, x_1, \ldots, x_n)$

That is, given a predicate p, for any set of entities $x_1, \ldots, x_n$, there is an eventuality e that is the eventuality of p being true of those entities. There is no guarantee that e exists in the real world. In fact, very very few of such eventualities will exist in the real world. In the Platonic universe we will be able to form eventualities referring to all sorts of nonsense, such as ideas being green and sets being members of themselves, but these eventualities will not exist in the real world.

Essentially, the Platonic universe is the place for all the things we can talk about. Whether these things exist in the real world requires an independent predication. Separating out things that can be talked about and things that really exist allows us to model language in its full complexity without getting tangled up in physics.

It may also be desirable to have an extensionality axiom schema, specifying the conditions under which two eventualities are identical. A good candidate would be

$$e_1 = e_2 \;\equiv\; (\forall\, x)[p'(e_1, x) \wedge p'(e_2, x)] \wedge [Rexist(e_1) \;\equiv\; Rexist(e_2)]$$

That is, two eventualities e_1 and e_2 are identical if and only if they are the first arguments of the same primed predication and one really exists if and only if the other does too. However, this axiom schema will play no further role in this book, and I will not commit myself to it one way or the other.

I have used words like "condition" and "event" in glossing primed predications, and the reader will find references to states, objects, entities, and so on, but these words, here and throughout the book, have no theoretical status. They are strictly for the convenience of the reader. It sounds less bizarre to talk of John's running as an event than as an individual. But in the semantics of the notation no such distinctions are made. There are only individuals. Conditions, events, situations, objects, entities, and so on, are simply individuals in the Platonic universe and possibly in the real world.

It is convenient, however, to have a term for individuals that occur as the first argument of a primed predication. These are referred to as "eventualities", following Bach's (1981) use of the term "eventuality" for possible states, conditions and events. The constants representing eventualities generally start with the letter E, and the variables with e.

It is sometimes convenient to know of an entity that it is an eventuality; hence, we have the axiom schema

$$(\forall\, e, x_1, \ldots) p'(e, x_1, \ldots) \;\supset\; eventuality(e)$$

It will be convenient to refer in a consistent way to the arguments of primed and unprimed predications. In both $p(x, y)$ and $p'(e, x, y)$, x will be referred to as the "first argument" and y as the "second argument". Rather than refer to e as the "zeroth argument", it will be referred to as the "self argument".

One's intuitions about the nature of eventualities can be strengthened in three steps.

1. First think of eventualities as partial mappings from bits of space-time to the bits of material that occupy that bit of space-time. For example, if I wave my arm, the domain of the partial mapping would be bits of the trajectory that my arm follows through space-time. The range would be bits of my arm. Lots of mappings could be defined. Only some of them will correctly describe the way the world is. Those are the ones that Rexist.

This is good in some instances as a model for eventualities, but it is not a completely adequate conception for linguistic purposes, because of adverbial modification. My running from SRI to Stanford might be fast, whereas my going from SRI to Stanford might be slow, and they correspond to the same partial mapping from space-time to material. The running and the going must be distinguished as eventualities. In this interpretation eventualities are not individuated finely enough.

2. Assign denotations in the real world to non-eventuality variable symbols. Then take the denotation of an eventuality variable symbol e where $p'(e, x, \ldots)$ to be an n-tuple whose first element is the intension of the predicate p, i.e., the mapping from possible worlds into the extension of the predicate p in that world, and whose subsequent elements are the denotations of the arguments x, ..., in the real world. Now my running and my going are different because "run" and "go" have different intensions.

This is also not quite right, because I would like to be able to talk about impossible eventualities, such as the eventuality of Clyde's being a unicorn and the eventuality of Clyde's being a phoenix. These would not be distinguished under this interpretation, since the extensions of both *unicorn* and *phoenix* are empty in all possible worlds (let's say), so they have the same intension.

3. Essentially, an eventuality is anything that can be nominalized—i.e., anything. John's running, John's being tall, John's not being tall, John's being John, Clyde's being a unicorn, the square root of 2's being 1, and so on.

If you accept the reality of such complex entities as the United States, then it is hard to see what ontological scruples you could have against this conception of eventualities.

2.2.4 Transparency and Opacity

A sentence in English typically asserts the existence of one or more eventualities in the real world, and this may or may not imply the existence of other individuals. The logical form of sentence (2.1) is

$$Rexist(E_1) \wedge worship'(E_1, JOHN, ZEUS)$$

This implies $Rexist(JOHN)$ but not $Rexist(ZEUS)$. Similarly, the logical form of "John wants to fly" is

$$Rexist(E_2) \wedge want'(E_2, JOHN, E_3) \wedge fly'(E_3, JOHN)$$

This implies $Rexist(JOHN)$ but not $Rexist(E_3)$. When the existence of the eventuality corresponding to some predication implies the existence of one of the arguments of the predication, we will say that the predicate is *transparent* in that argument, and *opaque* otherwise. Thus, *worship* and *want* are transparent in their first arguments and opaque in their second arguments.[6] In general, if a predicate p is transparent in its nth argument x, this can be encoded by the axiom

$$(\forall\, e, \ldots, x, \ldots) p'(e, \ldots, x, \ldots) \wedge Rexist(e) \supset Rexist(x)$$

That is, if e is p's being true of x and e really exists, then x really exists. Equivalently,

$$(\forall\, \ldots, x, \ldots) p(\ldots, x, \ldots) \supset Rexist(x)$$

In the absence of such axioms, predicates are assumed to be opaque. In practice, in this book, such axioms will not be stated. Where necessary, it will simply be stated that p is transparent or opaque in its nth argument. In a natural language processing system such axioms, as well as axiom schema (A1), would no doubt be implemented by a special-purpose mechanism, rather than being stated explicitly.

The following sentence illustrates the extent to which we must have a way of representing existent and nonexistent states and events in ordinary discourse.[7]

---

[6] Primed predicates are trivially transparent in their self argument.

[7] This sentence is taken from the *New Scientist*, June 3, 1982 (p. 632). I am indebted to Paul Martin for calling it to my attention.

(2.3) The government has repeatedly refused to deny that Prime Minister Margaret Thatcher vetoed the Channel Tunnel at her summit meeting with President Mitterand on 18 May, as *New Scientist* revealed last week.

In addition to the ordinary individuals Margaret Thatcher and President Mitterand and the corporate entity *New Scientist*, there are the intervals of time 18 May and "last week", the (at that time) nonexistent entity, the Channel Tunnel, an individual revealing event and the complex event of the summit meeting, which actually occurred, a set of real refusals distributed across time in a particular way, a denial event which did not occur, and a vetoing event which may or may not have occurred.

Let us take $Past(E_6)$ to mean that E_6 existed in the past and $Perfect(E_1)$ to mean what the perfect aspect means, roughly, that E_1 existed in the past and may not yet be completed. The representation of just the verb, nominalizations, adverbials and tenses of sentence (2.3) is as follows:

$$\begin{aligned}&Perfect(E_1) \wedge repeated(E_1) \wedge refuse'(E_1, GOVT, E_2)\\ &\quad \wedge deny'(E_2, GOVT, E_3) \wedge veto'(E_3, MT, CT) \wedge at'(E_4, E_3, E5)\\ &\quad \wedge meet'(E_5, MT, PM) \wedge on(E_5, 18MAY) \wedge Past(E_6)\\ &\quad \wedge reveal'(E_6, NS, E_3) \wedge last\text{-}week(E_6)^8\end{aligned}$$

Of the various entities referred to, the sentence, via unprimed predicates, asserts the existence of a typical refusal E_1 in a set of refusals and the revelation E_6. The existence of the refusal implies the existence of the government. It does not imply the existence of the denial; quite the opposite. It may suggest the existence of the veto, but certainly does not imply it. The revelation E_6, however, implies the existence of both the *New Scientist* NS and the *at* relation E_4, which in turn implies the existence of the veto and the meeting. These then imply the existence of Margaret Thatcher MT and President Mitterand PM, but not the Channel Tunnel CT. Of course, we know about the existence of some of these entities, such as Margaret Thatcher and President Mitterand, for reasons other than the transparency of predicates.

Sentence (2.3) shows that virtually anything can be embedded in a higher predication. This is the reason, in the logical notation, for flattening everything into predications about individuals.

[8] Actually, in the notation explained in Section 2.5.1, we should have $deny'(E_2, GOVT, E_3\&E_4)$ and $reveal'(E_6, NS, E_3\&E_4)$.

It may be objected[9] that this notation does not distinguish between the assertion, as in

> John ran.

and the nominalization, as in

> I doubt that John ran.

or

> That John ran surprised me.

In fact, in all of these sentences there is the eventuality E of John's running. In the first the real existence of the eventuality is asserted. In the second and third it is the content of a propositional attitude. The assertion and the nominalization both *describe* the eventuality; the assertion is distinguished by the fact that it also *makes a claim* about its real existence.

2.2.5 Sorts

One might think it desirable to distinguish individuals by sort in the notation itself and to restrict predicates as to the sorts to which they can apply. We could, for example, have the sorts $OBJECT$ and $SUBSTANCE$ and the restriction on the predicate *white* that it can only apply to individuals of these two sorts. This would avoid nonsensical expressions like $white(E) \wedge run'(E, JOHN)$, or "John runs whitely." I am not inclined to complicate the notation in this way. For one thing, the sorts that occur to one most naturally—"object", "substance", "state", "event", "action"—do not have clear boundaries. For example, is fog an object, substance, state, or event? Similar difficulties are likely in any scheme of types one might devise.

For another thing, it is not possible to classify things in the universe finely enough to prevent the construction of *all* nonsensical expressions. For example, we would not want to rule out $man(JOHN) \wedge dog(JOHN)$ on the basis of a sort violation. So why rule out anything on this basis? It may be that the sentences "John is a man and a dog" and "John runs whitely" fail to make sense for distinct reasons, but it is not clear that this distinction should be reflected in the increased complexity of the syntax of the formalism.

---

[9] Leonard Schubert, personal communication

Finally, any information that could be encoded by sorts in the notation can just as easily be encoded in the axioms. For example, if we want running to be an event, we can state this as an axiom,

$$(\forall\, e, x) run'(e, x) \supset event(e)$$

If e is x's running, then e is an event. If we want the classes of physical objects and events to be disjoint, we can state this as an axiom.

$$(\forall\, e) physical\text{-}object(e) \wedge event(e) \supset F$$

It is contradictory for e to be both a physical object and an event.

2.2.6 Views of Semantics

The reader may detect, in what has gone before and in what follows, an idiosyncratic view of semantics. Let me make my views explicit. There are three roles semantics may play in an enterprise like this one. First, it can really be semantics. It can be the attempted specification of the relation between language and the world. However, this requires a theory of the world. There is a spectrum of choices one can make in this regard. At one end of the spectrum – let's say the right end – one can adopt the "correct" theory of the world, the theory given by quantum mechanics and the other sciences. If one does this, semantics becomes impossible because it is no less than all of science, a fact that led Fodor (1980) to express some despair. There's too much of a mismatch between the way we view the world and the way the world really is. At the left end, one can assume a theory of the world that is isomorphic to the way we talk about it. This chapter, in fact, is an effort to work out the details in such a theory. In this case, semantics becomes very nearly trivial, as I show in Chapter 4. Most activity in semantics is just to the right of the extreme left end of this spectrum. One makes certain assumptions about the nature of the world that closely reflect language, and doesn't make certain other assumptions. Where one has failed to make the necessary assumptions, puzzles appear, and semantics becomes an effort to solve those puzzles. Nevertheless, it fails to move far enough away from language to represent significant progress toward the right end of the spectrum. The position adopted in this book is that there is no reason to make our task more difficult. We will have puzzles enough to solve when we get to encoding commonsense knowledge and interpreting discourse.

The second role for semantics is in conceptual guidance, and in this respect, ontological promiscuity is not very useful. It does not help us in axiomatizing domains in a consistent fashion, in a way that avoids unsuspected contradictions. In fact, I find that when I am axiomatizing a domain, I introduce sorts, functions, functionals, and higher operators all over the place when first working out the details, and then later flatten it into the notation presented here. The flat notation is intended for broad coverage of the content of discourse, and not as a conceptual tool for axiomatizing particular domains.[10]

Finally, semantics has a heuristic role. Standard model-theoretic semantics can be used as a conceptual tool to show the mutual consistency of a set of axioms. Here one's models are not at all the intended models, but set-theoretic or arithmetic constructions which are well-understood and convenient to work with. There are several examples of this in the rest of this chapter.

There is a range of attitudes the reader can take toward ontological promiscuity. At one extreme, he or she can view it as a way of cutting, in one swift stroke, through the Gordian knot presented by classical representational difficulties. At the other, more prudent extreme, the Platonic universe may be viewed as a notational halfway house on the road from language to meaning in the real world. Ontological promiscuity can be seen purely as a heuristic, not for solving the difficulties, but for bypassing them, in a way that will let us get on to the real problems of interest, and, hopefully, in a way that will survive future representational adjustments by the more ontologically scrupulous.

In summary, the logical notation is ontologically promiscuous in that it assumes a Platonic universe that has such disreputable entities as events, states, conditions, nonevents, nonexistent objects, and as we will see later, typical elements of sets. This maneuver allows us to keep our logical notation flat and simple, uncluttered by higher operators, and consequently easier for the inference process to manipulate.

2.3 Some Classical Notational Difficulties

In this section a number of classic problems in the representation of knowledge are considered. For each of them, notational devices and corresponding ontological assumptions are proposed that will allow us to slip past the difficulties. For the most part, they are devices that push the problem from the

[10] Ontological promiscuity, on the other hand, is essential in axiomatizing any interesting domain; events need to be reified.

notation into the axiomatization. Of all these, the key problems to be solved if the approach is to succeed are the problems of sets and quantification (Sections 2.4.5 and 2.4.6), negation and other opaque operators (Sections 2.5.1 and 2.5.3), and belief and other intensional contexts (Section 2.6).

2.3.1 Case Relations

First let us consider case relations, or the relations between predicates and their arguments. Generally, a sentence like

John builds the boat.

is represented as (ignoring tense and the definite article):

$$build(J, B)$$

or, from Section 2.2,

$$Rexist(E) \wedge build'(E, J, B)$$

The facts that John is the agent of the building event and the boat the object or "theme" are implicit in the order of the arguments. It is sometimes convenient, however, to be able to express these case relations explicitly. We can do so by means of the predicates *Agent* and *Object* related to the predicate *build* by the following axioms:[11]

$$(2.4) \quad (\forall\, e, x, y) build'(e, x, y) \supset Agent(x, e) \wedge Object(y, e)$$
$$(2.5) \quad (\forall\, e, x, y, z) build'(e, x, y) \wedge Agent(z, e) \supset x = z$$
$$(2.6) \quad (\forall\, e, x, y, z) build'(e, x, y) \wedge Object(z, e) \supset y = z$$

That is, if e is a building event by x of y, then x is the agent of e and y the object, and conversely, if z is the agent of a building event e by x of y, then z is x, and similarly for objects. We will not actually encode axioms like (2.4)–(2.6) for each predicate, but we will appeal to them when we need them.[12]

[11] The reader might well begin to see this approach as not only ontologically promiscuous, but also axiomatically promiscuous.

[12] This treatment of case labels means that we must be careful in individuating eventualities. The agent of a buying event is not the same as the agent of a selling event. See Castañeda (1967). The problem of individuating eventualities occurs in a more severe form in the treatment of adverbials.

It is often debated whether some type of phrase should appear as an argument or as an adverbial, where obligatory elements should be arguments and adverbials should be optional. In this framework, the issue evaporates. We could decide to represent the sentence

(2.7) John breaks the window with a hammer.

either as

$$Rexist(E) \wedge break'(E, J, W, H)$$

or as

$$(2.8) \quad Rexist(E) \wedge break'(E, J, W) \wedge Instrument(H, E)$$

In the former case, if we don't know the instrument, we can simply fill the argument position with an existentially quantified variable:

$$(\exists\, x)Rexist(E) \wedge break'(E, J, W, x)$$

In the latter case if we want the instrument to be obligatory, we can posit the axiom

$$(\forall\, e, x, y)break'(e, x, y) \supset (\exists\, z)Instrument(z, e)$$

Which way one decides the issue is of no real consequence. In this book sentences like (2.7) will generally be represented with fewer entities as arguments, as in (2.8). This is because there is virtually no limit to the things one might want to consider arguments of the predicate. Events may have instruments, times, locations, various manners, rates, sources, goals, paths, benefactors, and so on. Only a few of these will be relevant in any given context, and when we simply want to refer to John's leaving, it seems perverse to enforce a notation like

$$(\exists\, t_1, x_1, x_2, x_3, x_4, \ldots)leave(J, t_1, x_1, x_2, x_3, x_4, \ldots)$$

One reason, and perhaps the only reason, for incorporating case relations into one's semantic theory is to capture semantic generalizations. There may be things that are true of all agents or instruments, irrespective of what actions they are agents or instruments of. In our framework such generalizations would be captured by axioms of the form

$$(\forall\, e, x) Agent(x, e) \supset p(x, e, \ldots)$$

Very little use is made of such case labels in this book; my feeling is that the standard cases are very hard to apply in general—for example, what are the cases of the arguments for "outnumber"? But very many verbs follow the pattern described in Chapter 5 as the "ontological ascent", and for those verbs many of the standard case roles are simply labels for the role they play in that pattern.

Nevertheless, it is extremely useful to be able to talk about an entity being the nth *argument* of a predication (or more precisely, the corresponding "participant" in an eventuality), regardless of what case the argument is. It is thus convenient to have this axiom schema for every predicate p and its nth (non-self) argument x:

(A3) $(\forall\, e, \ldots, x, \ldots) p'(e, \ldots, x, \ldots) \supset Argn(x, n, e)$

In line with the convention in Section 2.2.3, the self argument e is the zeroth argument of itself:

(A4) $(\forall\, e, \ldots) p'(e, \ldots) \supset Argn(e, 0, e)$

Often it is sufficient to say simply that that an entity is an argument of a predication. The predicate Arg encodes this.

$$(\forall\, x, n, e) Argn(x, n, e) \supset Arg(x, e)$$

If x is the nth argument of e, then x is an argument of e.

It will also be useful to have a more general notion of argument, for things that are arguments, or arguments of arguments, or ... For this we will use the predicate $arg{+}$, defined recursively as follows:

$$(\forall\, x, e)[arg{+}(x, e) \equiv [arg(x, e) \vee (\exists\, e_1)[arg(e_1, e) \wedge arg{+}(x, e_1)]]]$$

Thus, in

John believes Bill is tall.

Bill is in an $arg{+}$ relation with John's belief. Think of the relation as "involved somehow in".

In general, in this book, the order of arguments in a predication will follow the order in the corresponding active English sentence, nonditransitive when there is a choice. Thus, A lifts B is represented $lift(A, B)$, A gives B to C is represented $give(A, B, C)$, A is strong is $strong(A)$, and A is the father of B is $father(A, B)$. The first argument is often referred to as the logical subject ($Lsubj$) and the second as the logical object ($Lobj$). Thus,

$$(\forall\, x, e) Argn(x, 1, e) \equiv Lsubj(x, e)$$
$$(\forall\, x, e) Argn(x, 2, e) \equiv Lobj(x, e)$$

These labels will be used in Chapter 4.

2.3.2 Context Parameters

Before launching into other specific semantic phenomena, it will be useful to introduce a device of broad applicability.

In thinking about a problematic phenomenon, we very often begin with simple, clear, and elegant intuitions. We then encounter problems and make various moves to get around them. By the time we have dealt with them, we have a formal treatment that is quite complex. It is my aim in this work to frontload this complexity, so that the representations and axiomatizations that result at the end preserve the original simplicity of the intuitions.

A common instance of this is the realization that some predicate needs more arguments than we originally thought, or that seem to be used in the corresponding natural language expression. For example, as we will see below, *tall* at first blush seems to be a property of an individual. A little analysis, however, suggests that we also need a parameter for the comparison set—tall in comparison with what group? Indexicals like "I" and "you" are treated as properties of individuals, but they also have to be parameterized on a speech situation—"I" refers to the speaker in a given speech situation. We can talk about one eventuality e_1 *enabling* another eventuality e_2, but as is argued in Chapter 5, we really require a specification of the particular *way* of achieving e_2. That is, e_1 enables e_2 relative to a particular "causal complex" s.

However, we would like to retain, as much as possible, the simplicity of having fewer arguments, and of having predicates whose arguments correspond to things that are most often explicit in discourse. We can do this by positing two predicates for such concepts—p and p_0. The predicate p_0 will have the extra context parameter; the predicate p will not. The relation between them can be expressed by axioms of the form

$$(\forall\, x)[p(x) \equiv (\exists\, c) p_0(x, c)]$$

That is, p is true of x just in case there is a context parameter c such that p_0 is true of x and c.

There is a problem with this formulation however. If we know $p(x)$ and $p(y)$, then we can infer $p_0(x, c_1)$ and $p_0(y, c_2)$, but the axiom gives us no way to establish the equality of c_1 and c_2. If I say that John is tall and Bill is tall, there is no guarantee that they are tall with respect to the same comparison group.

Nevertheless, in a single discourse it generally is true that the context variables are the same, and that we can safely use the lower-arity predicate. The method of interpretation by abduction presented in Chapter 3 aims for minimality in interpretations, where one contributor to minimality is assuming the identity of differently presented variables. Thus, the most common state in discourse interpretation will be one in which the corresponding context parameters are taken to be identical.

In this chapter, where such a context parameter is relevant, an axiom of the above form will be posited, and both predicates will be used—p_0 where the context parameter must be explicitly reasoned with and p where it needn't be.

Sometimes the assumption of identity is wrong. In these situations interpretations are often consciously problematic—"Oh, you meant short for a basketball player!" When this occurs, the axiom in the above form must be invoked so explicit reasoning about the context parameter can take place.

Following the style of the previous section, we could as easily have expressed the dependence on the context variable as a separate predication, using in place of $p_0(x, c)$ the expression

$$p'(e, x) \wedge \mathit{in\text{-}context}(e, c) \wedge \mathit{Rexists}(e)$$

That is, e is the eventuality of p's being true of x, e really exists, and e is true with respect to the context parameter c.

2.3.3 Time, Tense, and Aspect

In many proposals for a logical notation for events, the time of the event is built into the predication itself as an argument. The sentence

John left at 2 p.m.

would be represented

$$\mathit{leave}(J, 2PM)$$

or

$leave(J, t) \wedge 2pm(t)$

In favor of this notation is the fact that most states and events that we think of have a point or interval of time at or during which they occur.

For spatial information, we already need to be able to make an independent predication saying that an object is at a location.

John is at the store $\Longrightarrow at(J, S)$

We might as well represent the location of events in the same way.

John works at the store. $\Longrightarrow work'(E, J) \wedge at(E, S)$

The location of an event is simply one of its properties.

Similarly, the time of an event or state may be treated simply as one of its properties. Hence,

John left at 2 p.m.

may be represented

$leave'(E, J) \wedge at\text{-}time(E, 2pm)$

This parallels the treatment of "minor" arguments like instruments in Section 2.3.1. This approach allows us to regularize the large variety of ways temporal information is carried in English sentences. They all become predications about a state or event.

For tense and aspect, we assume there are predicates *Past*, *Future*, *Perfect* and *Progressive*. (Concerning the present tense, see below.) Then

(2.9) John left.

is represented

(2.10) $Past(E) \wedge leave'(E, J)$

That is, E is a leaving by John, and E occurred in the past. The sentence

John has been leaving.

is represented

$$Perfect(E_1) \wedge Progressive'(E_1, E_2) \wedge leave'(E_2, J)$$

That is, E_2 is a leaving event by John, E_1 is the eventuality that the property conveyed by the progressive aspect is true of E_2, and the property conveyed by the perfect aspect is true of E_1.

The semantic properties of the tenses and aspects are encoded in the knowledge base in axioms involving these predicates. Thus, $Past(E)$ implies that there is a time t that is before the time of utterance (now) and that E occurred at that time. $Progressive(E)$ implies that E is (being viewed as) an event with some duration. $Perfect(E)$ implies that E occurred before some reference point. How these predications are associated with their syntactic and morphological realizations is explicated in Chapter 4. How they are meshed with theories of time and event structure is explicated in Chapter 5.

Temporal adverbs like "yesterday" and temporal subordinate conjunctions like "before" are represented similarly. Hence, the following notations:

(2.11) John left yesterday. $\Longrightarrow Past(E) \wedge leave'(E, J) \wedge yesterday(E)$

(2.12) John left before Mary arrived. $\Longrightarrow$
$$Past(E) \wedge leave'(E, J) \wedge Past'(E_1, E_2) \wedge arrive'(E_2, M) \\ \wedge before(E, E_2)$$

In the case of indexicals like "yesterday" and the tenses, the predications will be related to the situation of utterance in the proper way by axioms (see Section 2.3.9). Note that the representation of (2.11) contains the redundant $yesterday(E) \wedge Past(E)$. This is because the sentence redundantly contains the past tense and the word "yesterday".

It is worth pointing out in (2.12) that *before* is not transparent in its second argument. In

John caught the glass before it broke.

we cannot assume the existence of the breaking event from the existence of the $before$ relation, but we can assume the existence of the catching event. For this reason, although example (2.12) conveys the possible eventuality E_1 of the pastness of the arriving E_2—$Past'(E_1, E_2)$—E_1 does not necessarily really exist, and we cannot write $Past(E_2)$.

As soon as we introduce time, we must ask what we really mean by an expression like

$$man(J)$$

or equivalently,

$$Rexist(E) \wedge man'(E, J)$$

Do we mean that John is a man at the present time; or that there is some condition sometime in the past, present or future when John was, is or will be a man; or that John is a man at all times past, present and future. This is not a question of fact, but of what we are going to take our notation to mean.

We will take $Rexist$ to mean existence *at the present time* and thus $man(J)$ to mean that John is a man *at the present time*.[13] When applied to the main predication of a sentence, it is equivalent to the present tense. For completeness, we introduce a predicate $Present$ that is conveyed by the present tense and is equivalent to $Rexist$:

$$(\forall\, e)Rexist(e) \equiv Present(e)$$

The simple notation $man(J)$ will be used primarily for simplifying logical forms when complicating them would serve no purpose. For sentence (2.9), instead of writing expression (2.10), we could have written

$$Rexist(E_1) \wedge Past'(E_1, E) \wedge leave'(E, J)$$

That is, the past-ness of John's leaving exists in the present, implying that John left in the past. For this sentence, this is more elaborate than necessary, because no use is made of the eventuality E_1 other than asserting its real existence. The tense of the highest predication in a sentence will generally use the unprimed predicate.

However, for serious analysis of complex discourse, the elaborated notation will be employed, and the reader, when spotting a possible flaw in representation or argumentation, should see whether it survives translation into to the more elaborated notation.

[13] Thus, $Rexist$ may not encode some people's intuitive notion of existence, which would include past existence as well, and perhaps future.

2.3.4 Events and Actions

In many approaches to semantics, a distinction is made between events and actions. In terms of logical form, in these approaches, an event is more like a fully saturated predication, whereas an action is more like a lambda expression waiting to take an agent as its argument. That is, we might represent the event of John's hitting the ball as $hit(J, B)$, whereas the action of hitting the ball would be represented as $\lambda x[hit(x, B)]$.

Being ontologically promiscuous, I have no objection in principle to allowing actions into the ontology. But it seems to me perfectly adequate to have actions as a subclass of events, namely those events that have agents.

Schubert (1994) raises several illegitimate objections to this conflation.

> Actions can be deliberate, devious, purposeful, unintentional, or hasty, while events cannot. Actions can be performed or carried out or done, while events cannot. For actions we can ask "Who (or what) did it?" while for events we cannot. In short, actions have agents, while events don't. (Clearly they are connected in that the PERFORMANCE of an action constitutes an event).

Note that all of the adjectives he lists are predicates of two arguments, an event/action and an agent. There is no problem with restricting their applicability to events with agents.

$$deliberate(e, a) \supset Agent(a, e)$$

That is, if event e is deliberate on the part of a, then a is the agent of e. The sense of "do" in "Who did it?" similarly selects for events with agents.

Schubert continues:

> Of course, if conflating these intuitively distinct notions could be shown to have no adverse consequences in computational semantics, then it may be justified (on grounds of simplicity). But the evidence seems to me to run in the other direction. In particular, treating manner adverbials as event predicates potentially leads to error. For instance, it seems reasonable to say that "John wrestled with Bill"and "Bill wrestled with John" can both characterize the SAME wrestling event (from slightly different perspectives). Now suppose it's true in addition that "John wrestled skillfully with Bill", while "Bill wrestled clumsily with John". On a Davidsonian analysis, we end up saying of one

> and the same event that it is skillful and that it is clumsy, a contradiction (and also a very odd way to talk about EVENTS). In a sense, the problem comes from obliterating (in the semantics) the SYNTACTIC fact that manner adverbs modify verb phrases, and the verb phrases are distinct in the two cases.

This argument is illegitimate, because *skillful* and *clumsy* are predicates of two arguments. John can be skillful in the wrestling match between John and Bill, while Bill can be clumsy in the same match. Even if they are the same event, represented as

$$wrestle'(E, J, B) \wedge wrestle'(E, B, J)$$

(same E), the *skillful* and *clumsy* predications have different first arguments:

$$skillful(J, E) \wedge clumsy(B, E)$$

John is skillful in the match and Bill is clumsy in it. We do not end up saying the match is both skillful and clumsy.

We will return to the issue of actions in Section **??**. The recovery of the correct arguments for manner adverbials in syntactic analysis is discussed in Chapter 4.

2.3.5 Manner and Other Transparent Adverbials

The treatment of transparent manner adverbials is now straightforward. Like temporal predications and "minor" case arguments, they are properties of the eventuality. Hence,

John runs slowly. $\Longrightarrow Rexist(E_1) \wedge run'(E_1, J) \wedge slow(E_1)$

John greets Mary with reluctance. $\Longrightarrow Rexist(E_2) \wedge greet'(E_2, J, M)$
$\wedge\, with(E_2, R) \wedge reluctance(R, J, E_2)$

John greets Mary reluctantly. $\Longrightarrow Rexist(E_2) \wedge greet'(E_2, J, M)$
$\wedge\, reluctant(J, E_2)$

R is the reluctance of John to do E_2. The predication $reluctant(J, E_2)$ says that John is reluctant to do E_2. The reader will not be surprised that the relation between "with reluctance" and "reluctantly" is something which must be extracted inferentially from the knowledge base.

It is important to note, as we saw in the previous section, that some manner adverbials take one argument (the event) while others take two arguments (the event and a participant in it). Thus, *slow* is a predicate with one argument, and *reluctant* is an argument with two.

Some linguists (??, 19??) have noticed that the interpretations of sentences can sometimes be influenced by the placement of adverbials, and have used this to argue for two different logical forms. In

John spoke to the queen rudely.

the sense is that it was appropriate for John to speak to the queen, but the manner in which he did it was rude. In

Rudely, John spoke to the queen.

the sense is that it was inappropriate for John to speak to the queen at all, no matter how polite his manner of speaking was. A common analysis of these two examples has the adverb attaching to the verb phrase, and thus the action, in the fist sentence, and to the clause as a whole, and hence to the event, in the second.

My view is that in both cases it is the event that is given the attribute. In both cases, the relevant fragment of logical form is $rude(J, E) \wedge speak'(E, J, Q)$. In the first case, the interpretation requires us to unpack the event (but certainly not by removing the agent to produce an action), to examine at a finer granularity its internal structure, and to note properties of parts of the utterance such as lexical choices and intonation. In the second case, we view the speaking event as an undecomposed whole and look only at its relation to the surrounding environment, such as the context in which the utterance occurs. It is not clear to me that this distinction is well captured by the distinction between actions and events. The two readings are just different views on an event.

This issue is briefly discussed again in Section 4.9.3.

2.3.6 Comparatives

In line with our decision to treat each morpheme as a predicate, it should come as no surprise that we will view comparatives as resulting from applying the predicate *more* to several arguments. The arguments will be the two entities being compared and the scale or partial ordering they are being compared with respect to.

John is taller than Bill. $\Longrightarrow$

(2.13) $more(J, B, E) \wedge tall'(E, X, S)$

The predicate *more* takes as its third argument an abstract condition which defines a scale, and takes as its first two arguments two entities that are being compared with respect to that scale.

More generally, whenever a comparative is used there is a scale corresponding to the property.

$$more(x, y, e) \supset scale\text{-}for(s, e) \wedge at(x, a, s) \wedge at(y, b, s) \wedge exceed(a, b, s)$$

Superlatives are represented similarly:

$$\begin{array}{ll} & \text{John is the tallest in the class.} \Longrightarrow \\ (2.14) & \quad most(J, C, E) \wedge tall'(E, X, S) \end{array}$$

The predicate *most* is like the predicate *more* except that its second argument is the set the first argument is being compared with.

One problem with comparatives is to find the appropriate associated scale, that is, a scale s such that we can infer from (2.13) that

$$at(x, a, s) \wedge at(y, b, s) \supset exceed(a, b, s)$$

that is, if x is at a on s and y is at b on s then a exceeds b on s. For the most common scalar adjectives we can assume there to be in the knowledge base axioms of the form

$$tall'(e, x) \supset (\exists\, s) Height\text{-}scale(s) \wedge at(x, y) \wedge in(y, s_1) \wedge Hi(s_1, s)$$

That is, if e is the tallness of x then there is a height scale s and e is the property of x being at y on s, and y is in the "high" region s_1 of s. From such axioms, the scale can be determined. In general, however, finding the intended scale can involve quite intricate inferential processing, as in the sentence

Faulkner is a better writer than Hemingway.

A theory of scales is developed in Chapter 5.

Hans Kamp (1975) suggests that the comparative involves an implicit quantification over reference sets. That is, if John is taller than Bill, then for all reference sets r if $tall(B, r)$ then $tall(J, r)$. This can be viewed as just one possible way to induce a scale. A counterexample would be where the reference sets are different, as in

John is better in English than Bill is in math.

Here the comparison set is good students in each subject, but they might be quite different.

To deal with "how" questions and nominalizations, like

How tall is John?

(2.15) I know how tall John is.

we introduce the predicate *measure*. The expression

measure(X,D,S)

means "X measures D on scale S" or "D is the measure of X on scale S". The word "how" is treated as we treat "what" and "who" in Section 2.6, as a context-dependent essential predicate, to be further specified by inference. Then the logical form for (2.15) is

$$know(I, H) \wedge how'(H, D) \wedge measure(J, D, T) \wedge tall'(T, X, S)$$

That is, I know H where H is the essential property, or the "how-ness", of D where John measures D on the tallness scale.

The predicate *measure* is related to *more* and *most* by axioms like the following:

$$(\forall\, x, y, s, d_1, d_2) more(x, y, s) \wedge measure(x, d_1, s) \wedge measure(y, d_2, s) \\ \quad \supset\ d_1 > d_2$$

$$(\forall\, x, c, y, s, d_1, d_2) most(x, c, s) \wedge measure(x, d_1, s) \wedge element(y, c) \\ \quad \wedge\, measure(y, d_2, s)\ \supset\ d_1 \geq d_2$$

If x is more than y on scale s and x and y measure d_1 and d_2, respectively, on s, then d_1 exceeds d_2 on s. If x is the highest of set c on scale s, x measures d_1 on s, y is an element of c, and y measures d_2 on s, then d_1 is greater than or equal to d_2 on s.

Measures are not necessarily numerical. The second argument of *measure* is not necessarily a number. Its essential property may be a loose sort of description. Thus, the above notations are adequate for adjectives like *happy*:

John is happier than Mary.

How happy is John?

2.3.7 Attributives

It has often been pointed out that in the two phrases "the big mosquito" and "the big elephant" the word "big" must be interpreted as referring to different ranges of size. There are three problems this observation raises.

First, should attributives like "big" have an explicit argument for the reference set, the set with respect to which the entity is big? This was answered in Section 2.3.2. We can introduce a predicate big_0 that has two arguments, on for the entity described and one for the comparison set. The predicate *big* has only the first of these arguments. They are related by the axiom

$$(\forall\, x) big(x) \;\equiv\; (\exists\, s) big_0(x, s)$$

That is, x is big if and only if it is "big_0" with respect to some comparison set s.

Where several comparisons are being made with respect to the same comparison set, we will rely on the preference for minimality in abduction to identify these. Where comparisons are made with respect to different comparison sets, we will expect specific features of the text and context to force this distinction. In analyzing a phrase like "big for a mosquito", we will expect the interpretation of "for" to induce the inference to big_0 and the comparison set.

The second issue is what role the rest of the noun phrase—"mosquito" or "elephant"—plays in determining that reference set. One might think that this is the job of compositional semantics. The comparison set is determined by the description in the rest of the noun phrase. The comparison set for "a big mosquito" is the set of all things satisfying the property "mosquito". But in

That mosquito is big.

compositional semantics will have a harder time doing the job. In

That nearby Alaskan mosquito is big.

how do we know that the comparison set should include all Alaskan mosquitoes and not just the ones that are nearby. Finally, I might point to a mosquito and say

That thing is big.

I don't intend the comparison set to be the set of all things.

In fact the determination of the comparison set is essentially a reference problem—resolve the reference of s—and hence part of the process of determining the best interpretation for the discourse as a whole. This is addressed further in Chapter 6.

The third problem is—to what range of values within the reference set is the attributive properly applied? What sizes count as big? This issue is discussed further in Section 5. A number of factors can come into play, including the place of the entity within the comparison set, the distribution of members of the comparison set across the scale, and functional relations between the measure of a quality or quantity and the achievement of certain goals.

2.3.8 Proper Names

In many of the examples in this book, purely as a notational convenience to avoid clutter, proper names in English sentences are translated into constants with suggestive names, like *JOHN* or J. If we were to take this as more than a convenience, the constants would designate specific possible individuals in the Platonic universe, which possessed a particular set of properties. They would thus convey semantic information, in violation of one of the decisions of Section 2.1. Therefore, when a proper name is a serious part of what we want to examine in a sentence, it is treated as a predicate applied to a constant or variable which itself has no semantic content. The representation of "John" will not simply be the semantically loaded constant *JOHN* but a semantically neutral constant (or existentially quantified variable), let's say X, together with the property $John(X)$. It is a property that is true of everyone named "John". It tells us no more than proper names in discourse tell us, namely, that someone is named "John". Sometimes the property is the main assertion of the sentence, as in introductions and identifications—

This is John.

My boss is John.

—in which case the namee is the referent of the subject of the sentence. Usually the property is grammatically subordinated, and the namee is the referent of the proper name itself. In these cases, the entity it identifies must be determined contextually, as with any other definite reference, on the basis of the information given – here, the property $John(X)$.

Where the reference has been resolved and we wish the name to designate a unique individual unambiguously, we can index the predicate — $John_1(X)$. The indexed predicate would be true of only that individual, and properties that are known about that individual would be encoded in axioms involving the indexed predicate. This treatment is exactly like other cases of lexical ambiguity.

Issues that are problems for philosophers of language may or may not be problems for one interested in defining a logical notation. In the next section we will see two issues that are. However, there are two questions about proper names that have played a large role in the philosophy of language and that are not relevant to our concerns. First is an issue raised by Kripke (1972). Kripke has argued that proper names are rigid designators, denoting the same individual in all possible worlds. There are a few situations in which names take priority over the entities to which they refer. For example, after the sudden death of Pope John Paul I, it was fairly clear that the next pope would choose the name "John Paul II". Thus, we can say

> If one cardinal had voted otherwise, Pope John Paul II would have been Italian.

Similarly, a couple may decide to have a baby and before it is conceived decide to call it Kim. Then one may say to the other

> I hope Kim is a girl.

But for the most part proper names are intended to function in a more stable manner. They are intended to fix reference, Kripke argues, in a way that is constant among possible worlds or situations. This might seem to indicate that predications like $John(X)$ are different in kind from other predications conveyed by sentences. It is just as reasonable, however, to say that, from our perspective, Kripke's argument amounts to no more than the point that the existential quantifier implicit in proper names ("There is an X such that X is John and ...") usually outscopes modal contexts. This is a fact about language use, and not a fact that in any way should shape or constrain our logical notation.

Another classical problem in the philosophy of language has been the question of whether a proper name is equivalent to a conjunction of properties (Frege, 1893; Searle, 1969, pp. 162-174). Is it possible to specify necessary and sufficient conditions for some entity to be Aristotle? It is the contention in this book, particularly in Chapter 5, that it is almost always impossible to give necessary and sufficient conditions for any concept, those encoded by proper names included. Thus, we can expect to have a large number of necessary conditions, expressed by axioms of the form

$$(\forall x) Aristotle(x) \supset p(x)$$

and a large number of sufficient conditions, expressed by axioms of the form

$$(\forall x) q(x) \supset Aristotle(x)$$

but no axioms of the form

$$(\forall x) Aristotle(x) \equiv r(x)$$

It is taken for granted that replacement of proper names is impossible in general, and, since we are introducing predicates like *Aristotle*, neither necessary nor desirable.

Names are put to many uses and therefore carry a great deal of information. Parents select names for their children for pleasing connotative or aesthetic qualities. Many women still change their names to signal a change in marital status. Movie stars, like Rock Hudson and Clint Eastwood, choose names that are in concord with the image they wish to present. There is a connection among the facts that President Carter ran as a "man of the people", that he wore cardigan sweaters during his fireside chats, and that he called himself "Jimmy". Immigrants often change their names to names more typical of their new country, "Golda Meyerson" to "Golda Meier", for example. A name change signals one's ascension as Pope or emperor. In other cultures, even more suggestive use is made of names. Japanese artists typically changed their names when they changed their artistic styles. Geertz (1973) reports that in Bali a man changes his name on the birth of his first son, and then again on the birth of his first grandson, signalling elevation of status in a society that honors age and family. I've drawn my examples only from names of people; place names can be put to an even wilder variety of uses. Proper names do not simply designate individuals. They frequently tell a great deal about these individuals. By treating names as predicates, we can hope to capture, as axioms in the knowledge base, some of the wealth of connotations associated with proper names.

2.3.9 Indexicals

The treatment of indexicals can be divided into two parts, a theory of discourse situations and a theory of the self and how one is embedded in the world. We can view most of the knowledge base as fairly stable across time. Generally, facts are learned and beliefs change at a much slower rate than texts are read or conversations are conducted. There is one exception to this, however. An important part of the knowledge base must be a continuously updated theory of "what's going on right now". This is the theory of how one is embedded in the world. In natural language understanding part of what's going on right now will always be the text itself and the discourse situation surrounding the text. Facts about these will be constructed out of a theory of discourse situations in general.

Within the discourse situation there are utterance situations—the situations surrounding and determined by the character of individual utterances. Among the many facts that characterize an utterance situation at any given moment, there are several kinds of facts that are always relevant: who the speaker is, who the hearers are, and the time and location of the utterance. Thus, an utterance situation is characterized by, among others, the following axioms:

$$(\forall\, s) Utterance\text{-}situation(s) \supset (\exists\, i) Speaker(i, s)$$
$$(\forall\, s) Utterance\text{-}situation(s) \supset (\exists\, u) Hearer(u, s)$$
$$(\forall\, s) Utterance\text{-}situation(s) \supset (\exists\, t) time\text{-}of(t, s)$$
$$(\forall\, s) Utterance\text{-}situation(s) \supset (\exists\, l) location\text{-}of(t, s)$$

$Speaker(i, s)$ means that i is the speaker in discourse situation s, and the other axioms are interpreted similarly.

Indexicals, such as "I", "you", "now", "hear" and "today", are, unsurprisingly, treated as predicates, like everything else. They are all with respect to an utterance situation, and we can include this as a second argument. Following the treatment of context parameters in Section 2.3.2, we will have subscripted predicates in which the utterance situation is explicit and unsubscripted ones in which it is not. The relevant axioms for the subscripted predicates are as follows.

$$(\forall\, i, s) I_0(i, s) \equiv Utterance\text{-}situation(s) \wedge Speaker(i, s)$$
$$(\forall\, u, s) you_0(u, s) \equiv Utterance\text{-}situation(s) \wedge Hearer(u, s)$$
$$(\forall\, t, s) now_0(t, s) \equiv Utterance\text{-}situation(s) \wedge time\text{-}of(t, s)$$
$$(\forall\, l, s) here_0(l, s) \equiv Utterance\text{-}situation(s) \wedge location\text{-}of(l, s)$$
$$(\forall\, d, s) today_0(d, s) \equiv (\exists\, t) Utterance\text{-}situation(s) \wedge time\text{-}of(t, s) \wedge in(t, d) \wedge day(d)$$

If an agent i is referred to as "I" in an utterance situation s, then i is the speaker in s. "Hearer" refers to intended hearers in s, not anyone who just happens along.

The axioms relating the subscripted and unsubscripted predicates are as follows:

$$(\forall\, i) I(i) \;\equiv\; (\exists\, s) I_0(i, s)]$$
$$(\forall\, u) you(u) \;\equiv\; (\exists\, s) you_0(u, s)]$$
$$(\forall\, t) now(t) \;\equiv\; (\exists\, s) now_0(t, s)]$$
$$(\forall\, l) here(l) \;\equiv\; (\exists\, s) here_0(l, s)]$$
$$(\forall\, d) today(d) \;\equiv\; (\exists\, s) today_0(t, s)]$$

The entity i is describable as "I" if and only if the subscripted predicate I_0 is true of i and some utterance situation s. The other axioms are interpreted similarly.

When we encounter an indexical in an utterance, we need to determine the utterance situation containing the indexical, and we are then able to identify the referent. Treating indexicals axiomatically like this, rather than metalinguistically, allows us to deal in a unified manner with uses of indexicals other than just those in the current situation, such as "now" referring to a time within the events of a narrative and "I" in direct quotes. The many uses of "you" can be analyzed in terms of envisioned utterance situations.

The word "this" is somewhat more complicated. At first cut, it refers to something nearby (under some granularity) and perceptually salient. It can thus be used to refer to something being pointed at or something that has just been mentioned in the discourse. We won't have built up the machinery to axiomatize this intuition until Chapter 5.

Now consider the simplest case of a context-dependent discourse rich in indexicals. I see scrawled on a wall this graffitti:

> I am now writing this sentence here for you.

The problem I face in interpreting this text is to piece together some idea of the utterance situation S. The only thing I know about the time of S, and hence the referent of "now", is that it was sometime after the last time the wall was washed and before my arrival on the scene. The referent of "this sentence", assuming this sentence is all that is written on the wall, can only be the sentence itself. The place of S, and hence the referent of "here", I can identify quite precisely if I can identify "this sentence". About the speaker in S, and hence the referent of "I", all I can know is that he wrote the sentence there at that time. It is quite uncertain who the hearers in S are; it might have been written for someone specific, or it might have been written for anyone who happens to read it, in which case I would be one of the hearers in S.

If I'm walking through a museum of Americana and I see a poster showing Uncle Sam pointing directly at me and saying "I want you!" I don't take it personally and join the U.S. Army. I understand that the utterance situation is some other time, and that in any case, the intended hearers, the referents of "you", are much younger and healthier than I.

In reading any text, there is a discourse situation S in which the speaker ("I") is the author, or the persona the author wishes to adopt, or the narrator. The hearers ("you") are the readers or imagined readers. "Now" can mean anything from this point in the reading of the text to the present era, depending on the granularity.

Many AI dialogue systems deal with indexicals by the simple rule: "I" means you and "you" means me. This has been considered a hack. But in fact, since the utterance situation in which the system finds itself is one in which the user is the speaker and the system is the hearer, this "hack" is precisely the right, and even principled, thing to do.

In addition to a theory of discourse and utterance situations, a language user A must have a theory of himself and how he is embedded in the world. Consonant with our decision to represent all semantic content via predicates, let us suppose there is in A's knowledge base the predicate Ego, which is true only of A, and a very large number of axioms of the form

$$(\forall x) Ego(x) \supset p(x)$$

encoding A's knowledge about himself.[14]

Most of A's beliefs about his embedding in the world can be represented in the same way as any other beliefs. However, since A is moving through time, there must be some way of anchoring him to the current time at any instant. The predicate $Rexist$ does just this. A's awareness that John is asleep can be encoded as

$$Rexist(E) \wedge asleep'(E, J)$$

[14] One might possess the predicate *Ego*, i.e. a concept of self, innately or because that has turned out to be the most economical way to organize one's beliefs. Either possibility seems perfectly reasonable to me.

If the next time A checks, John is no longer asleep, A's theory of the world must be updated to eliminate $Rexist(E)$ and to add $Past(E)$, or perhaps more precise temporal information about E, but no other properties must be added or deleted about E. In particular, E is still a sleeping event by John.[15]

Suppose T is an instant in time. Then we can say it is now time T by saying $Rexist(T)$. Otherwise, either $Past(T)$ or $Future(T)$ is true. Instants in time can be treated like eventualities.

When the discourse situation S is recognized to be the current situation, and the speaker is recognized to be one's self (this usually happens), the appropriate identifications can be made. In particular, it will be known that

$$Speaker(A, S) \wedge Ego(A)$$

so that "I" will be an appropriate way for A to refer to himself. It will also be known that

$$time\text{-}of(T, S) \wedge Rexist(T)$$

so that an appropriate way to refer to the current time T will be with the word "now".

Two problems of philosophical interest can be explicated in this approach to indexicals. Kaplan (1977) and Perry (1979) discuss the following problem: I'm walking along in a store and in a mirror I see the lower half of a man's body and I see that his pants are on fire. I think, "His pants are on fire." Suddenly I realize that the person reflected in the mirror is me, and I think, "My pants are on fire." These two beliefs differ radically in the actions they are likely to lead me to, but "his" and "my" refer to the same person in the two sentences. How can we characterize the distinction between the two beliefs? This is a case, like the two problems of Section 2.6, in which confusion results from assuming the objects of belief are natural language sentences. Put another way, it is a result of a failure to distinguish between the claim a sentence makes and the information it conveys in order to make that claim. If we view beliefs as atomic predications, the difficulty evaporates. When I have the first belief, I believe

$$on\text{-}fire(P) \wedge wear(X, P)$$

[15] Of course, many other eventualities will begin or cease to exist when the status of E changes. Determining these changes is the frame problem and beyond the scope of this book.

When I have the second belief, I believe

$$on\text{-}fire(P) \wedge wear(X, P) \wedge Ego(X)$$

The distinction is precisely that in the second case I believe that the man whose pants are on fire is myself.

Finally an example due to Geoff Nunberg: In the movie "Year of Living Dangerously", the hero, a reporter in Indonesia is going around trying to learn about a possible shipment of arms from China to the Indonesian Communists, who are planning a revolution. If the Communists discover his investigation, his life could be in danger. At one point he asks a total stranger about the shipment, and the man chastises him for asking, saying, "I could have been a Communist." This sentence is interesting because the word "I" does not rigidly designate the speaker. The man is not saying that he himself could have been a Communist, say, if his economic circumstances had been a bit different, but that a different man in the same discourse situation could have been. The modal "could" outscopes the indexical "I". Since we have channelled the interpretation of "I" through the theory of utterance situations, rather than taking it as referring directly to the speaker in the current utterance situation, we are free to take the relevant utterance situation to be not the current one but an abstracted version of a situation in which the reporter asks *somebody*, whom he knows no better than the speaker, about the shipment.

The treatment of both of these problems illustrates a strength of the ontologically promiscuous approach to logical notation. Viewing all information encoded in natural language utterances as combinations of simple predications gives us a tool for analysis sufficiently fine-grained to ferret out and capture the distinctions that natural language texts depend upon.

2.4 Quantities and Properties

2.4.1 Motivation

The sentence

> In most democratic countries most politicians can fool most of the people on almost every issue most of the time.

has 120 readings. Moreover, they are distinct, in that for any two readings one can find a model under which one is true and the other isn't. Yet when people hear this sentence, they have the impression they understand it. They do not compute the 120 possible readings and then choose the best among them. Rather, they use world knowledge to constrain some of the dependencies among quantified expressions and leave other dependencies unresolved. For example, for me, the sets of politicians and the sets of people depend on the country, but I have no view on whether or not the politicians outscope the people. A representation is needed that allows this underspecification of meaning and in which learning more about scoping relations is captured simply as new conjoined properties.

The style of representation presented in this chapter enables just that. The principle stated in Section 2.1,

> All morphemes are created equal.

indicates that information about plurals and quantification should be encoded in the form of predications consisting of a predicate applied to one or more arguments, just as all other information conveyed by a sentence is represented. My aim here is to show this is as possible for quantifiers as it is for every other morpheme.

It is desirable to have sets or aggregates[16] among the first-class individuals in our Platonic universe, since natural language discourse talks about sets. Sentences with plural noun phrases under collective interpretations make predications about sets:

> The men lifted the piano.
> The men agreed.

Many quantifiers and adjectives express properties of sets or relations between sets—"most", "three", "numerous", "various".

Thus, we will allow individual constants and variables referring to sets in our notation. Sets will have no special ontological status. They will just be individuals. They will have other individuals as their elements, and the *element-of* relation is just a relation between individuals, no different in kind from other relations.

It will also be convenient to assume that a set has a "typical element", because most instances of plural noun phrases are in the service of a predication made about all the members of the set. In

[16] I won't distinguish here between the abstract notion of sets and the physical assemblages often called aggregates. For every set there is a corresponding aggregate, and for every aggregate under a description there is a corresponding set.

The men ran.

it is not that the set of men ran; sets can't run. Each individual member of the set of men ran. We will represent this by saying that the typical element of the set ran, and formulate an axiom that says that real elements inherit the properties of typical elements. In a sense, the typical element is a reified universally quantified variable (cf. McCarthy, 1977).

The logical form for a plural noun phrase will then make reference to both a set and its typical element.

Part of the linguistic motivation for this idea is that one can use singular pronouns and definite noun phrases as anaphors for plurals. Consider

> Each node that has a bit set in the MARKED field is not reclaimed. The bit indicates that it is accessible from other nodes.

There are many nodes and many bits, yet "the bit" and "it" are singular. We can view "it" as referring to the typical node and "the bit" as referring to the typical node's bit. Definite and indefinite generics can also be understood as referring to the typical element of a set.

In the spirit of ontological promiscuity, we simply assume that typical elements of sets are things that exist, and we encode in axioms the necessary relations between a set's typical element and its real elements.

In brief, the approach to quantifiers presented here consists of four elements:

1. Sets are individuals. Quantifiers are relations between sets.

2. Sets have typical elements. Ordinary elements inherit the properties of typical elements.

3. Functional dependencies are expressed as relations between typical elements.

4. Disambiguating scope is done by learning functional dependencies.

The first three elements of this approach are discussed in some detail in this chapter. The last one is explored in Chapter 6.

This simple and appealing picture runs into technical difficulties. The aim of the rest of Section 2.4 is to solve the difficulties in a way that allows us to return to the original simplicity. In a sense, we are frontloading the complexity. We will have to involve ourselves in substantial complications in order to return to where we started on a basis that is more solid formally.

As noted, the principal property that typical elements should have is that their properties should be inherited by the ordinary elements of the set. A first cut at expressing this property is the following axiom schema:

$$\begin{aligned} (2.16) \qquad & (\forall\, x, s)[typelt(x, s) \\ & \quad \supset [p(x) \equiv (\forall\, y)[y \in s \supset p(y)]]] \end{aligned}$$

That is, if x is the typical element of set s, then p is true of x if and only if p is true of every ordinary element y of s.

We would also like to have means for expressing what in ordinary set notation is expressed by the formula

$$(\exists\, s)\; s = \{x \mid p(x)\}$$

For this we introduce a predicate *dset*, where $dset(s, x, e)$ means that s is a set whose typical element is x and whose defining property is e, corresponding to the $p(x)$ in ordinary set notation. The key property we would like is that an entity is in the set if and only if it satisfies the property. Thus, the key axiom schema for *dset* should be something like

$$\begin{aligned} (2.17) \qquad & (\forall\, s)[[(\exists\, x, e) dset(s, x, e) \wedge p'(e, x)] \\ & \quad \equiv (\forall\, y)[y \in s \equiv p(y)]] \end{aligned}$$

That is, s is the defined set whose typical element is x and whose defining condition is the eventuality e of p being true of x if and only if for all y, y is in the set if and only if p is true of y.

There are three problems concerning typical elements that now must be dealt with.

1. Because of the Law of the Excluded Middle, it would seem that for any predicate p, either $p(x)$ or $\neg p(x)$ would be true of the typical element x. Then by (5), the elements of s could not differ on any properties. They all would inherit either p or $\neg p$ from x.

2. There is a question as to whether the typical element of a set is itself an element of the set. Both choices seem to lead to difficulties.

3. The third difficulty arises because of the flat notation we are using. This forces us into a first-order axiomatization of substitution.

The next section deals with the first two of these problems. The third is addressed in the following section.

2.4.2 The Nature of Typical Elements

There are three ways one might try to view typical elements:

1. The typical element of a set is one of the ordinary elements, but we will never know which one, so that anything we learn about it will be true of all.

2. The typical element is not an element of the set, and only special kinds of predicates are true of typical elements.

3. The typical element is not an element of the set, and ordinary predicates are true of them, except in set-theoretic axioms, which must be formulated carefully.

The first alternative is similar to the stance one takes toward instantiations of universally quantified variables in proofs. In proving $(\forall\, x \in s)p(x)$, one might consider an element a of s and show $p(a)$ while relying only on properties of a that are true for all elements of s. This alternative seems dangerous, however. The set consisting of John and George would have as its typical element either John or George, so by the desired properties (2.16) and (2.17), any property one has the other has too. The variable a in the proof is used only in a very limited context and in a very constrained way, whereas we want typical elements to exist in a persistent fashion in the Platonic universe and sometimes in the real world as well.

The second approach was taken in Hobbs (1983). The problem that arises when the typical element is assumed to be something other than an element of a set is that if the property p in Axiom Schema (2.16) is taken to be $\lambda x[x \notin s]$, then we can conclude that none of the members of the set are members of the set. This difficulty was handled in Hobbs (1983) by introducing a complex scheme of indexing predicates according to the kinds of arguments they would take. Essentially, for every predicate p, there was a basic level predicate p_0 that applied to ordinary individuals that are not typical elements, and a number of other predicates p_s that applied to the typical element of set s. More precisely, if x is the typical element of s, then

$p_s(x)$ was defined to be true if and only if $p(y)$ was true for every y in s, and otherwise p_s was equivalent to p_0.

Axiom schema (2.16) could then be stated

$$(\forall\, x, s)[typelt(x, s) \\ \supset [p_s(x) \equiv (\forall\, y)[y \in_0 s \supset p_0(y)]]]$$

This solves the first difficulty with typical elements. It is true that either $p_s(x)$ or $\neg(p_s)(x)$ holds, but this does not imply that all elements of s have all the same properties. That would hold only if either $p_s(x)$ or $(\neg p)_s(x)$ were true, but this is not what the Law of the Excluded Middle entails. The difference is the same as the difference between having negation outscope universal quantification and having universal quantification outscope negation.

The second difficulty with typical elements is solved as well. Suppose x is the typical element of s. We can simply stipulate that $x \not\in_0 s$, and since this is a basic level rather than an indexed predicate, no consequences follow for real elements. To determine whether $x \in_s s$ is true, by the indexed version of Axiom Schema (2.16), we have to ask whether

$$(\forall\, y)[y \in_0 s \supset y \in_0 s]$$

and this of course is trivially true. So $x \in_s s$ is true. In fact, it is equivalent to saying that x is the typical element of s.

This solution is inconvenient, however, because it forces us to carry around complex indices in many contexts where they are irrelevant to the content being expressed. For example, the axiom

$$(\forall\, x)[man(x) \supset person(x)]$$

is true regardless of whether x is an ordinary individual or a typical element of a set. If all the members of a set are men, they are all persons. We would not like to have to specify indices in such axioms, and most axioms are exactly of this nature.

The primary place where the indices must be attended to is in set theoretic axioms. If x is the typical element of s, then $x \not\in_0 s$ but $x \in_s s$. Thus, axioms that depend crucially on whether an entity is or is not in a set must be stated in terms of indexed predicates.

This leads to the third alternative, which we will adopt. We can avoid the complexity of indices by considering a bit how they are actually used in discourse processing. One must reintroduce the unindexed predicate p to use in the logical form of sentences, before interpretation, that is, before quantifier scope ambiguities are resolved. The relation between the indexed and unindexed predicates can be expressed, *inter alia*, by the following axiom schemas:

$$(\forall\, x)[p_0(x) \;\supset\; p(x)]$$
$$(\forall\, s, x)[p_s(x) \;\supset\; p(x)]$$

That is, the indexed predicates are specializations or strengthenings of the unindexed predicates, and in the course of discourse interpretation by abduction, one of the things that happens is that, as the existentially quantified variables are resolved to ordinary entities or to typical elements, the predicates that apply to them are specialized to the corresponding indexed predicate.

In this context of use, the indexing of the predicate is uniquely determined by the nature of its arguments. This would hold if constraints such as the following were stipulated:

$$\begin{aligned}&(\forall\, x, s)[p(x) \;\wedge\; typelt(x, s)\\ &\qquad\supset\; [p_s(x) \;\wedge\; \neg p_0(x) \;\wedge\; (\forall\, s_1)[s_1 \neq s \;\supset\; \neg p_{s_1}(x)]]]\end{aligned}$$

That is, if p is true of the typical element x of a set s, then the specialization p_s of p is true of x, and no other indexing of p is true of x.

A more thorough development of this idea depends on a treatment of functional dependencies, and is addressed in Section 2.4.7 below.

It is worth noting that the consistency of the formulation I have given of typical elements can be demonstrated by taking as a model one in which the denotation of the typical element of a set is the set itself. In this case, *typelt* is simply identity. However, I wish to admit as well interpretations in which the set and its typical element are distinct, since there are a number of contexts in which this distinction is a useful one to make, including representing the difference between collective and distributive readings.

For the remainder of this paper, only the unindexed predicates are used.

2.4.3 Substitution

Consider

> John believes men work.

The logical form of this sentence is (approximately)

(2.18) $\quad (\exists\, e_1, m, s, e_2) believe(J, e_1) \;\wedge\; work'(e_1, m)$
$\qquad\qquad \wedge\, dset(s, m, e_2) \;\wedge\; man'(e_2, m)$

That is, John believes the eventuality e_1 to obtain where e_1 is the eventuality of m working, where m is the typical element of a set s defined by the property e_2 of m being a man.

Suppose John believes George is a man and thus in the set s. We would like to conclude that John believes George works. But this does not follow from Axiom Schemas (2.16) and (2.17). The entity m is the typical element of s, John believes m works, and so John should believe that George works. The predication $p(x)$ in Axiom (2.16) would have to be "John believes m works". If p is restricted to be an atomic predicate, the axiom schemas won't do the job, because "John believes m works" is not represented by an atomic predicate. Suppose p can be an arbitrary lambda expression. Then given that m is the typical element of s, Axiom (2.16) implies that any property of m must also hold of G, specifically, for the property

$$\lambda m[believe(J, e_1) \;\wedge\; work'(e_1, m)]$$

Thus it would follow from (2.18) that

$$believe(J, e_1) \;\wedge\; work'(e_1, G)$$

But this is the wrong result. The problem is that e_1 is the eventuality of men working, not the distinct eventuality of George's working. If Sam is also a man, then this approach leads to e_1's also being the eventuality of Sam's working.

To get around this difficulty, we can introduce a predicate *Subst* that expresses substitution relations among expressions directly. In a way, it mimics in the flat notation what substitution does in conventional notations, and one may thus suspect it is just a formal trick. However, I think that substitution itself is one particular formalization of an intuitive, commonsense concept—that of "playing the same role". $Subst(a, e_1, b, e_2)$ can be read as saying that a plays the same role in e_1 that b plays in e_2. (*Subst* differs from "playing the same role" in one aspect noted below.)

In conventional notations, the first important property of substitution is the following:

$$p(t_1, \ldots, t_n)|_b^a = p(t_1|_b^a, \ldots, t_n|_b^a)$$

That is, the substitution of a predicate applied to a number of terms is the predicate applied to the substitution of the terms.

We can remain maximally noncommittal about the identity conditions among eventualities if we translate this schema into the following four axiom schemas, where p is now restricted to atomic predicates.

$$
\begin{array}{ll}
(2.19) & (\forall\, a, b, e_1, e_2, \ldots, u_i, \ldots)[Subst(a, e_1, b, e_2) \\
& \qquad \wedge\, p'(e_1, \ldots, u_i, \ldots) \\
& \quad \supset\ (\exists\, \ldots, v_i, \ldots)[p'(e_2, \ldots, v_i, \ldots) \\
& \qquad \wedge\ \ldots\ \wedge\ Subst(a, u_i, b, v_i)\ \wedge\ \ldots]]
\end{array}
$$

This says that if a plays the same role in e_1 that b plays in e_2, p is the predicate of e_1, and the arguments of e_1 are u_i, then e_2 also is an eventuality with predicate p and arguments v_i where a plays the same role in each u_i that b plays in the corresponding v_i. This allows us to proceed in substitution from predications to their arguments.

$$
\begin{array}{ll}
(2.20) & (\forall\, a, b, e_1, \ldots, u_i, v_i, \ldots)[\ldots\ \wedge\ Subst(a, u_i, b, v_i)\ \wedge\ \ldots \\
& \qquad \wedge\, p'(e_1, \ldots, u_i, \ldots) \\
& \quad \supset\ (\exists\, e_2)[p'(e_2, \ldots, v_i, \ldots)\ \wedge\ Subst(a, e_1, b, e_2)]]
\end{array}
$$

This says that if e_1 is an eventuality with predicate p and arguments u_i, where a plays the same role in each u_i that b plays in a corresponding v_i, then there is an eventuality e_2 whose predicate is p and whose arguments are v_i and a plays the same role in e_1 that b plays in e_2. This allows us to proceed from arguments to predications involving the arguments.

Two more axiom schemas are required because eventualities are not necessarily uniquely determined by their predicates and arguments. $p'(E_1, X)$ and $p'(E_2, X)$ can both be true without E_1 being identical to E_2. Axiom Schemas (2.19) and (2.20) guarantee a "substitution" eventuality of the right structure. The next two axiom schemas say that an eventuality is of the right structure if and only if it is a substitution eventuality.

$$
\begin{array}{ll}
(2.21) & (\forall\, a, b, e_1, e_2, \ldots, u_i, v_i, \ldots)[Subst(a, e_1, b, e_2) \\
& \qquad \wedge\, p'(e_1, \ldots, u_i, \ldots) \\
& \quad \supset\ [p'(e_2, \ldots, v_i, \ldots) \\
& \qquad \equiv\ \ldots\ \wedge\ Subst(a, u_i, b, v_i)\ \wedge\ \ldots]]
\end{array}
$$

This says that if a plays the same role in e_1 that b plays in e_2, p is the predicate of e_1, and the arguments of e_1 are u_i, then e_2 also is an eventuality with predicate p and arguments v_i if and only if a plays the same role in each u_i that b plays in the corresponding v_i.

$$(2.22) \quad (\forall\, a, b, e_1, e_2, \ldots, u_i, v_i, \ldots)[\ldots \wedge\ Subst(a, u_i, b, v_i) \wedge \ldots \\ \wedge\, p'(e_1, \ldots, u_i, \ldots) \\ \supset\ [p'(e_2, \ldots, v_i, \ldots) \equiv\ Subst(a, e_1, b, e_2)]]$$

This says that if e_1 is an eventuality with predicate p and arguments u_i, where a plays the same role in each u_i that b plays in a corresponding v_i, then the eventuality e_2 has predicate p and arguments v_i if and only if a plays the same role in e_1 that b plays in e_2.

The next two axioms enable substitution to bottom out.

$$(2.23) \quad (\forall\, a, b) Subst(a, a, b, b)$$

That is, a plays the same role in a that b plays in b.

$$(2.24) \quad (\forall\, a, b, c) \neg eventuality(c) \wedge\ c \neq a\ \supset\ Subst(a, c, b, c)$$

That is, if c is not an eventuality and not equal to a, then a plays the same role in c that b plays in c.

Notice that Axiom (2.24) allows c to be b. Substituting b for a in b results in b. This is the one asymmetry in the *Subst* predicate, and the reason that *Subst* is really more like substitution than like playing the same role. This asymmetry will allow us to draw from the fact that everyone in a set including John likes John the conclusion that John likes himself. That is, from $typelt(x, s)$, $p(x, y)$, and $y \in s$, we can conclude $p(y, y)$. The one constraint on *Subst* is that the first and fourth arguments cannot be the same. Substitution for the first argument would have eliminated such occurrences.

$$(\forall\, a, b, t_1, t_2)[a \neq b\ \wedge\ Subst(a, t_1, b, t_2)\ \supset\ a \neq t_2]$$

That is, substituting b for a will never result in a.

In addition to its role in expressing the properties of quantification, *Subst* turns out to be a useful concept in discourse interpretation wherever the similarity of two entities must be established.

2.4.4 Lambda Abstraction

There has been a strong tradition since the time of Montague (1971) of using lambda abstraction in natural language semantics. Verb phrases, for example, are often taken to denote a one-argument function, to be applied to the representation of the subject of the sentence to produce a proposition. Thus, the representation of the verb phrase

builds a boat

would be

$$\lambda x[(\exists\, y) build(x, y) \wedge boat(y)]$$

In the simplest case, if John is represented by the constant J, then this function applied to J yields

$$(\exists\, y) build(J, y) \wedge boat(y)$$

for the sentence "John builds a boat."

Having axiomatized substitution, we have developed something equivalent to, or slightly more powerful than, lambda abstraction. If e is the eventuality for which

$$and'(e, e_1, e_2) \wedge build'(e_1, x, y) \wedge boat'(e_2, y) \tag{2.25}$$

and e_0 is such that John plays the same role in e_0 that x plays in e—

$$Subst(J, e_0, x, e)$$

—then the following holds as well:

$$and'(e_0, e_3, e_2) \wedge build'(e_3, J, y) \wedge boat'(e_2, y) \tag{2.26}$$

If x is viewed as the lambda-abstracted variable, then we have in effect applied the function corresponding to (2.25) to the representation of John, resulting in the desired (2.26).

Nothing was said, however, about x's being a variable, and in fact it need not be. It does not have to be a typical element or a reified universally quantified variable. It can represent an ordinary individual. John plays the same role in John's working that Bill plays in Bill's working. If

$$work'(e_1, J) \wedge work'(e_2, B)$$

then

$$Subst(J, e_1, B, e_2)$$

The predicate *Subst* thus gives us a way of doing all the work of lambda abstraction, with somewhat more power, in a very intuitive way.

2.4.5 Sets and Plurals

We will have a predicate *set* that says of an individual that it is a set. Its elements are related to the set by the *element-of* relation. Only sets have elements:

$$(\forall\, x, s)[element\text{-}of(x, s) \supset set(s)]$$

In Section 2.4.6 we will have occasion to refer to subsets of sets. The predicate *subset* is defined in the usual way. Both its arguments are sets.

$$(\forall\, s_1, s_2) subset(s_2, s_1) \supset set(s_1)$$
$$(\forall\, s_1, s_2) subset(s_2, s_1) \supset set(s_2)$$
$$(\forall\, s_1, s_2)[subset(s_2, s_1) \equiv (\forall\, x)[element\text{-}of(x, s_2) \supset element\text{-}of(x, s_1)]]$$

In addition, sets have typical elements:

$$(\forall\, s)[set(s) \supset (\exists\, x) typelt(x, s)]$$

Only sets have typical elements:

$$(\forall\, x, s)[typelt(x, s) \supset set(s)]$$

Although we will not require this property, we will later be able to show that distinct sets must have distinct typical elements.

We will soon write a correct axiom stating that the real elements of a set inherit the properties of the typical element. That is, we will be able to infer a property from set membership. However, with the *typelt* relation alone, we will not be able to infer set membership from a property. That is, the fact that p is true of a typical element of a set s and p is true of an entity y, does not imply that y is an element of s. After all, we will want "three men" to refer to a set, and to be able to infer from y's being in the set the fact that y is a man. But we do not want to infer from y's being a man that y is in the set. The phrase may have occurred in a sentence like

Three men walked into a theatre.

This does not entail that no other men also walked into the theatre.

Nevertheless, as indicated in Section 2.4.1, we will need a notation for expressing this stronger relation among a set, a typical element, and a defining condition. In particular, we need it for representing "every man".

Let us develop the notation from the standard notation for intensionally defined sets,

$$(2.27) \qquad S = \{x \mid p(x)\},$$

by performing a fairly straightforward, though ontologically promiscuous, syntactic translation on it. First, instead of viewing x as a universally quantified variable, let us treat it as the typical element X of S. Next, as a way of getting a handle on $p(x)$, we will use the nominalization operator $'$ to reify it, and refer to the condition E of p's being true of the typical element X of S—$p'(E, X)$. Expression (2.27) can then be translated into the following flat predicate-argument form:

$$dset(S, X, E) \wedge p'(E, X)$$

This should be read as saying that S is a set whose typical element is X and which is defined by condition E, which is the condition of p's being true of X.

The first argument of *dset* is a set, and the third argument is an eventuality:

$$dset(s, x, e) \supset set(s) \wedge eventuality(e)$$

Its second argument is the typical element of the set. That is, the relation between the predicates *dset* and *typelt* is expressed by the following axiom:

$$(2.28) \qquad (\forall\, s, x, e)[dset(s, x, e) \supset typelt(x, s)]$$

If s is the defined set whose typical element is x and whose defining condition is the eventuality e, then x is the typical element of s. The predicate *dset* is thus a specialization of the predicate *typelt*, a fact that will play an important role below in the treatment of monotone decreasing quantifiers.

There should probably not be a rule of the form

$$(\forall\, x, s)[typelt(x, s) \supset (\exists\, e) dset(s, x, e)]$$

since this would entail that every set is definable by some eventuality. This strikes me as an undesirable property. Some linguistically described sets, such as the set referred to by "all men", have natural defining properties. Others, such as the set referred to by "many men", do not, but they are sets nevertheless. In any case, we will not need this property.

Recall that the principal property of typical elements is that their real elements inherit the properties of their typical elements (Axiom Schema (2.16)), and the principal additional property of defined sets is that everything that has it defining property is in the set (Axiom Schema (2.17)). Having axiomatized substitution with the predicate $Subst$, we can now recast Axiom Schemas (2.16) and (2.17) as Axioms (2.29) and (2.30), respectively.

$$
\begin{aligned}
(2.29)\quad & (\forall\, x, s, e)[typelt(x, s) \\
& \quad \supset [(\exists\, e_1)[Subst(x, e, x, e_1) \wedge Rexists(e_1)] \\
& \quad\quad \equiv (\forall\, y)[y \in s \supset (\exists\, e_2)[Subst(x, e, y, e_2) \\
& \quad\quad\quad \wedge Rexists(e_2)]]]]
\end{aligned}
$$

This property is now expressed as an axiom rather than an axiom schema. The explicit specification of the structure $p'(e, x)$ has been eliminated here. Instead, the eventuality e represents that pattern and the predicate $Subst$ is used to stipulate that other eventualities exhibit the same pattern. This axiom says that if e is such a pattern and x is the typical element of s, then there is a really existing eventuality e_1 involving x exhibiting that pattern if and only if for every ordinary element of s, there is a corresponding eventuality e_2 exhibiting the same pattern that really exists.

Suppose, in (2.29), that x is the typical element of s. If e is not an eventuality, then it is either x or something else. If it is x, then $e = e_1 = x$ and $e_2 = y$, so the axiom is valid. If it is something else, then $e = e_1 = e_2$, and the axiom is valid. Suppose e is an eventuality and $p'(e, x)$ holds. Then $p(x)$ is equivalent to $(\exists\, e_1)p'(e_1, x) \wedge Rexists(e_1)$, which is equivalent to $(\exists\, e_1)Subst(x, e, x, e_1) \wedge Rexists(e_1)$. Similarly, $p(y)$ is equivalent to $(\exists\, e_2)Subst(x, e, y, e_2) \wedge Rexists(e_2)$. Thus, Axiom (2.29) captures the intent of Axiom Schema (2.16).

Replacing Axiom Schema (2.17) is Axiom (2.30):

$$(2.30) \quad \begin{aligned}&(\forall\, s,x,e)[eventuality(e)\\ &\quad \supset [(\exists\, e_1)[dset(s,x,e_1) \,\wedge\, Subst(x,e,x,e_1)]\\ &\qquad \equiv (\forall\, y)[y \in s \,\equiv\, (\exists\, e_2)[Subst(x,e,y,e_2)\\ &\qquad\quad \wedge\, Rexists(e_2)]]]]\end{aligned}$$

That is, if e is an eventuality (representing a pattern expressed in terms of the typical element x of a set s), then there is an eventuality e_1 of the same pattern that is the defining eventuality for s if and only if for every ordinary element y of s there is a corresponding eventuality e_2 of the same pattern that really exists. Here it is necessary to express the constraint that e be an eventuality, because the third argument of $dset$ must be an eventuality.

Let us return to example (2.18). If $dset(s,m,e_2)$ and $man'(e_2,m)$ hold and George is a man, then we have

$$\begin{aligned} man(G) &\equiv man'(e_3,G) \,\wedge\, Rexists(e_3) && \text{by (A1)}\\ &\equiv Subst(m,e_2,G,e_3) \,\wedge\, Rexists(e_3) && \text{by (2.22)}\\ &\equiv G \in s && \text{(by (2.30))}\end{aligned}$$

Suppose $Rexists(e_0)$, $believe'(e_0,J,e_1)$, and $work'(e_1,m)$ all hold. Since $typelt(m,s)$ holds, and letting e and e_1 in (2.29) both be e_0, there is, by (2.29), an e_4 such that

$$Subst(m,e_0,G,e_4) \,\wedge\, Rexists(e_4)$$

By (2.19) there is an e_5 such that

$$believe'(e_4,J,e_5) \,\wedge\, Subst(m,e_1,G,e_4) \,\wedge\, Rexists(e_4)$$

By (2.22),

$$believe'(e_4,J,e_5) \,\wedge\, work'(e_5,G) \,\wedge\, Rexists(e_4)$$

By (A1),

$$believe(J,e_5) \,\wedge\, work'(e_5,G)$$

That is, John believes George works. (I ignore here the problem of what inferences it is legitimate to draw inside belief contexts. Think of this expression as saying that, merely by virtue of the fact that George is a man, John believes George, whoever he may be, works.)

Next we want an axiom that will guarantee a defined set for any legitimate definition. That is, for any eventuality and any one of its arguments (or arguments of arguments ...), there is a set with a corresponding eventuality as its definition.

$$(2.31) \quad (\forall\, y, e_2)[eventuality(e_2) \wedge arg+(y, e_2) \\ \supset (\exists\, s, x, e_1)[Subst(x, e_1, y, e_2) \wedge dset(s, x, e_1)]]$$

That is, given an eventuality e_2 and an entity y that is involved in e_2 at some level of embedding, there is an eventuality e_1 with the same pattern and a defined set s whose typical element is x and whose definition is e_1. Or put more simply, for any description there is a set of entities, in the Platonic universe, satisfying that description.

We have said nothing yet about whether defined sets really exist or not. We can say that a set really exists if and only if something is a member of the set exactly when it satisfies the description.

$$(2.32) \quad (\forall\, s, x, e_1)[dset(s, x, e_1) \\ \supset [Rexists(s) \\ \equiv (\forall\, y)[member(y, s) \\ \equiv (\exists\, e_2)[Subst(y, e_2, x, e_1) \wedge Rexists(e_2)]]]]$$

This is the place we avoid Russell's paradox. If the property is something that nothing satisfies, we have the null set and it exists. However, if we get a contradiction from the description of the set and the membership property, as we do with the property $\lambda x[x \notin x]$, the set does not exist in the real world.

2.4.6 Quantifiers as Descriptions of Sets

A determiner like "most" can be viewed as expressing a relation between sets. The expression $most(s_2, s_1)$ says that set s_2 is a subset of s_1 consisting of more than half the elements of s_1. Then the sentence

Most men work.

can be represented as follows:

$$(2.33) \qquad (\exists\, s)[most(s_2, \{x \mid man(x)\}) \wedge (\forall\, y)[y \in s_2 \supset work(y)]]$$

That is, there is a set s_2 that is most of the set of all men (i.e., it is a subset with more than half the elements), and for every entity y in s_2, y works.

We can use the predicates *typelt* and *dset* to unwind this into a flat notation. The set $\{x \mid man(x)\}$ is a set s_1 such that $dset(s, x, e)$ where e is the eventuality of x's being a man, $man'(e, x)$. The set s_2 is a subset of s_1 that consists of most of s_1. We can represent this by treating *most* as a relation between the two sets—$most(s_2, s_1)$. The set s_2 has a typical element y—$typelt(y, s_2)$. To say that all elements of s_2 work, we say that y works.

With this machinery, we can now rewrite logical form (2.33) as follows:

$$\begin{array}{ll}(2.34) & (\exists\, s_2, s_1, x, e, x, y)[most(s_2, s_1) \wedge dset(s_1, x, e) \\ & \quad \wedge\, man'(e, x) \wedge typelt(y, s_2) \wedge work(y)]\end{array}$$

That is, there is a set s_1 defined by the property e of its typical element x being a man, there is a set s_2 which is most of s_1 and has y as its typical element, and y works. It is straightforward to show that (2.34) is equivalent to (2.33).

It is easy to see how a logical form like (2.34) could be generated compositionally in a strictly local fashion. The common noun "men" introduces a set, its typical element, and its defining property, generating the conjuncts $dset(s_1, x, e) \wedge man'(e, x)$. The determiner "most" introduces another set and its typical element, along with the conjuncts $most(s_2, s_1) \wedge typelt(y, s_2)$. The latter typical element becomes the logical subject of the predication of the main verb, which generates the conjunct $work(y)$. This is explicated in detail in Chapter 4.

Viewing determiners as predicates on sets allows us to express as axioms more refined properties of the determiners than can be captured by representing them in terms of the standard quantifiers. This point will be illustrated for a number of quantifiers in the remainder of this section, although a proper analysis of what they mean is deferred until Chapter 5.

First let us note that, with the proper definitions of "every" and "some",

$$(\forall\, s_1, s_2) every(s_2, s_1) \equiv s_1 = s_2$$

$$(\forall\, x, s_1) some(x, s_1) \equiv element(x, s_1)$$

the formulas corresponding to expression (2.33) reduce to the standard notation. (This can be seen, by the way, as explaining why the restriction is implicative in universal quantification and conjunctive in existential quantification.)

Suppose we adopt a common view that the phrase "every N" implies the existence of at least one N. We can encode this in the following axiom:

$$(\forall\, s_1, s_2) every(s_2, s_1) \;\supset\; |s_2| > 0$$

That is, if set s_2 is every element of set s_1, then the cardinality of s_2 is greater than zero. (I don't take a position on whether or not this axiom is appropriate.)

ANY –¿ RANDOM ???

An axiom that characterizes *most* in perhaps too mathematical a manner is

$$(\forall\, s_1, s_2) most(s_2, s_1) \;\supset\; |s_2| > 1/2\, |s_1|$$

That is, if s_2 is most of s_1, then the cardinality of s_2 is more than half the cardinality of s_1.

Somewhat more complicated is an axiom encoding the meaning of the determiner "the most", as in "Of the five candidates, John got the most votes."

$$\begin{aligned}&(\forall\, s_1, s_2) \textit{the-most}(s_2, s_1) \;\supset\; (\exists\, u) partition(u, s_1)\\&\quad \wedge s_2 \in u \,\wedge\, (\forall\, s_i)(s_i \in u \,\wedge\, s_2 \neq s_i \;\supset\; |s_1| > |s_i|)\end{aligned}$$

That is, s_1 is a element of a partition of s_1 and is larger than any other element of that partition. (The partition is of course contextually determined.)

Determiners like "many", which cannot be decomposed into more mathematically tractable relations between sets, can nevertheless enter in complex ways into axiomatizations of various domains, if it is represented as a relation between sets, and thereby receive the appropriate interpretations in context in the course of inferencing.

Some determiners, such as "the", "this", "a" and "any", are relations between entities and eventualities. For example, "the" expresses a relation between the entity referred to by the whole noun phrase and the property expressed by the remainder of the noun phrase, and says about that entity that it is mutually identifiable in context by means of that property. These determiners are discussed further in Section 4.?? on determiners and Section 5.?? on mutual identifiability.

In the noun phrase "numerous tall men" the property *numerous* applies to the set of men. The property *tall* can only apply to an individual man, not to a set, and thus we represent it as taking the typical element of the set of men as its argument.

$$typelt(X,S) \wedge numerous(S) \wedge tall(X) \wedge man(X)$$

In the sentence

The men agree.

the collective predicate *agree* takes the set as its argument. In

The men run.

the predicate *run* is applied to the typical element.

For collective predicates such as *agree* and *meet*, the main predication of the clause would apply to the set rather than to each of its elements.

$$typelt(X,S) \wedge agree(S)\ typelt(X,S) \wedge run(X)$$

We can similarly represent the distinction between the distributive and collective readings of a sentence like

The men lifted the piano.

For the collective reading the representation would include $lift(S,P)$ where S is the set of men. For the distributive reading, the representation would have $lift(M,P)$, where M is the typical element of the set S.

During semantic composition, a generic noun phrase would be taken to refer to some entity about which some predication is made. If the sentence is properly interpreted, then during interpretation, it would be inferred or assumed that the entity is in fact the typical element of some set.

A difficulty is presented by monotone decreasing quantifiers, such as "few" and "no" (cf. Barwise and Cooper, 1981). A monotone increasing quantifier, like "most", is "monotone increasing" beacuse when the predicate in the body of the quantified expression is made less restrictive, the truth value is preserved. Thus,

Most men work hard.

entails

(2.35) Most men work,

By contrast, for monotone decreasing quantifiers, when the predicate in the body of the quantified expression is made less restrictive, the truth value is not necessarily preserved. Quite the opposite. It is preserved when the body is made *more* restrictive.

(2.36) Few men work.

entails

Few men work hard.

Since "x works hard" entails "x works", a flat, scope-free representation for "few men work hard" runs into problems, because it would seem to allow the incorrect inference "few men work".

In Hobbs (1983) I suggested very briefly a logical form for such sentences in which the quantifier "few" is translated into a predicate that means "all but a few" and the predication of the body of the quantified expression is negated. Thus, sentence (2.36) would be interpreted as if it were

All but a few men don't work.

This solves the entailment problem. "x doesn't work" entails "x doesn't work hard." Thus, "Few men work" would be equivalent to "all but a few men don't work", which entails "all but a few men don't work hard," which would be equivalent to "few men work hard." This approach is similar to that of van Eijck (1983).

However, this is not a felicitous solution, since the negation of the main verb makes the compositional semantics of the quantifier nonlocal, in that information from the noun phrase other than its referent is required in the interpretation of the rest of the sentence.

In Hobbs (1995) and in this book I propose a different analysis of monotone decreasing quantifiers, one in which the right interpretation arises from a combination of a single rule for interpreting quantifiers, both monotone increasing and monotone decreasing, together with the pragmatic process of specializing or strengthening interpretations that is the basis of the abduction approach, and a reinterpretation via coercion of what is asserted by the sentence. The result is a picture wherein syntactic analysis and semantic translation yields a representation that makes fewer distinctions than we might wish, but is strictly locally compositional, and strengthening and coercing to the desired representation is done by pragmatic processes that are independently motivated.

Specifically, the semantic interpretation produced by syntactic analysis and compositional analysis, as described in Section 4.??, for "Few men work" is

$$few'(e, s_2, s_1) \wedge dset(s_1, x, e_1) \wedge man'(e_1, x) \wedge typelt(y, s_2) \wedge \\ work'(e_2, y) \wedge Rexists(e_2)$$

The eventuality e is the relation of "few-ness" between the sets s_1 and s_2. The man' predication derives from the morpheme "man", the $dset$ predication derives from the pluralization of "man", the few and $typelt$ predications derive from the word "few", the $work'$ predication derives from the word "work", and the $Rexists$ predicate derives from the fact that the working is the top-level assertion of the sentence. The treatment of "most" is exactly parallel.

For "few", in the course of interpretation by abduction, as described in Chapter 3, the $typelt$ predication is strengthened a $dset$ predication, saying that s_2 is the set of men who work.

$$few'(e, s_2, s_1) \wedge dset(s_1, x, e_1) \wedge man'(e_1, x) \wedge dset(s_2, y, e_3) \wedge \\ and'(e_3, e_4, e_2) \wedge Subst(y, e_4, x, e_2) \wedge work'(e_2, y) \wedge Rexists(e_2)$$

The eventuality e_4 is the eventuality of y's being a man, and e_3 is the conjunction of that eventuality with the eventuality of y's working.

Finally, there is a coercion of what is asserted by the sentence, that is, of the argument of $Rexists$, from e_2 to e, using the coercion relation

$$work'(e_2, y) \wedge typelt(y, s_2) \wedge few'(e, s_2, s_1)$$

as described in Section 4.??. This yields the expression

$$few'(e, s_2, s_1) \wedge dset(s_1, x, e_1) \wedge man'(e_1, x) \wedge dset(s_2, y, e_3) \wedge \\ and'(e_3, e_4, e_2) \wedge Subst(y, e_4, x, e_2) \wedge work'(e_2, y) \wedge Rexists(e)$$

This now is a representation of the content of the sentence "The men who work are few," which is the desired interpretation.

This entails the content of the sentence, "The men who work hard are few," since the men who work hard constitute a subset of the men who work. So the right inference results in the case of monotone decreasing quantifiers.

Like other quantifiers, *most* and *few* can be viewed as expressing relations (e.g., comparing cardinalities) between two sets, which can be expressed in axioms. For example, one property of "few" and "most" is that they pick out subsets:

$$(2.37)\qquad (\forall\, s_1, s_2)[most(s_2, s_1)\ \supset\ subset(s_2, s_1)]$$

$$(2.38)\qquad (\forall\, s_1, s_2)[few(s_2, s_1)\ \supset\ subset(s_2, s_1)]$$

The monotone increasing and monotone decreasing properties can also be expressed as axioms:

$$(2.39)\qquad (\forall\, s_1, s_2) most(s_2, s_1)\ \wedge\ subset(s_2, s)\ \wedge\ subset(s, s_1) \\ \supset\ most(s, s_1)$$

$$(2.40)\qquad (\forall\, s_1, s_2) few(s_2, s_1)\ \wedge\ subset(s, s_2)\ \wedge\ \neg null(s) \\ \supset\ few(s, s_1)$$

That is, if s_2 is most of s_1 and s_2 is a subset of s which in turn is a subset of s_1, then s is also most of s_1. This is the monotone increasing property. If s_2 constitutes few members of s_1, then so does a non-null subset s of s_2. This is the monotone decreasing property.

Some determiners, such as numbers, are neither monotone increasing or monotone decreasing. The phrase "at least three" is monotone increasing and "at most three" is monotone decreasing, but "exactly three" is neither. Barwise and Cooper (19??) suggested that such determiners can be viewed as a conjunction of a monotone increasing and a monotone decreasing quantifier. Thus, "exactly three" means "at most three and at least three".

For the determiner "exactly three" we should not object to the extra conjuncts since the sentence has an extra word. Nevertheless, a one-word determiner like "three" is a simple concept in English, and it seems unaesthetic to represent it as a conjunction of two determiners. An alternative is to interpret "three" as "at least three". This certainly seems to be what it conveys in sentences such as

> One evening three men walked into a movie theatre.
>
> Suppose we have three lines a, b, and c parallel to line d.

We are not saying that *only* three men walked into a movie theatre, and if we are talking about Euclidean geometry, we had better not mean there are *only* three lines parallel to line d. In those cases where it does convey "three and only three", as in

Q: How many people came to your party last night.
A: Three people came to my party.

we can see it as a result of a Gricean implicature, or pragmatic strengthening.

Specifically, the sentence "Three men arrive," has the logical form, before pragmatic strengthening,

$$three'(e_1, s) \wedge man'(e_2, x) \wedge typelt(x, s) \wedge arrive'(e_3, x) \wedge Rexists(e_3)$$

This does not commit us to "at most three men arrive". We strengthen this by strengthening the *typelt* predication to a *dset* relation, and coercing the assertion from the arriving to the "three-ness", just as we did in interpreting monotone decreasing quantifiers. This gives us the logical form

$$three'(e_1, s) \;\wedge\; man'(e_2, x) \;\wedge\; dset(s, x, e_4) \;\wedge\; and'(e_4, e_2, e_3) \;\wedge\; arrive'(e_3, x) \;\wedge\; Rexists(e_1)$$

The meaning of this is "The men who arrive are three," which means exactly three.

2.4.7 Functional Dependencies

There have been a number of proposals in recent years for underspecified or scope-neutral representations of natural language sentences involving quantification. Most of these proposals involve special purpose formalisms. However, further specification of the meaning of the sentences in context requires inference, so the first requirement on these formalisms is that there be an inference procedure that supports the full range of commonsense reasoning. This requirement is generally not met.

By contrast, I proposed an underspecified representation (Hobbs, 1983, henceforth ITQ), in which natural language sentences involving quantification are translated into "underspecified" expressions in first-order predicate calculus. All that is required to do this is a moderate liberalization in the ontology one is willing to accept. The advantage of this move is that the representation comes with a theory of inference that is already worked out, whether standard deduction, weighted abduction (Hobbs et al., 1993 and this book), or any of the various nonmonotonic logics for commonsense reasoning.

ITQ fell short of an adequate treatment in that the treatment of the functional dependencies that represented scope information was insufficiently precise. Here we address that problem—the representation of functional dependencies among variables.

Briefly, the solution is this: Universally quantified variables are reified as typical elements of sets. Universally quantified statements about members of a set become statements without universal quantification about the typical element of the set. Axiom (2.29) and (2.30) enable us to conclude properties of individual members from the same property of the typical element. When in ordinary notation, a universal quantifier outscopes an existential quantifier, the dependence between their variables is expressed as a functional dependency between two typical elements. When scope information is absent, e.g., when only compositional semantics has applied, then there are no predications of functional dependency. As scope information is acquired during abductive interpretation, it is represented as predications of functional dependency.

We thus have a representation that is scope-neutral initially and during the course of pragmatic processing becomes more specific merely through the conjunction of further properties.

Consider

$$(\exists\, s)(\forall\, u)[u \in s \;\supset\; (\exists\, v)p(u,v)]$$

If we skolemize v, we have

$$(\exists\, f,s)(\forall\, u)[u \in s \;\supset\; p(u,f(u))]$$

Replacing u by the typical element x of s gives us

$$(2.41) \qquad (\exists\, f,x,s)[typelt(x,s) \;\wedge\; p(x,f(x))]$$

We define FD_0 by

$$(\forall\, x,y)[FD_0(y,x,f) \;\equiv\; y = f(x)]$$

That is, y is functionally dependent on x via function f if y equals $f(x)$.

Then (2.41) becomes

$$(2.42) \qquad (\exists\, x,s,y,f)typelt(x,s) \;\wedge\; p(x,y) \;\wedge\; FD_0(y,x,f)$$

That is, p is true of x and y where x is the typical element of the set s and y is functionally dependent upon x.

A sentence with a quantifier scope ambiguity is then represented along the lines of

$$(\exists\, x, s, y) typelt(x, s) \wedge p(x, y)$$

This is scope-neutral. When scoping information is discovered by inference during pragmatic processing, the predication $FD(y, x, f)$ is conjoined to produce the disambiguated form (2.42).

The predication $FD_0(y, x, f)$ carries information about what the precise relation between y and x is, namely, f. Normally, we will not learn this. We will only learn that *some* functional dependency exists, and that will be enough to constrain the scoping possibilities in the ordinary representation of quantification. Essentially, what we learn is an FD relation, where FD is defined as

$$FD(y, x) \equiv (\exists\, f) FD_0(y, x, f)$$

In many cases the functional dependencies are never discovered and do not matter. For me, in the sentence we began the discussion of quantifiers with,

> In most democratic countries (c), most politicians (p) can fool most of the people (x) on almost every issue (i) most of the time (t).

the sets of politicians, people, and issues are dependent on the countries, but I have no opinion on what other dependencies there are. I would thus infer $FD(p, c)$, $FD(x, c)$, and $FD(i, c)$, but no other FD predications. (And the 120 readings are reduced to only 30 readings.)

TO BE DONE: Axiomatize the properties of the FD relation. Technical difficulties. Several examples showing ways in which functional dependencies can be derived during pragmatic processing.

The logical form for

> Most men like several women.

is

$$\begin{aligned}
&(\exists\, s_2, s_1, x, e, x, y, z, s_3)[most(s_2, s_1) \wedge dset(s_1, x, e) \\
&\quad \wedge man'(e, x) \wedge typelt(y, s_2) \wedge like(y, z) \wedge several(s_3) \\
&\quad \wedge typelt(z, s_3) \wedge woman(z)]
\end{aligned}$$

That is, there is a set s_1 defined by the property e of its typical element x being a man, there is a set s_2 which is most of s_1 and has y as its typical element, and y likes z, where z is the typical element of a set s_3, z is a woman, and s_3 has several members.

This is the scope-neutral representation. In the course of further processing, we may discover that s_3 is an *actual* set of several women, corresponding to wide scope for "several", or we may discover that s_3 is functionally dependent upon s_2, in which case s_3 is the typical element of a set of sets of women, one for each man in s_2, corresponding to the narrow scope.

This treatment of functional dependencies is similar to the ordering constraints of Allen (1987) and Poesio (1991).

2.4.8 Abstractions and Their Instances

Frequently predicates are used not to make predications but to name a concept which then functions in the sentence as an individual. The problem this poses is relating the two uses in the right way. For a concrete example, consider the following three sentences:

(2.43) The boat is red.
(2.44) Red is the color of the boat.
(2.45) Red is my favorite color.

Intuitively, "red" refers to the same entity in (2.44) and (2.45), and sentences (2.43) and (2.44) mean the same thing. We would like a representation of abstractions like the concept "red" in which "red" is represented in the same way in (2.44) and (2.45) and for which (2.43) and (2.44) are, with the appropriate axioms, logically equivalent.

We approach this problem in two steps, first appealling to set theory for conceptual guidance. Let us, for the moment, say that the concept "red" is in fact the set of all red things, $\{x \mid red(x)\}$, and that the concept "color" is the set of all such sets, $\{\{x \mid red(x)\}, \{x \mid blue(x)\}, \ldots\}$. Then the representation of (2.43) is $red(B)$ and of (2.44) is $color\text{-}of(\{x \mid red(x)\}, B)$. These are equivalent, given the following axiom:

(2.46) $(\forall\, x, y) color\text{-}of(x, y) \equiv color(x) \wedge y \in x$

That is, x is the color of y if and only if x is a color (i.e., in the set of colors) and y is in the set x.

Let us introduce a new predicate *Red* (with the first letter capitalized) which is true only of the set of all red things, i.e. of the color red. Thus the English word "red" is ambiguous between its noun sense, as in

> My favorite color is red.

where *Red* is predicated of some color X, and its adjective sense, as in

> My favorite boat is red.

where *red* is predicated of a physical object.

Switching from standard set-theoretic notation to the notation we have developed for defined sets, we now have the following axiom relating *Red* and *red*:

$$(2.47) \quad (\forall\, r)Red(r) \equiv (\exists\, x, e)[dset(r, x, e) \wedge red'(e, x)]$$

$$(2.48) \quad (\forall\, r)Red(r) \supset color(r)$$

That is, the color red is a set r whose typical element is x and whose defining property e is x's redness, and if r is the color red, then r is a color.

Now the second step: We would like to have abstractions like "Red" with all these properties, without committing ourselves to the somewhat unintuitive notion that a concept is a set. We can do this by inventing a new predicate *concept*, which is exactly parallel to the predicate *dset*. Corresponding to any set defined by a property e there is a concept defined by the same property. Something is an element of the set if and only if it is an instance of the concept.

$$(2.49) \quad (\forall\, x, e)[(\exists\, s)dset(s, x, e) \equiv (\exists\, c)concept(c, x, e)]$$

$$(2.50) \quad (\forall\, s, c, x, d)[dset(s, x, e) \wedge concept(c, x, e) \\ \supset (\forall\, z)[element(z, s) \equiv instance(z, c)]]$$

Properties of concepts and their instances can be proved by translating them into properties of defined sets and their members.

Note that no concept equalities follow from set equalities. The set of unicorns and the set of centaurs may both be empty and thus equal, but that does not imply that the concept of being a unicorn and the concept of being a centaur are the same.

Ontologically, a concept is just an individual in the Platonic universe, like any other.

Specific families of concepts, like color, can be axiomatized in terms of the predicate *concept*. We illustrate this with the concept "Red". First we would like to be assured that the concept "Red" is related to the predicate *red* in the proper way, with our new version of axiom (2.47):

(2.51) $(\forall\, r) Red(r) \equiv (\exists\, x, e)[concept(r, x, e) \wedge red'(e, x)]$

If we wish the concept "Red" to be unique, we could do so with an axiom of the standard sort. The existence of the concept "Red" can be assured with the axiom

(2.52) $(\exists\, r) Red(r)$

The rule that red is a color is

(2.48) $(\forall\, r) Red(r) \supset color(r)$

That is, if r is the color red, then r is a color.

Finally we rewrite axiom (2.46) about colors:

(2.53) $(\forall\, x, y) color\text{-}of(x, y) \equiv color(x) \wedge instance(y, x)$

Now the representations of sentences (2.43) - (2.45) become, respectively,

(2.43') $red(B)$
(2.44') $color\text{-}of(R, B) \wedge Red(R)$
(2.45') $color(R) \wedge favorite(R) \wedge Red(R)$

The proof of the equivalence of (2.43') and (2.44') is as follows: From Axiom Schema (A2) we know

(2.54) $\quad (\exists\, e_2, x) red'(e_3, x_1)$

and from this and Axiom (2.31) we know

(2.55) $\quad (\exists\, s, x, e_2) dset(s, x, e_2) \wedge red'(e_2, x)$

Replace the existentially quantified variables s, x, and e_2 by the constants S, X, and E_2, respectively. From this and Axiom (2.49) we know

(2.56) $\quad (\exists\, r)[concept(R, X, E_2) \wedge red'(E_2, X)$

Call this r "R". From this and Axiom (2.51) we know

(2.57) $\quad Red(R)$

From this and Axiom (2.48) we know

(2.58) $\quad color(R)$

Now the following expressions are equivalent:

(2.59)	$red(B)$	(2.43')
(2.60)	iff $red'(e, B) \wedge Rexists(e_1)$	from (A1)
(2.61)	iff $member(B, S)$	from (2.55), and (2.30)
(2.62)	iff $instance(B, R)$	from (2.55), (2.56), and (2.50)
(2.63)	iff $color\text{-}of(R, B)$	from (2.58) and (2.53)
(2.64)	$color\text{-}of(R, B) \wedge Red(R)$	(2.44')

2.5 Opaque Contexts

2.5.1 Logical Operators, Especially Negation

Consider

> Because Mary and her husband both work, they can afford expensive vacations.

It is not just Mary's working or just her husband's working that enables them to afford expensive vacations. It is that *both* conditions are true. Thus, we must be able to refer to a *conjunction* of eventualities, as well as the eventualities themselves. All of the other tests for whether something should be reified also hold for conjunctions; they can be modified adverbially, be specified as to time and place, be referred to pronominally, and be the object of a propositional attitude.

(2.65) Last year, Mary and her husband both worked.
John and Mary both work, and that surprises George.
Rudely, Mary and her husband both declined the invitation. One of them should have gone.

What is true of conjunction is true of the other logical operators as well.

(2.66) Mary doesn't work, and that surprised George.
(2.67) George believes that either Mary or her husband works.
(2.68) If Mary doesn't work, then her husband does, because their house is expensive.

Just as we need "handles" for events and conditions, we need them also for the results of applying logical operators to events and conditions.

We therefore introduce the predicates *and*, *or*, *not* and *imply*, and their corresponding primed predicates. The following translations of sentences (2.65) - (2.68) illustrate their use:

(2.65′) $Rexist(A) \wedge and'(A, WM, WH) \wedge work'(WM, M)$
$\wedge\, work'(WH, H) \wedge surprise(A, G)$

(2.66′) $Rexist(EN) \wedge not'(EN, WM) \wedge work'(WM, M)$
$\wedge\, surprise(EN, G)$

(2.67′) $believe(G, EO) \wedge or'(EO, WM, WH) \wedge work'(WM, M)$
$\wedge\, work'(WH, H)$

(2.68′) $cause(EE, EI) \wedge imply'(EI, EN, EWH) \wedge not'(EN, EWM)$
$\wedge\, work'(EWM, M) \wedge work'(EWH, H)$
$\wedge\, expensive'(EE, HOU)$

The predicates *and*, *or*, *imply* and *not* can be related to the logical connectives by means of axioms. First *and*:

$$(2.69) \qquad (\forall\, e_1, e_2) and(e_1, e_2) \equiv Rexist(e_1) \wedge Rexist(e_2)$$

Using this axiom and axioms A1, we can infer $p(X) \wedge q(Y)$ from $and(E_1, E_2)$ $\wedge\, p'(E_1, X) \wedge q'(E_2, Y)$. Conversely, from $p(X) \wedge q(Y)$ we can infer $and(E_1, E_2)$ $\wedge\, p'(E_1, X) \wedge q'(E_2, Y)$ for some E_1 and E_2.

In negation we encounter, for the first time in a serious way, an opaque predicate. From

$$Rexist(E_1) \wedge not'(E_1, E_2)$$

we cannot conclude

$$Rexist(E_2)$$

Quite the opposite. Hence, we should take some care in formulating the representation.

The predicate *not* is related to the logical operator $\neg$ by the following axiom:

$$(2.70) \qquad (\forall\, e) not(e) \equiv \neg Rexist(e)$$

Using this and axiom schema (A1), we can infer from $\neg p(X)$ that there is no E such that $p'(E, X) \wedge Rexist(E)$, or equivalently, for all E for which $p'(E, X)$ is true, $not(E)$ is true.

However, the converse does not necessarily hold. It does not follow from $not(E) \wedge p'(E, X)$ that $\neg p(X)$ holds, for the latter is a much stronger statement. The former denies a particular condition of p's being true of X, whereas the latter denies all such conditions. If we leave things like this, negation will pack no punch. It will be too weak. When I say "John didn't do his homework", I would mean something like "John didn't do his homework at 8:15 last night," and would not exclude the possibility that John did his homework at 9:15 last night.

The predicate *not* can be strengthened to equivalence with the logical operator $\neg$ by means of the following axiom for one-argument eventualities:

$$(2.71) \quad (\forall e_1, e_2, e_3, x)[not'(e_2, e_1) \wedge Subst(x, e_3, x, e_2) \\ \supset \neg[Rexists(e_3) \wedge Rexists(e_2)]]$$

That is, if e_2 is the negation of condition e_1 and e_3 is any other eventuality with the same pattern as e_1, then e_2 and e_3 cannot both exist in the real world. This axiom means that by negating any condition of a predicate p's being true of x we negate all of them.

Similar axioms can be written for two- and three-argument eventualities.

This however is too strong. When we say "John didn't do his homework," we are not saying he has never done his homework. We generally have a particular context in mind, for example, all possible doings of his homework one particular evening. Thus, when we write $not(e)$ we intend e to be not a particular eventuality but the typical element of some set of similar eventualities—$typelt(e, s)$. The set may be the set of all eventualities of that description—$dset(s, e, e)$. But more likely, it will be a subset of this set.

Condoravdi et al. (19??) propose a similar analysis of "prevent". When we say, "John's closing the barn door prevented the horse from leaving," we don't mean to deny every possible leaving by the horse from the barn, but just those within a particular temporal context. They describe this as preventing a subtype of the type, rather than preventing a token. This is equivalent to saying that the thing prevented is the typical element of some subset of the set defined by the property.

It is hopeless to prove the consistency of an entire large knowledge base. But it will allow us to proceed with more confidence if we can establish the consistency of the theory comprised of Axioms (2.70) and (2.71) and axiom schema (A1). The following model does just that: Let the domain be the set of integers. Let $Rexist(X)$ be true if and only if X is even, let $not(X)$ be true if and only if X is odd, and, for any p, including $Rexist$ and not, let $p'(E, X)$ be true if and only if E is even and $p(X)$ is true. That this is a model for the theory is left as an exercise for the reader.

With this treatment of negation, we have reified not only events and conditions, but also the nonoccurrence or nonexistence of events and conditions. Thus, a condition of John's not running is a condition E_1 such that there is a condition E_2 such that

$$not'(E_1, E_2) \wedge run'(E_2, J)$$

The axioms corresponding to (2.69) and (2.70) for *or* and *imply* are as follows:

$$(2.72) \quad (\forall\, e_1, e_2)[or(e_1, e_2) \equiv [Rexist(e_1) \vee Rexist(e_2)]]$$
$$(\forall\, e_1, e_2)[imply(e_1, e_2) \equiv [Rexist(e_1) \supset Rexist(e_2)]]$$

The commutativity of *and* and *or* follows from axioms (2.69) and (2.72). The usual relations among *and*, *or*, *imply* and *not* can be encoded in such axioms as the following:

$$\begin{aligned}&(\forall\, e_1, e_2)[or(e_1, e_2)\\ &\quad \equiv (\exists\, e_3, e_4, e_5)[not(e_3) \wedge and'(e_3, e_4, e_5)\\ &\qquad \wedge not'(e_4, e_1) \wedge not'(e5, e_2)\end{aligned}$$
$$(\forall\, e_1, e_2)[imply(e_1, e_2) \equiv (\exists\, e_3)[or(e_3, e_2) \wedge not'(e_3, e_1)]]$$

But there is a question: Are a conjunction and its commutation the *same* condition, or two different conditions that always exist at the same time? Should we write the strong axiom

$$(\forall\, e, e_1, e_2)and'(e, e_1, e_2) \equiv and'(e, e_2, e_1)$$

or the weaker axiom

$$\begin{aligned}&(\forall\, e, e_1, e_2)Rexist(e) \wedge and'(e, e_1, e_2) \equiv (\exists\, e_3)and'(e_3, e_2, e_1) \wedge\\ &\quad Rexist(e_3)\end{aligned}$$

On a parallel issue in their own logics, Church (1951) and Levesque (1984) both argue for the former, stronger position, essentially contending that the equivalence is too obvious to have to deduce and that the distinction is only forced upon us as a result of our linear representation. A similar problem faces us for negation. Should the equivalence $P \equiv \neg\neg P$ be stated with the strong axiom

$$(\forall\, e_1, e_2)not'(e_1, e_2) \equiv not'(e_2, e_1)$$

(i.e., P and $\neg\neg P$ are the *same* propositions), or with the weaker axioms

$$\begin{aligned}&(\forall\, e_1, e_2, e_3)[not'(e_2, e_1) \wedge not'(e_3, e_2)\\ &\quad \supset [Rexists(e_1) \equiv Rexists(e_3)]]\end{aligned}$$

(i.e., P and $\neg\neg P$ are different propositions that are always true at the same time)? Adopting the stronger axioms would make certain proofs go through more easily, for example, in reasoning about other agents' reasoning. However, I tend to favor the weaker forms, since it gives one greater control over what is inferred, and consequently, over what is "actively believed". I would not want to say, for example, that because someone knows $P \wedge Q$ is true, he also knows whether $(P \vee \neg Q) \supset \neg(\neg P \vee \neg Q)$ is true. Where language provides the means to make a distinction, occasions are likely to arise in which the distinction is significant, and we should not destroy our ability to deal with it at the earliest stage of the enterprise—in defining the logical notation.

Obviously, *and* is transparent in both its arguments, and *not*, *or* and *imply* are not transparent.

Our dual notation allows us to move freely between representations of logical operations as predicates and as operators. Although in our axiomatizations we use the standard logical operators freely, generally logical words occurring in English sentences are translated into the predicates. This is in line with our initial decision to represent all of the content of sentences in predicates, rather than in other logical symbols. The logical operator $\wedge$ will be used only for the conjunction implicit in syntactic structure.

It is convenient to introduce a notational convention at this point. We frequently have occasion to write

$$p(E) \wedge and'(E, E_1, E_2) \wedge q'(E_1, X) \wedge r'(E_2, X) \tag{2.73}$$

That is, p is a property of the conjunction of $q(X)$ and $r(X)$. Because this is cumbersome, we introduce a constant with the name $E_1\&E_2$. Then when we write

$$p(E_1\&E_2) \wedge q'(E_1, X) \wedge r'(E_2, X)$$

it is to be understood as an abbreviation for (2.73). That is, $E_1\&E_2$ refers to an entity which has the property $and'(E_1\&E_2, E_1, E_2)$. It is the eventuality of both E_1's and E_2's obtaining.

This notation allows us to finesse a problem in representing the propositional content of sentences. In

John dreams he is running slowly.

John is dreaming of both the running and the slowness of the running. We may represent this sentence as

$$dream(J, E_1\&E_2) \wedge run'(E_1, J) \wedge slow'(E_2, E_1)$$

The more clumsy alternative would be

$$dream(J, E) \wedge and'(E, E_1, E_2) \wedge run'(E_1, J) \wedge slow'(E_2, E_1)$$

involving the predicate *and'* when no word "and" occurs in the sentence. Similarly, in unembedded sentences with transparent adverbial modifiers, like

John runs slowly.

it seems incomplete to say that the slowness of the running is by itself the main assertion, and representing the conjunction explicitly seems cumbersome and unmotivated. With the notational convention, we can represent the sentence as

$$Rexist(E_1\&E_2) \wedge run'(E_1, J) \wedge slow'(E_2, E_1)$$

The noun phrase

a former eminent scientist,

with the opaque adjective "former", may refer to a person who is both formerly eminent and formerly a scientist. We can represent this as an X such that

$$former(E_1\&E_2) \wedge eminent'(E_1, X) \wedge scientist'(E_2, X)$$

2.5.2 Nonexistent Objects

Briefly, the treatment of nonexistent objects, such as unicorns and Santa Claus, is to assume they exist in the Platonic universe but do not exist in the real world. There might even be the following axioms in the knowledge base:

$$(\forall\, x)unicorn(x) \supset \neg Rexist(x)$$
$$(\forall\, x)Santa\text{-}Claus(x) \supset \neg Rexist(x)$$

Recall from Section 2.2 that the existential quantifier asserts existence in the Platonic universe of possible individuals, not in the real world of actual individuals. We can now see one reason for this. Suppose we could assert nonexistence in the real world by the negation of an existential statement.

$$\neg(\exists\, x)unicorn(x)$$
$$\neg(\exists\, x)Santa\text{-}Claus(x)$$

Then it would be vacuously true that Santa Claus is a unicorn, which we all know is false.

In the ontologically promiscuous approach, we may write axioms stating the well-known facts about nonexistent unicorns, that they are horse-like and have a single horn, and about Santa Claus that he is fat and jolly.

$$(\forall\, x)unicorn(x) \supset single\text{-}horned(x)$$
$$(\forall\, x)unicorn(x) \supset horse\text{-}like(x)$$
$$(\forall\, x)Santa\text{-}Claus(x) \supset fat(x)$$
$$(\forall\, x)Santa\text{-}Claus(x) \supset jolly(x)$$

Existence in the real world is only one possible "mode of existence". Existence in various counterfactual contexts is possible, and these contexts can be reified either as predicates or as entities that can be in a kind of "exists-in" relation with possible individuals. Negation and implication are the simplest such contexts. The representation of the sentence "If he is Santa Clause, he is jolly," is

$$imply(e_1, e_2) \wedge Santa\text{-}Claus'(e_1, x) \wedge jolly'(e_2, x)$$

The property of x's being Santa Claus and x's being jolly exist in the world created by the antecedent of the implication, and x exists in that world too.

A more complex example is the context created by an extended counterfactual. An assumption is made. Conditions and entities are inferred to exist in the world created by that assumption. It is shown a contradiction follows, and thus the assumed condition and some of the other conditions and entities do not exist in the real world.

We can similarly say that in fiction, a fictional world is created. Entities can exist in that fictional world, but may not exist in the real world. For example, the expression

$$Leopold\text{-}Bloom'(e, x) \wedge world\text{-}of(w, Ulysses) \wedge exist\text{-}in(e, w) \wedge \\ exist\text{-}in(x, w)$$

says that in the Platonic universe there is an entity x that is Leopold Bloom, and an eventuality e of x's *being* Leopold Bloom. There is a world w created by the novel *Ulysses*. Both e and x exist in that world, but not in the real world.

Fictional worlds cannot be applied in any simple-minded way. We cannot simply fix the world for the interpretation of a particular sentence, because many sentences mix worlds in complex ways, as in

Who do you think I am, Santa Claus?

Here the speaker is relying on several facts about Santa Claus, but his existence in the real world is not one of them. At the very least it would be necessary to say more specifically what such worlds look like, how they are related to the real world, and how several such worlds could be called upon in the interpretation of single sentences.

The relation of fictional worlds to the real world is in general quite complex. In some cases, the invocation of a single fictional entity calls with it an entire world, as Zeus may bring along with him the entire world of Greek mythology. In other cases, as with Santa Claus, the fictional world is just the real world with a few new characters and a few events of questionable plausibility. The nature of the fictional world that accompanies unicorns is not at all clear. Each of these fictional worlds is constructed by means of a transformation of the real world, involving perhaps just its augmentation by certain characters and improbable incidents. Often there is a different knowledge base in the fictional world from that used in the real world, with respect to which sentences mentioning such entities are interpreted. Perhaps more drastically, there can be changes in physical laws, as in the book *Flatland*. Just how these fictional worlds are constructed is a complex process that would have to be subjected to case-by-case analysis.

The problem of entities with contradictory properties presents a somewhat more difficult case, but yields to a similar solution. This problem is the focus of Russell's criticism (1905) of an earlier approach to logical form—that of Meinong (1904). Russell is worth quoting:

> Of the possible theories which admit such constituents [nonexistent entities] the simplest is that of Meinong. This theory regards any grammatically correct denoting phrase as standing for an *object*. Thus "the present King of France," "the round square," etc., are supposed to be genuine objects. It is admitted that such objects do not *subsist*, but nevertheless they are supposed to be objects. This in itself is a difficult view; but the chief objection is that such objects, admittedly, are apt to infringe on the law of contradiction. It is contended, for example, that the existent present King of France exists, and also does not exist; that the round square is round, and also not round; etc. But this is intolerable; and if any theory can be found to avoid this result, it is surely to be preferred.

The difference between ontological promiscuity and the Meinongian approach is that in the former, one distinguishes clearly between the real world, where the law of contradiction holds, and the Platonic universe, where one has certain escape hatches. Meinong's objects are entities in the Platonic universe and his notion of subsistence is encoded by our predicate $Rexist$, establishing the relation between the Platonic universe and the real world.

Let us consider Russell's own sentence,

> It is contended that ... the round square is round, and also not round.

The first question we face in discourse interpretation theory is how to represent the information content in this sentence *before* any inferences are drawn. The representation must allow us to reason about what the sentence says, and in particular to draw the conclusion from it that Russell wants us to draw, that no such thing as a round square could exist.

First of all, what do we mean by a round square existing in the Platonic universe? We will take this to mean an entity X for which there are, in the Platonic universe, a condition E_1 of X's roundness and a condition E_2 of X's squareness. Then we want to show that these conditions cannot both exist in the real world. More precisely, given

$$round'(E_1, X) \wedge square'(E_2, X)$$

we want to show

$$(2.74) \qquad \neg[Rexist(E_1) \wedge Rexist(E_2)]$$

Suppose X's squareness E_2 exists. Then by the transparency of "square", X would exist. By (A1), X would be square. By means of an axiom (or more likely a theorem)

$$(\forall\, x)square(x) \supset \neg round(x)$$

we could conclude that X was not round. If X's roundness E_1 existed, then by (A1), X would be round, a contradiction. Thus, $Rexist(E_2)$ implies $\neg Rexist(E_2)$, establishing (2.74).

We needed a notation to do this. The notation needs a semantics. The Platonic universe is an invention that allows us to specify a semantics for the notation.

2.5.3 Opaque Adverbials and Modal Operators

It seems reasonably natural to treat transparent adverbials as properties of events. For opaque adverbials, like "almost", it seems less natural, and one is inclined to follow Reichenbach (1947) in treating them as functionals mapping predicates into predicates. Thus,

John is almost a man.

would be represented

$$almost(man)(J)$$

That is, *almost* maps the predicate *man* into the predicate *almost-a-man*.

This representation is undesirable for our purposes since it is not first-order. It would be preferable to treat opaque operators as we do transparent ones, as properties of events or conditions. The sentence would be represented

$$almost(E) \wedge man'(E, J)$$

But does this get us into difficulty?

First note that this representation does not imply that John is a man, for we have not asserted E's existence in the real world, and *almost* is opaque and does not imply its argument's existence. In fact, the opposite is true. We would want the axiom

$$(\forall\, e) almost(e) \supset not(e)$$

That is, something that almost occurs doesn't occur.

But is there enough information in E to allow one to determine the truth value of $almost(E)$ in isolation, without appeal to other facts? The answer is that there could be. We can construct a model in which for every functional F there is a corresponding equivalent predicate q, such that

$$(\forall\, p, x)[F(p)(x) \equiv (\exists\, e)[q(e) \wedge p'(e, x)]]$$

The existence of the model shows that this is not necessarily contradictory.

Let the domain of interpretation D be the class of finite sets built out of a finite set of urelements. The interpretation of a constant X will be some element of D; call it $I(X)$. The interpretation of a monadic predicate p will a subset of D; call it $I(p)$. Then if E is such that $p'(E, X)$, we define the interpretation of E to be $< I(p), I(X) >$.

Now suppose we have a functional F mapping predicates into predicates. We can define the corresponding predicate q to be such that

$q(E)$ is true iff there are a predicate p and a constant X where the interpretation of E is $< I(p), I(X) >$ and $F(p)(X)$ is true.

The fact that we can define such a predicate q in a moderately rich model means that we are licensed to treat opaque adverbials as properties of (possible) events and conditions.

The purpose of this exercise is only to show the viability of the approach. I am not claiming that a running event *is* an ordered pair of the runner and the set of all runners, although it should be mostly harmless for those irredeemably committed to set-theoretic semantics to view it like that.

Returning to "almost", the sentence

John was almost killed by a car.

would be represented

$Past(A) \wedge almost'(A, E) \wedge kill'(E, C, J)$

That is, an 'almost-ness' of a car's killing John occurred in the past.

This representation is of course very unrevealing about the meaning of "almost", but as always that sort of information must be built into the axioms. "Almost" is discussed further in Section 5.??.

Modal operators can be treated similarly. They are simply predicates which take other entities as their arguments and, generally, do not imply the existence of their arguments. Thus, the modal adverbial "possibly" is represented by the predicate *possible*:

Possibly, John works. $\Longrightarrow possible(E) \wedge work'(E, J)$

Modal auxilliaries can also be represented like this:

John can work. $\Longrightarrow can(E) \wedge work'(E, J)$

Possible and *can* are opaque predicates, so the existence of John's working is not implied. If we had had $Rexist(E)$, it would tell us that John's working exists in the real world; $possible(E)$ and $can(E)$ tells us something rather weaker about the existential status of John's working.

Finally, recall that in Section 2.3.7 we represented

This glass is full for a drink at this bar.

as

$$for(full, DS)(G)$$

We can now translate it into first-order notation:

$$for(E, DS) \wedge full'(E, G)$$

That is, the fullness E of glass G is *for* or *with respect to* the reference set DS. The predicate for is not transparent in its first argument, so we cannot conclude that the glass is full absolutely. If the sentence is negated, it is $for(E, DS)$ that is negated, rather than $full(G)$. We would thus be denying the fullness *with respect to* a specific reference set, and not in general.

It should be noted that this treatment of adverbials has consequences for the individuating criteria on eventualities. We can say "John is almost a man" without wishing to imply "John is almost a mammal," so we would not want to say that John's being a man is the *same* condition as his being a mammal. We are forced, though not unwillingly, into a position of individuating eventualities according to very fine-grained criteria.

2.6 Belief and Mutual Belief

2.6.1 Objects of Belief

Our representation for "John believes P" is $believe(J, P)$. It is necessary to say something about what the second argument of the predicate $believe$ is. What sort of entity is P? There seem to be four possibilities for, say, "John believes Mary is asleep".

1. P is an eventuality, the eventuality or condition in the Platonic universe of Mary's being asleep, that may or may not exist in the real world.

2. P is the proposition that is true when the eventuality of Mary's being asleep exists in the real world, that is, the proposition that Mary is asleep.

3. P is some sort of mental representation in John's head that represents the proposition that Mary is asleep or the eventuality of Mary's being asleep.

4. P is the English sentence "Mary is asleep."

It doesn't make very much difference which of the first three positions one adopts, for they are interchangeable via coercion functions like "the proposition that is true when the eventuality P exists" or "the eventuality that exists when the proposition P is true" or "the mental representation that represents proposition/eventuality P". Whatever they are, the class of them must be fine-grained enough that logically equivalent propositions can be distinguished, and rich enough to include logically impossible propositions.

I generally speak of P as an eventuality, but I am deliberately cavalier about the distinction between eventualities and propositions, if there is one. One sign of this is my use of both the letters E and P, as well as others, in the second argument position.

It is important, however, that we do not take P to be an English sentence.[17] An English sentence conveys a number of propositions, and it is necessary to be clear about just which of them are in fact believed. In line with our principle of allowing only predicates and not constants to convey information, we will want to say that the sentence "Mary is asleep" conveys two predications. There is some entity—call it M, but don't assume from that choice of letters that we know anything about M—about which we know two things: M is asleep and M is named "Mary". We can represent these $asleep(M)$ and $Mary(M)$. Generally when we say "John believes Mary is asleep" we mean that John believes both of these propositions. This is not necessarily the case, however. I might explain John's behavior at a reception with the sentence

> John knew Chomsky was famous, but couldn't remember his name.

The first clause conveys

$$know(J, P) \wedge famous'(P, C)$$

but not (at least after the second clause cancels it)

$$know(J, Q) \wedge Chomsky'(Q, C)$$

Two classes of problems arise in the representation of belief, the first introduced by AI researchers and the second by philosophers. The first class involves reasoning about other agents' beliefs. The second class includes the problem, introduced by Quine (1956), of distinguishing *de re* and *de dicto* belief reports, and the problem of identity, raised by Frege (1892). We will look at the AI problems first.[18]

[17] Moore and Hendrix (1982), who argue that belief should be viewed as a relation between an agent and an internal representation, make this same point when they say the internal "language of thought" cannot be identical with an external, natural language.

[18] Other proposals have been made in AI for representing knowledge and belief (e.g., Moore, 1980; Konolige, 1985). They have been developed specifically for this problem and are superior to the present proposal possibly in clarity and certainly in efficiency. But it is not clear how conveniently they will extend to a logical notation for representing discourse

2.6.2 Reasoning about Beliefs

The problem raised by artificial intelligence researchers involves drawing the appropriate inferences from belief reports. Given that an agent believes some propositions and is rational, how do we draw inferences about what else he believes? How do we reason about an agent's reasoning?

The approach taken here has been called the "syntactic" approach. The deductive processes of agents are modelled explicitly. This requires us to state two kinds of rules. First, agents know and use *modus ponens*. Second, they know and use *universal instantiation*.

The first of these is straightforward to state.

$$(2.75) \quad (\forall\, a, p, q, i) believe(a, p) \wedge believe(a, i) \wedge imply'(i, p, q) \\ \supset believe(a, q)$$

That is, if an agent a believes p and he believes i which is the proposition that p implies q, then he believes q.

Universal instantiation is not so simple to state, for we have to decide upon how the agent represents or understands universal quantification. There is a way to sneak past the problem by using universal instantiation outside of the belief contexts to do the work for us inside belief contexts. In this approach "John believes all men are mortal" is represented as follows:

$$(2.76) \quad (\forall\, p, x) man'(p, x) \\ \supset (\exists\, q, i) believe(J, i) \wedge imply'(i, p, q) \wedge mortal'(q, x)$$

That is, if p is x's being a man, then there is a q which is x's being mortal and an i which is the implication from p to q, and John believes i. If we also know that John believes Socrates is a man,

$$believe(J, P) \wedge man'(P, S)$$

then we can use universal and existential instantiation on (2.76) to derive

$$believe(J, I) \wedge imply'(I, P, Q) \wedge man'(P, S) \wedge mortal'(Q, S)$$

---

in general. The logical notation developed here is aimed toward discourse in general. In this section I merely want to show that it is "representationally adequate" for belief, in the sense that classical distinctions can be represented and that reasoning can be performed, neglecting questions of efficiency.

Then we can use (2.75) to conclude that John believes Socrates is mortal:

$$believe(J,Q) \wedge mortal'(Q,S)$$

The difficulty with proceeding in this way is that axiom (2.76) does not make a statement about John's general knowledge. It says rather of every particular entity (in the Platonic universe) that John believes that that particular entity is mortal if it is a man. Rather than having one general belief about man's mortality, John has infinitely many specific beliefs about the mortality of infinitely many particular possible men.

The alternative approach, and the one adopted here, is to reconstruct the syntax of first-order predicate calculus in agents' belief contexts.[19] The reader should be warned at the outset that this is not pretty. Corresponding to the universally quantified variable, there is something which, following McCarthy (1977), I call an "inner variable". Think of it as meaning "anything". An inner variable is not a variable in the logic but an entity in the Platonic universe, whose mysterious nature I will explicate only by listing the axioms it satisfies. It is not unrelated to typical elements; indeed, it may be thought of as a typical element of the class of all entities in the Platonic universe. The fact that something is an inner variable is expressed with the predicate iv; $iv(X)$ says that X is an inner variable. Specifically, if T is truth, that is, an eventuality that always exists in the real world—

$$eventuality(T) \wedge Rexists(T)$$

—and has no internal structure—

$$(\forall\, x,y)Subst(x,T,y,T)$$

then x is an inner variable if and only if it is the typical element of the set defined by the eventuality T.

$$(\forall\, x)[iv(x) \equiv (\exists\, s)dset(s,x,T)$$

In this scheme, "John believes all men are mortal" is represented

$$(2.77) \quad believe(J,I) \wedge imply'(I,P,Q) \wedge man'(P,X) \wedge mortal'(Q,X) \wedge iv(X)$$

[19] This is similar to approaches taken by McCarthy (1977), Moore (1980), and Konolige (1985). I follow McCarthy most closely, though perhaps not so closely that he would recognize it.

A gloss of this is "John believes the proposition I that P, X's being a man, implies Q, X's being mortal, where X is anything." Notice that inner variables are represented by constants in the notation.

Universal instantiation can now be stated as follows:

$$(2.78) \qquad (\forall\, x, y, p)[Rexists(p) \wedge iv(x) \supset (\exists\, q)[Subst(y, q, x, p) \wedge Rexists(q)$$

If an eventuality p exists in the real world and x is an inner variable, then for any y there is a q such that y plays the same role in q as x plays in p and q also exists in the real world. This is intended to be used when p somehow involves x—$arg+(x,p)$—but we don't need to state this in the axiom because if this doesn't hold, p and q are identical.

It is a straightforward consequence of this and the axioms for $Subst$ that agents who believe general statements use universal instantiation:

$$\begin{aligned}(2.79) \qquad & (\forall\, a, x, y, p)[believe(a, p) \wedge iv(x) \\ & \quad \supset (\exists\, q)[Subst(y, q, x, p) \wedge believe(a, q)\end{aligned}$$

If agent a believes p and x is an inner variable, then for any y, there is a q such that y plays the same role in q that x plays in p, then a believes q. We need not state that a believes x is an inner variable. We can simply assume that knowing a proposition involving an inner variable is knowing the general fact. A creature incapable of doing universal instantiation would not know a proposition involving an inner variable.

We can now work through a simple syllogism.[20] John believes that Socrates is a man,

$$(2.80) \qquad believe(J, PS) \wedge man'(PS, S)$$

and that all men are mortal,

$$\begin{aligned}(2.81) \qquad & believe(J, I) \wedge imply'(I, P, Q) \wedge man'(P, X) \\ & \quad \wedge mortal'(Q, X) \wedge iv(X)\end{aligned}$$

[20] It will be obvious that this is a place where a computer implementation would use special purpose techniques rather than blind deduction.

To show John believes Socrates is mortal, we must show

$$(\exists\, qs) believe(J, qs) \wedge mortal'(qs, S)$$

It follows from (2.81) and (2.79) that there is a IS such that

$$(2.82) \qquad Subst(S, IS, X, I) \wedge believe(J, IS)$$

From the definitional axioms for $Subst$ this means that there is a QS such that

$$(2.83) \qquad believe(J, IS) \wedge imply'(IS, PS, QS) \wedge man'(PS, S) \\ \wedge\, mortal'(QS, S)$$

From this, (2.80), and (2.75) (agents apply modus ponens), we conclude

$$(2.84) \qquad believe(J, QS) \wedge mortal'(QS, S)$$

There are several further properties and nonproperties of belief and knowledge that we need to axiomatize or avoid axiomatizing. The first relates belief and conjunction:

$$(\forall\, a, p, q, c) believe(a, p) \wedge believe(a, q) \wedge and'(c, p, q) \supset believe(a, c)$$

That is, if an agent a believes p and believes q and c is a conjunction of p and q, then a believes c. The converse of this is also true:

$$(2.85) \qquad (\forall\, a, c, p, q) believe(a, c) \wedge and'(c, p, q) \\ \supset believe(a, p) \wedge believe(a, q)$$

If an agent a believes the conjunction c of p and q, then a believes p and a believes q.

Notice that we do not want the axiom relating belief and *disjunction*, corresponding to (2.85):

$$* \; (\forall\, a, d, p, q)[believe(a, d) \wedge or'(d, p, q) \supset [believe(a, p) \vee believe(a, q)]]$$

The fact that an agent believes p or q does not imply that he believes p or he believes q. He may be genuinely uncertain which is true. As Moore (1980) has pointed out, this is one of the pitfalls of simple models of belief.

Another pitfall is the relation between belief and negation. We want to be able to conclude from the fact that an agent believes $\neg p$ that the agent does not believe p:

$$(\forall\, a, n, p, b) believe(a, n) \wedge not'(n, p) \wedge believe'(b, a, p) \supset not(b)$$

But we do not want the converse to be true.

$$*\ (\forall\, a, n, p, b) not(b) \wedge believe'(b, a, p) \wedge not'(n, p) \supset believe(a, n)$$

The agent may have no opinion as to whether p is true or not.

Finally, let us point out a difference in the properties of *believe* and *know*. If someone knows something, that something is true:

$$(\forall\, a, p) know(a, p) \supset Rexist(p)$$

But it is not the case that if someone *believes* something, it is true:

$$*\ (\forall\, a, p) believe(a, p) \supset Rexist(p)$$

It should be noted that by having axioms (2.75) and (2.78) (or (2.79)), we not only are able to reason about someone's beliefs. We are also able to show that an agent believes all the logical consequences of his beliefs. This is true of other models of belief as well (e.g., Moore, 1980). However, it violates common sense. We don't know all the theorems of mathematics just because we know the axioms of set theory. Konolige (1985) has developed a treatment of belief that is more intuitive in that it lacks this "deductive closure" property. My response to this difficulty is what it will be to many similar difficulties. In the abductive approach outlined in Chapter 3 and used for the rest of the book, axioms are defeasible and used only when they contribute to the best interpretation of a text. We will not be able to conclude that an agent knows all the consequences of his beliefs, because the discourse being interpreted will not license all of those inferences.

2.6.3 Mutual Belief

An extremely important concept in the study of discourse is "mutual belief" (McCarthy, 1975; Lewis, 1969; Schiffer, 1972; Clark and Marshall, 1981), since ideally it determines what can be presupposed by the speaker. A set of agents s mutually believes a proposition p when they not only all belief p individually, but also believe that the others believe p, and believe that the others believe that the others believe p, and so on, *ad infinitum.* This notion can be axiomatized with the two following axioms:

$$\begin{array}{ll}(2.86) & (\forall\, s,p,a)mb(s,p) \wedge member(a,s) \supset believe(a,p) \\ (2.87) & (\forall\, s,p,m)Rexist(m) \wedge mb'(m,s,p) \supset mb(s,m)\end{array}$$

The first axiom says that if a set s of agents mutually believe p and agent a is a member of s, then a believes p. This allows us to infer belief from mutual belief. The second axiom says that if s mutually believes p, then s mutually believes that s mutually believes p. This allows us to embed mutual belief in mutual belief contexts.

If in addition these rules are mutually believed, we can derive that if a set s of agents mutually believes p and mutually believes a is a member of s, then they will mutually believe that a believes p.

$$\begin{array}{ll}(2.88) & (\forall\, s,p,a,e,m)[mb(s,p) \wedge member'(e,a,s) \wedge mb(s,e) \\ & \quad \supset (\exists\, b)[mb(s,b) \wedge believe'(b,a,p)]]\end{array}$$

From this axiomatization it is straightforward to show, for example, that if a set s mutually believes p and John, Bill, and Mary are in s, then John believes that Bill believes that John believes that Mary believes p.

All of the axioms that have been stated and will be stated should be thought of as mutually believed by some set s of agents. Some of the axioms, such as axiom (2.75) about using modus ponens, are mutually believed by all rational beings. Other axioms will be much more culturally dependent. In fact, it's useful to think of agents' knowledge bases as being indexed by what social group each class of facts is shared with. Thus, there is some knowledge that I share with all Americans or all English speakers, some facts I share with other people raised in the Midwest, some I share with other computer scientists, and some I share with a few close friends.

Properly, then, all the axioms should be stated not in the form I have given them, but embedded in mutual belief contexts. That will not be done in this book, since that would make the axioms totally impenetrable, whereas now the reader may eventually come to be able to read them. However, just once, it should be shown that the axioms can be embedded within mutual belief contexts. To do this, one first translates the axiom into Skolemized form. One replaces each universally quantified variable by an inner variable, say X, and add the conjunct $iv(X)$. One replaces the terms of the form $f(x)$, where f is a Skolem function, by a constant, say Y, and adds a conjunct of the form $FD(Y, X)$. Logical operators connecting unprimed predications are translated into the corresponding logical operator predicates introduced in Section 2.5.1. The eventuality corresponding to the highest level logical operator becomes the second argument of the predicate mb.

As an example, consider an axiom of the form

$$(\forall\, x)p(x) \,\wedge\, q(x) \,\supset\, (\exists\, y)r(y, x)$$

Embedded in a mutual belief context, this becomes

$$\begin{aligned} &mb(S, I) \wedge imply'(I, EA, ER) \wedge and'(EA, EP, EQ) \wedge p'(EP, X) \\ &\quad \wedge q'(EQ, X) \,\wedge\, r'(ER, Y, X) \,\wedge\, iv(X) \,\wedge\, FD(Y, X) \end{aligned}$$

A final example: Consider axiom (2.75) that people use modus ponens, repeated here for the reader's convenience.

$$\begin{aligned} (2.75) \quad &(\forall\, a, i, p, q) believe(a, p) \,\wedge\, believe(a, i) \,\wedge\, imply'(i, p, q) \\ &\quad \supset\, believe(a, q) \end{aligned}$$

Corresponding to the universally quantified variables, we have the constants A, I, P, and Q, which are inner variables. The translation would be as follows:

$$\begin{aligned} (2.89) \quad &mb(S, I_1) \wedge imply'(I_1, C, B_3) \wedge and'(C, B_1, B_2) \wedge believe'(B_1, A, P) \\ &\quad \wedge believe'(B_2, A, I) \,\wedge\, imply'(I, P, Q) \,\wedge\, believe'(B_3, A, Q) \\ &\quad \wedge iv(A) \,\wedge\, iv(I) \,\wedge\, iv(P) \,\wedge\, iv(Q) \end{aligned}$$

This can be given the tortured paraphrase, the set S of agents mutually believes the implication I_1 from the conjunction C of A's belief B_1 that P and A's belief B_2 in the implication I from P to Q to A's belief B_3 that Q.[21]

[21] It is worth stating, though surely not worth demonstrating, that the "Three Wise Men" problem, posed by McCarthy (1975) as a challenge for the adequacy of representation schemes for knowledge and belief, can be solved in the formalism presented here.

2.6.4 *De Re* and *De Dicto* Belief Reports

The AI problems concerning belief involve reasoning about an agent's beliefs. The philosophers' problems concern the adequate expressivity of the notation. The first of these is the problem of distinguishing between *de re* and *de dicto* belief reports.

A belief report like

(2.90) John believes a man at the next table is a spy.

has two interpretations. The *de dicto* interpretation is likely in the circumstance in which John and some man are at adjacent tables and John observes suspicious behavior. The *de re* interpretation is likely if some man is sitting at the table next to the speaker of the sentence, and John is nowhere around but knows the man otherwise and suspects him to be a spy. A sentence that very nearly forces the *de re* reading is

John believes Bill's mistress is Bill's wife.[22]

whereas the sentence

John believes Russian consulate employees are spies.

strongly indicates a *de dicto* reading. In the *de re* reading of (2.90), John is not necessarily taken to know that the man is in fact at the next table, but he is normally assumed to be able to identify the man somehow. More on "identify" below. In the *de dicto* reading John believes there is a man who is both at the next table and a spy, but may be otherwise unable to identify the man. The *de re* reading of (2.90) is usually taken to support the inference

(2.91) There is someone John believes to be a spy.

whereas the *de dicto* reading supports the weaker inference

(2.92) John believes that someone is a spy.

[22] This example is due to Moore and Hendrix (1982).

As Quine has pointed out, as usually interpreted, the first of these sentences is false for most of us, the second one true. A common notational maneuver (though one that Quine rejects) is to represent this distinction as a scope ambiguity (e.g., Montague, 1974). Sentence (2.91) is encoded as (2.93) and (2.92) as (2.94):

(2.93) $(\exists\, x) believe(J, spy(x))$
(2.94) $believe(J, (\exists\, x) spy(x))$

If one adopts this notation and stipulates what the expressions mean, then there are certainly distinct ways of representing the two sentences. But the interpretation of the two expressions is not obvious. It is not obvious for example that (2.93) could not cover the case where there is an individual such that John believes him to be a spy but has never seen him and knows absolutely nothing else about him—not his name, nor his appearance, nor his location at any point in time—beyond the fact that he is a spy.

In fact, the notation we propose takes (2.93) to be the most neutral representation. Since quantification is over entities in the Platonic universe, (2.93) says that there is some entity in the Platonic universe such that John believes of that entity that it is a spy. Expression (2.93) commits us to no other beliefs on the part of John. When understood in this way, expression (2.93) is a representation of what is conveyed in a *de dicto* belief report. Translated into the flat notation and introducing a constant for the existentially quantified variable, (2.93) becomes

(2.95) $believe(J, P) \wedge spy'(P, S)$

Anything else that John believes about this entity must be stated explicitly. In particular, the *de dicto* reading of (2.90) would be represented by something like

(2.96) $believe(J, P \& Q) \wedge spy'(P, S) \wedge at'(Q, S, T)$

where T is the next table. That is, John believes S is a spy and that S is at the next table. John may know many other properties about S and still fall short of knowing *who* the spy is. There is a range of possibilities for John's knowledge, from the bare statements of (2.95) and (2.96) that correspond to a *de dicto* reading to the full-blown knowledge of S's identity that is normally present in a *de re* reading. In fact, an FBI agent would progress through just such a range of belief states on his way to identifying the spy.

To state John's knowledge of S's identity properly, we would have to state explicitly John's belief in a potentially very large collection of properties of the spy. To arrive at a succinct way of representing knowledge of identity in our notation, let us consider the two pairs of equivalent sentences:

> What is that?
> Identify that.
>
> The FBI doesn't know who the spy is.
> The FBI doesn't know the spy's identity.

The answer to the question "Who are you?" and what is required before we can say that we know *who* someone is or that we know their identity is a highly context-dependent matter. Several years ago, before I had ever seen Kripke, if someone had asked me whether I knew who Saul Kripke was, I would have said, "Yes. He's the author of 'Naming and Necessity'." Then once I was at a workshop which I knew was being attended by Kripke, but I didn't yet know what he looked like. If someone had asked me whether I knew who Kripke was, I would have had to say, "No." The relevant property in that context was not his authorship of some paper, but any property that distinguished him from the others present, such as "the man in the back row holding a cup of coffee".

Generally when someone asks me who I am, the appropriate answer is "Jerry Hobbs." If someone asks me who Jerry Hobbs is, the appropriate answer is "Me." But if I accidently walked into the board room of IBM during a meeting and the chairman of the board asked me, "Who are you?", it would not be an adequate answer for me to say, "Jerry Hobbs." I am being asked for some other property, one that will explain my presence, such as "I was looking for the cafeteria."

Knowledge of a man's identity is then a matter of knowing some context-dependent essential property that serves to identify him for present purposes—that is, a matter of knowing *who* he is.

Therefore, we need a kind of place-holder predicate to stand for this essential property, that in any particular context can be specified more precisely. It happens that English has a morpheme that serves just this function—the morpheme "wh". In line with our decision to use simple English morphemes for predicates where they are available and seem to mean the right thing, we posit a predicate *wh* that stands for the contextually determined property or conjunction of properties that would count as an identification in that particular context.

The *de re* reading of (2.90) is generally taken to include John's knowledge of the identity of the alleged spy. Assuming this, a *de re* belief report would be represented as a conjunction of two beliefs, one for the main predication and the other expressing knowledge of the essential property, the what-ness, of the argument of the predication.

$$believe(J, P) \wedge spy'(P, X) \wedge know(J, Q) \wedge wh'(Q, X)$$

That is, John believes S is a spy and John knows who S is.

This solves the *representation* problem, but it doesn't solve the real problem of how to determine what that context-dependent essential property is in any particular context. That's a pragmatics problem and, hence, is dealt with in Chapter 6. Depending on the context, we may be able to determine what specific properties of S John knows, or more likely, we will be able to draw the appropriate inferences about what John will now do, given his knowledge of the identity of the spy.

However, let us probe this distinction just a little more deeply and in particular call into question whether knowledge of identity is really part of the meaning of the sentence in the *de re* reading. The representation of the *de dicto* reading of (2.90), I have said, is

(2.97) $$believe(J, P) \wedge spy'(P, S) \wedge believe(J, Q) \wedge at'(Q, S, T)$$

Let us represent the *de re* reading as

(2.98a) $$believe(J, P) \wedge spy'(P, S) \wedge Rexist(Q) \wedge at'(Q, S, T)$$
(2.98b) $$\wedge\, know(J, R) \wedge wh'(R, S)$$

What is common to (2.97) and (2.98) are the conjuncts $believe(J, P)$, $spy'(P, S)$ and $at'(Q, S, T)$. There is a genuine ambiguity as to whether Q exists in the real world (*de re*) or is merely believed by John (*de dicto*). In addition, (2.98) includes the conjuncts $know(J, R)$ and $wh'(R, S)$ – line

(2.98b). But are these necessarily part of the *de re* interpretation of sentence (2.90)? The following example casts doubt on this. Suppose the entire Rotary Club is seated at the table next to the speaker of (2.90), but John doesn't know this. John believes that some member of the Rotary Club is a spy, but has no idea which one. Sentence (2.90) describes this situation, and only (2.98a) holds, not (2.98b) and not (2.97). Judgments are sometimes uncertain as to whether sentence (2.90) is appropriate in these circumstances, but it is certain that the sentence

John believes someone at the next table is a spy.

is appropriate, and that is sufficient for the argument.

It seems then that the conjuncts $know(J, R)$ and $wh'(R, S)$ are not part of what we want in the initial logical form of the sentence,[23] but only a very common conversational implicature. The reason the implicature is very common is that if John doesn't know that the man is at the next table, there must be some other description under which John is familiar with the man. The story I just told provides such a description, but not one sufficient for identifying the man.

This analysis of *de re* belief reports is particularly appealing since it allows us to see the *de re - de dicto* distinction as only one example of a more general problem—the existential status of the grammatically subordinated propositions conveyed by a sentence. This problem is addressed in Section 4.??.

Moore (1980) approached the problem of essential properties by assuming a set of "standard names" for entities and saying that someone knows the identity of something when he knows its standard name. Suppose every entity X in the Platonic universe had a standard name—call it $sn(X)$. In our approach to proper names this would be a predicate, true of only one individual, and what we would know when we knew the standard name would be an eventuality E such that $sn(X)'(E, X)$. The standard names approach would then be equivalent to our approach if we had the axiom

$$(2.99) \qquad (\forall\, e, x) wh'(e, x) \equiv sn(x)'(e, x)$$

That is, the context-dependent property wh, the *what-ness* of something, is always exactly the property conveyed by the standard name. Representing

[23] Another way of putting it: they are not part of the literal meaning of the sentence.

the essential property by means of the predicate *wh* is thus a generalization of the standard names approach, since we do not necessarily want to have axiom (2.99).

We certainly do not have standard names for everything, for example, the thumb tack that is currently highest on my bulletin board, and examples above show that knowing Kripke's name is not the same as knowing who Kripke is. Moore notes these difficulties and makes no claims for the standard names approach beyond the statement that it is a convenient idealization. It is not an unreasonable idealization. Proper names (and social security numbers too) can be viewed as attempts to assign standard names to large classes of entities. One reason that proper names often function as the essential property is that much of our knowledge is organized around proper names. When we learn someone's name, we usually gain access to a large class of other facts about that person, often including the relevant essential property.

But even in cases where standard names seem most appropriate, difficulties arise. Telephone numbers and combinations of safes are frequent examples in the literature. It is almost always the essential property of my telephone number that it is 859-2229, for knowing that is sufficient for dialing it, and that's almost all one ever has to do with a telephone number. But even with telephone numbers, it is possible to construct examples where knowing the sequence of numbers is not sufficient for the purposes at hand. Suppose I know that George's telephone number is 848-7465 and that Claudia's telephone number is VITRIOL. Both of these would be considered standard names. But if what is desired is to know whether the telephone numbers are the same, then if I don't know the correspondences between numbers and letters on the telephone pad, I don't know the essential property. For entities which are characterized by more complex combinations of properties, namely, nearly everything, it is much more difficult to specify a property which will almost always tell what is required for the specific situation and would thus function as a useful standard name.

It was stated above that the representation (2.95) for the *de dicto* reading conveys no properties of S other than that John believes him to be a spy. In particular, it does not convey S's existence in the real world. S thus refers to a possible individual, who may turn out to be actual if, for example, John ever comes to be able to identify the person whom he believes to be the spy, or if there is some actual spy who has given John good cause for his suspicions.

However, it may be that S is not actual, only possible. Suppose this is the case. One common objection to possible individuals is that they may seem to violate the Law of the Excluded Middle. Is S married or not married? Our intuition is that the question is inappropriate, and indeed the answer given in our formalism has this flavor. By axiom (A1), $married(S)$ is really just an abbreviation for $married'(E,S) \wedge Rexist(E)$. This is false, for the existence of E in the real world would imply the existence of S. So $married(S)$ is also false. But its falsity has nothing to do with S's marital status, only his existential status. The predication $unmarried(S)$ is false for the same reason. The primed predicates are basic, and for them the problem of the excluded middle does not arise. The predication $married'(E,S)$ is true or false depending on whether E is the condition of S's being married. An unprimed, transparent predicate carries along with it the existence of its arguments, and it can fail to be true of an entity either through the entity's being actual but not having that property or through the nonexistence of the entity.

2.6.5 Identity in Intensional Contexts

The second of the philosophers' problems arises in *de dicto* belief reports. It is the problem of identity in intensional contexts, raised by Frege (1892). One way of stating the problem is this. Why is it that if

(2.100) John believes the Evening Star is rising.

and if the Evening Star is identical to the Morning Star, it is not necessarily true that

(2.101) John believes the Morning Star is rising.

By Leibniz's principle, we ought to be able to substitute for an entity any entity that is identical to it.

This puzzle survives translation into the logical notation, if John knows of the existence of the Morning Star and if proper names are unique. The representation for (the *de dicto* reading of) sentence (2.100) is

(2.102) $$believe(J,P_1) \wedge rise'(P_1,ES) \wedge believe(J,Q_1) \\ \wedge Evening\text{-}Star'(Q_1,ES)$$

John's belief in the Morning Star would be represented

$$believe(J, Q_2) \wedge Morning\text{-}Star'(Q_2, MS)$$

The existence of the Evening Star and the Morning Star is expressed by

$$Rexist(Q_1) \wedge Rexist(Q_2)$$

The uniqueness of the proper name "Evening Star" is expressed by the axiom

$$(\forall\, x, y) Evening\text{-}Star(x) \wedge Evening\text{-}Star(y) \supset x = y$$

The identity of the Evening Star and the Morning Star is expressed

$$(\forall\, x) Evening\text{-}Star(x) \equiv Morning\text{-}Star(x)$$

From all of this we can infer that the Morning Star MS is also an Evening Star and hence is identical to ES, and hence can be substituted into $rise'(P_1, ES)$ to give $rise'(P_1, MS)$. Then we have

$$\begin{array}{l} believe(J, P_1) \wedge rise'(P_1, MS) \wedge believe(J, Q_2) \\ \quad \wedge Morning\text{-}Star'(Q_2, MS) \end{array}$$

This is a representation for the paradoxical sentence (2.101).

There are three possibilities for dealing with this problem. The first is to discard or restrict Leibniz's Law. The second is to deny that the Evening Star and the Morning Star are identical as entities in the Platonic universe; they only happen to be identical in the real world, and that is not sufficient for intersubstitutivity. The third is to deny that expression (2.102) represents sentence (2.100) because "the Evening Star" in (2.100) does not refer to what it seems to refer to.

The first possibility is the approach of researchers who treat belief as an operator rather than as a predicate, and then restrict substitution inside the operator. We cannot avail ourselves of this solution because of the flatness of the our notation. The predicate *rise* is surely referentially transparent, so if ES and MS are identical, MS can be substituted for ES in the expression $rise'(P_1, ES)$ to give $rise'(P_1, MS)$. Then the expression $believe(J, P_1)$ would not even require substitution to be a belief about the Morning Star.

In any case, this approach does not seem wise in view of the central importance played in discourse interpretation by the identity of differently presented entities, i.e. by coreference. Free intersubstitutibility of identicals seems a desirable property to preserve.

The second possible answer to Frege's problem is to say that in the Platonic universe, the Morning Star and the Evening Star are different entities. It just happens that in the real world they are identical. But it is not true that $ES = MS$, for equality, like quantification, is over entities in the Platonic universe. The fact that ES and MS are identical in the real world must be stated explicitly, say, by the expression

$$identical(ES, MS)$$

or more properly,

$$(\forall\, x, y) Morning\text{-}Star(x) \wedge Evening\text{-}Star(y) \supset identical(x, y)$$

The predicate *identical* would refer only to identity in the real world.

For reasoning about "identical" entities, that is, Platonic entities that are identical in the real world, we may take an approach that parallels our approach to instantiations of universally quantified variables. The predicate *identical* will play a role similar to that of *Subst*. Corresponding to axiom schemas (2.19)-(2.22), we would have an axiom schema

$$\begin{aligned}(2.103)\qquad & (\forall\, e_1, e_3, e_4, \ldots) p'(e_1, \ldots, e_3, \ldots) \wedge identical(e_4, e_3) \\ & \quad \supset (\exists\, e_2) p'(e_2, \ldots, e_4, \ldots) \wedge identical(e_2, e_1)\end{aligned}$$

where e_3 is the kth argument of p and p is referentially transparent in its kth argument. There would be an axiom for the first argument of *believe* but not its second argument. Substitution of *identical*s in a condition results not in the same condition but in an *identical* condition.

Axioms will express the fact that *identical* is an equivalence relation:

$$(\forall\, x) identical(x, x)$$

$$(\forall\, x, y) identical(x, y) \supset identical(y, x)$$

$$(\forall\, x, y, z) identical(x, y) \wedge identical(y, z) \supset identical(x, z)$$

Corresponding to the axiom (2.78) expressing universal instantiation, we would have the following axiom, which together with axiom (2.103) expresses Leibniz's Law:

$$(\forall\, e_1, e_2) identical(e_1, e_2) \supset (Rexist(e_1) \equiv Rexist(e_2))$$

From all of this we can prove that if the Evening Star rises then the Morning Star rises, but we cannot prove from John's belief that the Evening Star rises that John believes the Morning Star rises. If John knows the Morning Star and the Evening Star are identical, and he knows axiom (2.103), then his belief that the Morning Star rises can be proved as one would prove his belief in the consequences of any other syllogism whose premises he believed.

This solution is in the spirit of our whole representational approach in that it forces us to be painfully explicit about everything. The notation does no magic for us. In my view, it is the correct approach.

There is a significant cost associated with this solution, however. When proper names are represented as predicates and not as constants, the natural way to state the uniqueness of proper names is by means of axioms of the following sort:

$$(\forall\, x, y) Evening\text{-}Star(x) \wedge Evening\text{-}Star(y) \supset x = y$$

But since from the axioms for *identical* we can show that $Evening\text{-}Star(MS)$, it would follow that $MS = ES$. We must thus restate the axiom for the uniqueness of proper names as

$$(\forall\, x, y) Evening\text{-}Star(x) \wedge Evening\text{-}Star(y) \supset identical(x, y)$$

A similar modification must be made for functions. Since we are using only predicates, the uniqueness of the value of a function must be encoded with an axiom like

$$(\forall\, x, y, z) father(x, z) \wedge father(y, z) \supset x = y$$

If x and y are both fathers of z, then x and y are the same. This would have to be replaced by the axiom

$$(\forall\, x, y, z) father(x, z) \wedge father(y, z) \supset identical(x, y)$$

The very common problems involving reasoning about equality, which can be done efficiently, are thus translated into problems involving reasoning about the predicate *identical*, which is very cumbersome.

One way to view this second solution is as a fix of the first solution. For $=$ we substitute the relation *identical*, and by means of axiom schema (2.103), we force substitutions to propagate to the eventualities they occur in, and we force the distinction between referentially transparent and referentially opaque predicates to be made explicitly. It is thus an indirect way of rejecting Leibniz's Law.

The third solution is to say that "the Evening Star" does not really refer to the Evening Star, but to some abstract entity somehow related to the Evening Star. That is, sentence (2.100) is really an example of metonymy. This may seem counterintuitive, and even bizarre, at first blush. But in fact the most widely accepted classical solutions to the problem of identity are of this flavor. For Frege (1892) "the Evening Star" in sentence (2.100) does not refer to the Evening Star but to the *sense* of the phrase "the Evening Star". In a more recent approach, Zalta (1983) takes such noun phrases to refer to "abstract objects" related to the real object. In both approaches noun phrases in intensional contexts refer to senses or abstract objects, while other noun phrases refer to actual entities, and so it is necessary to specify which predicates are intensional. In a Montagovian approach, "the Evening Star" would be taken to refer to the *intension* of the Evening Star, not its *extension* in the real world, and noun phrases would *always* be taken to refer to intensions, although for nonintensional predicates there would be meaning postulates that make this equivalent to reference to extensions.

Thus, in all these approaches intensional and extensional predicates must be distinguished explicitly, and noun phrases in intensional contexts are systematically interpreted metonymically.

It would be easy enough in our framework to implement these approaches. We can define a function α of three arguments – the actual entity, the cognizer, and the condition used to describe the entity—that returns the sense, or intension, or abstract entity, corresponding to the actual entity for that cognizer and that condition. Sentence (2.100) would be represented, not as (2.102), but as

$$\begin{array}{ll}(2.104) & believe(J, P_1) \wedge rise'(P_1, \alpha(ES, J, Q_1)) \wedge believe(J, Q_1) \\ & \quad \wedge\, Evening\text{-}Star'(Q_1, ES)\end{array}$$

I tend to prefer to think of the value of $\alpha(ES, J, Q_1)$ as an abstract entity. Whatever it is, it is necessary that the value of $\alpha(ES, J, Q_1)$ be something different from the value of $\alpha(ES, J, Q_2)$ where $Morning\text{-}Star'(Q_2, ES)$. That is, different abstract objects must correspond to the condition Q_1 of being the Evening Star and the condition Q_2 of being the Morning Star. It is because of this feature that we escape the problem of intersubstitutivity of identicals, for substitution of MS for ES in (2.104) yields $\ldots \wedge rise'(P_1, \alpha(MS, J, Q_1)) \wedge \ldots$ rather than $\ldots \wedge rise'(P_1, \alpha(MS, J, Q_2)) \wedge \ldots$, which would be the representation of sentence (2.101).

The difficulty with this approach is that it makes the interpretation of noun phrases dependent on their embedding context:

Intensional context $\Longrightarrow$ metonymic interpretation
Extensional context $\Longrightarrow$ nonmetonymic interpretation

It thus violates the naive compositionality that I have been at so many pains to preserve in this chapter. Metonymy is a very common phenomenon in discourse. In Chapter 4 it will be used to solve several seemingly syntactic problems. In Chapter 6 we will develop means for dealing with it. But I prefer to think of it as occurring irregularly, and not as signalled systematically by other elements in the sentence.

Having laid out the three possible solutions and their shortcomings, I find that I would like to avoid the problem of identity altogether. The third approach suggests a ruse for doing so. We can assume that, in general, (2.102) is the representation of sentence (2.100). We invoke no extra complications where we don't have to. When, in interpreting the text, we encounter a difficulty resulting from the problem of identity, we can go back and revise our interpretation of (2.100), by assuming the reference must have been a metonymic one to the abstract entity and not to the actual entity. In those cases it would be as if we are saying, "John couldn't believe about the Evening Star itself that it is rising. The paradox shows that he is insufficiently acquainted with the Evening Star to refer to it directly. He must be talking about an abstract entity related to the Evening Star." My guess is that we will not have to resort to this ruse often, for I suspect the problem rarely arises in actual discourse interpretation.

The reader may be either delighted or disturbed that, rather than use Frege's problem to force us into a more complex logic or ontology, I have simply used available resources to patch up the difficulty.

Bibliography

[1] Bach, Emmon, 1981. "On Time, Tense, and Aspect: An Essay in English Metaphysics", in P. Cole, ed., *Radical Pragmatics*, pp. 63-81, Academic Press, New York.

[2] Barwise, Jon, and Robin Cooper, 1981. "Generalized Quantifiers and Natural Language",*Linguistics and Philosophy*, Vol. 4, No. 2, pp. 159-219.

[3] Bobrow, Daniel, and Terry Winograd, 1977. "An Overview of KRL, A Knowledge Representation Language", *Cognitive Science*, Vol. 1, pp. 3-46.

[4] Castañeda, Hector-Neri, 1967. "Comments on D. Davidson's 'The Logical Form of Action Sentences'", in N. Rescher, ed., *The Logic of Decision and Action*, pp. 104-112, University of Pittsburgh Press, Pittsburgh, Pennsylvania.

[5] Church, Alonzo, 1951. "A Formulation of the Logic of Sense and Denotation", in P. Henle, H. M. Kallen, and S. K. Langer, eds., *Structure, Method and Meaning*, Liberal Arts Press, New York.

[6] Clark, Herbert H., and Catherine R. Marshall, 1981. "Definite Reference and Mutual Knowledge", in A. Joshi, B. Webber, and I. Sag, eds., *Elements of Discourse Understanding*, pp. 10-63, Cambridge University Press, Cambridge, England.

[7] Davidson, Donald, 1967. "The Logical Form of Action Sentences", in N. Rescher, ed., *The Logic of Decision and Action*, pp. 81-95, University of Pittsburgh Press, Pittsburgh, Pennsylvania.

[8] Fodor, Jerry A., 1980. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", *The Behavioral and Brain Sciences*, Vol. 3, No. 1.

[9] Frege, Gottlieb, 1892. "On Sense and Nominatum", in H. Friegel and W. Sellars, eds., *Readings in Philosophical Analysis*, pp. 85-102, Appleton-Century-Crofts, Inc., New York, 1949.

[10] Frege, Gottlieb, 1893. *Grundgesetze der Arithmetik*, Jena.

[11] Geertz, Clifford, 1973. "Person, Time, and Conduct in Bali", in *The Interpretation of Cultures: Selected Essays by Clifford Geertz*, pp. 360-411, Basic Books, New York.

[12] Hawkins, John, 1978. *Definiteness and Indefiniteness*, Croom Helm, London.

[13] Hayes, Patrick J., 1979. "The Logic of Frames", in D. Metzing, ed., *Frame Conceptions and Text Understanding*, pp. 46-61, Walter de Gruyter and Company.

[14] Hendrix, Gary G., 1975. "Extending the Utility of Semantic Networks through Partitioning", *Advance Papers, International Joint Conference on Artificial Intelligence*, Tbilisi, Georgian SSR, pp. 115-121, September 1975.

[15] Hobbs, Jerry R., 1983. "An Improper Treatment of Quantification in Ordinary English", *Proceedings of the 21st Annual Meeting, Association for Computational Linguistics*, pp. 57-63. Cambridge, Massachusetts, June 1983.

[16] Kamp, J. A. W., 1975. "Two Theories about Adjectives", in E. Keenan, ed., *Formal Semantics of Natural Language*, pp. 123-155, Cambridge University Press, Cambridge, England.

[17] Kaplan, David, 1977. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals", unpublished manuscript.

[18] Konolige, Kurt, 1985. "Belief and Incompleteness", in J. Hobbs and R. Moore, eds., *Formal Theories of the Commonsense World*, pp. 359-403, Ablex Publishing Corp., Norwood, New Jersey.

[19] Kripke, Saul A., 1972. "Naming and Necessity", in D. Davidson and G. Harman, eds., *Semantics of Natural Language*, pp. 253-355, D. Reidel Publishing Co., Dordrecht, Holland.

[20] Levesque, Hector J., 1984. "A Logic of Implicit and Explicit Belief", *Proceedings of the National Conference on Artificial Intelligence*, pp. 198-202, Austin, Texas.

[21] Lewis, David K., 1969. *Convention: A Philosophical Study*, Harvard University Press, Cambridge, Massachusetts.

[22] McCarthy, John, 1975. "An Axiomatization of Knowledge and the Example of the Wise Man Puzzle", unpublished manuscript.

[23] McCarthy, John, 1977. "Epistemological Problems of Artificial Intellegence", *Proceedings, International Joint Conference on Artificial Intelligence*, pp. 1038-1044, Cambridge, Massachusetts, August 1977.

[24] McCarthy, John, 1979. "First Order Theories of Individual Concepts and Propositions", in J. Hayes, D. Michie, and L. Mikulich, eds., *Machine Intelligence 9*, pp. 129-147, John Wiley & Sons, New York.

[25] Meinong, Alexius, 1904. "Kinds of Being", in R. M. Chisolm, ed., Realism and the Background of Phemenology, The Free Press.

[26] Minsky, Marvin, 1975. "A Framework for Representing Knowledge", in P. Winston, ed., The Psychology of Computer Vision, pp. 211-277, McGraw-Hill, New York.

[27] Montague, Richard, 1974. "The Proper Treatment of Quantification in Ordinary English", in R. H. Thomason, ed., *Formal Philosophy: Selected Papers of Richard Montague*, pp. 247-270, Yale University Press, New Haven, Connecticut.

[28] Moore, Robert C., 1980. "Reasoning about Knowledge and Action", SRI International Artificial Intelligence Center Technical Report 191, October 1980.

[29] Moore, Robert C., and Gary G. Hendrix, 1982. "Computational Models of Belief and the Semantics of Belief Sentences", in S. Peters and E. Saarinen, eds., *Processes, Beliefs, and Questions*, pp. 107-127, D. Reidel Publishing Company, Dordrecht, Netherlands.

[30] Perry, John, 1979. "The Problem of the Essential Indexical", *NÔUS*, Vol. 13, pp. 3-21.

[31] Quillian, M. Ross, 1968. "Semantic Memory", in M. Minsky, ed., *Semantic Information Processing*, pp. 227-270, MIT Press, Cambridge, Massachusetts.

[32] Quine, Willard V., 1953. "On What There Is", in *From a Logical Point of View*, pp. 1-19, Harvard University Press, Cambridge, Massachusetts.

[33] Quine, Willard V., 1956. "Quantifiers and Propositional Attitudes", *Journal of Philosophy*, Vol. 53.

[34] Reichenbach, Hans, 1947. *Elements of Symbolic Logic*, The MacMillan Company.

[35] Russell, Bertrand, 1905. "On Denoting", in H. Feigel and W. Sellars, eds., *Readings in Philosophical Analysis*, pp. 85-102, Appleton-Century-Crofts, Inc., New York, 1949.

[36] Schiffer, Stephen R., 1972. *Meaning*, Oxford University Press, Oxford.

[37] Schmolze, James G., and Robert J. Brachman, 1982. "Summary of the KL-ONE Language", in *Proceedings, 1981 KL-ONE Workshop*, pp. 231-257, Fairchild Laboratory for Artificial Intelligence Research, Palo Alto, California.

[38] Searle, John R., 1969. *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge, England.

[39] Simmons, Robert F., 1973. "Semantic Networks: Their Computation and Use for Understanding English Sentences", in R. Schank and K. Colby, eds., *Computer Models of Thought and Language*, pp. 63-113, W. H. Freeman, San Francisco.

[40] Zalta, Edward N., 1983. *Abstract Objects: An Introduction to an Axiomatic Metaphysics*, D. Reidel Publishing Company, Dordrecht, Netherlands.