1-1-1981

An introduction to a theory of abstract objects.

Edward N. Zalta
University of Massachusetts Amherst

AN INTRODUCTION TO A THEORY
OF ABSTRACT OBJECTS

A Dissertation Presented

By

EDWARD NOURI ZALTA

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1981

Philosophy

E R R A T A

The corrected line appears after the page and line indication.

p. 19, l. 14: D_1 being abstract ("A!") $=_{dfn}$ $[\lambda x \sim E!x]$

p. 24, l. 16: $\mathcal{P}(\mathcal{D}^n)$ ("the power set of $\mathcal{D}^n$"), i.e., ext_R: $R_n \rightarrow \mathcal{P}(\mathcal{D}^n)$.

We call $ext_R(r^n)$ the exemplification exten-/

p. 25, l. 4,5: $PLUG_1$ maps $(R_2 \cup R_3 \cup \ldots) \times \mathcal{D}$ into $(R_1 \cup R_2 \cup \ldots)$. $PLUG_j$, for each j, j>1, maps $(R_j \cup R_{j+1} \cup \ldots) \times \mathcal{D}$ into $(R_{j-1} \cup R_j \cup \ldots)$. $PLUG_i$ is subject to the condition:/

p. 25, l. 17,18: $PROJ_1$ maps $(R_2 \cup R_3 \cup \ldots)$ into $(R_1 \cup R_2 \cup \ldots)$. $PROJ_j$, for each j, j>1, maps $(R_j \cup R_{j+1} \cup \ldots)$ into $(R_{j-1} \cup R_j \cup \ldots)$. $PROJ_i$ is subject to the condition:/

p. 27, l. 18: function, ext_A, which maps each $r^1 \in R_1$ into $\mathcal{P}(\mathcal{D})$, i.e., ext_A: $R_1 \rightarrow \mathcal{P}(\mathcal{D})$.

p. 35, l. 11: λ-EQUIVALENCE: for any propositional formula ϕ, the/

p. 68, l. 16,17: The behavior of our new descriptions is governed by the following proper axiom schema ("DESCRIPTIONS"):/

p. 76, l. 19-22: The first two clauses in the definition of satisfaction must be redefined. For example, clause one should read:

If $\phi = \rho^n o_1 \ldots o_n$, then $\mathfrak{f}$ satisfies ϕ iff $(\exists o_1)\ldots(\exists o_n)(\exists r^n)$ $(o_1 = d_{I,\mathfrak{f}}(o_1) \& \ldots \& o_n = d_{I,\mathfrak{f}}(o_n) \& r^n = d_{I,\mathfrak{f}}(\rho^n) \& \langle o_1, \ldots, o_n \rangle \in ext_R(r^n))$

Also, we need to restrict λ-EQUIVALENCE and RELATIONS - you can't use formulas ϕ which contain descriptions./

p. 80, l. 29: D_1 being abstract ("A!") $=_{dfn}$ $[\lambda x\ \Box \sim E!x]$

p. 83, l. 2: $\mathcal{R}_n \times \mathcal{W}$ into $P(\mathcal{D}^n)$, where n≥1. and which maps $\mathcal{R}_0 \times \mathcal{W}$ into {T,F}. $ext_{\mathcal{W}}(r^n)$/

p. 86, l. 10: member of I, ext_A, is a function which maps $\mathcal{R}_1$ into $P(\mathcal{D})$. $ext_A(r^1)$ is/

p. 91, l. 3-6: If $\phi = \rho^n o_1 \ldots o_n$, f satisfies ϕ with respect to w iff $(\exists o_1)\ldots(\exists o_n)(\exists r^n)(o_1 = d_{I,f}(o_1)$ &...& $o_n = d_{I,f}(o_n)$ & $r^n = d_{I,f}(\rho^n)$ & $\langle o_1,\ldots,o_n\rangle \in ext_w(r^n))$

If $\phi = o\rho^1$, f satisfies ϕ with respect to w iff $(\exists o)(\exists r)$ $(o = d_{I,f}(o)$ & $r^1 = d_{I,f}(\rho^1)$ & $o \in ext_A(r^1))$

p. 92, l. 23,24: (This is a proper axiom)

p. 92, l. 25: λ-EQUIVALENCE: for any propositional formula ϕ which has no descriptions, the/

p. 93, l. 23: isn't free and ϕ has no descriptions, the universal closure of the following is/

p. 94, l. 17: PROPOSITIONS: where ϕ is any propositional formula where F^0 isn't free and which has no descriptions, the/

p. 112, l. 18: D_{22} x is a correlate of z at w ("Cor(x,z,w)") $=_{df}$

p. 112, l. 20: That is, x is a correlate of z at w iff x exemplifies F at w iff z/

p. 129, l. 8: (4) Necessarily, the teacher of Alexander is a teacher

p. 142, l. 26: D_1 being abstract$^{t/p}$ ("$A!^{t/p}$") $=_{df}$ $[\lambda x^t \square{\sim}E!^{t/p}x]$

p. 148, l. 16: (a) ext_w: $\mathcal{R}_{(t_1,\ldots,t_n)/p} \times \mathcal{W} \rightarrow P(\mathcal{D}_{t_1} \cup \mathcal{D}_{t_2} \cup \ldots \cup \mathcal{D}_{t_n})$

p. 152, l. 19: type t, ext_A maps $\mathcal{R}_{t/p}$ into $P(\mathcal{D}_t)$. ext_A assigns each higher order/

p. 158, l. 11-14: (Redefine these clauses as in correction for page 91 above)

p. 159, l. 13-15: (DESCRIPTIONS is a proper axiom)

p. 159, l. 16,17: λ-EQUIVALENCE: for any propositional formula which has no descriptions, the universal closure of the following is an axiom:/

AN INTRODUCTION TO A THEORY

OF ABSTRACT OBJECTS

A Dissertation Presented

By

EDWARD NOURI ZALTA

Approved as to style and content by:

Terence Parsons, Chairperson of Committee

Edmund Gettier, Member

Gary Hardegree, Member

Barbara Partee, Outside Member

Edmund Gettier, Acting Department Head
Department of Philosophy

PREFACE

Alexius Meinong and his student, Ernst Mally, were the two most influential members of a school of philosophers and psychologists working in Graz in the early part of the twentieth century. They investigated psychological, abstract, and nonexistent objects--a realm of objects which weren't being taken seriously by Anglo-American philosophers in the Russell tradition. I first took the views of Meinong and Mally seriously in a course on metaphysics taught by Terence Parsons in the Fall of '78. Parsons had developed an axiomatic version of Meinong's naive theory of objects. The theory with which I was confronted in the penultimate draft of Parsons' book, Nonexistent Objects, had a profound impact upon me. I was convinced that Parsons' work would serve as a new paradigm for philosophical investigations.

While canvassing the literature during my research for Parsons' course, I discovered, indirectly, that Mally, who had originated the nuclear/extranuclear distinction among properties (a seminal distinction adopted by both Meinong and Parsons), had had another idea which could be developed into an alternative axiomatic theory. This discovery was the result of reading both a brief description of Mally's theory in J.N. Findlay's book, Meinong's Theory of Objects and Values (pp. 110-112) and what appeared to be an attempt to reconstruct Mally's theory by W. Rapaport in his paper "Meinongian Theories and a Russellian Paradox." With the logical devices Parsons had used in his book, plus others that I had learned from my colleagues, I began elaborating and applying the alternative theory in a series of unpublished papers written between November 1978 and August 1979. These papers were then assimilated into the first draft of this work in Fall 1979.

The entire project could not have been carried off without the inspiration and aid of teachers and colleagues. Throughout the project, Parsons served as a sharp critic. Our conversations every couple of weeks always left me with an idea for improving what I had done or with an outline of a problem which had to be tackled and solved. It is to his credit that he was such a great help despite the fact that our theories offered rival explanations to certain pieces of data.

Barbara Hall Partee graciously gave of her time in weekly discussions during the writing of the first draft. Her enthusiasm, encouragement, and suggestions were invaluable.

My colleague, Alan McMichael, also deserves special mention. Besides teaching me the techniques of algebraic semantics, and discovering a paradox within the theory, McMichael served as my first critic. Whenever I discovered a new application of the theory or got stuck on a point of logic, I frequently presented it to Alan. His criticisms and suggestions helped me to sharpen up many of the intricate details.

I'd also like to thank Mark Aronszajn, Blake Barley, Cynthia Freeland, Edmund Gettier, Gary Hardegree, Herbert Heidelberger, Larry Hohm, Michael Jubien, and Robert Sleigh. Spirited discussions with these individuals forced me to think deeply about a variety of

issues. They were some of the many people who helped to make the philosophy department here such a stimulating one.

Finally, thanks goes to Nancy Scott for her dedication in typing unfriendly looking manuscripts

Ed Zalta
October, 1980

University of Massachusetts/Amherst

ABSTRACT

An Introduction to a Theory of Abstract Objects

(February 1981)

Edward Nouri Zalta, B.A., Rice University,

Ph.D., University of Massachusetts

Directed by: Professor Terence Parsons

An axiomatic theory of abstract objects is developed and used to construct models of Plato's Forms, Leibniz's Monads, Possible Worlds, Frege's Senses, stories, and fictional characters. The theory takes six primitive metaphysical notions: <u>object</u> ($x,y,\ldots$); <u>n-place relations</u> ($F^n,G^n,\ldots$); $x_1,\ldots,x_n$ <u>exemplify</u> F^n ("$F^n x_1 \ldots x_n$"); x <u>exists</u> ("$E!x$"); it is <u>necessary</u> that ϕ ("$\Box\phi$"); and x <u>encodes</u> F^1 ("xF^1"). <u>Properties</u> and <u>propositions</u> are one place and zero place relations, respectively. <u>Abstract</u> objects ("$A!x$") are objects which necessarily fail to exist ("$\Box{\sim}E!x$"). The two most important proper axioms are that (1) no possibly existing object encodes any properties ($(x)(\Diamond E!x \rightarrow {\sim}(\exists F)xF)$), and (2) for every expressible condition on properties, there is an abstract object which encodes just the properties satisfying the condition ($(\exists x)(A!x \;\&\; (F)(xF \equiv \phi))$, where ϕ has no free x's). Semantically, an abstract object encodes a property iff the property is an element of the set of properties correlated with the object. Two abstract objects will be <u>identical</u> just in case they encode the same properties. Abstract objects may exemplify properties as

well--being non-red, not having spatial location, etc.

The models of Forms, Monads, and Possible Worlds consist in: (a) identifying these philosophical entities as different species of abstract objects through definitions of the object language, and (b) either showing that there are consequences of the theory which are reasonable facsimiles of assertions made by important philosphers (for example, Plato's One Over the Many Principle and Leibniz's principle that monads mirror their worlds are theorems) or showing that there are many interesting theorems which capture our own intuitions about these entities (for example, we prove that every world is maximal, every world is consistent, there is a unique actual world, and that a proposition is true at a world iff everything exemplifies at that world being such that that proposition is true). The model for Frege's Senses consists in: (a) developing a language in which the abstract objects generated by the theory serve as the senses of names and descriptions, and (b) showing that the data (propositional attitude triads, identity statements, etc.) may be consistently translated into the language preserving truth. Finally, the model for fictional characters consists in: (a) assigning abstract objects to serve as the denotations of names of stories and characters, and (b) showing that such an assignment allows us to understand how the intuitively true sentences about stories and characters have the truth value that they do.

The material is structured as follows. In Chapters I, III, and V, the elementary, modal, and typed theories of abstract objects are developed, respectively. In Chapters II, IV, and VI, these theories are applied, respectively. Each of the theoretical chapters is subdivided as follows: §1--The Language; §2--The Semantics; §3--The Logic; and §4--The Proper Axioms. Thus, the proper axioms of the theory are couched in a precisely defined and interpreted language, and the logic (logical axioms and rules of inference) allow us to prove the consequences of the theory.

TABLE OF CONTENTS

PREFACE iv

ABSTRACT vii

INTRODUCTION 1

§1. Theory, Data, and Explanation 1
§2. The Origins of the Theory 6
Endnotes 12

I. ELEMENTARY OBJECT THEORY 15

§1. The Language 16
§2. The Semantics 23
§3. The Logic 34
§4. The Proper Axioms 39
§5. Avoiding Known Paradoxes 44
§6. An Auxiliary Hypothesis 48
Endnotes 51

II. APPLICATIONS OF THE ELEMENTARY THEORY 59

§1. Modelling Plato's Forms 59
§2. Modelling the Round Square, etc. 66
§3. The Problem of Existence 71
Endnotes 76

III. THE MODAL THEORY OF ABSTRACT OBJECTS
(WITH PROPOSITIONS) 79

§1. The Language 79
§2. The Semantics 82
§3. The Logic 91
§4. The Proper Axioms 95
Endnotes 99

IV. THE APPLICATIONS OF THE MODAL THEORY 103

§1. Truth 103
§2. Modelling Possible Worlds 105
§3. Modelling Leibniz's Monads 112
§4. Modelling Stories and Native Characters 116
§5. Modelling the Fountain of Youth 127
Endnotes 132

V. THE TYPED THEORY OF ABSTRACT OBJECTS 137

§1. The Language 140
§2. The Semantics 147
§3. The Logic 159
§4. The Proper Axioms 161
Endnotes 163

VI. APPLICATIONS OF THE TYPED THEORY 166

§1. Modelling Frege's Senses (I) 166
§2. Modelling Frege's Senses (II) 184
§3. Modelling Impossible and Fictional Relations . . . 191
Endnotes 194

. . . .

BIBLIOGRAPHY 198

APPENDICES

A. CLARK'S PARADOX 204

B. McMICHAEL'S PARADOX 206

C. INCORPORATING NON-RIGID DESIGNATORS 207

D. A PROOF 210

E. MODELLING NOTIONS 211

F. AN OBJECTION AND A REPLY 217

INTRODUCTION

§1. Theory, Data, and Explanation

In this book, we shall discuss the things there are. In addition to existing objects (like you, me, my desk, sub-atomic particles, etc.) and the properties and relations they exemplify, we theorize that there are abstract entities as well. Among the abstract entities, we find abstract objects, abstract properties, and abstract relations. For the major part of this book, the theory of abstract objects is developed and applied (Chapters I-IV). At the end, the theory of abstract properties and relations is developed and applied, using the resources of a new kind of type theory (Chapters V-VI). We refer to the overall theory as the theory of abstract objects, and the first principles of this theory tell us not only the conditions under which there are particular such objects, but also the conditions under which any two of them are identical. The first principles serve as the cornerstone of the theory and a good reason for accepting them is that they will help us to construct explanations of data which have yet to be assimilated in a natural way into the current philosophical paradigm.

The current philosophical paradigm is a metaphysical theory that many philosophers have attributed to Russell. It can be stated roughly as follows: existing (actual, or real) objects, first level properties (and relations) exemplified by these objects, second level properties (and relations) exemplified by these first level properties, and so forth, are the only things that there are.[1] This metaphysical theory gives us a background ontology and philosophers working in the tradition usually supplement the theory with certain views about the history of philosophy and about language. A supplementary view about the history of philosophy is that if earlier philosophers who postulated theoretical entities were describing anything at all, they must have been describing entities (i.e., objects, properties, relations) which can be found in the background ontology. The supplementary view about language is a closely associated one: if the terms of a natural language denote anything at all, they must denote entities found in the background ontology. These supplementary views help to organize the data to be explained and help to ground the methodological principles of investigation and explanation that are part of this philosophical tradition. For example, here are two methodological principles grounded by these supplementary views:

A. Try to construe the philosophical discussions of earlier philosophers who described theoretical entities (like Forms, Monads, Possible Worlds, Senses) as discussions about existing objects or properties and relations.
B. Show that there are no true sentences of natural language which contain terms that denote entities not found in the ontology.

In what follows, we shall not argue that there are data which are not, or cannot, be accounted for by the current paradigm. Enough *ad hoc* auxiliary hypotheses could be added to any theory to enable it to handle the data. Instead, we shall try to establish one thesis--namely, that the alternative metaphysical theory developed here does help us to explain certain pieces of data. Then, any philosopher or linguist who agrees that the data we've chosen to explain are important and currently lack natural explanations should either consider our theory as a serious alternative or be prompted to find natural explanations from within the current paradigm. Let's then describe the data we shall try to explain.

The data consists of true sentences which we suppose to be philosophically important, for one reason or another. Some of these true sentences are *a priori*. For example, Plato and Leibniz developed *a priori* hypotheses about Forms and Monads, respectively:

(i) If there are two distinct F-things, then there is a Form of F in which they both participate (Plato, Parmenides, 132a)

(ii) . . . each simple substance (i.e., monad) . . . is a perpetual living mirror of the universe (Leibniz, Monadology, §56)

It turns out that certain abstract objects display features resembling those of Platonic Forms, while others display features resembling Leibnizian Monads. By saying, for example, that certain abstract objects display features resembling those of Platonic Forms, we mean three things: (1) a definition such as the following,

$$x \text{ is a Form } =_{dfn} \ldots x \ldots,$$

can be given using only the primitive and defined notions of the theory, (2) features of the Forms that Plato describes are definable in the theory as well, and (3) it follows from the first principles of the theory that the abstract objects which satisfy the definition of "Form" have the features Plato says *his* Forms are supposed to have. Consequently, we shall suppose that data like (i) and (ii) have been explained if there are reasonable facsimiles of them which turn out to be consequences of the theory.

There is another group of *a priori* truths which we take to be data. Here are some examples:

(iii) The round square is round.

(iv) The fountain of youth is a fountain.

(v) The set of all sets which aren't members of themselves is a set of all sets which aren't members of themselves.

We shall try to show that abstract objects can serve as the denotations of the descriptions in (iii)-(v). And as with all *a priori* data, we shall suppose that (iii)-(v) have been explained if we can deduce them as consequences of the first principles of our theory.

The *a posteriori* data also fall into two major groups. The first group consists of statements we ordinarily make about fictional characters, mythical figures, dream objects, and the like. Here are some examples:

(vi) Santa Claus doesn't exist.

(vii) Stephan Dedalus is a fictional character.

(viii) In the myth, Achilles fought Hector.

(ix) Some Greeks worshipped Dionysus.

(x) Ponce de Leon searched for the fountain of youth.

(xi) Franz Kafka wrote about Gregor Samsa.

To construct a *prima facie* case for thinking that fictional characters, mythical figures, etc., just are abstract objects, we shall focus on a formal language that we develop in Chapter III. Certain definitions tell us the conditions under which a given sentence of the formal language is true. We then translate (vi)-(xi) into our formal language, using names and descriptions in the language which denote abstract objects (as well as existing ones). Thus, an explanation of (vi)-(xi), and others like them, consists in showing that they can be translated into sentences which preserve their intuitive truth value.

The other group of *a posteriori* data contains triads of sentences. These are sentences which involve verbs of propositional attitude and the "is" of identity:

(xii) S believes that Socrates taught Plato.
(xiii) S doesn't believe that the son of Phaenarete taught Plato.
(xiv) Socrates is the son of Phaenarete.

(xv) S believes that Woodie is a woodchuck.
(xvi) S doesn't believe that Woodie is a groundhog.
(xvii) Being a woodchuck just is being a groundhog.

We follow Frege in supposing that the English terms inside propositional attitude contexts do not have their ordinary denotations and that they denote their senses instead. This was Frege's explanation of why (xii) and (xiv) don't imply the negation of (xiii), and why (xv) and (xvii) don't imply the negation of (xvi). However, the senses of terms denoting objects will be construed as abstract objects and the senses of terms denoting properties (relations) will be construed as abstract properties (relations). To do this, we again focus on a formal language developed in Chapter V. (xii)-(xvii) are translated into our language, using names and descriptions of the language which denote abstract entities that serve as the senses of the English terms. Thus, our Fregean explanation of the consistency of each triad lies in showing that the sentences which translate the members of a given triad are consistent.

These, then, will be the kinds of data and explanation which shall occupy our attention. But in the course of setting up our theory, many other philosophical issues will be confronted. For example, we will end up developing a full-fledged theory of relations (where properties and propositions turn out to be one-place and zero-place relations, respectively). This includes a definition which tells us when any two relations (properties, propositions) are in fact the same. Semantics for our formal languages are developed in which we may consistently suppose that logically equivalent relations are distinct. The resulting metaphysical system should be attractive not only because it might handle important kinds of data which seem problematic for the current tradition, but also because it exhibits many interesting and philosophically satisfying qualities in its own right.

§2. The Origins of the Theory

The theory we develop has its origins directly in the naive theory of nonexistent objects which Meinong and Mally investigated at the turn of the century. A very simple statement of the theory upon which Meinong seemed to be relying in his early work is the following, which we call Naive Object Theory:

(NOT) For every set of properties, there is an object which exemplifies just the members of the set.

To make this theory precise, we could capture it in a second order predicate calculus as follows:

(NOT') $(\exists x)(F)(Fx \equiv \phi)$, where ϕ has no free x's.

(NOT') yields an object with just two properties, roundness ("R") and squareness ("S") as follows:

$(\exists x)(F)(Fx \equiv F{=}R \vee F{=}S)$

(NOT') also gives us an object which exemplifies just the properties that Socrates ("s") exemplifies:

$(\exists x)(F)(Fx \equiv Fs)$

By Leibniz's Law, it follows that this object just is Socrates.

Unfortunately, we can derive both falsehoods and contradictions from (NOT'). Consider the following instance, where "E!" stands for existence, "G" stands for goldenness, and "M" stands for mountainhood:

$(\exists x)(F)(Fx \equiv F{=}E! \vee F{=}G \vee F{=}M)$

This asserts the falsehood that there is an existing golden mountain. (NOT') also implies the falsehood that Russell never thought about the round square (as generated above), since this latter object (provably) doesn't exemplify the property of being thought about by Russell (on the assumption that the property of being thought about by Russell is distinct from both the property of being round and the property of being square).

A contradiction that (NOT') entails comes about as follows. Let "$[\lambda x\ Rx \;\&\; {\sim}Rx]$" denote the property of being-red-and-not-being-red. We then have:

$(\exists x)(F)(Fx \equiv F{=}[\lambda x\ Rx \;\&\; {\sim}Rx])$

Call an arbitrary such object a_0. It follows that $[\lambda x\ Rx \;\&\; {\sim}Rx]a_0$, i.e., a_0 exemplifies the property of being-red-and-not-being-red. But by ordinary λ-conversion, it follows that Ra_0 and ${\sim}Ra_0$. So (NOT') is inconsistent.

One suggestion by Mally to refine (NOT) was to distinguish two general types of properties--nuclear and extranuclear.[2] The nuclear properties an object has are more central to its being and identity than are its extranuclear properties. Terence Parsons follows up on this suggestion.[3] He adds this distinction as one new primitive to a standard second order predicate calculus. He develops the theory and logic associated with nuclear and extranuclear relations. He restricts the range of the property quantifier in (NOT') so that it ranges just over nuclear properties. His theory is therefore based on the following three principles, where "F^n" ranges over extranuclear n-place relations and "f^n" ranges over nuclear n-place relations:[4]

(I) $(\exists x)(f^1)(f^1 x \equiv \phi)$, where ϕ has no free x's.
For every set of nuclear properties, there is an object which exemplifies just the members of the set.

(II) $x=y \equiv (f^1)(f^1 x \equiv f^1 y)$
Two objects are identical iff they exemplify the same nuclear properties.

(III) $(\exists F^n)(x_1)\ldots(x_n)(Fx_1\ldots x_n \equiv \phi)$, where ϕ has no free F^n's.
For every set of objects, there is an extranuclear property which just those objects exemplify.

So on Parsons' theory, there is one kind of object and two kinds of properties. Also, each extranuclear property is associated with some (not necessarily distinct) nuclear property which serves as its "watered-down version."

Consequently, Parsons avoids generating the above falsehoods and inconsistencies. The existent golden mountain exemplifies watered-down existence, not the more important kind of extranuclear existence. The object which exemplifies just nuclear roundness and nuclear squareness could, and in fact did, exemplify the extranuclear property of being thought about by Russell. Also, "[λx Rx & ~Rx]" denotes an extranuclear property. So his theory doesn't imply that there is an object which exemplifies being-red-and-not-being-red. And with these obstacles out of the way, Parsons finds interesting applications for his theory. In particular, he models fictional characters (and the like), Leibnizian Monads, and suggests how to model Plato's Forms.[5] These models served as prototypes for the models we have constructed in our alternative object theory.

Our theory of abstract objects is based on a different suggestion of Mally's, however. He distinguished two relationships which relate objects to their properties. On Mally's view, properties can *determine* objects which don't in turn *satisfy* the properties.[6] For example, the properties roundness and squareness can determine an abstract object which satisfies neither roundness nor squareness. The properties of existence, goldenness, and mountainhood can determine an abstract object which doesn't satisfy any of these properties. The properties which determine an abstract object are central to its identity. For a recent attempt to reconstruct Mally's theory, see W. Rapaport's discussion in "Meinongian Theories and a Russellian Paradox."[7]

In what follows, we construct languages capable of representing the distinction between satisfying and being determined by a property. However, we shall employ different terminology. We shall say that an object <u>exemplifies</u> a property instead of satisfying it. We shall say that an object <u>encodes</u> a property instead of saying that the object is determined by the property. The distinction between exemplifying and encoding a property is a primitive one and will be represented by a distinction in atomic formulas of the languages we construct. The primitive metaphysical notions that we shall need in order to state the first principles of the theory are:

<u>object</u> $(x,y,z,\ldots)$
<u>n-place relation</u> $(F^n, G^n, H^n, \ldots)$
$x_1,\ldots,x_n$ <u>exemplify</u> F^n ("$F^n x_1 \ldots x_n$")
x <u>encodes</u> F^1 ("xF^1")
x <u>exists</u> ("$E!x$")
<u>necessarily</u>, ϕ ("$\Box\phi$")

Using these basic notions, we define a <u>property</u> to be a one-place relation and define an <u>abstract</u> ("$A!x$") object to be an object which fails to exemplify existence (in the modal theory, abstract objects necessarily fail to exemplify existence).[8] We also say that two objects are <u>identical</u>$_E$ ("$x=_E y$") iff they both exemplify existence and exemplify the same properties (in the modal theory, two objects are identical$_E$ iff they both possibly exemplify existence and exemplify the same properties).

In the context of the work of Meinong, Mally, Parsons, Findlay, and Rapaport, the following two principles seem to be the cornerstones for a theory of abstract objects:[9]

(I) For every expressible set of properties, there is an abstract object which encodes just the properties in the set
$(\exists x)(A!x \;\&\; (F)(xF \equiv \phi))$, where ϕ has no free x's.

(II) Two objects are identical iff they are identical$_E$ or they are both abstract and encode the same properties
$x=y \equiv x=_E y \vee (A!x \;\&\; A!y \;\&\; (F)(xF \equiv yF))$

Principle (I) gives us "being" conditions for abstract objects. Intuitively, to say that x encodes F^1 is to say that F^1 is an element of the set of properties corresponding to x. Principle (II) gives us identity conditions for objects in general. On our theory, in contrast with Parsons, there are two kinds of objects and one kind of property.

Given these two principles, we will find an abstract object which encodes just roundness and squareness $((\exists x)(A!x \; \& \; (F)(xF \equiv F{=}R \vee F{=}S)))$. This object won't exemplify these properties, but may have exemplified the property of being thought about by Russell. We also find an abstract object which encodes just existence, goldenness, and mountainhood $((\exists x)(A!x \; \& \; (F)(xF \equiv F{=}E! \vee F{=}G \vee F{=}M)))$. By definition, this abstract object fails to exemplify existence. Also, we will find an abstract object which encodes being-red-and-not-red $((\exists x)(A!x \; \& \; (F)(xF \equiv F{=}[\lambda x\; Rx \; \& \sim Rx])))$. But we won't get the contradiction discussed above because this object provably doesn't exemplify this property.

In Chapter I, we shall couch principles (I) and (II) as an axiom schema and definition, respectively, of an interpreted formal language. We also axiomatize the other logical and non-logical principles which round out the theory. This should make the details of an ontology rich with abstract objects sharp and accurate. Once our background ontology is set, we too shall adopt corresponding supplementary views about the history of philosophy and the philosophy of language. Our methods should therefore be familiar to most analytic philosophers.

INTRODUCTION ENDNOTES

[1]I shall not attribute this theory to Russell, though I will call this view "Russellian" because so many philosophers seem to make the attribution. Russell explicitly maintained ([1918], p. 269) that the ordinary things we speak of as existing (you, my desk, sub-atomic particles) are logical fictions. For Russell, there are no such objects. Moreover, some objects that there are, are ones we would ordinarily consider to be nonexistent, for example, phantoms, hallucinations, and their constituents ([1918], pp. 274-276).

Today, many philosophers would supplement the "Russellian" view with either propositions or possible worlds (I'd like to thank Mark Aronszajn for helping me to get straight about these details).

[2]See J.N. Findlay [1933], p. 176. Findlay references Meinong [1915], pp. 175-177.

[3]There are a host of good papers by Parsons on the subject: [1974], [1975], [1978], [1979b], and [1979c]. However, the most important statement of his theory is in his book Nonexistent Objects [1980].

[4]Parsons [1980], Ch. IV.

[5]Parsons [1980], Chs. VII, VIII.

[6]Findlay describes Mally's theory of determinates in his [1933], pp. 110-112 and pp. 183-184. He cites Mally [1912], pp. 64, 76.

[7]Rapaport [1978], pp. 153-180. For some reason, Rapaport doesn't attribute the theory he is working on to Mally. He calls it a reconstruction of Meinong's theory. His Meinongian objects clearly seem to be Mally's determinations. Although Mally uses the word "determiniert," he also uses "Konstitutiven" ([1912], p. 64). Compare Rapaport's use of "being constituted by."

[8]In what follows, I use the word "abstract" purely as a piece of technical terminology. Also, I take the words "existing," "actual," and "real" to be synonymous.

We could make a long list of entities which, at some time or another, philosophers have supposed to be abstract. It is *not* to be presupposed that the set of abstract objects which we will investigate is (intended to be) identical with the set of objects which some other philosopher pretheoretically intuits to be abstract. Many philosophers have firm intuitions to the effect that certain objects are abstract. However, these intuitions are rarely supported by presenting precise conditions which tell when there are abstract objects or which tell us when any two abstract objects are identical.

There is both a perogative and an intellectual obligation to specify how one plans to use the word "abstract." This has been done informally with principles I and II which follow in the text, and will be done formally in Chapter I (§§ 1, 4). The notion we end up with may not correspond exactly with that of others, but at least it should be clear. And in the course of our investigations, we shall discover that certain objects that other philosophers have taken to be abstract are identifiable among our abstract objects.

[9]For a long time, I thought I had been the first to formulate these principles. But I've subsequently discovered that versions of these principles are embedded in Rapaport's dissertation [1976]. His thesis T7a (p. 190) is similar to principle I, and his identity conditions (p. 184) form part of principle II.

CHAPTER I

ELEMENTARY OBJECT THEORY

The full presentation of the elementary theory of abstract objects shall occupy the first four sections of this chapter. In each of these sections, we concentrate on the following major groups of definitions:

§1--The Language

§2--The Semantics

§3--The Logic

§4--The Proper Axioms

The proper axioms are stated in the language. Since the semantics contains a definition which tells us the conditions under which an arbitrary sentence of the language is true, we will know what is being asserted by our proper axioms. The logic we associate with the language allows us to prove the consequences of the proper axioms.

In the course of the definitions which follow, we frequently provide examples and make extended remarks to explain and motivate unusual features. In the remarks, we frequently define (with the help of underlines and lists) certain syntactic or semantic concepts which will help us to single out classes of expressions or entities which have certain properties. We use quotation marks to mention expressions of the language. We generally omit these standard devices (quotation marks, corner marks) for mentioning and describing pieces of language when the intent is clear. We use quotation marks inside parentheses (". . ."), to give readings and/or abbreviations of formulas. All definitions of the object language appear with the label "D_n."

With the exception of "λ" and "ι," we always use lower case Greek letters as variables ranging over classes of expressions of the object language. In particular, we always use:

ϕ, ψ, χ to range over formulas

o's to range over object terms

ρ^n's to range over relation terms

α, β, γ to range over all variables

τ's to range over all terms

ν's to range over object variables

π^n's to range over relation variables

μ, ξ, ζ to range over λ-expressions

Finally, we note that in most of Chapter I, we shall not give the intuitive readings in natural language of the formulas and complex terms of the object language. That's because our aim is to focus on the expressive capacity of a formal language, without prejudice as to how English sentences and terms are to be translated into the language. However, it will be useful to provide some examples in natural language, since this will help the reader to picture what the language and theory can, and ultimately will, say.

§1. The Language

We shall utilize a slightly modified second order language. One interesting modification is that new atomic formulas have been added--they express the fact that an object encodes a property. These new atomic formulas are called "encoding formulas," and whereas the ordinary "exemplification formulas" (which we shall still have around) have n object terms to the right of an n-place relation term, encoding formulas have a single object term to the left of a *one*-place relation term. These atomic encoding formulas can combine with other formulas to make molecular and quantified formulas. The complex formulas which result may be constructed solely out of atomic exemplification subformulas, solely out of atomic encoding subformulas, or may be of mixed construction (many of the interesting definitions, axioms, and theorems are mixed formulas).

A second modification of the standard second order language is that there will be *two* identity signs--one primitive and one defined. "$=_E$" shall denote the primitive relation of identity among existing objects. A proper axiom will guarantee that two objects exemplify the relation "$=_E$" denotes just in case they both exemplify existence and exemplify the same properties. However, a more universal identity among objects, and identity among relations and properties will be defined with the help of encoding formulas. We shall use the symbol "=," as distinct from "$=_E$," for these definitions (in the metalanguage, we shall distinguish "identity$_E$" from "identity").

These two modifications result in a language which has more expressive capacity than the normal second order language. We subdivide the definitions for our new language as follows:

A. Primitive symbols
B. Formulas and terms
C. Identity definitions

A. Primitive symbols. We have two kinds of primitive object terms: names and variables. Officially we use the subscripted letters a_1, a_2, a_3, ... as primitive object names, but unofficially, we use a, b, c, ... for convenience. Officially, we use the subscripted letters x_1, x_2, x_3, ... as primitive object variables, but unofficially we use x, y, z, ... There are also two kinds of primitive relation terms: names and variables. Officially, we use the superscripted and subscripted letters P^n_1, P^n_2, ... , $n \geq 1$, as primitive relation names (unofficially: P^n, Q^n, ...) and F^n_1, F^n_2, ..., $n \geq 1$, as primitive relation variables (unofficially: F^n, G^n, ...). E! is a distinguished one-place relation name; $=_E$ is a distinguished two-place relation name.

In addition we use two connectives, ~, and &; a quantifier: $\exists$; a lambda: λ; and we avail ourselves of parentheses and brackets to disambiguate.

B. Formulas and terms. We present a simultaneous inductive definition of (propositional) formula, object term, and n-place relation term. The definition contains six clauses:

1. All primitive object terms are object terms and all primitive n-place relation terms are n-place relation terms

2. Atomic exemplification: If ρ^n is any n-place relation term, and $o_1,\ldots,o_n$ are any object terms, $\rho^n o_1 \ldots o_n$ is a (propositional) formula (read: "$o_1,\ldots,o_n$ exemplify relation ρ^n")

3. Atomic encoding: If ρ^1 is any one-place relation term and o is any object term, $o\rho^1$ is a formula (read: "o encodes property ρ^1")

4. Molecular: If ϕ and ψ are any (propositional) formulas, then $(\sim\phi)$ and $(\phi \ \& \ \psi)$ are (propositional) formulas

5. Quantified: If ϕ is any (propositional) formula, and α is any (object) variable, then $(\exists\alpha)\phi$ is a (propositional) formula

6. Complex n-place relation terms: If ϕ is any propositional formula with n-free object variables $\nu_1,\ldots,\nu_n$, then $[\lambda\nu_1\ldots\nu_n \ \phi]$ is an n-place relation term

We rewrite atomic exemplification formulas of the form $=_E o_1 o_2$ as $o_1 =_E o_2$. We drop parentheses to facilitate reading complex formulas whenever there is little potential for ambiguity. We utilize the standard abbreviations: $(\phi \rightarrow \psi)$, $(\phi \equiv \psi)$, $(\phi \vee \psi)$, and $(\alpha)\phi$. And we define:

D_1 x is abstract ("A!x") $=_{dfn}$ ~E!x

Here then are some examples of formulas: P^3axb ("a, x, and b exemplify relation P^3"); aG ("a encodes property G"); $\sim(\exists x)(xQ \ \& \ Qx)$ ("no object both encodes and exemplifies Q"); $(x)(E!x \rightarrow \sim(\exists F)xF)$ ("every object which exemplifies existence fails to encode any properties"); and $(\exists x)(A!x \ \& \ (F)(xF \equiv Fa))$ ("some abstract object encodes exactly the properties a exemplifies").

By inserting all the parenthetical remarks when reading the above definition, we obtain a definition of <u>propositional formula</u>. In effect, a formula ϕ is propositional iff ϕ has no encoding subformulas and ϕ has no subformulas with quantifiers binding relation variables.[1] Only propositional formulas may occur in λ-expressions. λ-expressions allow us to name complex relations.[2] We read $[\lambda\nu_1\ldots\nu_n \ \phi]$ as "being objects $\nu_1,\ldots,\nu_n$ such that $\phi(\nu_1,\ldots,\nu_n)$," or as "being a first thing, second thing, ..., and n^{th}-thing such that ϕ."

For example: $[\lambda x\ {\sim}Rx]$ ("being an object x such that x fails to exemplify R"); $[\lambda x\ Px\ \&\ Qx]$ ("being an object x such that x exemplifies both P and Q"); $[\lambda x\ x{=}_E b]$ ("being identical$_E$ with b"); $[\lambda xy\ Px\ \&\ Syx]$ ("being objects x and y such that x exemplifies P and y bears S to x"); $[\lambda x\ (\exists y)Fxy]$ ("being an x such that x bears F to something"); $[\lambda xyz\ Gzx\ \&\ E!y]$ ("being a first, second, and third thing such that the third bears G to the first and the second exists").[3]

Since arbitrary formulas ϕ cannot appear after λ's, the following expressions are ill-formed: $[\lambda x\ xP]$, $[\lambda y\ yP\ \&\ Py]$, $[\lambda x\ (\exists F)Fx]$, and $[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]$. The first two are ill-formed because the formula after the λ has an encoding subformula; the third because the formula contains a quantifier binding a relation variable ("relation quantifier"); the fourth fails both "restrictions" on propositional formulas.

Although a more detailed discussion is reserved for §5, we indicate briefly why these restrictions appear. The "no encoding subformulas" restriction is essential--it serves to avoid paradoxes in the presence of the proper axioms (see §5). The "no relations quantifiers" may not be essential, however, it allows us to effect a huge simplification of the semantics. Since we shall not critically need to use λ-expressions with relation quantifiers in the applications of the theory, we choose not to complicate the semantics any further than necessary. We shall eliminate this latter restriction once we move to the typed theory of abstract objects (Chapter V). The semantics for the language which couches the typed theory more easily assimilates the interpretation of λ-expressions with "higher order" quantifiers. The net result of these restrictions is that no relation denoting expression not already found in the standard second order language can be constructed. Intuitively, this means we'll be working with familiar sorts of complex properties and relations.

These λ-expressions widen the possibilities for atomic and complex formulas: $[\lambda xy\ Px\ \&\ Qy]ab$ ("a and b exemplify being two objects x and y such that x exemplifies P and y exemplifies Q"); $x[\lambda y \sim Ry]$ ("x encodes failing to exemplify R"); $(\exists x)(A!x\ \&\ (F)(xF \equiv (\exists G^2)((Gab\ \&\ F=[\lambda y\ Gyb])\ v\ (Gba\ \&\ F=[\lambda y\ Gby]))))$ ("some abstract object encodes just the relational properties a exemplifies with respect to b").

Finally, we say that τ is a <u>term</u> iff τ is an object term or there is an n such that τ is an n-place relation term.

<u>C. Identity definitions</u>. The identity definitions are presented without the use of metavariables:

D_2 $x=y =_{dfn} x=_E y\ v\ (A!x\ \&\ A!y\ \&\ (F)(xF \equiv yF))$

In the presence of the proper axiom governing identity_E among objects (i.e., $x=_E y \equiv E!x\ \&\ E!y\ \&\ (F)(Fx \equiv Fy)$), our first definition tells us that objects x and y are identical iff either x and y both exist and exemplify the same properties or x and y are both abstract and encode the same properties. Note that $[\lambda xy\ x=y]$ is not well-formed--"x=y" abbreviates a formula with both encoding subformulas and relation quantifiers.

D_3 $F=G =_{dfn} (x)(xF \equiv xG)$

Two properties are identical iff they are encoded by the same objects. At first glance, it will not be apparent why this should be a good, insightful definition of property identity. We justify this definition in §4.

D_4 $F^n = G^n =_{dfn}$

$(x_1)\ldots(x_n)((x_1[\lambda y\ F^n yx_2\ldots x_n] \equiv x_1[\lambda y\ G^n yx_2\ldots x_n])$

$\&\ (x_1[\lambda y\ F^n x_2 yx_3\ldots x_n] \equiv x_1[\lambda y\ G^n x_2 yx_3\ldots x_n])$

$\&\ldots\&\ (x_1[\lambda y\ F^n x_2\ldots x_n y] \equiv x_1[\lambda y\ G^n x_2\ldots x_n y]))$

This definition may be read in the following, intuitive manner: relations F^n and G^n are identical iff the one place properties which result no matter how n-1 objects are "plugged" into them (provided F^n and G^n are plugged up in the same way) are identical (encoded by the same objects). This is meant to "generalize" the definition of property identity.

All of the definienda in these identity definitions have the form $\alpha=\beta$. The proper axiom governing substitutions of identicals operates on these defined forms (see §4).

In the definitions which follow in §§2, 3, and 4, it shall be useful to have precise definitions for certain syntactic concepts which up until now, we have used on an intuitive basis:[4] all and only formulas and terms are <u>well-formed expressions</u>. An occurrence of a variable α in a well-formed expression is <u>bound</u> (<u>free</u>) iff it lies (does not lie) within a formula of the form $(\exists\alpha)\phi$ or a term of the form $[\lambda\nu_1\ldots\alpha\ldots\nu_n\ \phi]$ within the expression. A variable is <u>free</u> (<u>bound</u>) iff it does (does not) have a free occurrence in that expression. A <u>sentence</u> is a formula having no free variables.

Furthermore, a term τ is said to be <u>substitutable for</u> a variable α in a formula ϕ iff for every variable β free in τ, no free occurrence of α in ϕ occurs either in a subformula of the form $(\exists\beta)\psi$ in ϕ or in a term $[\lambda\nu_1 \ldots \beta \ldots \nu_n\ \psi]$ in ϕ. Intuitively, if τ is substitutable for α in ϕ, no free variable β in τ gets "captured" when τ is substituted for α, by a quantifier or λ in ϕ which binds β. We write $\phi(\alpha_1, \ldots, \alpha_n)$ to designate a formula which may or may not have $\alpha_1, \ldots, \alpha_n$ occurring free. Finally, we write $\phi^{\tau_1, \ldots, \tau_n}_{\alpha_1, \ldots, \alpha_n}$ to designate the formula which results when, for each i, $1 \leq i \leq n$, τ_i is substituted for each free occurrence of α_i in ϕ.

§2. The Semantics

The definitions which help to determine the conditions under which the formulas of the language are true may be grouped as follows:

A. Interpretations

B. Assignments and Denotations

C. Satisfaction

D. Truth under an interpretation

In the definitions which follow, we use script letters as names and variables for sets, entities, and functions which are all peculiarly associated with the semantics.

A. Interpretations.[5] An interpretation, $\mathcal{I}$, of our language is any 6-tuple $\langle \mathcal{D}, \mathcal{R}, ext_{\mathcal{R}}, \mathcal{L}, ext_{\mathcal{A}}, \mathcal{F} \rangle$ which meets the conditions described in this subsection. The first two members, $\mathcal{D}$ and $\mathcal{R}$, must be non-empty classes--they provide entities for the primitive and complex names of the language to denote and they serve as the domains of quantification.

$\mathcal{D}$ is called the domain of objects, and we use o's as metalinguistic variables ranging over members of this domain. $\mathcal{R}$ is called the domain of relations and it is the union of a sequence of classes $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, ...; i.e., $\mathcal{R} = \bigcup_{n\geq 1} \mathcal{R}_n$. Each $\mathcal{R}_n$ is called the class of n-place relations. We use "r^n" as a metalinguistic variable ranging over the elements of $\mathcal{R}_n$. $\mathcal{R}$ must be closed under all of the logical functions specified in the fourth member of the interpretation ($\mathcal{L}$).

Intuitively, the third, fourth, and fifth members of any interpretation are functions (or classes of functions) which impose a certain structure on the elements of $\mathcal{D}$ and $\mathcal{R}$. We suppose that for each n-place relation in $\mathcal{R}_n$, there is a set of n-tuples drawn from $\mathcal{D}$ which serves as the exemplification extension ("extension$_{\mathcal{R}}$") of the relation. Each n-tuple of the set represents an ordered group of objects which exemplify (bear, stand in) the relation. The third member of an interpretation is therefore a function, $ext_{\mathcal{R}}$, which maps each $r \in \mathcal{R}_n$ into $\mathcal{D}^n$, i.e., $ext_{\mathcal{R}}\colon \mathcal{R}_n \rightarrow \mathcal{D}^n$. We call $ext_{\mathcal{R}}(r^n)$ the exemplification extension of r^n.

The fourth member of any interpretation, $\mathcal{L}$, is a class of logical functions which operate on the members of $\mathcal{R}_n$ and $\mathcal{D}$ to produce the complex relations which serve as the denotations for the λ-expressions. Each complex relation receives an exemplification extention$_{\mathcal{R}}$ which must mesh, in a natural way, with the extensions$_{\mathcal{R}}$ of the simpler relations it may have as parts.

There are six elements in $\mathcal{L}$--the first four are each families of indexed logical functions:[6] $PLUG_i$ ("i-plug"), $PROJ_i$ ("i-projection"), $CONV_{i,j}$ ("i,j-conversion"), and $REFL_{i,j}$ ("i,j-reflection"), where i and j are elements of the set of natural numbers. The other two members of L are particular functions $CONJ$ ("conjunction") and NEG ("negation"). These six elements of L work as follows:

(a) $PLUG_i$, for each i, $1 \leq i \leq n$, is a function mapping $(R_{i+1} \cup R_{i+2} \cup \ldots) \times D$ into $(R_i \cup R_{i+1} \cup \ldots)$ subject to the condition: $ext_R(PLUG_i(r^n, o)) = \{\langle o_1, \ldots, o_{i-1}, o_{i+1}, \ldots, o_n\rangle \mid \langle o_1, \ldots, o_{i-1}, o, o_{i+1}, \ldots, o_n\rangle \in ext_R(r^n)\}$

This basically says that the extension$_R$ of the new relation, $PLUG_i(r^n, o)$ ("the i^{th}-plugging of r^n by o"), includes just those n-1 tuples which result by deleting the object o from the i^{th} place of every n-tuple in the extension$_R$ of the original relation r^n which has o in its i^{th} place. This ensures, for example, that an object o_1 which falls in the extension$_R$ of the property $PLUG_2(r^2, o_2)$ is such that $\langle o_1, o_2\rangle$ is in the extension$_R$ of r^2. Also, if $\langle o_1, o_2\rangle$ is in the extension$_R$ of $PLUG_2(gives,\ o_3)$, then $\langle o_1, o_3, o_2\rangle$ is in the extension$_R$ of *gives*.

(b) $PROJ_i$, for each i, $1 \leq i \leq n$, is a function mapping $(R_{i+1} \cup R_{i+2} \cup \ldots)$ into $(R_i \cup R_{i+1} \cup \ldots)$ subject to the condition: $ext_R(PROJ_i(r^n)) = \{\langle o_1, \ldots, o_{i-1}, o_{i+1}, \ldots, o_n\rangle \mid (\exists o)(\langle o_1, \ldots, o_{i-1}, o, o_{i+1}, \ldots, o_n\rangle \in ext_R(r^n))\}$

This tells us that $PROJ_i(r^n)$ ("the i^{th}-projection of r^n") is an n-1 place relation which has in its extension all those n-1 tuples which result by deleting the i^{th} object of every n-tuple in the extension$_R$ of r^n. Intuitively, $PROJ_2(r^2)$ is a property an object o_1 has just in case o_1 bears r^2 to something. If o_1 exemplifies this property, then there is at least one object o_2 such that $\langle o_1, o_2\rangle \in ext_R(r^2)$.

(c) $CONV_{i,j}$, for each i,j, $1 \leq i < j \leq n$, is a function mapping $(R_j \cup R_{j+1} \cup \ldots)$ into $(R_j \cup R_{j+1} \cup \ldots)$ subject to the condition: $ext_R(CONV_{i,j}(r^n)) =$

$\{\langle o_1, \ldots, o_{i-1}, o_j, o_{i+1}, \ldots, o_{j-1}, o_i, o_{j+1}, \ldots, o_n \rangle |$

$\langle o_1, \ldots, o_i, \ldots, o_j, \ldots, o_n \rangle \in ext_R(r^n)\}$

This says that $CONV_{i,j}(r^n)$ ("the conversion of r^n about its i^{th} and j^{th} places") is an n-place relation which has in its $extension_R$ all those n-tuples which result by switching the i^{th} and j^{th} members of every n-tuple in $ext_R(r^n)$. So $\langle o_1, o_2 \rangle \in ext_R(CONV_{1,2}(r^2))$ iff $\langle o_2, o_1 \rangle \in ext_R(r^2)$.

(d) $REFL_{i,j}$, for each i,j, $1 \leq i < j \leq n$, is a function mapping $(R_j \cup R_{j+1} \cup \ldots)$ into $(R_{j-1} \cup R_j \cup \ldots)$ subject to the condition: $ext_R(REFL_{i,j}(r^n)) =$

$\{\langle o_1, \ldots, o_i, \ldots, o_{j-1}, o_{j+1}, \ldots, o_n \rangle |$

$\langle o_1, \ldots, o_i, \ldots, o_j, \ldots, o_n \rangle \in ext_R(r^n)$ and $o_i = o_j\}$

When given place numbers i and j, $REFL_{i,j}(r^n)$ ("the i,j^{th} reflection of r^n") is an n-1 place relation which has in its $extension_R$ all those n-1 tuples which result by deleting the j^{th} member from every n-tuple in the $extension_R$ of r^n which has identical i^{th} and j^{th} members. This ensures that any object o which falls in the $extension_R$ of $REFL_{1,2}(r^2)$ is such that o bears r^2 to itself, i.e., $\langle o_1, o_1 \rangle \in ext_R(r^2)$.

(e) $CONJ$ is a function from $(R_1 \cup R_2 \cup \ldots) \times (R_1 \cup R_2 \cup \ldots)$ into $(R_2 \cup R_3 \cup \ldots)$ subject to the condition: $ext_R(CONJ(r^n, s^m)) =$

$\{\langle o_1, \ldots, o_n, o'_1, \ldots, o'_m \rangle \mid \langle o_1, \ldots, o_n \rangle \in ext_R(r^n)$ and

$\langle o'_1, \ldots, o'_m \rangle \in ext_R(s^m)\}$

$CONJ$ maps any n-place relation r^n and m-place relation s^m to an n+m place relation which has in its extension$_R$ any n+m-tuple which has an n-tuple from $ext_R(r^n)$ as its first n members and an m-tuple from $ext_R(s^m)$ as its second m members. So $\langle o_1, o_2 \rangle \in ext_R(CONJ(r^1, s^1))$ iff $o_1 \in ext_R(r^1)$ and $o_2 \in ext_R(s^1)$.

(f) NEG is a function from $(R_1 \cup R_2 \cup \ldots)$ into $(R_1 \cup R_2 \cup \ldots)$ subject to the condition: $ext_R(NEG(r^n)) =$
$\{\langle o_1, \ldots, o_n \rangle \mid \langle o_1, \ldots, o_n \rangle \notin ext_R(r^n)\}$

$NEG(r^n)$ is an n-place relation which has in its extension$_R$ all of the n-tuples not in the extension$_R$ of r^n.

This completes the definitions of the logical functions. They guarantee that the domain of relations, R, houses a rich variety of complex relations.

The fifth member of an interpretation is the last function which imposes a structure on the domains $\mathcal{D}$ and R. We suppose that every property in R_1 has an encoding extension ("extension$_A$"). The encoding extension of a property is a set of members of $\mathcal{D}$ which encode the property. The fifth member of an interpretation is therefore a function, ext_A, which maps each $r^1 \in R_1$ into $\mathcal{D}$, i.e., $ext_A \colon R_1 \rightarrow \mathcal{D}$.

The final member of an interpretation, the F function, maps the simple names of the language to elements of the appropriate domain. For each object name a_i, $F(a_i) \in \mathcal{D}$. For each relation name P^n_i, $F(P^n_i) \in R_n$. Since "E!" is a simple property name, $F(E!) \in R_1$, and so $ext_R(F(E!)) \subseteq \mathcal{D}$. We call this subset of $\mathcal{D}$ the set of <u>existing</u> objects ("E"). We call the complement of E on $\mathcal{D}$ (i.e., $ext_R(NEG(F(E!)))$) the set of <u>abstract</u> objects ("A").

B. Assignments and denotations. As usual, an assignment with respect to an interpretation I will be any function, f_I, which assigns to each primitive variable an element of the domain over which the variable ranges. And, a denotation function with respect to an interpretation I and an I-assignment f_I, will be any function, d_{I,f_I}, defined on the terms of the language, which: (1) agrees with F_I on the primitive names, (2) agrees with f_I on the primitive variables, and (3) assigns to the complex terms on the basis of the denotations of their parts and the way in which they are arranged. But consider a complex term like "[λx Px & Syx]." Suppose that F_I(P) is the property of being a painting and F_I(S) is the study relation. Our λ-expression would then read: "being an object x such that x exemplifies paintinghood and y bears the study relation to it." This λ-expression might serve to translate the English phrase "is a painting which ( ) studied." The denotation of this λ-expression will be assigned in terms of the denotations of "P," "S," and "y," and the way in which these parts of the expression are arranged.

Since the denotation function $d_{I,f}$ (for convenience, we drop the subscript on the f) must agree with F_I, we know:

$d_{I,f}$(P) = *paintinghood*

$d_{I,f}$(S) = the *study* relation

$d_{I,f}$ will also agree with f on its assignment to "y"; so let's suppose that $d_{I,f}$(y) = o. However, there are three ways to construct a complex property which might serve to interpret the way in which these simple parts are arranged in the λ-expression. One alternative is to first plug $d_{I,f}$(y) into the first place of $d_{I,f}$(S), conjoin $d_{I,f}$(P) with the one place property which results, and then reflect the first and second places of the 2-place relation resulting from the conjunction. This would give us:

$$REFL_{1,2}(CONJ(d_{I,\mathcal{f}}(P),\ PLUG_1(d_{I,\mathcal{f}}(S),\ d_{I,\mathcal{f}}(y))))$$

On the other hand, we might first conjoin $d_{I,\mathcal{f}}(P)$ with $d_{I,\mathcal{f}}(S)$ to get a three place relation, reflect its first and third places to get a 2-place relation, and then plug $d_{I,\mathcal{f}}(y)$ into the second place of the result. This would give us:

$$PLUG_2(REFL_{1,2}(CONJ(d_{I,\mathcal{f}}(P),\ d_{I,\mathcal{f}}(S))),\ d_{I,\mathcal{f}}(y))$$

Finally, we might conjoin $d_{I,\mathcal{f}}(P)$ with $d_{I,\mathcal{f}}(S)$, then plug $d_{I,\mathcal{f}}(y)$ into the second place of this 3-place relation, and then reflect the first and second places of the result. This would give us:

$$REFL_{1,2}(PLUG_2(CONJ(d_{I,\mathcal{f}}(P),\ d_{I,\mathcal{f}}(S)),\ d_{I,\mathcal{f}}(y)))$$

That is, the following three properties are all sitting around in R_1 and could equally well serve as the denotation of [λx Px & Syx] with respect to I and $\mathcal{f}$:

$$REFL_{1,2}(CONJ(\mathit{paintinghood},\ PLUG_1(\mathit{study},\ o)))$$
$$PLUG_2(REFL_{1,3}(CONJ(\mathit{paintinghood},\ \mathit{study})),\ o)$$
$$REFL_{1,2}(PLUG_2(CONJ(\mathit{paintinghood},\ \mathit{study}),\ o))$$

The claim that these three complex properties are in fact the *same* property is a metaphysical thesis of great interest. The idea is that these complicated looking script expressions which are displayed immediately above just represent different decompositions of the same property. Of course, such a thesis needs to be supported, preferably with a (mathematical) theory which predicts when any two such properties or relations are identical.[7] But such a theory has yet to be devised.[8]

Consequently, we face the question, which of the above three properties should be assigned as the denotation$_{I,\mathfrak{f}}$ of our λ-expression [λx Px & Syx]? In order to answer this question, we shall develop a mechanical procedure which selects one of the above properties and which makes a similar kind of selection for each of the other λ-expressions. This mechanical procedure is embodied primarily in a definition which partitions the λ-expressions into seven syntactic equivalence classes. Six of these classes will correspond to the logical functions found in *L*.[9] The seventh houses all of the "simple" λ-expressions. [λx Px & Syx] will be categorized as a 1,2-reflection of the expression [λxu Px & Syu], which in turn will be categorized as the conjunction of the two expressions [λx Px] and [λu Syu]. The first of these is simple and the second will be categorized as the 1^{st}-plugging of [λuw Swu] by term y. Once the λ-expressions have been partitioned, it will be straightforward to define <u>*I*-assignment</u> and <u>denotation</u>$_{I,\mathfrak{f}}$ so that [λx Px & Syx] denotes the first of the above three properties. The definitions of <u>*I*-assignment</u> and <u>denotation</u>$_{I,\mathfrak{f}}$ follow the partitioning.

<u>Partitioning the λ-expressions</u>. We use μ, ξ, ζ as metavariables ranging over λ-expressions. Suppose μ is an arbitrary λ-expression. Then $\mu = [\lambda\nu_1 \ldots \nu_n\ \phi]$, for some $\phi, \nu_1, \ldots, \nu_n$. Utilizing the following five major rules, we then define: μ is the <u>i,jth-conversion of</u> ξ, μ is the <u>negation of</u> ξ, μ is the <u>disjoint conjunction of</u> ξ <u>and</u> ζ, μ is the <u>i,jth-projection of</u> ξ, μ is the <u>i,jth-reflection of</u> ξ, μ is the <u>ith-plugging of</u> ξ <u>by</u> o, and μ is <u>elementary</u>.

1. If $(\exists i)(1 \leq i \leq n$ and ν_i is not the i^{th} free object variable in ϕ and

i is the least such number), then where ν_j is the i^{th} free object variable in ϕ, μ is the <u>i,j^{th}-conversion of</u> $[\lambda\nu_1\ldots\nu_{i-1}\nu_j\nu_{i+1}\ldots\nu_{j-1}\nu_i\nu_{j+1}\ldots\nu_n\ \phi]$

2. If μ is not the i,j^{th}-conversion of any λ-expression, then:
   (a) if $\phi = \sim\psi$, μ is the <u>negation of</u> $[\lambda\nu_1\ldots\nu_n\ \psi]$
   (b) if $\phi = (\psi\ \&\ \chi)$, and ψ and χ have no free object variables in common, then where $\nu_1,\ldots,\nu_p$ are the variables in ψ and $\nu_{p+1},\ldots,\nu_n$ are the variables in χ, μ is the <u>(disjoint) conjunction of</u> $[\lambda\nu_1\ldots\nu_p\ \psi]$ <u>and</u> $[\lambda\nu_{p+1}\ldots\nu_n\ \chi]$
   (c) if $\phi = (\exists\nu)\psi$, and ν is the i^{th} free object variable in ϕ, then μ is the <u>i^{th}-projection of</u> $[\lambda\nu_1\ldots\nu_{i-1}\nu\nu_i\nu_{i+1}\ldots\nu_n\ \psi]$

3. If μ is none of the above, then if $(\exists i)(1\leq i\leq n$ and ν_i occurs free in more than one place in ϕ and i is the least such number), then where:
   (a) k is the number of free object variables between the first and second occurrences of ν_i,
   (b) ϕ' is the result of replacing the second occurrence of ν_i with a new variable ν, and
   (c) $j = i+k+1$,

   μ is the <u>i,j^{th}-reflection of</u> $[\lambda\nu_1\ldots\nu_{i+k}\nu\nu_j\ldots\nu_n\ \phi']$.

4. If μ is none of the above, then if o is the left most object term occurring in ϕ, then where:
   (a) j is the number of free variables occurring before o,
   (b) ϕ' is the result of replacing the first occurrence of o by a new variable ν, and
   (c) $i = j+1$,

   μ is the <u>i^{th}-plugging of</u> $[\lambda\nu_1\ldots\nu_j\nu\nu_{j+1}\ldots\nu_n\ \phi']$ <u>by</u> o.

5. If μ is none of the above, then

   (a) ϕ is atomic,

   (b) $\nu_1,\ldots,\nu_n$ is the order in which these variables first occur in ϕ,

   (c) $\mu = [\lambda\nu_1\ldots\nu_n\ \rho^n\nu_1\ldots\nu_n]$, for some relation term ρ^n, and

   (d) μ is called elementary.

Rules (1)-(5) partition the class of λ-expressions into seven equivalence classes. The reader should verify that: [λx Rxb] is the 2nd-plugging of [λxy Rxy] by b; [λx (Px & Skx)] is the 1,2-reflection of [λxy (Px & Sky)]; [λxy ($\exists$w)Bwxy] is the 2nd-projection of [λxwy Bxwy]; and [λxy (Rxx & Syy)] is the conjunction of [λx Rxx] and [λy Syy]; among other examples.

I-assignments.[10] If given an interpretation *I* of our language, an *I*-assignment, $\mathfrak{f}$, will be any function defined on the primitive variables of the language which satisfies the following two conditions:

1. where ν is any object variable, $\mathfrak{f}(\nu) \in \mathcal{D}$
2. where π^n is any relation variable, $\mathfrak{f}(\pi^n) \in \mathcal{R}_n$

Denotations. If given an interpretation *I* of our language, and an *I*-assignment $\mathfrak{f}$, we recursively define the denotation of term τ with respect to interpretation *I* and *I*-assignment $\mathfrak{f}$ ("$d_{I,\mathfrak{f}}(\tau)$") as follows:

1. where κ is any primitive name, $d_{I,\mathfrak{f}}(\kappa) = F_I(\kappa)$
2. where ν is any object variable, $d_{I,\mathfrak{f}}(\nu) = \mathfrak{f}(\nu)$
3. where π^n is any relation variable, $d_{I,\mathfrak{f}}(\pi^n) = \mathfrak{f}(\pi^n)$
4. where $[\lambda\nu_1\ldots\nu_n\ \rho^n\nu_1\ldots\nu_n]$ is any elementary λ-expression, $d_{I,\mathfrak{f}}([\lambda\nu_1\ldots\nu_n\ \rho^n\nu_1\ldots\nu_n]) = d_{I,\mathfrak{f}}(\rho^n)$

5. where μ is the i^{th}-plugging of ξ by o,

   $d_{I,\mathfrak{f}}(\mu) = PLUG_i(d_{I,\mathfrak{f}}(\xi), d_{I,\mathfrak{f}}(o))$

6. where μ is the i^{th}-projection of ξ,

   $d_{I,\mathfrak{f}}(\mu) = PROJ_i(d_{I,\mathfrak{f}}(\xi))$

7. where μ is the i,j^{th}-conversion of ξ,

   $d_{I,\mathfrak{f}}(\mu) = CONV_{i,j}(d_{I,\mathfrak{f}}(\xi))$

8. where μ is the i,j^{th}-reflection of ξ,

   $d_{I,\mathfrak{f}}(\mu) = REFL_{i,j}(d_{I,\mathfrak{f}}(\xi))$

9. where μ is the disjoint conjunction of ξ and ζ,

   $d_{I,\mathfrak{f}}(\mu) = CONJ(d_{I,\mathfrak{f}}(\xi), d_{I,\mathfrak{f}}(\zeta))$

10. where μ is the negation of ξ,

   $d_{I,\mathfrak{f}}(\mu) = NEG(d_{I,\mathfrak{f}}(\xi))$

Here are some examples of λ-expressions and their denotations:

$d_{I,\mathfrak{f}}([\lambda x\ Rxa]) = PLUG_2(d_{I,\mathfrak{f}}(R), d_{I,\mathfrak{f}}(a))$

$d_{I,\mathfrak{f}}([\lambda x\ Sxbd]) = PLUG_2(PLUG_3(d_{I,\mathfrak{f}}(S), d_{I,\mathfrak{f}}(d)), d_{I,\mathfrak{f}}(b))$

$d_{I,\mathfrak{f}}([\lambda x\ Px \ \&\ Skx]) = REFL_{1,2}(CONJ(d_{I,\mathfrak{f}}(P), PLUG_1(d_{I,\mathfrak{f}}(S),$
$d_{I,\mathfrak{f}}(k))))$

$d_{I,\mathfrak{f}}([\lambda xy\ (\exists w)Bxwy]) = PROJ_2(d_{I,\mathfrak{f}}(B))$

$d_{I,\mathfrak{f}}([\lambda xy\ Rxx \ \&\ Syy]) = CONJ(REFL_{1,2}(d_{I,\mathfrak{f}}(R)), REFL_{1,2}(d_{I,\mathfrak{f}}(S)))$

$d_{I,\mathfrak{f}}([\lambda x\ Bx \ \&\ (\exists y)(Wyx \ \&\ Lmy)]) = REFL_{1,2}(CONJ(d_{I,\mathfrak{f}}(B),$
$PROJ_1(REFL_{1,3}(CONJ(d_{I,\mathfrak{f}}(W), PLUG_1(d_{I,\mathfrak{f}}(L), d_{I,\mathfrak{f}}(m)))))))$

C. Satisfaction.[11] If we're given an interpretation I, and an assignment $\mathfrak{f}$, we may define $\mathfrak{f}$ satisfies ϕ, recursively, as follows:

1. If $\phi = \rho^n o_1 \ldots o_n$, $\mathfrak{f}$ satisfies ϕ iff $\langle d_{I,\mathfrak{f}}(o_1), \ldots, d_{I,\mathfrak{f}}(o_n)\rangle$ $\varepsilon\ ext_R(d_{I,\mathfrak{f}}(\rho^n))$

2. If $\phi = o\rho^1$, $\mathfrak{f}$ satisfies ϕ iff

   $d_{I,\mathfrak{f}}(o) \in ext_A(d_{I,\mathfrak{f}}(\rho^1))$

3. If $\phi = (\sim\psi)$, $\mathfrak{f}$ satisfies ϕ iff $\mathfrak{f}$ fails to satisfy ϕ

4. If $\phi = (\psi\&\chi)$, $\mathfrak{f}$ satisfies ϕ iff $\mathfrak{f}$ satisfies both ψ and χ

5. If $\phi = (\exists\alpha)\psi$, $\mathfrak{f}$ satisfies ϕ iff

   $(\exists\mathfrak{f}')(\mathfrak{f}' \overline{\overline{\alpha}} \mathfrak{f}$ and $\mathfrak{f}'$ satisfies $\phi)$, where:

   $\mathfrak{f}' \overline{\overline{\alpha}} \mathfrak{f} =_{dfn} \mathfrak{f}'$ is an I-assignment just like $\mathfrak{f}$ except perhaps for what it assigns to α.[12]

D. Truth under an interpretation. ϕ is true under interpretation I iff every I-assignment $\mathfrak{f}$ satisfies ϕ. Using this definition, we say that ϕ is valid (logically true) iff ϕ is true under all interpretations. The logical axioms which follow in the next section are all valid. We say that an interpretation I is a model of elementary object theory iff all the proper axioms of the theory (§4) are true under I.

§3. The Logic

The logic for our interpreted language consists of:

A. Logical axioms

B. Rules of Inference

A. The logical axioms. There are an infinite number of formulas which are logically true (valid). Some of these are designated as logical axioms and they, together with the rules of inference, store the analytical power of the theory. The logical axioms are introduced by schemata, which indicate that all formulas of a certain form are to be axioms. The schemata fall into three groups: the propositional schemata, the quantificational schemata, and a schema governing λ-expressions.[13]

<u>Propositional schemata</u>.

LA1: $\phi \rightarrow (\psi \rightarrow \phi)$

LA2: $(\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \chi))$

LA3: $(\sim\phi \rightarrow \sim\psi) \rightarrow ((\sim\phi \rightarrow \psi) \rightarrow \phi)$

<u>Quantificational schemata</u>.

LA4: $(\alpha)\phi \rightarrow \phi^{\tau}_{\alpha}$, where τ is substitutable for α

LA5: $(\alpha)(\phi \rightarrow \psi) \rightarrow (\phi \rightarrow (\alpha)\psi)$, provided α isn't free in ϕ

<u>Lambda schema</u>.

λ-EQUIVALENCE: For any formula ϕ where F^n isn't free, the universal closure of the following is an axiom:

$(x_1)\ldots(x_n)([\lambda\nu_1\ldots\nu_n\ \phi]x_1\ldots x_n \equiv \phi^{x_1,\ldots,x_n}_{\nu_1,\ldots,\nu_n})$, provided $x_1,\ldots,x_n$ are substitutable for $\nu_1,\ldots,\nu_n$, respectively.

For example, $(x)([\lambda y\ \sim Ry]x \equiv \sim Rx)$ is an instance of λ-EQUIVALENCE which says that an arbitrary object x exemplifies failing-to-exemplify-R iff x fails to exemplify R. If we let "P" denote the property of being a painting and "S" denote the study relation, then

$$(x_1)(x_2)([\lambda x_5 x_3\ Px_5\ \&\ Sx_3x_5]x_1x_2 \equiv Px_1\ \&\ Sx_2x_1)$$

is an instance of λ-EQUIVALENCE which asserts that arbitrary objects x_1 and x_2 exemplify being two objects such that the first exemplifies paintinghood and the second bears the study relation to the first iff x_1 exemplifies paintinghood and x_2 bears the study relation to x_1.

B. Rules of inference.[14]

1. Arrow Elimination ("$\rightarrow$E"): If $\vdash \phi$ and $\vdash \phi \rightarrow \psi$, then $\vdash \psi$

2. Universal Introduction ("UI"): If $\vdash \phi$, then $\vdash (\alpha)\phi$

Officially, these are all the rules we'll need. In the usual manner, a proof will be any finite sequence of wffs which is such that each wff in the sequence is either an axiom (logical or proper) or follows from earlier members of the sequence by a rule of inference. A theorem is any wff that appears as the last member of a proof. We distinguish the logical theorems from the proper theorems: the former are derivable using only logical axioms as premises, whereas the latter are derivable from some proper axiom.

It is convenient to employ other standard, derived rules of inference. For example, we call the rule of inference derivable from $\rightarrow$E and LA4 universal elimination ("UE"). Standard formulations of the existential introduction and elimination rules ("EI" and "EE"), the quantifier negation rules ("QN"), and the introduction and elimination rules for ~, &, v, and $\equiv$ are also employed. We avail ourselves of conditional and indirect proof techniques. The proofs which follow (as well as those in the appendices) are usually constructed with the aid of these derived rules.

By using UE on the universal quantifiers of the instances of λ-EQUIVALENCE, we obtain biconditionals. Rules of inference governing the biconditional allow us to introduce (eliminate) λ-expressions into proofs when the right (left) side of the biconditional is added as a premise. We may shorten this process by formulating two rules of inference derived from λ-EQUIVALENCE, $\equiv$I, and $\equiv$E:

where ϕ is any propositional formula with object terms $o_1,\ldots,o_n$, and $\nu_1,\ldots,\nu_n$ are object variables substitutable for $o_1,\ldots,o_n$, respectively, then the following are rules of inference:

1. λ-Introduction ("λI"): If $\vdash \phi$, then

   $\vdash [\lambda\nu_1\ldots\nu_n\ \phi^{\nu_1,\ldots,\nu_n}_{o_1,\ldots,o_n}]o_1\ldots o_n$

2. λ-Elimination ("λE"): If $\vdash [\lambda\nu_1\ldots\nu_n\ \phi^{\nu_1,\ldots,\nu_n}_{o_1,\ldots,o_n}]o_1\ldots o_n$, then $\vdash \phi$

Also, since $[\lambda\nu_1\ldots\nu_n\ \phi]$ is an n-place relation term, it is subject to existential introduction. We get an important logical theorem schema by applying EI to λ-EQUIVALENCE:

THEOREM(S) ("RELATIONS"): where ϕ is a propositional formula which has no free F^n's, but has $x_1,\ldots,x_n$ free, the universal closure of the following is a theorem:

$(\exists F^n)(x_1)\ldots(x_n)(F^nx_1\ldots x_n \equiv \phi)$

The instances of this schema tell us what complex properties and relations there are. Here are some examples:

(a) $(\exists F)(x)(Fx \equiv x{=}_E x)$

(b) $(\exists F)(x)(Fx \equiv x{\neq}_E x)$

(c) $(\exists F)(x)(Fx \equiv {\sim}Gx)$

(d) $(\exists F)(x)(Fx \equiv Gx\ \&\ Hx)$

(e) $(\exists F)(x)(Fx \equiv Gx \vee Hx)$

(f) $(\exists F)(x)(y)(Fxy \equiv Gyx)$

(g) $(\exists F)(x)(Fx \equiv (\exists y)Gxy)$

Axioms (a) and (b) assert that there is a universal property and empty property among the existing objects, respectively; (c)-(f) assert, respectively, that every property has a negation, every two properties have a (non-disjoint) conjunction and disjunction, every two place relation has a converse, and every two place relation has a projection on its second place.

RELATIONS and D_4 constitute a full-fledged theory of n-place relations ($n \geq 1$). We no longer need to suppose that relations are "creatures of darkness." They have precise "existence" and identity conditions. It is not a consequence of our theory that logically equivalent relations are identical--we cannot prove that relations which have the same exemplification extension are identical. For example, it does not follow from the fact that being a rational animal and being a featherless biped are logically equivalent properties that they are identical.

If we restrict ourselves to properties, we should note that there are two senses of "F is logically equivalent to G." One sense is that F and G have the same exemplification extension. The second sense is that F and G have the same encoding extension. If F and G are logically equivalent in this latter sense, then, by definition, they are identical. In what follows, we always use "logical equivalence" in the former sense.

We call our modified second order language, together with its semantics and its logic, the <u>object calculus</u>. The object calculus is the metaphysical system in which the proper axioms of elementary object theory are stated.

§4. The Proper Axioms[15]

We state the theory of abstract objects using two axioms and two schemata. Though these axioms are not logically true, we nevertheless suppose them to be true *a priori*. The first two axioms express truths about existing objects. The first schema tells us about any objects or relations which satisfy the definitions for identity. The second schema tells us what abstract objects there are. Since the schemata indicate that all sentences of a certain form are to be axioms, we end up with a denumerably infinite number of proper axioms. These axioms, plus the three definitions of identity in §1, constitute the first principles of abstract object theory.

The first axiom tells us that two objects bear the identity$_E$ relation to one another iff they both exist and exemplify the same properties:[16]

AXIOM 1 ("E-IDENTITY"):

$(x)(y)(x{=_E}y \equiv E!x \;\&\; E!y \;\&\; (F)(Fx \equiv Fy))$

With this axiom, we prove the following theorem schema for identity:

THEOREMS ("IDENTITY INTRODUCTION"):

$(\alpha)(\alpha{=}\alpha)$, where α is any variable

Proof: Clearly, if α is an object variable x, and E!x, then since we have $(F)(Fx \equiv Fx)$ from propositional logic and UI, we may use E-IDENTITY to prove $x{=_E}x$. So x=x, by D_2 (§1). If ~E!x, then x is abstract and similar techniques get us the right hand disjunction of D_2. If α is a one place property variable F^1, we easily get $(x)(xF \equiv xF)$. So by D_3, $F^1 = F^1$. And a generalized version of this procedure gets us

$\alpha = \alpha$, when α is any n-place relation variable. ⊠

In what follows, we abbreviate "IDENTITY INTRODUCTION" as "=I."

The second axiom tells us that no existing object encodes any properties.

AXIOM 2 ("NO-CODER"):

$(x)(E!x \rightarrow \sim(\exists F)xF)$

The third axiom is the schema for identity:

AXIOM 3 ("IDENTITY"):

$\alpha = \beta \rightarrow (\phi(\alpha,\alpha) \rightarrow \phi(\alpha,\beta))$, where $\phi(\alpha,\beta)$ is the result of replacing some, but not necessarily all, free occurrences of α by β, provided β is substitutable for α in the occurrences of α it replaces.[17]

This identity schema governs both primitive and defined identities. It clearly governs the latter. It also governs the primitive $x =_E y$ formulas since if $x =_E y$, it follows by D_2 that $x = y$. The rule of inference derivable from $\rightarrow$E and IDENTITY is called <u>identity elimination</u> ("=E").

The schema for abstract objects generates the most important set of axioms of the theory. In effect, the schema guarantees that for every expressible set of properties, there is an abstract object which encodes just the members of the set.[18] However, the schema does this without a commitment to sets. We generally use open formulas with one free property variable with this axiom. The formulas express conditions on properties. Metalinguistically, it is legitimate to talk about the set of properties satisfying a given condition, but in the object language, our schema says something more like: for every condition on properties, there is an abstract object which encodes just the properties which meet the condition:[19]

AXIOM(S) ("A-OBJECTS"): For any formula ϕ where x isn't free, the universal closure of the following is an axiom:

$$(\exists x)(A!x \ \& \ (F)(xF \equiv \phi))$$

Some examples will help. If we let "F=R v F=S" be our formula ϕ, and suppose that "R" denotes roundness and "S" denotes squareness, then our axiom guarantees that there is a "round square" as follows:

$$(\exists x)(A!x \ \& \ (F)(xF \equiv F{=}R \vee F{=}S))$$

Suppose a_0 is such an object. It is easy to see that a_0 must be unique. For suppose some other abstract object a_1, $a_1 \neq a_0$, encoded exactly roundness and squareness. By D_2, it would follow that either a_1 encoded a property a_0 didn't, or vice versa, contrary to hypothesis. In fact, given D_2, we have the following theorem schema:

THEOREM(S): ("UNIQUENESS") For every formula ϕ where x isn't free, the universal closure of the following is a theorem:

$$(\exists! x)(A!x \ \& \ (F)(xF \equiv \phi))$$

Proof: An arbitrary instance of A-OBJECTS says that there is an abstract object which encodes *exactly* the properties which satisfy the given formula. But there couldn't be distinct such objects, since distinct abstract objects must differ with respect to at least one of the properties they encode. $\boxtimes$

Another instance of the schema for objects says there is an "existent golden mountain." Suppose "G" denotes goldenness and "M" denotes mountainhood. We then have:

$$(\exists x)(A!x \ \& \ (F)(xF \equiv F{=}E! \vee F{=}G \vee F{=}M))$$

It follows that there is an abstract object which encodes a property (existence) it fails to exemplify.

By letting ϕ = $\ulcorner F \neq F \urcorner$, we obtain the empty object--it fails to encode any properties. By letting ϕ = $\ulcorner F=F \urcorner$, we obtain the universal object--it encodes every property.

Suppose "a_5" denotes Socrates. Then the following instance of A-OBJECTS yields an A-object which encodes exactly the properties Socrates exemplifies:

$$(\exists x)(A!x \ \& \ (F)(xF \equiv Fa_5))$$

We might call this object Socrates' <u>blueprint</u>, and call Socrates the <u>correlate</u> of the blueprint.[20] We define these terms as follows:[21]

D_5 x is the blueprint of y and y is the correlate of x

("Blue(x,y)" and "Cor(y,x)") $=_{dfn}$ $(F)(xF \equiv Fy)$

A-OBJECTS guarantees that every object, existing or abstract, has a unique blueprint:

$$(y)(\exists !x)(A!x \ \& \ (F)(xF \equiv Fy))$$

This is an instance of the theorem schema since it is the universal closure when "Fy" is the formula ϕ.

Given any object b, A-OBJECTS yields an object which encodes all the properties b fails to exemplify. Given any two objects b and c, A-OBJECTS yields an object which encodes (1) just the properties b and c have in common, (2) just the properties exemplified by either b or c, and (3) just the relational properties b has with respect to c. This last object is yielded by the following instance:

$$(\exists x)(A!x \ \& \ (F)(xF \equiv (\exists G^2)((Gbc \ \& \ F=[\lambda x \ Gxc]) \vee$$
$$(Gcb \ \& \ F=[\lambda x \ Gcx]))))$$

These examples give one a pretty good idea of what A-OBJECTS says.

We use A-OBJECTS to justify the definition we've proposed for property identity. Suppose that instead of defining identity between properties as in D_3, we had added primitive identity formulas between property terms. Then the following would have been a consequence of A-OBJECTS:

$$(G)(H)((x)(xG \equiv xH) \rightarrow G=H)$$

To see this, suppose arbitrary properties P and Q were encoded by exactly the same objects, but were distinct. By A-OBJECTS, it would have followed that there is an object which encodes just P, without encoding Q, contrary to hypothesis.

Since we would have had this consequence had property identity been primitive, there was every reason to just define identity among properties. Semantically, our definition ensures that two properties which have the same $\text{extension}_{\mathcal{A}}$ are identical. But our theory doesn't commit us to the view that properties exemplified by the same objects (i.e., which have the same $\text{extension}_{\mathcal{R}}$) are identical. An overriding reason for choosing the style of semantics we've employed is that properties and relations are not identified with their $\text{extensions}_{\mathcal{R}}$. The semantics doesn't force upon us a view to which the theory is not committed.

E-IDENTITY, NO-CODER, IDENTITY, and A-OBJECTS jointly are called the <u>elementary theory of abstract objects</u> and I believe that the theory is consistent.[22,23] In the next section, we shall see that it is protected from the known paradoxes.

§5. Avoiding Known Paradoxes

As in set theory, unrestricted abstraction schemata lead to paradox. The λ-EQUIVALENCE schema has been restricted indirectly (and RELATIONS has been restricted directly) so as to avoid these paradoxes. There are two paradoxes that have been discovered--one by Romane Clark, the other by Alan McMichael.[24] In Appendices A and B, we provide strict derivations of these paradoxes, demonstrating how they result from λ-expressions without restrictions on the formulas which appear behind them. It is important, however, to sketch the proofs here.

Suppose we dropped the two major restrictions on λ-formation and RELATIONS (i.e., the restrictions imposed by the definition of propositional formula). We could then form the following two λ-expressions: $[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]$ ("encoding a property that's not exemplified"), and $[\lambda x\ (F)(xF \rightarrow Fx)]$ ("exemplifying every property that's encoded"). Alternatively, we could produce instances of RELATIONS as follows:

$$(\exists F)(x)(Fx \equiv (\exists G)(xG\ \&\ {\sim}Gx))$$

$$(\exists F)(x)(Fx \equiv (G)(xG \rightarrow Gx))$$

But then consider the following argument ("Clark's paradox"):

Consider the abstract object a_0 which encodes just $[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]$, and suppose it exemplifies $[\lambda x\ (F)(xF \rightarrow Fx)]$. By λE, it follows that $(F)(a_0F \rightarrow Fa_0)$, so a_0 must exemplify $[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]$ as well as encode it. Again, by λE, $(\exists F)(a_0F\ \&\ {\sim}Fa_0)$, i.e., ${\sim}(F)(a_0F \rightarrow Fa_0)$. But then a_0 must fail to exemplify $[\lambda x\ (F)(xF \rightarrow Fx)]$ (by λ-EQUIVALENCE and ≡E), contrary to hypothesis.

So suppose a_0 fails to exemplify $[\lambda x\ (F)(xF \rightarrow Fx)]$. Then $\sim(F)(a_0F \rightarrow Fa_0)$, i.e., $(\exists F)(a_0F\ \&\ \sim Fa_0)$. Call this property "R" and note also that by λI, a_0 exemplifies $[\lambda x\ (\exists F)(xF\ \&\ \sim Fx)]$. Since a_0 encodes just one property, R must be $[\lambda x\ (\exists F)(xF\ \&\ \sim Fx)]$. But by definition of R, a_0 fails to exemplify R, i.e., $\sim[\lambda x\ (\exists F)(xF\ \&\ \sim Fx)]a_0$, contradiction. ⊠

A second contradiction would also be provable because we could form the following λ-expression: $[\lambda y\ y{=}x]$ ("being identical to x"), where this abbreviates a much longer λ-expression with encoding subformulas and relation quantifiers. Again, by RELATIONS, we would know that there is such a property. But then consider the following argument ("McMichael's paradox"):

By A-OBJECTS, we have that $(\exists x)(A!x\ \&\ (F)(xF \equiv (\exists u)(F{=}[\lambda y\ y{=}u]\ \&\ \sim uF)))$. Call this object a_1 and consider the property $[\lambda y\ y{=}a_1]$. Assume that a_1 encodes $[\lambda y\ y{=}a_1]$. By definition of a_1, we know $(\exists u)([\lambda y\ y{=}a_1] = [\lambda y\ y{=}u]\ \&\ \sim u[\lambda y\ y{=}a_1])$. Call this object a_2. So, $[\lambda y\ y{=}a_1] = [\lambda y\ y{=}a_2]\ \&\ \sim a_2[\lambda y\ y{=}a_1]$. By =I, we know $a_1{=}a_1$, and by λI, we know $[\lambda y\ y{=}a_1]a_1$. Since $[\lambda y\ y{=}a_1] = [\lambda y\ y{=}a_2]$, it follows by =E that $[\lambda y\ y{=}a_2]a_1$. So by λE, $a_1{=}a_2$. But then, $\sim a_1[\lambda y\ y{=}a_1]$ (from the definition of a_2 and =E), contrary to hypothesis.

So suppose that $\sim a_1[\lambda y\ y{=}a_1]$. By definition of a_1, $\sim(\exists u)([\lambda y\ y{=}a_1] = [\lambda y\ y{=}u]\ \&\ \sim u[\lambda y\ y{=}a_1])$. That is, $(u)([\lambda y\ y{=}a_1] = [\lambda y\ y{=}u] \rightarrow u[\lambda y\ y{=}a_1])$. But since $[\lambda y\ y{=}a_1] = [\lambda y\ y{=}a_1]$, it follows that $a_1[\lambda y\ y{=}a_1]$. Contradiction. ⊠

It is doubtful that the source of these paradoxes lies with the presence of relation quantifiers in λ-expressions. Logicians have not found any special trouble with the second order predicate calculus, in which one finds relations defined with quantification over other relations. For example, here's a standard instance of the relations schema of the second order predicate calculus:

$(\exists F)(x)(Fx \equiv (\exists G)Gx)$

This asserts that there is a property of "having a property." This property would be denoted using "[λx (∃G)Gx]." Properties such as these don't seem to cause any special consistency problems. The *only* reason for adding the "no relation quantifiers" restriction on λ-expressions is that given the style of semantics we have employed, it is rather complicated to interpret such expressions without the resources of type theory (in type theory, we suppose that "[λx (∃G)Gx]" abbreviates "[λx (∃G)ExGx]," where "Ex" is a predicate which denotes the exemplification relation between a property and an object which exemplifies it. We then interpret this latter λ-expression as $PROJ_1(d_{I,\delta}(\text{Ex}))$, i.e., as the first projection of this exemplification relation).[25]

Consequently, the elimination of encoding subformulas from λ-expressions appears to us to be the most theoretically satisfying way of avoiding the paradoxes. This move makes [λx (∃F)(xF & ~Fx)], [λx (F)(xF → Fx)], and [λxy x=y] all ill-formed. It affords us the convenience of not having to interpret λ-expressions containing such subformulas. It leaves us with a rich variety of properties and relations--all the ones we're familiar with from the second order predicate calculus.[26]

The theory with which this move leaves us is rather powerful, and there is one interesting consequence to which the reader should be alerted. It turns out that some complex properties and relations do not have unique constituents.[27] It is provable, for example, that for some objects a_3 and a_4, $a_3 \neq a_4$, that $[\lambda y\ Rya_3] = [\lambda y\ Rya_4]$. Here's how:

Let R be any two place relation. By A-OBJECTS, $(\exists x)(A!x\ \&\ (F)(xF \equiv (\exists u)(F=[\lambda y\ Ryu]\ \&\ {\sim}uF)))$. Call this object a_3 and suppose a_3 doesn't encode $[\lambda y\ Rya_3]$. By reasoning similar to the second part of McMichael's paradox, it follows that a_3 does encode $[\lambda y\ Rya_3]$. So supposing a_3 encodes $[\lambda y\ Rya_3]$, it follows by definition of a_3 that for some object, say a_4, $[\lambda y\ Rya_4] = [\lambda y\ Rya_3]$ and ${\sim}a_4[\lambda y\ Rya_3]$. Since a_3 encodes $[\lambda y\ Rya_3]$ and a_4 doesn't, it follows that $a_3 \neq a_4$.[28]

There is a weaker, though fairly useful, version of our theory in which this is not a consequence. That is the theory which results by banishing encoding subformulas from the formulas ϕ used in the A-OBJECTS axiom.[29] By doing this, the instance of A-OBJECTS which begins the preceding proof will be ill-formed. I think one could then consistently suppose (add axioms which guarantee) that complex relations and properties do have unique constituents.[30] With only a few exceptions, almost all of the applications of the present theory could be preserved.[31] Officially, however, we shall not work with this version of the theory--we prefer to work with the most powerful version of the theory which has a reasonable degree of probability of being consistent.

§6. An Auxiliary Hypothesis

In Chapter II, we shall put the theory we've now formulated to work. For these applications, we add to our primitive vocabulary abbreviations of the "gerundive versions" of standard English transitive verbs, intransitive verbs, predicate adjectives, and predicate nouns.[32] By the "gerundive version" of these words, I mean the phrases constructed out of English gerunds which can appear in the subject places or direct object places of English sentences. Here are some examples:

	Gerundive Version
A. Transitive verbs	
kick	(the) kicking (relation)
worship	(the) worshipping (relation)
hate	(the) hating (relation)
B. Intransitive verbs	
run	(the property of) running
walk	(the property of) walking
C. Predicate adjectives	
red	(the property of) being red
courageous	(the property of) being courageous
happy	(the property of) being happy
D. Predicate nouns	
horse	(the property of) being a horse
person	(the property of) being a person
building	(the property of) being a building

We abbreviate these gerundive versions in our language with single letters, appropriately chosen. For example, "K" might abbreviate "the kicking relation," etc. (Also, we shorten our readings of λ-expressions so that they more closely resemble their natural language counterparts.

So instead of reading "b[λx Cx & Px]" as "b encodes the property an object x exemplifies iff x exemplifies being courageous and x exemplifies being a person," we read it as "b encodes being a courageous person.")

These additions to our primitive vocabulary are supposed to reveal our pretheoretic conceptions about what properties and relations there are. By adding these properties and relations to our system, A-OBJECTS provides us with a rich variety of abstract objects which encode familiar sorts of simple and complex properties.

These additions also make it possible to state an auxiliary hypothesis of the elementary theory--an hypothesis to which we shall appeal on occasion in the applications. Despite its rather vague character, it grounds a wide range of intuitions some of us may share about abstract objects. Pretheoretically, we have a pretty good idea of what properties existing objects exemplify. And the theory tells us the conditions under which both existing and abstract objects encode these properties. But other than being abstract (i.e., [λx ~E!x]), we haven't said anything about which properties abstract objects exemplify.

Some of us may share the following intuitions. Abstract objects *do not exemplify* the following properties: being round, having a shape, being red, having a color, being large, having a size, being soft, having a texture, having mass, having spatio-temporal location, being visible, being capable of thought (this is *not* to say that they aren't thought of), being capable of feeling, etc. In addition, it might seem that no two abstract objects could ever meet each other, kick each other, kiss each other, etc. I'm sure the reader can provide many more examples.

These properties and relations are "ordinary" properties and relations of existing objects. Mally, Meinong, Findlay, Parsons, and others call them "nuclear" relations. They are to be distinguished from "extranuclear" relations such as being abstract, being thought about, being written about, being worshipped, being dreamed of, being possible, being more famous than, etc. We can easily imagine that abstract objects exemplify these extranuclear relations.[33]

We shall not be interested in pursuing this distinction among relations in any detail. We mention it because there will be occasion to appeal to the above intuitions and it would be nice to ground them all in some general principle. Consequently, we suppose that no abstract objects exemplify nuclear relations.[34]

We incorporate this hypothesis into elementary object theory by supposing that we can divide the n-place relation terms of our language according to whether they are nuclear or extranuclear.[35] We then add:

AUXILIARY HYPOTHESIS: Where ρ^n is a nuclear n-place relation term, the following is an axiom:

$$(x_1)\ldots(x_n)(A!x_1 \ \& \ \ldots \ \& \ A!x_n \rightarrow {\sim}\rho^n x_1 \ldots x_n)$$

We trouble the reader with this hypothesis because it seems possible that some such set of truths like these govern abstract objects *a priori*.

CHAPTER I ENDNOTES

[1]Every formula is a subformula of itself. If $\phi = (\sim\psi)$, $(\psi \,\&\, \chi)$, or $(\exists\alpha)\psi$, then ψ is a subformula of ϕ. If ψ is a subformula of χ, and χ is a subformula of ϕ, then ψ is a subformula of ϕ.

[2]See Alonzo Church [1941].

[3]For convenience, we will read "E!y" as "y exists" and "$[\lambda x\ x=_E b]$" as "being identical$_E$ to b," instead of using the more cumbersome readings.

[4]These definitions are all standard.

[5]Compare Parsons [1980], Ch. IV, §3.

[6]Most of these logical functions can be traced back to Moses Schonfinkel [1924]. Quine picked up the trail in his [1960]. McMichael, after seeing my work with *PLUG* (a function I picked up from Parsons [1980]), developed the other logical functions, using Quine's operators as the prototypes. Quine has no need for *PLUG* since his project was to explain away singular terms. I learned these algebraic techniques from Alan, and together we wrote [1979b]. We used these logical functions to validate λ-EQUIVALENCE, the relations schema I had been using informally along with the formal monadic theory developed in my [1979a].

Since [1979b], I have made several improvements to these logical functions. The place numbers have been indexed to those functions which operate on specific places in a relation. The definitions are now sharpened up so that we don't have an infinite number of empty relations like the plugging of a two place relation in its 15^{th} place, the projection of a 3-place relation in its 300^{th} place, etc. Also, I've added *VAC* and *NEC* in Chapter III.

[7] I'm indebted to T. Parsons for pointing this problem out to me.

[8] One possibility I've yet to explore is a reference made by Quine in a footnote in [1960]. He says that Bernays had developed a system which included axioms. So maybe there is such a theory. Also, Tarski's cylindrical algebra or a polyadic algebra might be relevant here.

[9] I was motivated to construct these definitions after reading Bealer [1981] (in manuscript), Ch. 3. The definition he had constructed to partition his complex terms seemed too complicated. I then had the idea that by subscripting the place numbers to the relevant logical functions and by ordering the rules for assignments, a much simpler procedure for partitioning λ-expressions could be found. Thanks goes to Alan McMichael for his valuable help in working out many of the details in the following definition.

[10] I'd like to thank Michael Jubien for pointing out a flaw in an earlier version of the definition of I-assignment.

[11] See Alfred Tarski [1944].

[12] With this definition of satisfaction, we may define what it is for a property or relation of a given interpretation to be expressible:

Property r^1 of I is expressible $=_{dfn}$ $(\exists\phi)(\phi$ has one free object variable ν and (f) $(f$ satisfies $\phi \equiv d_{I,f}([\lambda\nu\ \phi]) = r^1)))$.

Relation $\mathscr{r}^n$ of I is expressible $=_{dfn}$ $(\exists\phi)(\phi$ has n free variables $\nu_1,\ldots,\nu_n$ and $(\mathscr{f})$ $(\mathscr{f}$ satisfies $\phi \equiv d_{I,\mathscr{f}}([\lambda\nu_1\ldots\nu_n\ \phi])$ $= \mathscr{r}^n)))$.

[13]I follow Eliot Mendelson [1964], p. 57. Also, see Robert Rogers [1971], pp. 87-88.

[14]I've adopted E.J. Lemmon's procedure of stating rules of inference as introduction and elimination rules. We use "→E" instead of "MP," and "UI" instead of "Gen."

[15]The axioms which follow represent the culmination of the process of axiomatization which first began in my [1979a]. The axioms found here are basically the same as those found in that paper, with the exception of λ-EQUIVALENCE, which first appeared in [1979b]. The only real difference is that they are now stated in a language which doesn't sort terms denoting abstract objects from terms denoting existing objects. The most important axiom, A-OBJECTS, was visualized after reading Parsons, Findlay, and Rapaport.

[16]The reader might wonder here why we haven't just defined $x{=_E}y$ instead of taking "$=_E$" as primitive. The reason is as follows: We shall want to be able to form λ-expressions like $[\lambda xy\ x{=_E}y]$. Had we defined $x{=_E}y$ as $E!x\ \&\ E!y\ \&\ (F)(Fx \equiv Fy)$, $[\lambda xy\ x{=_E}y]$ would be ill-formed, due to the presence of the relation quantifier. Recall that we've banished relation quantifiers to simplify the semantics. The slight loss of elegance which results by having to add a non-logical axiom governing $x{=_E}y$ is minor compared to the complexity which would result from having to add the technical apparatus required to interpret λ-expressions with relation quantifiers.

In Chapter V, we will eliminate the primitive identity$_E$, since the resources of the type theory more easily handle the interpretation of λ-expressions with "higher order" quantifiers. It is for this reason that we do not mention identity$_E$ in the list of primitives provided in the Introduction.

[17]In the standard second order predicate calculus, where identity is defined ($x=y =_{dfn} (F)(Fx \equiv Fy)$), a restricted version of this proper axiom would be a logical theorem. If we were given that for object terms o_1 and o_2, $o_1=o_2$, we could show that for a formula ϕ with one free object variable ν, $\phi^{o_1}_{\nu} \rightarrow \phi^{o_2}_{\nu}$. Here's how:

Since $o_1=o_2$, $(F)(Fo_1 \equiv Fo_2)$. In the second order predicate calculus, *every* formula ϕ with one free object variable can be turned into a property denoting expression $[\lambda\nu\ \phi]$. Thus we may instantiate the universal F quantifier to get $[\lambda\nu\ \phi]o_1 \equiv [\lambda\nu\ \phi]o_2$. But by λ-abstraction, $[\lambda\nu\ \phi]o_1 \equiv \phi^{o_1}_{\nu}$ and $[\lambda\nu\ \phi]o_2 \equiv \phi^{o_2}_{\nu}$. So $\phi^{o_1}_{\nu} \rightarrow \phi^{o_2}_{\nu}$.

Note that no such proof could be carried out in the object calculus since *not* every formula ϕ with one free variable ν can be turned into a property denoting expression $[\lambda\nu\ \phi]$. Therefore, our identity schema is necessary, since it is not derivable. Our identity schema has even greater significance since it governs identities between relation terms as well. You generally don't even find identity among relation terms defined in the standard second order calculus.

[18]We define:

a set S of I properties is <u>expressible</u> $=_{dfn}$ $(\exists\phi)(\phi$ has exactly one free F^1-variable and $(\mathfrak{f})(\mathfrak{f}$ satisfies ϕ iff $d_{I,\mathfrak{f}}(F^1) \in S))$.

[19] Compare Parsons [1980], Ch. IV, §2. Also compare Rapaport [1976], p. 190, T7a. Also compare Castañeda [1974], pp. 15-21, C*.1-.7, and *C.1-.7. And compare Routley [1979], p. 263.

[20] I've adapted these terms from Rapaport [1978].

[21] We use abbreviations like "Notion (x,y)" to remind the reader that these formulas do not abbreviate formulas which can appear behind λ's.

[22] I have yet to discover the proof, however. Although I've looked hard for paradoxes, this is no guarantee that there aren't any lurking around. I think, however, that the probability of finding a proof is higher than the probability of finding a contradiction. But I would be delighted if the logician who discovers either a proof or an inconsistency would send me a copy of the result.

[23] You can't just model abstract objects as sets of properties. That's because sets of properties can't exemplify the very same properties which serve as their elements. This would be a violation of type (just as in ZF, no set of sets can be an element of one of its members). However, in object theory, A-objects may exemplify the very properties they encode. For example, the A-object which encodes $[\lambda x \sim E!x]$ also exemplifies this property. And the A-object which encodes not-being-red may very well exemplify this property.

Nor can you: (1) model existing objects as individuals, (2) model A-objects as sets of nuclear properties (or as sets of sets of individuals), (3) model properties as extranuclear properties (or as sets of sets of nuclear properties), (4) map down the extranuclear properties so that they become correlated with nuclear, "watered down" versions, and (5) define "x encodes F" as "the nuclear version of F is an element of x." That's because distinct extranuclear properties sometimes get mapped down to the same nuclear, watered-down version. In object theory, if $P \neq Q$, then the object encoding just P is distinct from the object encoding just Q. But on the above model, these two objects would be identified if the nuclear versions of P and Q were identical.

These two suggestions exhaust the obvious first moves for modelling the theory, and I suspect that the consistency proof will not be straightforward. I also suspect that we may have to construe A-objects as sets of extranuclear properties, in some careful manner, and somehow "cook up" the right relations necessary to define exemplifications and encoding. I've yet to discover how to do this.

[24] The Clark Paradox is reported by Rapaport in both [1976] and [1978]. McMichael's paradox is unpublished, though it is described in a footnote of our paper [1979b].

[25] See Chapter V, §§1, 2.

[26] I'm assuming here that our type theory is one where λ-expressions with higher order quantifiers are interpreted successfully.

[27] For a similar result in Parsons' theory, see [1980], Ch. IX, §1.

[28] The reason we don't get McMichael's paradox here is because we don't have the facts about R like we would have about the identity relation. For example, we don't get a contradiction from

$$(F)(a_3F \equiv (\exists u)(F = [\lambda y\ y =_E u]\ \&\ {\sim}uF))$$

and the supposition that $a_3[\lambda y\ y =_E a_3]$ because we don't know that $a_3 =_E a_3$.

[29] On this version, we would have to add F=G as a primitive to the language in order that the following instance of A-OBJECTS, and others like it, be well-formed:

$(\exists x)(A!x \ \& \ (F)(xF \equiv F{=}G))$

Note that by banishing encoding subformulas from all expressions in formulas used in A-OBJECTS, Clark's Paradox and McMichael's Paradox don't arise, even if we were to allow λ-expressions, in general, to have encoding formulas. Consequently, this move alone would be an alternative way of avoiding the paradoxes. Clearly, it is not as theoretically satisfying as just banishing encoding subformulas from λ-expressions--it may weaken the theory more than necessary, it requires the addition of the primitive identity formulas for relation terms, we would have to interpret λ-expressions with encoding subformulas in them, etc. This alternative is to be preferred over the one developed in the text only if the latter should prove to be inconsistent.

[30] For example, the following four axioms would guarantee that complex relations and properties had unique constituents:

AXIOM: $\mu \neq \xi$, where μ and ξ are λ-expressions in distinct equivalence classes of the partition.

AXIOM: $\mu=\xi \equiv \mu'=\xi'$, where $\mu(\xi)$ is the i,j^{th}-conversion (negation, i^{th} projection, i,j^{th}-reflection) of $\mu'(\xi')$.

AXIOM: $\mu=\xi \equiv (\mu'=\xi' \ \& \ \mu''=\xi'')$, where $\mu(\xi)$ is the conjunction of $\mu'(\xi')$ and $\mu''(\xi'')$.

AXIOM: $\mu=\xi \equiv (\mu'=\xi' \ \& \ o_1=o_2)$, where $\mu(\xi)$ is the i^{th}-plugging of $\mu'(\xi')$ by $o_1(o_2)$.

See Bealer [1981], Ch. 3, §16.

[31]In particular, we couldn't model stories in the manner suggested in Chapter IV, §4. However, alternative suggestions would work, though they are not quite as elegant.

[32]By "standard," I mean that the verbs are not propositional attitude verbs and that, intuitively, they do not denote higher order properties.

[33]For more on this distinction, see the previously cited works of Meinong, Mally, Findlay, and Parsons. Note that we differ with these authors on the property of existence. These authors suppose that it is extranuclear. This would mean that A-objects might exemplify this property. But, *by definition*, we suppose that A-objects fail to exemplify this property. We take existence to be nuclear.

[34]It may very well be true that the negations of nuclear properties are extranuclear. A-objects exemplify properties like not being red, not being colored, not having spatio-temporal location, etc. This agrees with Parsons' suggestion in [1979b], p. 658.

[35]Of course, this suggestion needs to be spelled out for the complex expressions which contain a mixture of simple nuclear and extranuclear terms. For a rough idea of what this involves, see Parsons [1979b], pp. 658-660.

CHAPTER II

APPLICATIONS OF THE ELEMENTARY THEORY

§1. Modelling Plato's Forms[1]

In this section, we construe certain assertions by Plato as consequences of the theory. Most philosophers today regard Plato's Forms as first level properties of some sort and view participation as just exemplification. But this view of Plato from within the Russellian background theory turns Plato's major principle about the Forms into a triviality.

Plato's major principle about the Forms is the One Over the Many Principle. It is stated principally in Parmenides (132a).[2] The following characterization is, I think, a faithful one:

(OMP) If there are two distinct F-things, then there is a Form of F in which they both participate.

According to the orthodox view,

The Form of F $=_{dfn}$ F

x participates in F $=_{dfn}$ Fx

So translating (OMP) into a standard second order predicate calculus, we would get:

$$x \neq y \ \& \ Fx \ \& \ Fy \rightarrow (\exists G)(G{=}F \ \& \ Gx \ \& \ Gy)$$

But the consequent of this conditional just follows from the antecedent by existential introduction. Clearly, we don't want to attribute such a triviality to Plato.[3] Yet it is difficult to conceive of it as an interesting metaphysical truth from within the Russellian framework.

In object theory, however, we may think of Forms as just a special kind of A-object. When (OMP) is translated into our language, it turns out to be an interesting theorem. To see this, consider the following series of definitions and proofs:[4]

D_6 x is a Form of G ("Form(x,G)") $=_{dfn}$ A!x & (F)(xF $\equiv$ F=G)

So a Form of G is any abstract object which encodes just G. So we have:

Theorem 1: (G)($\exists$x)Form(x,G)

Proof: By A-OBJECTS.

In fact, given UNIQUENESS, it also follows that:

Theorem 2: (G)($\exists$!x)Form(x,G)

Given Theorem 2, we may introduce a symbol to name the Form of G. Let's use "Φ_G" to denote it. Clearly,

Theorem 3: Φ_GG ("The Form of G encodes G")

Now we can define participation:

D_7 y participates in x ("Part(y,x)") $=_{dfn}$ ($\exists$F)(xF & Fy)

So something participates in the Form of G just in case there's a property the Form encodes which the object exemplifies. All objects which exemplify redness participate in the Form of Redness ("Φ_R").

These definitions validate (OMP). The translation of (OMP) into our language turns out to be a theorem:

Theorem 4: x$\neq$y & Fx & Fy $\rightarrow$ ($\exists$u)(u=Φ_F & Part(x,u) & Part(y,u))

Proof: By considering Φ_F and the definition of participation.

Another theorem quickly falls out of these definitions:

Theorem 5: Fx $\equiv$ Part(x,Φ_F)

<u>Proof</u>: ($\rightarrow$) Assume Fx. By Theorem 3, Part(x,Φ_F). ($\leftarrow$) Assume Part(x,Φ_F). Call the property Φ_F encodes and x exemplifies, G. Since Φ_F encodes just F, it must be that G=F. So Fx. ⊠

So in our system, the notions of exemplification and participation are distinct (unlike the orthodox view) though nonetheless equivalent. This should preserve at least some of the intuitions of orthodox theorists.

On our theory, some Forms participate in other Forms, and indeed, some Forms participate in themselves. Consider the property [λx ~E!x] ("$\bar{E}$!"). Let's call this property: Platonic existence. Since all A-objects fail to exist, they all exemplify Platonic existence. In particular, we have:

<u>Theorem 6</u>: $(x)((\exists F)(x=\Phi_F) \rightarrow \bar{E}!x)$

So the Forms exemplify a kind of existence which is different from the existence exemplified by actual objects.[5] But now consider $\Phi_{\bar{E}!}$, which we may call "Platonic Being," or "Reality." From Theorems 5 and 6 it follows that:

<u>Theorem 7</u>: $(x)((\exists F)(x=\Phi_F) \rightarrow Part(x,\Phi_{\bar{E}!}))$

So all Forms participate in Platonic Being.[6] In particular, $\Phi_{\bar{E}!}$ participates in itself, justifying our earlier assertion.

To reach this conclusion, we might also have used the AUXILIARY HYPOTHESIS and the assumption that being blue, for example, is a nuclear property. It would follow that all A-objects fail to exemplify this property. So all A-objects would exemplify [λx ~Bx] ("$\bar{B}$"), where "B" denotes being blue. Then by Theorem 5, all Forms participate in $\Phi_{\bar{B}}$. So would $\Phi_{\bar{B}}$.

Consider now the Third Man Argument. This is a puzzle which commentators say Plato produces in the Parmenides (132aff.).[7] The puzzle is that several of Plato's principles about the Forms seem to be jointly inconsistent. We've seen two of these principles: (OMP) (Theorem 4) and the Uniqueness Principle (Theorem 2). There are two others: the Self-Predication Principle and the Non-Identity Principle:

(SP) The Form of F is F

(NI) If something participates in the Form of F, then it is not identical with that Form.

We can prove a contradiction if we assume that there are two distinct F-things x and y. By (OMP), there is a Form of F in which x and y participate. By (SP), the Form of F is an F-thing. By (NI), it is distinct from x and y. But then, (OMP) guarantees that there is another Form of F in which x and the first Form participate. Then (NI) yields the conclusion that the latter Form must be distinct from the first. But this violates the uniqueness principle, which says that the Form of F is unique.

On the theory we've presented, NI must be false. We can derive its negation as a theorem:

Theorem 8: $\sim(x)(Part(x,\Phi_F) \rightarrow x \neq \Phi_F)$

Proof: Consider $\Phi_{\overline{E!}}$.

So by rejecting NI, we dissolve the puzzle.

However, it is worthwhile to examine (SP). If we translate (SP) into our language as Φ_F exemplifies F, then it must be false. This time, consider $\Phi_{E!}$. But if we translate (SP) into our language as Φ_F encodes F, then it turns out to be Theorem 3. Does the word "is" in the (SP) principle mesh the distinction between exemplifying and encoding a property?

Of course we can't generalize on this one example, but we can look for further evidence for thinking that the "is" of English is ambiguous. Maybe we have an option of translating a sentence involving the predicative "is" as an exemplification or as an encoding formula. And in case there is such an ambiguity, let us now stipulate that whenever we use the word "is" in what follows, we shall mean "exemplifies."

We may conclude, with respect to the Third Man Argument, that our theory rules that (OMP) and (U) (Uniqueness Principle) are true, that (NI) is false, and that (SP) has a true reading and a false one. Since we abandon the (NI) principle, further research should be directed toward the question of how deeply Plato was committed to it.

Finally, we discuss the Sophist. The four assertions by Plato in that work that we discuss are ones which, taken together, are somewhat mysterious. Many scholars regard Plato's theory of Forms as his attempt to reconcile two major philosophical schools of thought. The first was the school of Parmenides, founded as the view that the world had to be considered as a whole without parts, without motion and change, and without generation and decay. The opposing school (Thales, Anaximander, Anaximenes, Heraclitus, Empedocles, and the Atomists) denied this and attempted to isolate the elementary parts of the world, the interaction of which was responsible for motion, change, generation, and decay. Plato's Forms were entities he postulated to capture certain truths of the Parmenidean school--they were changeless, motionless, and eternal. Yet Plato allowed that there were ordinary objects which moved, changed, came into being, and passed away. But, apparently, he supposed them to have a lesser degree of reality.

Plato's attempt to capture the Parmenidean truths wasn't completely successful. Some Forms gave him trouble, especially the ones which reflected some of the more mundane things in the world. He could never quite accept the fact that there were Forms with respect to hair, dirt, or mud. And the Form of motion--did it move? If so, how could it remain a Form? Forms were supposed to be motionless. Given the (SP) principle, how could there be a real Form of Motion if it didn't move? And how do the Forms of Motion and Rest interact with each other?

In this context, the following four assertions by Plato in the Sophist seem mysterious:

(1) Rest and Motion are completely opposed to one another (250a)

(2) Rest and Motion are real (250a)

(3) Reality must be some third thing (250b)

(4) In virtue of its own nature, then, reality is neither at rest nor in movement (250c)

To analyze these assertions, we need the following definitions and (reasonable) assumptions, where "M" denotes being in motion.

D_8 Being at rest ("R") $=_{dfn}$ $[\lambda x\ {\sim}Mx]$

A_1 Being in motion is a nuclear property

A_2 Being in motion, being at rest, and Platonic existence are distinct properties

(1) may be interpreted as a true statement about the Forms. Consider (1a):

(1a) $(x)(Part(x,\Phi_M) \equiv {\sim}Part(x,\Phi_R))$

This is provable, given D_8, λ-EQUIVALENCE, and Theorem 5. That is, by

D_8 and λ-EQUIV, something exemplifies being in motion iff it fails to exemplify being at rest. So by Theorem 5, something participates in Φ_M iff it fails to participate in Φ_R.

There is also an uncharitable way to interpret (1) as a statement about the Forms. Consider (1b):

(1b) $\sim R\Phi_M$ & $\sim M\Phi_R$ ("The Form of Motion doesn't exemplify being at rest and the Form of Rest doesn't exemplify being in motion")

This is false, since by A1, D_8, the AUXILIARY HYPOTHESIS, and λ-EQUIV, the Form of Motion does exemplify the negation of a nuclear property.

Consider (2), "Rest and motion are real." (2a) seems to be a good candidate for translating it:

(2a) $\bar{E}!\Phi_R$ & $\bar{E}!\Phi_M$ ("The Forms of Rest and Motion exemplify Platonic existence")

(2a) is a theorem. We also know that both Φ_R and Φ_M participate in $\Phi_{\bar{E}!}$. If we define "blend with" as "participate in," we get that both of these Forms blend with Being or Reality. (3) could be read as (3a):

(3a) $\Phi_R \neq \Phi_{\bar{E}!}$ & $\Phi_M \neq \Phi_{\bar{E}!}$

This is provable from assumption A2.

Finally, we consider (4), "In virtue of its own nature, reality is neither at rest nor in motion." (4) is another example of a sentence which turns out false when we read the copula "is" as exemplification and true when read as encoding. Consider (4a):

(4a) $\sim R\Phi_{\bar{E}!}$ & $\sim M\Phi_{\bar{E}!}$

Since we've defined Platonic Being, or Reality, as $\Phi_{\bar{E}!}$, (4a) captures (4) when "is" is read as "exemplifies." (4a) is false since $\Phi_{\bar{E}!}$ exemplifies being at rest. But consider (4b):

(4b) $\sim\Phi_{\bar{E}!}R$ & $\sim\Phi_{\bar{E}!}M$

The key to seeing that this might be right comes from the following definition:

D_9 The nature of $\Phi_F =_{dfn} F$

The nature of a Form is the property it encodes. Thus, we read "in virtue of its own nature" as a clue to thinking that Plato is going to conclude something about the fact that $\bar{E}!$ is central to the identity of $\Phi_{\bar{E}!}$. Assumption A_2 tells us that the nature of $\Phi_{\bar{E}!}$ is distinct from the natures of Φ_R and Φ_M. So (4b) is a theorem.

Assertion (4) has always been rather puzzling to me, and I think it is interesting that the distinction between exemplifying and encoding a property has helped us to find a true reading for it.

Is there, after all, some unity to the history of philosophy? Do we have here a *prima facie* link between Plato, Meinong, Mally, and the theory of abstract objects? Maybe further investigations along the above lines will help us to answer these questions.

§2. Modelling the Round Square, etc.

In our first encounter translating certain theoretical statements of natural language into the language of the theory, we discovered a few of them containing the copula "is" which turned out true when translated using an encoding formula yet which turned out false when translated using an exemplification formula. In this section, we look at a class of English sentences which exhibit this feature. These sentences can be recognized by the facts that: (1) they have the form "The $F_1, F_2, \ldots, F_n$ is F_i" ($1 \le i \le n$), and (2) there isn't (or couldn't be) an object which jointly exemplifies $F_1, F_2, \ldots, F_n$. Here are some examples:

(1) The set of sets which aren't members of themselves is a set of sets which aren't members of themselves (The F is F)

(2) The round square is round (The F,G is F)

(3) The existent golden mountain is existent, golden, and a mountain. (The F,G,H is F,G,H)

These sentences seem to be true *a priori*. But if we translate the description in (2), for example, as "the object which exemplifies roundness and squareness," then the description would fail to denote. It would then be hard to see how to account for the intuitive truth value of the sentence. And similar remarks apply to (1) and (3).

However, if we translate the description in (2) as "the object which encodes just roundness and squareness," and read the "is" as "encodes," we end up with the truth: the object which encodes just roundness and squareness encodes roundness. In a similar manner, we read (1) as: the object which encodes just being a set of sets which aren't self-members encodes being a set of sets which aren't self-members. And we do something similar for (3). The suggestion, then, is to translate "The $F_1, \ldots, F_n$ is F_i" as "the object which encodes just $F_1, \ldots, F_n$ encodes F_i."

To make this suggestion precise, we need to incorporate definite descriptions into our language and focus on a certain subset of them--those which can be used specifically to implement our suggestion. After investigating the interpretation and logic of our descriptions, we should be in a good position to translate the data properly.

The first thing to do is to specify that where ϕ is any formula with one free x-variable, $(\iota x)\phi$ ("the object x such that ϕ") is to be a complex object term of our language. Some examples are: $(\iota y)(E!y \ \& \ Typ)$ ("the object y such that y exists and bears the teaching relation to p"), and $(\iota x)(A!x \ \& \ (F)(xF \equiv \psi))$, where ψ is any formula where x isn't free ("the A-object which encodes just the properties which satisfy ψ").

Semantically, we interpret these descriptions as denoting the unique object which satisfies ϕ, if there is one, and as not denoting anything if there isn't one. We can make this precise as follows. Recalling the definitions of interpretation and assignment, we assign a denotation to these descriptions with respect to an interpretation and an assignment by the following stipulation:[8]

$$d_{I,\mathfrak{f}}((\iota x)\phi) = \begin{cases} o \text{ iff } (\exists\mathfrak{f}')(\mathfrak{f}' \overset{=}{_x} \mathfrak{f} \ \& \ \mathfrak{f}'(x)=o \ \& \ \mathfrak{f}' \text{ satisfies } \phi \ \& \\ \quad (\mathfrak{f}'')(\mathfrak{f}'' \overset{=}{_x} \mathfrak{f} \ \& \ \mathfrak{f}''(x)=o' \ \& \ \mathfrak{f}'' \text{ satisfies } \phi \rightarrow o'=o)) \\ \text{undefined, otherwise} \end{cases}$$

Consequently, the logical behaviour of our new descriptions is governed by the following logical axiom schema ("DESCRIPTIONS"):

$$\psi_{\nu}^{(\iota x)\phi} \equiv (\exists!y)(\phi_x^y \ \& \ \psi_{\nu}^y) \ \& \ (\exists!y)\phi_x^y, \text{ where } \psi \text{ is any atomic formula with one free object variable } \nu$$

This ensures that any closed atomic formula, ψ, containing a definite description $(\iota x)\phi$ is true iff there's a unique object y which satisfies both ϕ and ψ and there is a unique object satisfying ϕ. For example:

$$E!(\iota x)Txp \equiv (\exists!y)(Typ \ \& \ E!y) \ \& \ (\exists!y)Typ$$

This might say: the object x which exemplifies being the teacher of p exists iff there's a unique object y which exemplifies being the teacher of p and which exists, and there is a unique teacher of p.

With this rather normal understanding of definite descriptions, a certain subset of them prove to be interesting. These are the descriptions of the form: $(\iota x)(A!x \ \& \ (F)(xF \equiv \chi))$. We call this class of descriptions A-object descriptions, and the reason they are interesting is because when χ is any formula with no free x's, the resulting description always has a denotation. This is a result of the UNIQUENESS theorem schema for objects. In fact, UNIQUENESS and DESCRIPTIONS allow us to prove an interesting set of theorems governing the A-object descriptions:

THEOREM(S) ("A-DESCRIPTIONS"): $\chi_F^G \equiv (\iota x)(A!x \ \& \ (F)(xF \equiv \chi))G$

Proof: $(\rightarrow)$ Suppose G satisfies χ. By UNIQUENESS, there is a unique A-object, say b, which encodes exactly the properties which satisfy χ. So b encodes G. So there is a unique A-object which encodes exactly the properties satisfying χ and which encodes G. So by DESCRIPTIONS, the A-object which encodes exactly the properties which satisfy χ encodes G.[9] $(\leftarrow)$ By reversing the reasoning. ⊠

Using this theorem schema, it now becomes possible to prove certain facts regarding the objects denoted by A-descriptions. Consider $(\iota x)(A!x \ \& \ (F)(xF \equiv F{=}R \vee F{=}S)$, where "R" denotes roundness and "S" denotes squareness. If we let $\chi = (F{=}R \vee F{=}S)$, then χ_F^R and χ_F^S. So by A-DESCRIPTIONS, $(\iota x)(A!x \ \& \ (F)(xF \equiv F{=}R \vee F{=}S))$ encodes both R and S, as we might have expected. In general, when $\chi = (G{=}F_1 \vee G{=}F_2 \vee \ldots \vee G{=}F_n)$, it is provable that:

$$(\iota x)(A!x \ \& \ (G)(xG \equiv G{=}F_1 \vee \ldots \vee G{=}F_n))F_i,$$

where $1 \leq i \leq n$, given A-DESCRIPTIONS.

This provides us with the key to the proper translation of our data. For simplicity, let's shorten A-descriptions by using restricted variables to range over A-objects. In fact, throughout the remainder of this work, we use z-variables to range over A-objects.[10] So our A-object descriptions now have the form: $(\iota z)(F)(zF \equiv \chi)$. Where "S" denotes being a set, "ε" denotes the membership relation, and where the other abbreviations are obvious, we may translate the descriptions in (1)-(3) as (a)-(c), respectively:

(a) $(\iota z)(F)(zF \equiv F=[\lambda x\ Sx \ \&\ (y)(y\varepsilon x \equiv Sy \ \&\ y\not\varepsilon y)])$[11]

(b) $(\iota z)(F)(zF \equiv F=R \vee F=S)$

(c) $(\iota z)(F)(zF \equiv F=E! \vee F=G \vee F=M)$

And in general, the descriptions in the class of English sentences we've singled out are to be translated as:

$$(\iota z)(G)(zG \equiv G=F_1 \vee G=F_2 \vee \ldots \vee G=F_n)$$

In the metalanguage, we will signal the fact that we intend this reading of the English definite article by writing "the_A."

Now let $(\iota z)\psi_1$, $(\iota z)\psi_2$, and $(\iota z)\psi_3$ abbreviate the descriptions in (a)-(c), respectively. We then translate (1)-(3) into our language as (1)'-(3)', respectively:

(1)' $(\iota z)\psi_1[\lambda x\ Sx \ \&\ (y)(y\varepsilon x \equiv Sy \ \&\ y\not\varepsilon y)]$

(2)' $(\iota z)\psi_2 R$

(3)' $(\iota z)\psi_3 E! \ \&\ (\iota z)\psi_3 G \ \&\ (\iota z)\psi_3 M$

(1)'-(3)' are all theorems, hence the *a priori* character of the English. In general, our translation of "The $F_1,\ldots,F_n$ is F_i," where there isn't (or couldn't be) an object which jointly exemplifies $F_1,\ldots,F_n$, will always be a theorem of the following form:

$$(\iota z)(G)(zG \equiv G=F_1 \vee \ldots \vee G=F_n)F_i$$

There is a closely related use of the English definite article. Here are some examples:

(4) The fountain of youth is a fountain

(5) The set of all non-self-membered sets is a set

(6) The existent golden mountain is colored

(7) Necessarily, the teacher of Aristotle is a teacher

These sentences will be represented with the help of a slightly modified version of our A-descriptions. For example, "the fountain of youth" shall be translated as "the A-object which encodes being a fountain of youth or any property implied by this property." To represent and interpret this reading of the definite article, we must define "F implies G" as necessarily, everything exemplifying F exemplifies G. So we postpone further investigation until the modal theory has been developed.

§3. The Problem of Existence

The property of existence has puzzled philosophers for years. The assertion that some particular thing fails to exemplify existence (or being) strangely carries with it a commitment to the existence (or being) of the very thing which serves as the subject of the assertion. This is partly a result of trying to keep the theory of language as simple as possible--we try to account for the truth of a sentence by supposing that the objects denoted by the object terms are in an extension of the relation denoted by the relation term. But when we have a true non-existence claim, talk about "the object denoted by the object name" seems illegitimate.

The theory we've developed so far is rather flexible with respect to this issue. Although we've defined "abstract" to mean "nonexistent," we didn't have to do it this way. We could either have all objects, abstract or otherwise, exemplify existence, or we could have called the property of nonexistence by a less puzzling name. We can develop alternative "versions" of our theory based on these suggestions. But before we do this, let's examine the present course in more detail.

For one thing, keeping things as they are preserves some intuitive data about objects. Before we are uncorrupted by philosophy, it is perfectly natural to say "the round square doesn't exist" or "Santa Claus doesn't exist." If we read the description in the former sentence as we did in the last section, then it is provable that the$_A$ round square doesn't exemplify existence.[12] In Chapter IV, §4, we make it plausible to think that the proper name in the latter sentence denotes an abstract object. Consequently, we have an account of sentences like these in accordance with a simple theory of language--the objects denoted by the object terms fail to be in an extension of the relation denoted by the relation term.

Another reason for preferring the present formulation of our theory is that it leaves us with a formal language which can be used to investigate the claim that there is a distinction in natural language between the quantifiers "there is" and "there exists." Some philosophers, myself included, believe that there is an exploitable difference in meaning between these two quantifiers of English. Our view can be made precise by investigating a language in which this difference in meaning might be represented. The language we have now is such a language. We could use "$(\exists x)\phi$" to express the fact that *there is* an x such that ϕ, and use "$(\exists x)(E!x \ \& \ \phi)$" to express the fact that *there exists* an x such that ϕ. In a theory which supposed that all the things there are exist, it would not be possible to do this.

Of course, other philosophers fail to see a distinction between "there is" and "there exists"--for them, everything there is exists. They will claim that the A-objects we've allowed into our ontology must *really be there*. So they must exist.

The following suggestions should show that our theory is flexible enough to accommodate the views of these philosophers. We could use the primitive predicate "A!" instead of "E!" (i.e., we could have taken the property of being abstract as primitive). We could then define:

$$x \text{ is concrete ("}Cx\text{")} =_{dfn} \sim A!x$$

$$x \text{ exists ("}E!x\text{")} =_{dfn} A!x \vee Cx$$

We could then revise NO-CODER as follows:

$$(x)(Cx \rightarrow \sim(\exists F)xF)$$

Finally, we could relabel "$=_E$" as "$=_C$," change E-IDENTITY to C-IDENTITY (i.e., $Cx \ \& \ Cy \ \& \ (F)(Fx \equiv Fy) \rightarrow x=_C y$), and redefine general identity (i.e., $x=y =_{dfn} x=_C y \vee A!x \ \& \ A!y \ \& \ (F)(xF \equiv yF)$). We would leave A-OBJECTS as it stands and call the theory which would result: VERSION 2.

On VERSION 2, it's provable that everything exists. VERSION 2 can do all the work the original theory can do. That's because the exemplification/encoding distinction, and the distinction between two kinds of objects, remain intact. Of course, we have to reapproach the analysis of data like "The round square doesn't exist" and "Santa

Claus doesn't exist." We still suppose that the object terms denote abstract objects, but we interpret "doesn't exist" as meaning that there are (exist) no concrete objects which exemplify the properties encoded by these abstract objects.

I think that philosophers who insist that VERSION 2 is the only correct version are mistaken. The theory remains useful no matter which of the two versions you choose. Meinong used to say that "the Object as such stands beyond being and non-being" and that "the Object is by nature indifferent to being."[13] I'm not a Meinong scholar, so I don't claim to know what Meinong meant by this "doctrine of ausser-sein," and I don't suppose that he had these two versions of our theory before his mind when he said things like this. Nevertheless, something like these cryptic utterances of Meinong are relevant here. It just doesn't matter whether you conceive of A-objects as existing or failing to exist. It mostly ends up a question of how you prefer to use the word "exists."

Maybe the word "exists" is an ambiguous word, one of the senses of which is a property which has a negation that also turns out to be a sense of the word.[14] To make this idea plausible, we could stick with the original version of the theory, and read "E!" as "real existence" and "$[\lambda x \sim E!x]$" as "Platonic existence." Now we have two kinds of existence, with A-objects exemplifying the latter kind. This reading of $[\lambda x \sim E!x]$, besides working to our advantage in §1, is further justified by the fact that in the modal theory which follows, A-objects end up having being in every possible world and the class of A-objects stays fixed from world to world. Platonic beings are necessary beings, and A-objects turn out to be necessary beings.[15] They, therefore, exhibit a more perfect kind of existence.[16]

So talking in terms of two kinds of existence is yet a third way of approaching the problems of existence. This means that we really don't have to commit ourselves on the question: Do A-objects fail to exist? Three equivalent versions of the theory decide the question in different ways. The version one prefers to go with will be mostly a result of a decision about which of the various senses of the word "exist" one prefers to use.[17]

CHAPTER II ENDNOTES

[1]The results in this section were detailed principally in two early papers [1979c] and [1979d]. I'd like to thank Cynthia Freeland for her assistance in locating the relevant passages in Plato's works.

[2]See also Phaedo, 100c7-e2, 101a.

[3]An orthodox theorist might suggest that Plato discovered that existential generalization (introduction) on predicate terms was a valid rule of inference. This would turn Plato into a language theorist, whereas on our view, he was doing metaphysics.

[4]Compare Parsons [1980], Ch. VIII, §5; also Castañeda [1974].

[5]See Timaeus, 52c.

[6]See Timaeus, 51e, 52a. Also, see The Republic, 518ff.

[7]See Vlastos' [1954]; and Strang [1971].

[8]Given the following definition, there will be terms in the language which might fail to denote. Consequently, LA4 (one of the logical axioms governing quantification) must be revised to cover such cases. It should now read as follows:

LA 4: $(\alpha)\phi \rightarrow ((\exists\beta)\beta{=}\tau \rightarrow \phi^{\tau}_{\alpha})$, where τ is substitutable for α

Also, the biconditionals in the first two clauses of the definition of satisfaction (I., §2, C.) must be understood as follows: the clauses will be false iff the sides of the biconditional have opposite truth value or the right side is undefined.

[9]For those who think more syntactically, let ϕ = A!x &

$(F)(xF \equiv \chi)$. Let $\psi = \nu G$. Then $\psi_{\nu}^{(\iota x)\phi} = (\iota x)(A!x \;\&\; (F)(xF \equiv \chi))G$. We then deduced the right side of DESCRIPTIONS using UNIQUENESS and the fact that G satisfies χ.

[10]As usual, with restricted variables:

(i) $(\exists z)\phi_x^z$ abbreviates $(\exists x)(A!x \;\&\; \phi)$

(ii) $(z)\phi_x^z$ abbreviates $(x)(A!x \rightarrow \phi)$

[11]One suggestion for understanding the ontological status of mathematical objects is to say explicitly which objects exist when formulating the relevant set of axioms. So, for example, we formulate axioms for set theory as follows:

NULL: $(\exists x)(E!x \;\&\; Sx \;\&\; (y)(y \not\in x))$

UNIONS: $(x)(Sx \rightarrow (\exists y)(E!y \;\&\; Sy \;\&\; (w)(w\in y \equiv (\exists u)(u\in x \;\&\; w\in u))))$

POWER: $(x)(Sx \rightarrow (\exists y)(E!y \;\&\; Sy \;\&\; (w)(w\in y \equiv w\subseteq x)))$

SUBSET: For any formula ϕ with one free variable u, the following is an axiom: $(x)(Sx \rightarrow (\exists y)(E!y \;\&\; Sy \;\&\; (z)(z\in y \equiv z\in x \;\&\; \phi_u^z)))$

It should be clear how to then formulate INFINITE, REGULARITY, and REPLACEMENT. On this kind of formulation, it's provable that there *doesn't exist* an object which exemplifies being a set of all non-self-membered sets, though A-OBJECTS guarantees that some objects encode this property.

Note also that we could get much the same effect by formulating the axioms of set theory in the usual way with the addition of the extra axiom that all sets exist.

[12]Note that it is also provable that the$_A$ existent golden mountain doesn't exist. We cannot prove the contingent truth that there doesn't exist an object which exemplifies all the properties that the$_A$

existent golden mountain encodes.

[13]See Meinong [1904], p. 86 (§4 of "Über Gegenstandstheorie").

[14]When we talk about the various senses of an ambiguous property name, we mean the various properties it denotes. We are not referring to its "Fregean" sense.

[15]Necessary beings exist in every possible world or fail to exist in every world--they do not go in and out of existence from world to world. In the next chapter, we redefine A-objects as objects which necessarily fail to exist.

[16]Some philosophers may hesitate because they prefer to reserve the term "Platonic existence" to describe properties, relations, and propositions. But we've seen that a certain class of A-objects behave like the Forms and this is how we justify calling the kind of existence A-objects exemplify "Platonic." Those who now hesitate probably used "Platonic" in connection with properties, etc., in the first place because of the orthodox view that Plato's Forms just are properties.

Those philosophers who still wish to preserve "Platonic existence" for properties, etc., would at least agree that on this usage, the term denotes a (higher order) property of properties. But that won't have bearing on the important question we're now facing--whether it's plausible to think of the negation of the first level property of existence as some *special* kind of existence.

[17]Those philosophers who believe both that properties exist and that sets exist may wonder why we can't dispense with A-objects by modelling them as sets of properties. For the reasons why we can't do this, see footnote 23, Chapter I.

C H A P T E R I I I

THE MODAL THEORY OF ABSTRACT OBJECTS (WITH PROPOSITIONS)

§1. The Language

A. Primitive symbols. To the language of Chapter I, we add the "$\Box$"-operator (to express the English sentential adverb "necessarily") and names of (and variables ranging over) propositions. By allowing the superscripts on the primitive relation terms to reach zero, we obtain names and variables for propositions. For convenience, we use P^0, Q^0, R^0, ... and F^0, G^0, H^0, ... as names and variables, respectively, for propositions. Officially, however, our new list of primitive symbols is as follows:

1. Primitive object terms

   Names: $a_1, a_2, \ldots$

   Variables: $x_1, x_2, \ldots$

2. Primitive n-place relation terms

   Names: $P^n_1, P^n_2, \ldots, =_E, E!$ $\quad n \geq 0$

   Variables: $F^n_1, F^n_2, \ldots$

3. Connectives: ~, &
4. Quantifier: $\exists$
5. Lambda: λ
6. Iota: ι
7. Box: $\Box$
8. Parentheses and brackets: (,), [,]

B. Formulas and terms. We simultaneously define (propositional) formula, object term, and n-place relation term, inductively, as follows:

1. All primitive object terms are object terms and all primitive n-place relation terms are n-place relation terms
2. If ρ^0 is any zero-place relation term, ρ^0 is a (propositional) formula
3. Atomic exemplification: If ρ^n is any n-place relation term and $o_1,\ldots,o_n$ are any object terms, $\rho^n o_1 \ldots o_n$ is a (propositional) formula
4. Atomic encoding: If ρ^1 is any one-place relation term, and o is any object term, $o\rho^1$ is a formula
5. Molecular, Quantified, and Modal: If ϕ and ψ are any (propositional) formulas, and α is any (object) variable, then $(\sim\phi)$, $(\phi \;\&\; \psi)$, $(\exists\alpha)\phi$, and $(\Box\phi)$ are (propositional) formulas
6. Object descriptions: If ϕ is any formula with one free object variable x, then $(\iota x)\phi$ is an object term
7. Complex n-place relation terms: If ϕ is any propositional formula, and $\nu_1,\ldots,\nu_n$ are any object variables which may or may not be free in ϕ, then $[\lambda\nu_1\ldots\nu_n\ \phi]$ is an n-place relation term ($n\geq 1$) and ϕ itself is a zero-place relation term

In addition to the standard abbreviations for the connectives and quantifiers, we use $\Diamond\phi$ to abbreviate $\sim\Box\sim\phi$. However, we now define:

D_1 x is abstract ("A!x") $=_{dfn}$ $\Box\sim E!x$

D_2 x is a possibly existing object $=_{dfn}$ $\Diamond E!x$

Here are some examples of schemas and formulas: $\Box Q$ ("it is necessary that Q"); $\Box(\exists x)(A!x \;\&\; (F)(xF \equiv \phi))$ ("necessarily, some abstract object encodes exactly the properties satisfying ϕ"); $\Diamond(\exists y)(F)(xF \rightarrow Fy)$ ("possibly, there is an object which exemplifies every property x encodes"); and $(x)(\Diamond E!x \rightarrow \sim(\exists F)(xF)$ ("possibly existing objects fail to encode any properties").

We say that a formula ϕ <u>necessarily implies</u> a formula ψ ("$\phi \Rightarrow \psi$") iff $\Box(\phi \rightarrow \psi)$. ϕ is necessarily equivalent to ψ ("$\phi \Leftrightarrow \psi$") iff $\Box(\phi \equiv \psi)$.

There are two kinds of complex terms--object descriptions and complex n-place relation terms. Modal formulas may appear in both. For example, $(\iota x)(A!x \ \& \ (F)(xF \equiv F{=}R)$ is an object description which reads: the abstract object which encodes just R.[1] The inductive clause for complex n-place relation terms differs from its counterpart in the elementary theory in three important respects: (1) it allows modal formulas to appear after λ's if the formula is propositional, (2) it allows λ's to bind variables which aren't free in the ensuing formula, and (3) it allows propositional formulas themselves to be relation terms. Here are some examples of new complex n-place relation terms: $[\lambda xy \ \Box Qb]$ ("being a first thing and a second thing such that necessarily, b exemplifies Q"); $[\lambda x \ \Box(E!x \rightarrow Px)]$ ("being an x such that necessarily, if x exists, x exemplifies P"); $\Box Gb$ ("b exemplifies G essentially").

As before, τ is a <u>term</u> iff either τ is an object term or τ is an n-place relation term.

C. Identity definitions.

D_3 $\quad x{=}y =_{dfn} x{=}_E y \ v \ (A!x \ \& \ A!y \ \& \ \Box(F)(xF \equiv yF))$

D_4 $F^1 = G^1 =_{dfn} \Box(x)(xF^1 \equiv xG^1)$

D_5 $F^n = G^n =_{dfn} (x_1)\ldots(x_{n-1})([\lambda y\ F^n yx_1\ldots x_{n-1}] = [\lambda y\ G^n yx_1\ldots x_{n-1}]$ & $[\lambda y\ F^n x_1 yx_2\ldots x_{n-1}] = [\lambda y\ G^n x_1 yx_2\ldots x_{n-1}]$ & ... & $[\lambda y\ F^n x_1\ldots x_{n-1}y] = [\lambda y\ G^n x_1\ldots x_{n-1}y])$ $(n>1)$

D_6 $F^0 = G^0 =_{dfn} [\lambda y\ F^0] = [\lambda y\ G^0]$

That is, F^0 and G^0 are identical iff the property of being such that F^0 is identical with the property of being such that G^0. We shall see the usefulness and insightfulness of this definition when we prove that there is a unique actual world.

As usual, these definitions for identity will be governed by the proper axiom for identity (and the rule of inference, =E, derivable from it).

§2. The Semantics

A. Interpretations. An interpretation, I, of our modified second order modal language is any octuple, $\langle W, w_0, D, R, ext_w, L, ext_A, F\rangle$, which meets the conditions described in this subsection. The first member of I is a non-empty class, W, called the class of possible worlds.[2] The second member of I, w_0, is chosen from W and is called the actual world. The third member, D, is a non-empty class and is called the domain of objects. The fourth member, R, is also a non-empty class, and is called the domain of relations. R is the union of a sequence of non-empty classes $R_0, R_1, R_2, \ldots$, i.e., $R = \bigcup_{n\geq 0} R_n$. Each R_n is called the class of n-place relations (we call R_1 the class of properties, and R_0 the class of propositions). R must be closed under all the logical functions specified in the sixth member of the interpretation (L).

The fifth, sixth, and seventh members of I impose a structure on W, D, and R. The fifth member of I, ext_w, is a function which maps $R_n \times W$ into D^n, where $n \geq 1$, and which maps $R_0 \times W$ into $\{T,F\}$. $ext_w(r^n)$ is the exemplification extension ("extension$_w$" in the metalanguage) of relation r^n at world w.

The sixth member of I, L, is a class of logical functions which operate in a manner similar to their counterparts in the semantics of the elementary theory. However, we: (1) add two additional functions, VAC_i ("i-vacuous expansion") and NEC ("necessitation"), (2) constrain the extensions$_w$ of the complex relations resulting from all the logical functions at every possible world, and (3) allow $PLUG_i$ and $PROJ_i$ to operate on properties, allow $CONJ$ and NEG to operate on propositions, and allow VAC_i and NEC to operate on all relations. The definitions which make these three major changes precise go as follows:

(a) $PLUG_i$, for each i, $1 \leq i \leq n$, is a function mapping $(R_i \cup R_{i+1} \cup \ldots) \times D$ into $(R_{i-1} \cup R_i \cup \ldots)$ subject to the conditions:

(1) for $n>1$, $ext_w(PLUG_i(r^n,o)) =$
$\{\langle o_1,\ldots,o_{i-1},o_{i+1},\ldots,o_n\rangle \mid$
$\langle o_1,\ldots,o_{i-1},o,o_{i+1},\ldots,o_n\rangle \in ext_w(r^n)\}$

(2) for $n=1$, $ext_w(PLUG_1(r^1,o)) =$
$$\begin{cases} T \text{ iff } o \in ext_w(r^1) \\ F \text{ otherwise} \end{cases}$$

(b) $PROJ_i$, for each i, $1 \leq i \leq n$, is a function mapping $(R_i \cup R_{i+1} \cup \ldots)$ into $(R_{i-1} \cup R_i \cup \ldots)$ subject to the conditions:

(1) for $n>1$, $ext_w(PROJ_i(r^n)) =$

$\{<o_1,\ldots,o_{i-1},o_{i+1},\ldots,o_n> \mid$

$(\exists o)(<o_1,\ldots,o_{i-1},o,o_{i+1},\ldots,o_n>$

$\varepsilon\ ext_w(r^n))\}$

(2) for $n=1$, $ext_w(PROJ_1(r^1)) =$

$$\begin{cases} T \text{ iff } (\exists o)(o\ \varepsilon\ ext_w(r^1)) \\ F \text{ otherwise} \end{cases}$$

(c) $CONV_{i,j}$, for each i,j, $1\leq i<j\leq n$, is a function mapping $(R_j \cup R_{j+1} \cup \ldots)$ into $(R_j \cup R_{j+1} \cup \ldots)$ subject to the condition:

$ext_w(CONV_{i,j}(r^n)) =$

$\{<o_1,\ldots,o_{i-1},o_j,o_{i+1},\ldots,o_{j-1},o_i,o_{j+1},\ldots,o_n> \mid$

$<o_1,\ldots,o_i,\ldots,o_j,\ldots,o_n>\ \varepsilon\ ext_w(r^n)\}$

(d) $REFL_{i,j}$, for each i,j, $1\leq i<j\leq n$, is a function mapping $(R_j \cup R_{j+1} \cup \ldots)$ into $(R_{j-1} \cup R_j \cup \ldots)$ subject to the condition:

$ext_w(REFL_{i,j}(r^n)) =$

$\{<o_1,\ldots,o_i,\ldots,o_{j-1},o_{j+1},\ldots,o_n> \mid$

$<o_1,\ldots,o_i,\ldots,o_j,\ldots,o_n>\ \varepsilon\ ext_w(r^n)$ and $o_i=o_j\}$

(e) VAC_i, for each i, $1\leq i\leq n+1$, is a function mapping $(R_{i-1} \cup R_i \cup \ldots)$ into $(R_i \cup R_{i+1} \cup \ldots)$ subject to the conditions:

(1) for $n\geq 1$, $ext_w(VAC_i(r^n)) =$

$\{<o_1,\ldots,o_{i-1},o,o_i,o_{i+1},\ldots,o_n> \mid$

$<o_1,\ldots,o_i,\ldots,o_n>\ \varepsilon\ ext_w(r^n)\}$

(2) for n=0, $ext_w(VAC_1(r^0)) = \{o \mid ext_w(r^0) = T\}$

(f) *CONJ* is a function mapping $(R_0 \cup R_1 \cup \ldots) \times (R_0 \cup R_1 \cup \ldots)$ into $(R_0 \cup R_1 \cup \ldots)$ subject to the following conditions:

(1) for n≥1, m≥1, $ext_w(CONJ(r^n, s^m)) =$

$\{<o_1, \ldots, o_n, o_1', \ldots, o_m'> \mid <o_1, \ldots, o_n>$
$\in ext_w(r^n)$ and $<o_1', \ldots, o_m'> \in ext_w(s^m)\}$

(2) for n=0, m≥1, $ext_w(CONJ(r^0, s^m)) =$

$\{<o_1, \ldots, o_m> \mid ext_w(r^0) = T$ and
$<o_1, \ldots, o_m> \in ext_w(s^m)\}$

(3) for n≥1, m=0, $ext_w(CONJ(r^n, s^0)) =$

$\{<o_1, \ldots, o_n> \mid <o_1, \ldots, o_n>$
$\in ext_w(r^n)$ and $ext_w(s^0) = T\}$

(4) for n=0, m=0, $ext_w(CONJ(r^0, s^0)) =$

$$\begin{cases} T \text{ iff } ext_w(r^0) = T \text{ and } ext_w(s^0) = T \\ F \text{ otherwise} \end{cases}$$

(g) *NEG* is a function mapping $(R_0 \cup R_1 \cup \ldots)$ into $(R_0 \cup R_1 \cup \ldots)$ subject to the conditions:

(1) for n≥1, $ext_w(NEG(r^n)) =$

$\{<o_1, \ldots, o_n> \mid <o_1, \ldots, o_n>$
$\notin ext_w(r^n)\}$

(2) for n=0, $ext_w(NEG(r^0)) =$

$$\begin{cases} T \text{ iff } ext_w(r^0) = F \\ F \text{ otherwise} \end{cases}$$

(h) NEC is a function mapping $(R_0 \cup R_1 \cup \ldots)$ into $(R_0 \cup R_1 \cup \ldots)$ subject to the conditions:

(1) for $n \geq 1$, $ext_w(NEC(r^n)) =$
$\{<o_1, \ldots, o_n> \mid (w')(<o_1, \ldots, o_n> \varepsilon\ ext_{w'}(r^n))\}$

(2) for $n=0$, $ext_w(NEC(r^0)) =$
$\begin{cases} T \text{ iff } (w')(ext_{w'}(r^0) = T) \\ F \text{ otherwise} \end{cases}$

This completes the definitions of the logical functions.[3] The seventh member of I, ext_A, is a function which maps R_1 into D. $ext_A(r^1)$ is called the encoding extension ("extension$_A$" in the metalanguage) of r^1.

The final member of I, the F function, maps the simple names of the language to elements of the appropriate domain. For each object name a_i, $F(a_i)\ \varepsilon\ D$. For each relation name P^n_i, $F(P^n_i)\ \varepsilon\ R_n$. We call $ext_w(F(E!))$ the set of existing objects at w ("E_w"). We call $ext_{w_0}(F(E!))$ the set of existing objects (i.e., $E =_{dfn} E_{w_0}$). We call $\{o \mid (\exists w)(o\ \varepsilon\ ext_w(F(E!)))\}$ the set of possibly existing objects ("PE"). The complement of PE on D is called the set of abstract objects ("A").

B. Assignments and denotations.

Partitioning the λ-expressions. Since we have λ-expressions in the modal language which were ill-formed in the elementary language, we must incorporate rules to classify the new possibilities. These new rules correspond to VAC_i and NEC--they help to classify λ-expressions with vacuously bound λ-variables and with $\Box$'s.

The following six major rules partition the class of λ-expressions into nine equivalence classes. If μ is an arbitrary λ-expression, $[\lambda\nu_1 \ldots \nu_n\ \phi]$, μ is defined as follows:

1. If $(\exists i)(1 \leq i \leq n$ and ν_i doesn't occur free in ϕ and i is the least such number), then μ is the <u>i^{th}-vacuous expansion of</u> $[\lambda\nu_1 \ldots \nu_{i-1}\nu_{i+1} \ldots \nu_n\ \phi]$.
2. If μ is not an i^{th}-vacuous expansion, then if $(\exists i)(1 \leq i \leq n$ and ν_i is not the i^{th} free object variable in ϕ and ν_i is the least such number), then where ν_j is the i^{th} free object variable in ϕ, μ is the <u>i,j^{th}-conversion of</u> $[\lambda\nu_1 \ldots \nu_{i-1}\nu_j\nu_{i+1} \ldots \nu_{j-1}\nu_i\nu_{j+1} \ldots \nu_n\ \phi]$.
3. If μ is neither of the above, then
   (a) if $\phi = (\sim\psi)$, μ is the <u>negation of</u> $[\lambda\nu_1 \ldots \nu_n\ \psi]$
   (b) if $\phi = (\psi\ \&\ \chi)$, and ψ and χ have no free object variables in common, then where $\nu_1, \ldots, \nu_p$ are the variables in ψ and $\nu_{p+1}, \ldots, \nu_n$ are the variables in χ, μ is the <u>disjoint conjunction of</u> $[\lambda\nu_1 \ldots \nu_p\ \psi]$ <u>and</u> $[\lambda\nu_{p+1} \ldots \nu_n\ \chi]$.
   (c) if $\phi = (\exists\nu)\psi$ and ν is the i^{th} free object variable in ϕ, then μ is the <u>i^{th}-projection of</u> $[\lambda\nu_1 \ldots \nu_{i-1}\nu\nu_i\nu_{i+1} \ldots \nu_n\ \psi]$
   (d) if $\phi = (\Box\psi)$, then μ is the <u>necessitation of</u> $[\lambda\nu_1 \ldots \nu_n\ \psi]$.
4. If μ is none of the above, then if $(\exists i)(1 \leq i \leq n$ and ν_i occurs free in more than one place in ϕ and i is the least such number), then where:

(a) k is the number of free object variables between the first and second occurrences of ν_i,

(b) ϕ' is the result of replacing the second occurrence of ν_i with a new variable ν, and

(c) $j = i+k+1$,

μ is the <u>i,jth-reflection of</u> $[\lambda\nu_1 \ldots \nu_{i+k}\nu\nu_j \ldots \nu_n\ \phi']$.

5. If μ is none of the above, then if o is the left most object term occurring in ϕ, then where:

(a) j is the number of free variables occurring before o,

(b) ϕ' is the result of replacing the first occurrence of o by a new variable ν, and

(c) $i = j+1$,

μ is the <u>ith-plugging of</u> $[\lambda\nu_1 \ldots \nu_j\nu\nu_{j+1} \ldots \nu_n\ \phi']$ <u>by</u> o.

6. If μ is none of the above, then

(a) ϕ is atomic

(b) $\nu_1, \ldots, \nu_n$ is the order in which these variables first occur in o,

(c) $\mu = [\lambda\nu_1 \ldots \nu_n\ \rho^n\nu_1 \ldots \nu_n]$, for some relation term ρ^n, and

(d) μ is called <u>elementary</u>.

In addition to the examples we saw from the elementary theory, we now have: $[\lambda x\ \Box(Px\ \&\ {\sim}Qx)]$ is the necessitation of $[\lambda x\ Px\ \&\ {\sim}Qx]$; $[\lambda xwy\ Px\ \&\ Qy]$ is the 2^{nd}-vacuous expansion of $[\lambda xy\ Px\ \&\ Qy]$; $[\lambda xvwy\ Px\ \&\ Qy]$ is the 2^{nd}-vacuous expansion of $[\lambda xwy\ Px\ \&\ Qy]$; etc.

<u>*I*-assignments</u>. If given an interpretation I of our language, an I-assignment, f, will be any function defined on the primitive variables of the language satisfying the following two conditions:[4]

1. where ν is any object variable, $f(\nu) \in \mathcal{D}$

2. where π^n is any relation variable, $f(\pi^n) \in R_n$

Denotations. If given an interpretation I of our language, and an I-assignment $\mathfrak{f}$, then we recursively define the denotation of term τ with respect to interpretation I and I-assignment $\mathfrak{f}$ ("$d_{I,\mathfrak{f}}(\tau)$") as follows:

1. where κ is any primitive name, $d_{I,\mathfrak{f}}(\kappa) = F(\kappa)$
2. where α is any primitive variable, $d_{I,\mathfrak{f}}(\alpha) = \mathfrak{f}(\alpha)$
3. where $(\iota x)\phi$ is any object description,

$$d_{I,\mathfrak{f}}((\iota x)\phi) = \begin{cases} o \text{ iff } (\exists \mathfrak{f}')(\mathfrak{f}' \overset{}{\underset{x}{=}} \mathfrak{f} \;\&\; \mathfrak{f}'(x)=o \;\&\; \mathfrak{f}' \text{ satisfies } \phi \\ \quad \text{with respect to } w_0 \;\&\; (\mathfrak{f}'')(\mathfrak{f}'' \underset{x}{=} \mathfrak{f} \;\&\; \mathfrak{f}''(x)=o' \;\&\; \\ \quad \mathfrak{f}'' \text{ satisfies } \phi \text{ with respect to } w_0 \rightarrow o'=o)) \\ \text{undefined, otherwise,} \end{cases}$$

where satisfaction is defined as in subsection C.

4. where $[\lambda\nu_1\ldots\nu_n\ \rho^n\nu_1\ldots\nu_n]$ is any elementary λ-expression,

$d_{I,\mathfrak{f}}([\lambda\nu_1\ldots\nu_n\ \rho^n\nu_1\ldots\nu_n]) = d_{I,\mathfrak{f}}(\rho^n)$

5. where μ is the i^{th}-plugging of ξ by o,

$d_{I,\mathfrak{f}}(\mu) = PLUG_i(d_{I,\mathfrak{f}}(\xi),\ d_{I,\mathfrak{f}}(o))$

6. where μ is the i^{th}-projection of ξ,

$d_{I,\mathfrak{f}}(\mu) = PROJ_i(d_{I,\mathfrak{f}}(\xi))$

7. where μ is the i,j^{th}-conversion of ξ,

$d_{I,\mathfrak{f}}(\mu) = CONV_{i,j}(d_{I,\mathfrak{f}}(\xi))$

8. where μ is the i,j^{th}-reflection of ξ,

$d_{I,\mathfrak{f}}(\mu) = REFL_{i,j}(d_{I,\mathfrak{f}}(\xi))$

9. where μ is the i^{th}-vacuous expansion of ξ,

$d_{I,\mathfrak{f}}(\mu) = VAC_i(d_{I,\mathfrak{f}}(\xi))$

10. where μ is the disjoint conjunction of ξ and ζ,

$d_{I,\mathfrak{f}}(\mu) = CONJ(d_{I,\mathfrak{f}}(\xi),\ d_{I,\mathfrak{f}}(\zeta))$

11. where μ is the negation of ξ,

$$d_{I,f}(\mu) = NEG(d_{I,f}(\xi))$$

12. where μ is the necessitation of ξ,

$$d_{I,f}(\mu) = NEC(d_{I,f}(\xi))$$

13. where ϕ is any propositional formula, $d_{I,f}(\phi)$ is defined as follows:

(a) if ϕ is a primitive zero-place term, $d_{I,f}(\phi)$ is defined above

(b) if $\phi = \rho^n o_1 \ldots o_n$, $d_{I,f}(\phi) =$

$PLUG_1(PLUG_2(\ldots(PLUG_n(d_{I,f}(\rho^n), d_{I,f}(o_n)), \ldots),$
$d_{I,f}(o_2)), d_{I,f}(o_1))$

(c) if $\phi = (\sim\psi)$, $d_{I,f}(\phi) = NEG(d_{I,f}(\psi))$

(d) if $\phi = (\psi \& \chi)$, $d_{I,f}(\phi) = CONJ(d_{I,f}(\psi), d_{I,f}(\chi))$

(e) if $\phi = (\exists\nu)\psi$, $d_{I,f}(\phi) = PROJ(d_{I,f}([\lambda\nu\ \psi]))$

(f) if $\phi = \Box\psi$, $d_{I,f}(\phi) = NEC(d_{I,f}(\psi))$

Here are some examples of λ-expressions and their denotations with respect to a given I and f:

$$d_{I,f}([\lambda x\ \Box(Px \ \&\ {\sim}Qx)]) = NEC(d_{I,f}([\lambda x\ Px \ \&\ {\sim}Qx]))$$

$$d_{I,f}([\lambda xvwy\ Px \ \&\ Qy]) = VAC_2(d_{I,f}([\lambda xwy\ Px \ \&\ Qy]))$$

$$d_{I,f}([\lambda xwy\ Px \ \&\ Qy]) = VAC_2(d_{I,f}([\lambda xy\ Px \ \&\ Qy]))$$

$$d_{I,f}([\lambda xy\ Px \ \&\ Qy]) = CONJ(d_{I,f}(P), d_{I,f}(Q))$$

$$d_{I,f}([\lambda y\ \Box Gb]) = VAC_1(NEC(PLUG_1(d_{I,f}(G), d_{I,f}(b))))$$

$$d_{I,f}(\Box(E!b \rightarrow Gb)) = NEC(NEG(CONJ(PLUG_1(d_{I,f}(E!), d_{I,f}(b)), NEG(PLUG_1(d_{I,f}(G), d_{I,f}(b)))))).$$

C. Satisfaction. If we're given an interpretation I, and an I-assignment f, we may define f satisfies ϕ with respect to world w as follows:

1. If ϕ is any primitive zero-place term, $\mathfrak{s}$ satisfies ϕ with respect to w iff $ext_w(d_{I,\mathfrak{s}}(\phi)) = T$
2. If $\phi = \rho^n o_1 \ldots o_n$, $\mathfrak{s}$ satisfies ϕ with respect to w iff $<d_{I,\mathfrak{s}}(o_1), \ldots, d_{I,\mathfrak{s}}(o_n)> \varepsilon\ ext_w(d_{I,\mathfrak{s}}(\rho^n))$
3. If $\phi = o\rho^1$, $\mathfrak{s}$ satisfies ϕ with respect to w iff $d_{I,\mathfrak{s}}(o)\ \varepsilon\ ext_A(d_{I,\mathfrak{s}}(\rho^1))$
4. If $\phi = (\sim\psi)$, $\mathfrak{s}$ satisfies ϕ with respect to w iff $\mathfrak{s}$ fails to satisfy ψ with respect to w
5. If $\phi = (\psi \& \chi)$, $\mathfrak{s}$ satisfies ϕ with respect to w iff $\mathfrak{s}$ satisfies both ψ and χ with respect to w
6. If $\phi = (\exists\alpha)\psi$, $\mathfrak{s}$ satisfies ϕ with respect to w iff $(\exists\mathfrak{s}')(\mathfrak{s}' \overset{}{=}_{\alpha} \mathfrak{s}$ and $\mathfrak{s}'$ satisfies ψ with respect to $w)$
7. If $\phi = (\Box\psi)$, $\mathfrak{s}$ satisfies ϕ with respect to w iff $(w')(\mathfrak{s}$ satisfies ψ with respect to $w')$

D. Truth under an interpretation. ϕ is true under interpretation I iff every I-assignment $\mathfrak{s}$ satisfies ϕ with respect to w_0. The definitions of validity and model remain the same.

§3. The Logic

A. Logical axioms. Each of the following logical axioms should be valid. The logical theorems which are provable from these axioms and the rules of inference should also be valid.

We retain LA1-LA5 intact from the elementary theory. They have greater significance, however, since they cover a wider variety of formulas. For example, we now have modal formulas in the language and so $\Box P \to (Q \to \Box P)$ will be an instance of LA1. Also, LA4 is modified to cover descriptions which may fail to denote:

LA1: $\phi \to (\psi \to \phi)$

LA2: $\phi \to (\psi \to \chi) \to ((\phi \to \psi) \to (\phi \to \chi))$

LA3: $(\sim\phi \to \sim\psi) \to ((\sim\phi \to \psi) \to \phi)$

LA4: $(\alpha)\phi \to ((\exists\beta)\beta{=}\tau \to \phi^{\tau}_{\alpha})$, where τ is substitutable for α

LA5: $(\alpha)(\phi \to \psi) \to (\phi \to (\alpha)\psi)$, provided α isn't free in ϕ

In addition to these propositional and quantificational schemata, we also have the modal axioms. There are three standard modal axioms of S5, plus one other interesting modal consequence of our semantics:

LA6: $\Box\phi \to \phi$

LA7: $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$

LA8: $\Diamond\phi \to \Box\Diamond\phi$

LA9: $(x)(F)(\Diamond xF \to \Box xF)$

LA9 is a consequence of the fact that the encoding extension of a property is not relativized to a world. The conditions for satisfaction with respect to encoding formulas (§2, C., 3) are totally independent of the worlds. So if an encoding formula is true at some world, it is true at every world.

Finally, we have two schema which govern our complex terms--the first governs definite descriptions and the second governs λ-expressions.

DESCRIPTIONS: $\psi^{(\iota x)\phi}_{\nu} \equiv (\exists!y)(\phi^{y}_{x} \ \& \ \psi^{y}_{\nu}) \ \& \ (\exists!y)\phi^{y}_{x}$, where ψ is any atomic (or identity) formula[5]

λ-EQUIVALENCE: For any formula ϕ where F^n isn't free, the universal closure of the following is an axiom:

$(x_1)\ldots(x_n)\Box([\lambda\nu_1\ldots\nu_n\ \phi]x_1\ldots x_n \equiv \phi^{x_1,\ldots,x_n}_{\nu_1,\ldots,\nu_n})$,
provided $x_1,\ldots,x_n$ are substitutable for $\nu_1,\ldots,\nu_n$, respectively, in ϕ

λ-EQUIVALENCE also has greater significance due to the fact that among the λ-expressions there are vacuous expansions and necessitations.

B. Rules of inference. In addition to $\rightarrow$E and UI, the only other official rule of inference is the box introduction rule ("$\Box$I"):

If $\vdash \phi$, then $\vdash \Box\phi$, where ϕ has no definite descriptions[6]

Unofficially, we use the many standard, derived rules of the elementary object calculus, as well as the obvious derived rules and logical theorems of S5 (for example, diamond introduction, "$\Diamond$I"). Our derived rules of λ-introduction and λ-elimination are formulated as in Chapter I.

Using these logical axioms and rules, the Barcan formulas are derivable.[7] However, they are unobjectionable since the quantifiers in them range over all objects, and not just over the objects which exist. So $\Box$'s commute with universal quantifiers and $\Diamond$'s commute with existential quantifiers.

Finally, we note that the RELATIONS theorem schema (I, §3) is now derivable without the restriction that $x_1,\ldots,x_n$ be free in ϕ:

RELATIONS: where ϕ is any propositional formula where F^n isn't free, the universal closure of the following is a logical theorem:

$(\exists F^n)(x_1)\ldots(x_n)\Box(F^n x_1\ldots x_n \equiv \phi)$

In addition to the examples of this schema offered in Chapter I, we now have further examples:

(a) $(\exists F)(x)\Box(Fx \equiv \Box Gx)$

(b) $(\exists F)(x)\Box(Fx \equiv \Box(E!x \rightarrow Gx))$

(c) $(\exists F)(x)\Box(Fx \equiv Gb)$

(d) $(\exists F)(x)\Box(Fx \equiv \Box Gb)$

(e) $(\exists F)(x)(y)\Box(Fxy \equiv \Box Gb)$

(a) tells us that for any property G, there is a property of exemplifying G essentially; (b) tells us that for any property G, there is a property of necessarily exemplifying-G-if-existing; (c) and (d) assert, respectively, that there is a property objects exemplify just in case b exemplifies G and just in case necessarily b exemplifies G; (e) asserts that there is a two-place relation objects bear to one another just in case necessarily b exemplifies G.

Note also that while the following is, strictly speaking, not an instance of RELATIONS, it is nevertheless easily derivable:

PROPOSITIONS: Where ϕ is any propositional formula, the universal closure of the following is a logical theorem: $(\exists F^0)\Box(F^0 \equiv \phi)$

RELATIONS, PROPOSITIONS, D_4, D_5, and D_6 jointly comprise a complete modal theory of n-place relations ($n \geq 0$). It is an important feature of this theory that relations with the same exemplification extensions at each possible world may nevertheless be distinct. So, for example, it is consistent with our theory that the properties of being an equilateral triangle and being an equiangular triangle are distinct, even though they have the same exemplification extensions at each world.

And the properties of being-blue-or-not-blue and being-green-or-not-green may be distinct, though logically equivalent.

We call the metaphysical system which consists of the interpreted modal language, together with its logic, the modal object calculus (with propositions).

§4. The Proper Axioms

We assert that the following four axioms are true *a priori*:

AXIOM 1 ("E-IDENTITY"): $(x)(y)(x{=_E}y \equiv \Diamond E!x \;\&\; \Diamond E!y \;\&\; \Box(F)(Fx \equiv Fy))$

AXIOM 2 ("NO-CODER"): $(x)(\Diamond E!x \rightarrow \sim(\exists F)xF)$

AXIOM 3 ("IDENTITY"): $\alpha{=}\beta \rightarrow (\phi(\alpha,\alpha) \rightarrow \phi(\alpha,\beta))$, where $\phi(\alpha,\beta)$ is the result of replacing some, but not necessarily all, free occurrences of α by β, provided β is substitutable for α in the occurrences of α it replaces

AXIOM 4 ("A-OBJECTS"): For any formula ϕ where x isn't free, the universal closure of the following is an axiom: $(\exists x)(A!x \;\&\; (F)(xF \equiv \phi))$

Given our discussion of the axioms and theorems of the elementary theory, these axioms should be straightforward. Intuitively, each possible world looks somewhat like a model of elementary object theory--there are objects which exist at that world, and objects which fail to exist at that world. But from the point of view of a given world, say the actual world, the objects which fail to exist divide up into two mutually exclusive classes--the objects which necessarily fail to exist and the objects which exist at some other possible world. So from the point of view of the actual world, E-IDENTITY and NO-CODER govern the objects which either exist at this world or exist at some other world.

The IDENTITY axiom has greater significance than its counterpart in the elementary theory because of the many new kinds of terms that have been added to the modal language. For example, $F^0=G^0 \rightarrow (\Box F^0 \rightarrow \Box G^0)$ is an instance of IDENTITY. So is: $b=(\iota x)\phi \rightarrow (\psi^b_\nu \rightarrow \psi^{(\iota x)\phi}_\nu)$.

A-OBJECTS also has greater significance since it now yields objects which encode vacuous properties and modal properties. Given our rule of necessitation, $\Box$I, the following is a consequence of A-OBJECTS:

$\Box(\exists x)(A!x \,\&\, (F)(xF \equiv \phi))$, where ϕ has no free x's.

Semantically, this tells us that given a world w and wff ϕ, there is an abstract object at w which encodes just the properties satisfying ϕ with respect to w. A formula like "Fs" ("Socrates exemplifies F") is satisfied by different properties at different worlds. At each world, then, there is an A-object which encodes just the properties Socrates exemplifies at that world. A formula like "F=R v F=S" is satisfied by the same two properties, roundness and squareness, at each world. Given LA9, and the definition of identity, it follows that the_A round square of one world is identical with the_A round square of any other world. Intuitively, all of the A-objects from each of the worlds can be grouped into one set, the set of A-objects, which stays fixed from world to world. In the future, when we use restricted z-variables, they will range over this set.

E-IDENTITY, NO-CODER, IDENTITY, and A-OBJECTS jointly constitute the <u>modal theory of abstract objects</u> and I believe that the theory is consistent.[8] It is provable that some propositions as well as some complex relations don't have unique constituents.[9] However, it is possible to weaken the theory in the manner described in Chapter I, guarantee the unique constituency of all relations, and still preserve most of the applications which follow.

In these applications, it will be important to distinguish three senses of the phrase "possible object." On one sense of this phrase, objects which satisfy D_2 (§1) are possible objects, whereas abstract objects are not. We always use "possibly existing object" to indicate this sense of "possible object."

The other two senses of the phrase are ones in which abstract objects are possible objects. Consider D_7, where "z" is a restricted variable ranging over abstract objects:

D_7 z is strongly possible ("SPoss(z)") $=_{dfn} \Diamond(\exists x)\mathrm{Blue}(z,x)$

We always use the phrase "strongly possible object" to indicate this sense of "possible object." For example, Socrates' blueprint is strongly possible, and so is the blueprint of Socrates' blueprint.[10]

The third sense of "possible object" we distinguish requires a preliminary definition.

D_8 x is weakly correlated with z ("WCor(x,z)") $=_{dfn}$ $(F)(zF \rightarrow Fx)$

For example, abstract objects which encode just some of the properties a given object exemplifies are incomplete blueprints of the object--the object is weakly correlated with them. We now have,

D_9 z is weakly possible ("WPoss(z)") $=_{dfn} \Diamond(\exists x)\mathrm{WCor}(x,z)$

Weakly possible A-objects are "possible objects" in the sense that they do not encode any contradictory properties. F and G are contradictory properties iff it's not possible that some object exemplify both of them. We shall keep these distinctions straight in the applications which follow. To prepare for these applications, we add to our primitive vocabulary the usual abbreviations of standard English gerunds. Also, we adopt a modal version of our AUXILIARY HYPOTHESIS--A-objects necessarily fail to exemplify nuclear relations.

CHAPTER III ENDNOTES

[1]Now that we are in a modal theory, we have to face the question of whether our descriptions will vary in denotation from world to world or be "rigid designators." On the first alternative, they would denote at a world w, the unique object satisfying ϕ with respect to w (if there is one). On the second alternative, they would denote at a world w, the unique object satisfying ϕ at w_0 (the base world). We could have two kinds of descriptions in our language--rigid and non-rigid descriptions. However, we shall employ just one type of description, and suppose that all our descriptions are rigid designators.

We do this for two reasons. One is that we will not need non-rigid descriptions in any of our applications. Instead, we shall try to show that rigid descriptions have interesting, heretofore undiscovered, applications. Secondly, by having just rigid descriptions in the language, we can simplify the definition of <u>denotation</u>$_{I,\delta}$. Since all of the terms of the language will be rigid designators, we need not define <u>the denotation</u>$_{I,\delta}$ <u>of</u> term τ <u>with respect to</u> world w. Were we to allow descriptions which might change denotations from world to world, we'd have to define "$d_{I,\delta}((\iota x)\phi, w)$." This would force us to revise the entire definition of denotation so that it becomes a binary function. See Appendix C for a definition of <u>denotation</u>$_{I,\delta}(\tau,w)$ which is required by languages in which there are non-rigid designators.

[2]The technique here is due to Saul Kripke [1963], pp. 83-94.

[3]Note that one can consistently maintain that logically equivalent propositions (i.e., propositions r^0 such that for all worlds, $ext_w(r^0) = T$) need not be identical. For example, the proposition that either Carter is President or it's not the case that Carter is President (i.e., *NEG(CONJ(PLUG*$_1$*(being President, Carter), NEG(PLUG*$_1$*(being President, Carter)))))* need not be identical with the proposition that either Nixon is President or it's not the case that Nixon is President (i.e., *NEG(CONJ(PLUG*$_1$*(being President, Nixon), NEG(PLUG*$_1$*(being President, Nixon)))))*.

Given our statement of the axioms of set theory as in footnote 10, II, §3, we need not believe that there is only one mathematical proposition.

[4]Recall that, for convenience, we drop the subscripts on *I*-assignment variables which relativize them to particular interpretations.

[5]By inserting the parenthetical remark when reading the axiom, it becomes, strictly speaking, a proper axiom. When you eliminate the parenthetical remark, the axiom is true under all interpretations. But since identity formulas are defined formulas, we have to guarantee, for example, that $\tau=(\iota x)\phi \rightarrow \phi^{\tau}_{x}$. To do this, we could add the following proper axiom:

$\psi^{(\iota x)\phi}_{\nu} \equiv (\exists! y)(\phi^{y}_{x} \& \psi^{y}_{\nu}) \& (\exists! y)\phi^{y}_{x}$, where ψ is any defined identity formula and ν is any object variable

For convenience, however, we've combined this proper axiom with the original logical axiom presented in Chapter II, §2. But if we were to be more careful, we would distinguish L-DESCRIPTIONS (the logical axiom) from P-DESCRIPTIONS (the above proper axiom) and add the latter to the set of proper axioms described in §4.

[6]Had we chosen to interpret our descriptions non-rigidly, this restriction would be unnecessary. However, it is an interesting fact about the logic of rigid descriptions that the rule of necessitation, $\Box$I, must be restricted. This prevents the following derivation of a logical theorem which is not valid:

(a)	$F(\iota x)Gx$	Assumption
(b)	$(\exists! y)Gy$	DESCRIPTIONS, (a)
(c)	$F(\iota x)Gx \rightarrow (\exists! y)Gy$	CP, (a)-(b)
(d)	$\Box(F(\iota x)Gx \rightarrow (\exists! y)Gy)$	$\Box$I, (c)
(e)	$\Box F(\iota x)Gx \rightarrow \Box(\exists! y)Gy$	LA7, (d)

(e) is not logically true when the description rigidly denotes the object in the base world, w_0, which satisfies Gx. Unrestricted $\Box$I seems to be the source of the trouble.

Unfortunately, this clearly leaves our logic for descriptions incomplete, and I've yet to figure out just which logical axioms or rules must be added in order to get a logic which might have a claim to completeness. For example, as it stands, we can't derive: $\Box(F(\iota x)\phi \rightarrow F(\iota x)\phi)$. Nor can we derive: $\Box(\iota z)(F)(zF \equiv F{=}P)P$. Nor: $\Box((x)(Px \rightarrow Qx) \;\&\; P(\iota x)\phi \rightarrow Q(\iota x)\phi)$.

The problem cuts across all axiomatic modal theories which employ rigid descriptions. It may be easier to solve in a natural deduction system. Whatever the case, more investigation into the logic of rigid descriptions is needed here (I'd like to thank Ed Gettier for pointing this problem out to me).

[7]See A.N. Prior, [1956].

[8]Again, the proof has yet to be discovered. However, I suspect that if the elementary theory is consistent, then so is the modal version.

[9]Here's the proof. Let R be any one-place property. By A-OBJECTS, we have $(\exists x)(A!x \ \& \ (F)(xF \equiv (\exists u)(F=[\lambda y\ Ru] \ \& \ {\sim}uF)))$. Call this object a_5 and suppose ${\sim}a_5[\lambda y\ Ra_5]$. Then, by definition of a_5, $(u)([\lambda y\ Ra_5] = [\lambda y\ Ru] \rightarrow u[\lambda y\ Ra_5])$. So $a_5[\lambda y\ Ra_5]$, contrary to hypothesis.

So suppose $a_5[\lambda y\ Ra_5]$. By definition of a_5, for some object, say a_6, $[\lambda y\ Ra_6] = [\lambda y\ Ra_5]$ and ${\sim}a_6[\lambda y\ Ra_5]$. By definition of proposition identity (§1, D_6), $Ra_5{=}Ra_6$. But since a_5 encodes $[\lambda y\ Ra_5]$ and a_6 doesn't, $a_5 \neq a_6$.

[10]We can distinguish the strong possibility of Socrates' blueprint from the strong possibility of its blueprint as follows:

$$z \text{ is strongly}_E \text{ possible} =_{dfn} \Diamond(\exists x)(E!x \ \& \ Blue(z,x))$$

Socrates' blueprint is strongly_E possible, whereas its blueprint isn't.

CHAPTER IV

THE APPLICATIONS OF THE MODAL THEORY

§1. Truth

By now, we should have a good grasp of what it means to say "it's necessary that" and "it's possible that." We took the notion of a possible world as a primitive and used it in our semantics so that we could interpret the modal, sentential operators "$\Box$" and "$\Diamond$" which represent these English expressions. But we are really interested in the object language. When speaking from the object language, we may suppose that everything we say is analyzable in terms of our six, primitive metaphysical notions: object, n-place relation, exemplification, encoding, existence, and possibility. In this chapter, all of our definitions will be constructed in terms of these notions. We begin with some preliminary observations and definitions.

Since propositional formulas are also terms which denote propositions, we shall follow Ramsey in supposing that the predicate "is true" and the operator "it is true that" are eliminable from our language.[1] The language we developed in the previous chapter allows us to make Ramsey's suggestion precise through the formulation of the following definitions:

D_{10} F^0 is true $=_{dfn}$ F^0

D_{11} It is true that F^0 $=_{dfn}$ F^0

With these definitions, we may translate "Everything Aristotle asserted is true" as follows:

(F^0)(Aristotle asserted $F^0 \rightarrow F^0$ is true)

This reduces to

(F^0)(Aristotle asserted $F^0 \rightarrow F^0$)

Both of these generalizations may be instantiated with complex propositional terms to yield intelligible sentences. And once we move to type theory, we shall be able to symbolize these generalizations as:

$(F^0)(AaF^0 \rightarrow F^0)$,

where "A" denotes the asserting relation between a person and a proposition.

Ramsey's idea works fine as long as we're interested in truths relative to the world we're in. A less mundane notion of truth is the notion of truth at a particular world. We shall produce a definition of this notion in the next section, once we've modelled possible worlds. But before we do so, we require a few more preliminary definitions.

We shall say that a property F^1 is constructed out of a proposition F^0 iff F^1 is the property of being such that F^0:

D_{12} F^1 is constructed out of F^0 ("Const(F^1,F^0)") $=_{dfn}$
$F^1 = [\lambda x\ F^0]$

We then define a vacuous property to be one which is constructed out of some proposition:

D_{13} F^1 is a vacuous property ("Vac(F^1)") $=_{dfn}$ $(\exists F^0)\mathrm{Const}(F^1,F^0)$

Examples of vacuous properties are: being such that Carter is President, being such that Fischer defeated Spassky, being such that Nixon did not resign the Presidency, being such that a Luxembourgian was the first man on the moon, being such that every man loves every fish, etc.

Finally, we say that if an abstract object encodes a vacuous property, then the proposition out of which the property is constructed is encoded in the abstract object:

D_{14} F^0 is encoded in z ("$\Sigma_z F^0$") $=_{dfn}$ $z[\lambda y\ F^0]$

§2. Modelling Possible Worlds[2]

Possible worlds will be abstract objects which encode only vacuous properties and which meet two other conditions. For one thing, they must be maximal, i.e., for every proposition F^0, either F^0 or the negation of F^0 must be encoded in them.

D_{15} z is maximal ("Max(z)") $=_{dfn}$ $(F^0)(\Sigma_z F^0 \vee \Sigma_z {\sim}F^0)$

So if an object z is maximal, it must encode, for every proposition F^0, either being such that F^0 or being such that it's not the case that F^0.

Secondly, inconsistent propositions must not be encoded in the same world. One way to make this requirement precise would be to stipulate that worlds must be weakly possible (i.e., as in III, §4, D_9). This would require that it's possible that some object exemplifies every property the world encodes. However, a more elegant way of ensuring that inconsistent propositions won't be encoded in the same world is to stipulate that if an object z is to be a world, then it must be possible that every proposition encoded in z is true.[3]

We can formalize all these conditions on worlds in the following definition:

D_{16} z is a possible world $=_{dfn}$ $(F^1)(zF \rightarrow Vac(F))$ & Max(z) & $\Diamond(F^0)(\Sigma_z F^0 \rightarrow F^0)$

Although this definition would serve us well, there is a more elegant definition which is equivalent:

D_{17} z is a possible world ("World(z)") $=_{dfn}$ (F)(zF $\rightarrow$ Vac(F)) & $\Diamond(F^0)(\Sigma_z F^0 \equiv F^0)$

That is, an object z is a world iff every property it encodes is vacuous and it's possible that all and only true propositions are encoded in z. From D_{17}, we can prove that worlds are maximal:

<u>Theorem 1</u>: (z)(World(z) $\rightarrow$ Max(z))

<u>Proof</u>: Suppose z_0 is an arbitrary world. By definition, $\Diamond(F^0)(\Sigma_{z_0}F^0 \equiv F^0)$. We want to conclude that for an arbitrary proposition Q^0, that $\Sigma_{z_0}Q^0$ or $\Sigma_{z_0}{\sim}Q^0$. We do this in two stages: in stage (A), we prove that $\Diamond(\Sigma_{z_0}Q^0 \vee \Sigma_{z_0}{\sim}Q^0)$, and in stage (B), we use a theorem of S_5 (which distributes a $\Diamond$ over a disjunction) and our new logical axiom LA9, to prove that $\Sigma_{z_0}Q^0$ or $\Sigma_{z_0}{\sim}Q^0$.

(A) In this stage, we rely on the following theorem of S_5: $\Box(\phi \rightarrow \psi) \rightarrow (\Diamond\phi \rightarrow \Diamond\psi)$. If we let $\phi = \ulcorner(F^0)(\Sigma_{z_0}F^0 \equiv F^0)\urcorner$ and $\psi = \ulcorner\Sigma_{z_0}Q^0 \vee \Sigma_{z_0}{\sim}Q^0\urcorner$, then by establishing that $\Box(\phi \rightarrow \psi)$, we can apply the S_5 theorem using the fact that $\Diamond(F^0)(\Sigma_{z_0}F^0{\equiv}F^0)$ and reap our initial result. So we first establish that $(\phi \rightarrow \psi)$, and then use $\Box$I. So assume ϕ, and instantiate the quantifier to both Q^0 and ${\sim}Q^0$. So $\Sigma_{z_0}Q^0 \equiv Q^0$ and $\Sigma_{z_0}{\sim}Q^0 \equiv {\sim}Q^0$. Since $Q^0 \vee {\sim}Q^0$, it follows that $\Sigma_{z_0}Q^0 \vee \Sigma_{z_0}{\sim}Q^0$, i.e., ψ. So $\Box(\phi \rightarrow \psi)$, by $\Box$I. And by the S_5 theorem, we have our initial result: $\Diamond(\Sigma_{z_0}Q^0 \vee \Sigma_{z_0}{\sim}Q^0)$.

(B) It is also a theorem of S_5 that $\Diamond(\phi \vee \psi) \rightarrow (\Diamond\phi \vee \Diamond\psi)$. By letting ϕ, ψ be the disjuncts of our initial result, it follows that $\Diamond\Sigma_{z_0}Q \vee \Diamond\Sigma_{z_0}{\sim}Q^0$. By LA9, it follows that $\Box\Sigma_{z_0}Q^0 \vee \Box\Sigma_{z_0}{\sim}Q^0$, since if possibly an object encodes a property, it does so necessarily. And by another theorem of S_5, $\Box(\Sigma z_0 Q^0 \vee \Sigma z_0 {\sim}Q^0)$. By LA6, $\Sigma z_0 Q^0 \vee \Sigma z_0 {\sim}Q^0$, i.e., $Max(z_0)$. So every D_{17}-world is maximal.$\boxtimes$[4]

Theorem 1 is instrumental for showing that D_{17} implies D_{16}. It is a good exercise to show that D_{16} implies D_{17}.[5]

Let us say that propositions F^0 and G^0 are <u>inconsistent</u> iff it's not possible that both F^0 and G^0 be true. We then have:

<u>Theorem 2</u>: $(z)(World(z) \rightarrow {\sim}(\exists F^0)(\exists G^0)({\sim}\Diamond(F^0 \;\&\; G^0) \;\&\; \Sigma_z F^0 \;\&\; \Sigma_z G^0))$

That is, inconsistent propositions are not encoded in any world.

<u>Proof</u>: Assume for reductio that $World(z_0)$, $\Sigma z_0 P^0$, $\Sigma z_0 Q^0$, and that ${\sim}\Diamond(P^0 \;\&\; Q^0)$. By D_{17}, $\Diamond(F^0)(\Sigma z_0 F^0 \equiv F^0)$. So let $\phi = \ulcorner(F^0)(\Sigma z_0 F^0 \equiv F^0)\urcorner$ and assume ϕ. If we let $\psi = \ulcorner P^0 \;\&\; Q^0\urcorner$, it is easy to see that $\phi \rightarrow \psi$. So by $\Box$I, $\Box(\phi \rightarrow \psi)$. By using our S_5 theorem that $\Box(\phi \rightarrow \psi) \rightarrow (\Diamond\phi \rightarrow \Diamond\psi)$, it follows that $\Diamond(P^0 \;\&\; Q^0)$, contrary to hypothesis.$\boxtimes$

The proofs of Theorems 1 and 2 should give us a good grasp of the second clause of D_{17}. They should help us to see that the following definition is justified:

D_{18} F^0 is <u>true at</u> z $=_{dfn}$ $World(z) \;\&\; \Sigma_z F^0$

So, whenever z is a world, the propositions encoded in the world are just the propositions true at that world. This definition suggests what it is for a world to be actual:

D_{19} z is an actual world ("$World_A(z)$") $=_{dfn}$ $World(z) \;\&\; (F^0)(\Sigma_z F^0 \rightarrow F^0)$

That is, an actual world is any world such that every proposition encoded in it is true. We now get the following result:

Theorem 3: $(z)(z')(World_A(z) \& World_A(z') \rightarrow z'=z)$

That is, there is at most one actual world.

Proof: Suppose, for reductio, that $World_A(z_1)$ and $World_A(z_2)$, where $z_1 \neq z_2$. Since z_1 and z_2 are distinct A-objects, they must differ with respect to at least one encoded property. Since they are both worlds, any such property must be vacuous. So without loss of generality, suppose $\Sigma_{z_1} Q^0 \& \sim\Sigma_{z_2} Q^0$. By Theorem 1, z_2 must be maximal. So $\Sigma_{z_2}\sim Q^0$. But since both z_1 and z_2 are actual, every proposition they encode must be true. Contradiction. $\boxtimes$

We also get:

Theorem 4: $(\exists z)World_A(z)$

Proof: By A-OBJECTS, there is an abstract object which encodes a property F iff it's a vacuous property constructed out of a true proposition, i.e., $(\exists z)(F)(zF \equiv (\exists F^0)(F^0 \& F=[\lambda y\ F^0]))$. Call this object z_0. To show z_0 is an actual world, we show that it satisfies both clauses of D_{19}. So we show (a) $World(z_0)$, and (b) $(F^0)(\Sigma_{z_0} F^0 \rightarrow F^0)$:

(a) Clearly, every property z_0 encodes will be vacuous. So we want to show that possibly, all and only the true propositions are encoded in z_0. Consider an arbitrary proposition Q^0. $(\rightarrow)$ Suppose $\Sigma_{z_0} Q^0$. Then by definition of z_0, $(\exists F^0)(F^0 \& [\lambda y\ Q^0] = [\lambda y\ F^0])$. Call this proposition R^0. Since $[\lambda y\ Q^0] = [\lambda y\ R^0]$, it follows from the definition of proposition identity (III, §1, D_6) that $Q^0 = R^0$. Since R^0 is true, so is Q^0. $(\leftarrow)$ Suppose Q^0 is true. Then $z_0[\lambda y\ Q^0]$, i.e., $\Sigma_{z_0} Q^0$.

Since we have established that $\Sigma_{z_0} Q^0 \equiv Q^0$, for an arbitrary proposition Q^0, it follows that $\Diamond(F^0)(\Sigma_{z_0} F^0 \equiv F^0)$.

(b) Clearly, every proposition encoded in z_0 is true, by definition of z_0. ☒

With Theorems 3 and 4, we have proven that there is a unique actual world (from *a priori* assumptions alone). We're entitled to name this object, and so we do so as follows: $w_A =_{dfn} (\iota z)World_A(z)$.

It should be interesting that the actual world is an abstract, and not an actual (existing) object. If we were to think that the actual world was an existing object like you, me, or some sub-atomic particle, it would fail to encode any properties (by NO-CODER). There would be no reason why its vacuous properties were any more crucial to its identity than the other properties it exemplified (like not being a cat, being non-red, etc.). As Wittgenstein said, the world is just all that is the case.[6] w_A does encode all and only that which is the case.

It is important to distinguish w_A from the world of existing things. Consider the following definition:

D_{20} z is a world of existing things ("$World_{E!}(z)$") $=_{dfn}$
$World(z)$ & $(\exists x)(E!x$ & $(F^0)(z[\lambda y\ F^0] \equiv [\lambda y\ F^0]x))$

So a world of existing things is any world such that some existing object exemplifies exactly the vacuous properties it encodes. From D_{20}, we get the following result:

<u>Theorem 5</u>: $(z)(z')(World_{E!}(z)$ & $World_{E!}(z') \rightarrow z'=z)$

<u>Proof</u>: Assume, for reductio, $World_{E!}(z_1)$, $World_{E!}(z_2)$, and $z_1 \neq z_2$. Call the objects which exemplify exactly the vacuous properties z_1 and z_2 encode b_1 and b_2, respectively. Since $z_1 \neq z_2$, there must be a vacuous property one encodes which the other doesn't. Without loss

of generality, suppose $z_1[\lambda y\ Q^0]$ & $\sim z_2[\lambda y\ Q^0]$. Since z_2 is maximal, $z_2[\lambda y \sim Q^0]$. So $[\lambda y\ Q^0]b_1$ and $[\lambda y \sim Q^0]b_2$. But by λE, Q^0 & $\sim Q^0$. ⊠

Consequently, there is at most one world of existing things. If we add the contingent assumption that something exists, it would follow that there is a world of existing things--just consider the world which encodes exactly the vacuous properties the existing thing exemplified. By Theorem 5, we would be entitled to talk about *the* world of existing things, and give it a name.

This is why we distinguish the actual world, w_A, from the world of existing things. It is a contingent matter that there is a world of existing things, whereas it is not a contingent matter that there is an actual world. Indeed, some philosophers take it as an *a priori* datum that there is a unique actual world. We may suppose this on our theory without having to worry that we've attributed *a priori* existence to a contingent object.

We now consider the implications of another definition which seems justified:

D_{21} x $\underline{\text{exists at}}$ z $=_{dfn}$ World(z) & $\Sigma_z E!x$

That is, x exists at z iff the proposition that x exists is encoded in world z. It is consistent with this definition that objects exist at more than one world. However, some philosophers apparently like to work with a notion of existing at a world on which objects can exist at at most one world.[7] We could accommodate the views of these philosophers were we to define existence at a world as follows:

x exists at z $=_{dfn}$ World(z) & $(F^0)(\Sigma_z F^0 \equiv [\lambda y\ F^0]x)$

Using this definition, we would get the result that individuals are world-bound. To see this, suppose b exists at both z_1 and z_2, $z_1 \neq z_2$. If $[\lambda y\ Q^0]$ was the vacuous property distinguishing z_1 and z_2, it would follow that b both exemplified and failed to exemplify this property. So b can't exist at both, on this definition of "exists at." Counterpart theorists will then prefer to use this latter definition in their investigations.

We conclude this section on worlds with a proof of a lemma which will be instrumental in §3. Let "w" be a restricted variable ranging over worlds.

<u>Lemma</u>: $(F^0)(w)(x)(\Sigma_w F^0 \equiv \Sigma_w[\lambda y\ F^0]x)$

That is, a proposition F^0 is true at w iff everything exemplifies being such that F^0 at w.

<u>Proof</u>: $(\rightarrow)$ Let Q^0, w_0, and b_0 be an arbitrary proposition, world, and object, respectively. We assume $\Sigma_{w_0} Q^0$ and try to show that $\Sigma_{w_0}[\lambda y\ Q^0]b_0$. Since w_0 is a world, $\Diamond(F^0)(\Sigma_{w_0} F^0 \equiv F^0)$. So if we can deduce $\Sigma_{w_0}[\lambda y\ Q^0]b_0$ from $(F^0)(\Sigma_{w_0} F^0 \equiv F^0)$, we can apply our standard S_5 theorem: $\Box(\phi \rightarrow \psi) \rightarrow (\Diamond\phi \rightarrow \Diamond\psi)$. So suppose $(F^0)(\Sigma_{w_0} F^0 \equiv F^0)$. Then $\Sigma_{w_0} Q^0 \equiv Q^0$, and since we've assumed $\Sigma_{w_0} Q^0$, we know Q^0. By λI, $[\lambda y\ Q^0]b_0$. So this must be a proposition encoded in w_0, i.e., $\Sigma_{w_0}[\lambda y\ Q^0]b_0$. By our S_5 theorem, $\Diamond\Sigma_{w_0}[\lambda y\ Q^0]b_0$; and by LA9 and LA6, $\Sigma_{w_0}[\lambda y\ Q^0]b_0$.

$(\leftarrow)$ Suppose $\Sigma_{w_0}[\lambda y\ Q^0]b_0$. We try to show $\Sigma_{w_0} Q^0$. Again, since w_0 is a world, $\Diamond(F^0)(\Sigma_{w_0} F^0 \equiv F^0)$. From $(F^0)(\Sigma_{w_0} F^0 \equiv F^0)$ and our assumption, it follows that $[\lambda y\ Q^0]b_0$. By λE, we get Q^0. So $\Sigma_{w_0} Q^0$. So $(F^0)(\Sigma_{w_0} F \equiv F^0)$ implies $\Sigma_{w_0} Q^0$, and since the former is possible, so is the latter. Consequently, the latter is true as well. $\boxtimes$

§3. Modelling Leibniz's Monads[8]

The investigation of monads is as philosophically satisfying as the definition of truth and the investigation of worlds. Although it is unclear what Leibniz intended his monads to be, they have traditionally been regarded as properties of some sort.[9] However, we model them here as abstract objects which are strongly possible.[10] Strongly possible abstract objects have correlates "at" some possible world. For example, Socrates' blueprint is a monad since it has a correlate at the actual world. Intuitively, "compossible" monads have correlates at the same world. So your blueprint and my blueprint are compossible. A monad "mirrors" the world at which it has a correlate by encoding the vacuous properties the correlate exemplifies--properties constructed out of the propositions true at that world.

To make these ideas precise, we utilize the following definitions. As with the previous lemma, we use "w" as a restricted variable ranging over the abstract objects which satisfy the definition of a world:

D_{22} x is the correlate of z at w ("Cor(x,z,w)") $=_{dfn}$ $(F^1)(\Sigma_w Fx \equiv zF)$

That is, x is the correlate of z at w iff x exemplifies F at w iff z encodes F.

D_{23} z appears at w ("Appear(z,w)") $=_{dfn}$ $(\exists x)$Cor(x,z,w)

D_{24} z is a monad ("Monad(z)") $=_{dfn}$ $(\exists w)$Appear(z,w)

D_{25} z mirrors w ("Mirror(z,w)") $=_{dfn}$ $(F^0)(\Sigma_w F^0 \equiv \Sigma_z F^0)$

Using the lemma at the end of §2, we now get the following result:

Theorem 6: $(z)(w)(\text{Monad}(z) \,\&\, \text{Appear}(z,w) \rightarrow \text{Mirror}(z,w))$

That is, every monad mirrors any world where it appears.

Proof: Suppose z_0 is a monad and z_0 appears at w_0. We want to show that for an arbitrary proposition, Q^0, $\Sigma_{w_0}Q^0 \equiv \Sigma_{z_0}Q^0$. ($\rightarrow$) Suppose $\Sigma_{w_0}Q$. Since z_0 appears at w_0, call its correlate there b_0. z_0 encodes exactly the properties b_0 exemplifies at w_0. In particular z_0 encodes $[\lambda y\ Q^0]$ iff b_0 exemplifies $[\lambda y\ Q^0]$ at w. By our assumption, $\Sigma_{w_0}Q^0$. So $\Sigma_{w_0}[\lambda y\ Q^0]b_0$, by the lemma. So $\Sigma_{z_0}Q^0$.

($\leftarrow$) Suppose $\Sigma_{z_0}Q^0$. Again, let b_0 be the object which exemplifies at w_0 exactly the properties z_0 encodes. Clearly, b_0 must exemplify $[\lambda y\ Q^0]$ at w. So by the lemma, $\Sigma_{w_0}Q^0$. ⊠

Another interesting fact about monads is provable with the help of Theorem 6:[11]

Theorem 7: $(z)(\text{Monad}(z) \rightarrow (\exists!w)\text{Appear}(z,w))$

That is, every monad appears at a unique world.

Proof: Suppose z_0 is a monad. So there is a world, say w_1, where it appears. We want to show that w_1 is unique, so for reductio, suppose z_0 appears also at w_2, $w_2 \neq w_1$. Since the worlds are distinct, there must be some vacuous property which distinguishes them. With out loss of generality, suppose $\Sigma_{w_1}Q^0$ and $\sim\Sigma_{w_2}Q^0$. And since w_2 is maximal, $\Sigma_{w_2}\sim Q^0$. But by Theorem 6, z_0 mirrors both worlds. So $\Sigma_{z_0}Q^0$ and $\Sigma_{z_0}\sim Q^0$. But this is impossible, since Q^0 and $\sim Q^0$ would both be true in any world where z_0 has a correlate. ⊠

Since we know that every monad appears at a unique world, we're entitled to talk about *the* world where it appears. Let us use "m" as a restricted variable ranging over the objects satisfying the definition of monad. We then define:

D_{26} $w_m =_{dfn} (\iota w)\mathrm{Appear}(m,w)$

Theorems 6 and 7 allow us to say that every monad mirrors its world.[12] Here now is a definition of compossibility.

D_{27} m_1 is compossible with m_2 ("$\mathrm{Comp}(m_1,m_2)$") $=_{dfn}$
$(\exists w)(\mathrm{Appear}(m_1,w)\ \&\ \mathrm{Appear}(m_2,w))$

With these definitions, we have the following lemma:[13]

Lemma: $(m_1)(m_2)(\mathrm{Comp}(m_1,m_2) \equiv w_{m_1}=w_{m_2})$

That is, two monads are compossible iff the worlds where they appear are identical.

Proof: ($\rightarrow$) Since m_1 and m_2 are compossible, call the world where they both appear w_0. By Theorem 7, $w_1=w_{m_1}$ and $w_0=w_{m_2}$. So $w_{m_1}=w_{m_2}$. ($\leftarrow$) Clearly, if the worlds where they appear are identical, there is a world where they both appear. ⊠

With the help of this lemma, we get the following result:[14]

Theorem 8: $(m_1)(m_2)(m_3)(\mathrm{Comp}(m_1,m_1)\ \&\ (\mathrm{Comp}(m_1,m_2) \rightarrow \mathrm{Comp}(m_2,m_1))\ \&\ (\mathrm{Comp}(m_1,m_2)\ \&\ \mathrm{Comp}(m_2,m_3) \rightarrow \mathrm{Comp}(m_1,m_3)))$

That is, compossibility is an equivalence notion among the monads.

Proof: Clearly, compossibility is reflexive and symmetrical. To show transitivity, suppose $\mathrm{Comp}(m_1,m_2)$ and $\mathrm{Comp}(m_2,m_3)$. By the previous lemma, $w_{m_1}=w_{m_2}$ and $w_{m_2}=w_{m_3}$. So $w_{m_1}=w_{m_3}$. ⊠

It should also be clear that by defining "embedding" as follows:

D_{28} z_1 is embedded in z_2 ("$\mathrm{Embed}(z_1,z_2)$") $=_{dfn} (F^1)(z_1F \rightarrow z_2F)$,

we can prove that every monad has the world where it appears embedded in it:

<u>Theorem 9</u>: $(m)\text{Embed}(w_m, m)$

<u>Proof</u>: m mirrors its world w_m by encoding exactly the vacuous properties w_m encodes. So w_m must be embedded in m since the vacuous properties exhaust the properties w_m encodes. ⊠

Consequently, every monad will be maximal with respect to the propositions encoded in it. But an even stronger claim is warranted--monads are complete:

D_{28} z is complete ("Com(z)") $=_{dfn} (F)(zF \vee z\bar{F})$,
where $\bar{F} =_{dfn} [\lambda x \sim Fx]$.

<u>Theorem 10</u>: $(z)(\text{Monad}(z) \rightarrow \text{Com}(z))$

<u>Proof</u>: Clearly, if m_0 is a monad, then some object is its correlate in the world where it appears. That object must exemplify, for every property, either it or its negation. Consequently, m_0 will encode, for every property, either it or its negation. ⊠

Theorems (1)-(10) outline a certain picture of objects, monads, and worlds. In addition to being informative in its own right, this picture may be useful for deciding hard questions in Leibnizian scholarship. Our notion of encoding a property is very similar to Leibniz's notion of concept containment. We could regard our notion of mirroring a world as representing both Leibniz's notion of mirroring (in the Monadology) and his notion of expression (in his letters to Arnauld). Leibniz says repeatedly in the letters to Arnauld that every individual substance of this universe expresses (in its concept) the universe into which it enters.[15]

In §56 of the Monadology, Leibniz calls monads both living and perpetual. On our understanding of monads, we can see how they could be perpetual. By appealing to our AUXILIARY HYPOTHESIS, we suppose that monads don't have spatio-temporal location and therefore aren't subject to the laws of generation and decay. But it is difficult to understand how monads could be thought of as "living." Has Leibniz confused the blueprints of living persons with the persons themselves? Could we find support for this idea in the fact that the word "Adam" is used sometimes to talk about the existing person ("the actual Adam") and used sometimes to talk about the complete individual concept of Adam (one of "the many possible Adams") in the correspondences with Arnauld? These questions, and others, seem to be obvious points of departure for future investigations.

§4. Modelling Stories and Native Characters

By adding a few primitives to the language of Chapter III, we may model stories, and certain characters in them, as A-objects. First we add abbreviations for any proper name of English which denotes an object which, pretheoretically, we judge to be a story (for example, novels, myths, legends, plays, dreams, etc.), or an author or character thereof (where we take characters to be any story object, not necessarily animate). So we shall have object names in our language which abbreviate "*The Tempest*," "Shakespeare," "Prospero," "*The Brothers Karamazov*," "Alyosha," "*The Clouds*," "Strepsiades," "Socrates," "*Ulysses*," "Joyce," "Bloom," "Dublin," etc.

Secondly, we add the name of a new primitive relation which is of central importance to our investigations--the authorship relation.

The formula "Axy" shall say that x authors y, and we trust that our readers have at least an intuitive grasp on what it is to author something.

Consequently, we may define:

D_{30} z is a story ("Story(z)") $=_{dfn}$
$(F)(zF \rightarrow Vac(F)) \ \& \ (\exists x)(E!x \ \& \ Axz)$

That is, stories are abstract objects which encode only vacuous properties and which are authored by some existing thing. Hence, it is a contingent matter that there are any stories. Lots of A-objects might have been stories, however. To say this is to say that they encode just vacuous properties and that *possibly* there exists an object which authors it.

Stories don't have to be consistent, nor do they have to be maximal. If every proposition encoded in the story is true, then the story is a "true story"; if it's possible that every proposition encoded in the story is true, then the story is a "possibly true story."[16] But stories and worlds do have something in common--they encode only vacuous properties. It is therefore appropriate to use our defined operator "Σ" to talk derivatively about the propositions encoded in the story. In fact, if z is a story, then we may utilize "Σ_z" as our translation for the English prefix "according to (in) the story z." So when z is a story, "$\Sigma_z F^0$" says that F^0 is true according to z. This allows us to prove an interesting consequence of D_{30} which helps us to identify a given story: a story z just is that abstract object which encodes exactly the properties F which are constructed out of propositions true according to the story. That is,

THEOREM ("STORIES"): $(z)(Story(z) \rightarrow z=(\iota z')(F)(z'F \equiv (\exists F^0)(\Sigma_z F^0 \ \& \ F=[\lambda y\ F^0])))$[17]

For example, Little Red Riding Hood is a story, so it is that abstract object which encodes exactly the vacuous properties constructed out of propositions true according to Little Red Riding Hood. Although this is not a definition of "Little Red Riding Hood," we can identify this story in so far as we have a good pretheoretical idea about which propositions are true according to it. Fortunately, the data begins where our suggestion ends, for we suppose that the data are intuitively true English sentences of the form "according to the story, ..." For example,[18]

(1) According to The Tempest, Prospero had a daughter.

(2) According to The Iliad, Achilles fought Hector.

(3) In the Brothers Karamazov, everyone that met Alyosha loved him.

(4) In The Clouds, Strepsiades converses with Socrates.

(5) In Joyce's Ulysses, Bloom journeys through Dublin.

Thus, STORIES helps us to understand which A-objects might be denoted by the underlined terms in the above sentences. We next try to identify the denotations of some of the other terms.

We can say what it is to be a character of a story. Let us use "s" variables as restricted variables ranging over stories:

D_{31} x is a character of s ("Char(x,s)") $=_{dfn} (\exists F)\Sigma_s Fx$

That is, the characters of a story are the objects which exemplify properties according to it. As we noted previously, the characters of a story are any story objects, not just real or imaginary persons or animals. Note also that this definition allows existing objects to be characters of stories--we can tell stories (true or false) about existing objects, just as we can about non-existent ones.[19]

Of the non-existent characters in a given story, some will have originated entirely from that story. We call these the "native" characters, and they are to be distinguished from the other non-existent characters which may have been borrowed or imported from other stories. But the non-native non-existent characters are nevertheless "fictional," since, presumably, they are native to (originate from) some other story.

We may define the notions of being native and being fictional by utilizing a higher order primitive relation--one which could be analyzed in the context of some other work. This is the relation that two propositions F^0 and G^0 bear to one another just in case F^0 occurs (obtains, takes place) before G^0. We shall represent the fact that F^0 occurs before G^0 as "$F^0 < G^0$." This relation helps us to be more specific about what it is to originate in a story:

D_{32} x originates in s ("Origin(x,s)") $=_{dfn}$
Char(x,s) & ~E!x & (y)(y')(s')(Ays & Ay's' & (Ay's' < Ays) → ~Char(x,s'))

That is, x originates in s iff x is a non-existent object which is a character of s and which is not a character of any earlier story. We then define being native and being fictional as follows:

D_{33} x is native to s ("Native(x,s)") $=_{dfn}$ Origin(x,s)

D_{34} x is fictional ("Fict(x)") $=_{dfn}$ (∃s)Native(x,s)

So fictional characters are native to (originate in) some story.

Clearly, fictional characters may be characters of stories to which they are not native. Sherlock Holmes is not native to The Seven Per Cent Solution.[20] Nor is the monster Grendel, in John Gardner's recent account of the Beowulf legend from the monster's point of view (Grendel). For simplicity, we shall suppose that Achilles and Hector are native to The Iliad, even though they may instead be native to some earlier epic of which no copies have survived. Also, in what follows, we shall suppose that Prospero is native to The Tempest, Alyosha and Raskolnikov are native to The Brothers Karamazov and Crime and Punishment, respectively, Bloom is native to Joyce's Ulysses, and Gregor Samsa is native to Kafka's Metamorphosis.

It would be a philosophical achievement of great importance were someone to discover a way of identifying fictional characters in general. The best we can accomplish here is to present a means of identifying the characters native to a given story. The identifying properties of native characters are exactly the properties exemplified by that character in the story. So we may utilize the following axiom which identifies the native characters of a story as specific A-objects:[21]

AXIOM ("N-CHARACTERS"): $(x)(s)(Native(x,s) \rightarrow$
$x=(\iota z)(F)(zF \equiv \Sigma_s Fx))$

For example, since Prospero is native to The Tempest, Prospero is that abstract object which encodes exactly the properties Prospero exemplifies according to The Tempest. This tells us an important fact about the Σ_s-operator and native characters--the Σ_s-operator "transforms" a property a native character exemplifies according to story s into one which the character encodes. That is, it is a theorem that:[22]

$$(x)(s)(\text{Native}(x,s) \rightarrow (F)(xF \equiv \Sigma_s Fx))$$

So if according to the play, Prospero had a daughter, it follows that he encodes having a daughter.

This theorem assumes greater significance in the presence of the following axiom schema which also should govern the Σ_s-operator:

AXIOM(S) ("Σ_s-SUBSTITUTION"): where ϕ is any propositional formula in which there occurs an object term o for which x is substitutable, the following is an axiom:

$$(s)(\Sigma_s \phi \rightarrow \Sigma_s [\lambda x\ \phi^x_o]o)$$

For example, in the myth, Achilles fought Hector. It therefore follows from Σ_s-SUB both that in the myth, Achilles exemplifies the property of fighting Hector and that in the myth, Hector exemplifies the property of being fought by Achilles. From the supposition that Achilles and Hector are both native to the myth in question, we may deduce that they encode these properties, respectively, by N-CHARACTERS.[23]

With these definitions, axioms, and consequences, we can translate a wide variety of data. We begin with (1)-(5) above. The translation procedure is straightforward--since the Σ_s-operator is defined only on proposition terms, we translate the English "in the story" using the operator, and translate the rest of the sentence just as we would into an ordinary predicate calculus:

(1)' $\Sigma_{Tempest}(\exists y)Dyp$

(2)' $\Sigma_{Iliad}Fah$

(3)' $\Sigma_{BK}(x)(Mxa \rightarrow Lxa)$

(4)' $\Sigma_{\text{Clouds}} Cs_1 s_2$

(5)' $\Sigma_{\text{Ulysses}} Jbd$

There is an interesting class of sentences relevantly similar to (1) which we should discuss briefly. These true sentences begin with the story prefix and involve the predicative copula "is." For example, (6) and (7):

(6) According to Crime and Punishment, Raskolnikov is a student.

(7) In the Conan Doyle novels, Holmes is a detective.

Frequently, there are contexts in which it is acceptable to drop the story prefix and just use the remainder of the sentence. We can think of the resulting sentences "Raskolnikov is a student," "Holmes is a detective," as true if we suppose that the English copula "is" is to be read as "encodes." We can therefore assimilate more phenomena consistent with our earlier discovery about the possibility of the ambiguity of "is."[24]

I think we can partially accommodate the views of philosophers who object to (4)' and (5)' by arguing that the real Socrates and the real Dublin aren't characters of The Clouds and Ulysses, respectively. We do this by supposing, instead, that the objects known as "the Socrates of The Clouds," and "the Dublin of Ulysses," are the relevant characters of these stories. We could suppose that these latter objects were native to these stories and use N-CHARACTERS to identify them. Such a procedure could be broadened to identify all non-native fictional characters. For example, we could say that the Sherlock Holmes of The Seven Per Cent Solution is native to that work, even though Sherlock Holmes isn't.

The problem with this procedure is that one is forced to say something about the relationships between the real Socrates and the Socrates of The Clouds, between the Sherlock Holmes native to the Conan Doyle novels and the Sherlock Holmes native to The Seven Per Cent Solution, etc. This is no easy task. Clearly, the notions of weak correlation or embedding won't be of much help--the Socrates of The Clouds exemplifies-according-to-The Clouds (and consequently, encodes) properties not exemplified by the real Socrates. A full discussion of the host of problems which arise here would take us too far afield. Much further investigation is warranted before this procedure is to be adopted.

Consider next the following data:

(8) Santa Claus doesn't exist.

(9) Franz Kafka wrote about Gregor Samsa.

(10) Some Greeks worshipped Dionysus.

(11) Prospero is a character of The Tempest.

(12) Raskolnikov is a fictional student.

Clearly, (8)-(10) are to be translated as follows:

(8)' $\sim E!sc$

(9)' Wks

(10)' $(\exists y)(Gy \; \& \; Wyd)$

Once we've identified Santa Claus using N-CHARACTERS, it is provable that he fails to exist. And it is a consequence that all fictional objects fail to exist.

We translate (9) and (10) using exemplification formulas because they involve extranuclear properties which abstract objects may exemplify. Being written about and being worshipped are extranuclear properties. They were not ascribed to (exemplified by) Samsa and Dionysus in the relevant stories. But given our AUXILIARY HYPOTHESIS, these are just the kind of properties Samsa and Dionysus would exemplify.

Given our work above, (11) should be translated as:

(11)' Char(Prospero, The Tempest)

However, (12) is a more subtle case. Being fictional is a notion we've defined--it may not be a property ($[\lambda x\ \text{Fictional}(x)]$ is ill-formed). But being a student is a property that Raskolnikov encodes, since he is native to Crime and Punishment and exemplifies that property in the novel. Consequently, we may define:

D_{35} x is a fictional student ("F-student(x)") $=_{dfn}$ $(\exists s)(\text{Native}(x,s)\ \&\ \Sigma_s Sx)$,

where "S" denotes being a student. Then from the assumptions that Raskolnikov is native to Crime and Punishment and that he is a student according to that story, we have (12)' as a consequence:

(12)' F-student(Raskolnikov)

In fact, we can generalize and suppose there is a whole group of notions, each one defined with respect to a given property G:

D_{36} x is a fictional G $=_{dfn}$ $(\exists s)(\text{Native}(x,s)\ \&\ \Sigma_s Gx)$

So Holmes is a fictional detective, Achilles is a fictional Greek warrior, etc., given the appropriate assumptions.[25]

Finally, we discuss definite descriptions. Consider (13) and (14):

(13) The detective who lived at 221 Baker St. of the Conan Doyle novels is more famous than any real detective.

(14) In Crime and Punishment, Porphyry arrested the student who killed an old moneylender.

It would be inappropriate to read the description in (13) as "the object which exemplifies detectivehood, exemplifies living at 221 Baker St., and exemplifies being a character of the Conan Doyle novels," since this description fails to denote. But we often use the description in (13) to refer to Holmes. The proper way to translate it is as "the object which according to the Conan Doyle novels exemplifies both detectivehood and living at 221 Baker St." Using "MFT" to abbreviate "more famous than," and other obvious abbreviations, we may read (13) as:

(13)' $(y)(Dy \mathbin{\&} E!y \rightarrow MFT(\iota x)\Sigma_{CD}(Dx \mathbin{\&} Lx)y)$

This says that every existing detective y is such that the object which according to the Conan Doyle novels exemplifies both detectivehood and living at 221 Baker St. is more famous than y.

A similar reading must be given to the definite description in (14). The following would be the *wrong* symbolization of (14):

$\Sigma_{CP}Ap(\iota x)(Sx \mathbin{\&} (\exists y)(OMLy \mathbin{\&} Kxy))$

The definite description fails to denote anything, even though it is entirely within the scope of the story operator. There may not be an object which exemplifies being a student and which killed an old moneylender. Or there may be two. But there is exactly one object which according to Crime and Punishment exemplifies being a student and killing an old moneylender. Consequently, (14) is properly read as (14)':

(14)' $\Sigma_{CP}Ap(\iota x)\Sigma_{CP}(Sx \; \& \; (\exists y)(OMLy \; \& \; Kxy))$

When we read (hear) definite descriptions in the context of a story, there is an implicit understanding that the description denotes a character of the story. This implicit understanding is captured by placing the appropriate Σ-operator immediately after the iota-operator of the description. This guarantees that the description, should it denote, denotes a character of the story.

To see this, consider the above example (14'). If we assume that Raskolnikov is the object which according to Crime and Punishment is a student who killed an old moneylender, we can show that Raskolnikov is a character of that story. So assume (15):

(15) $r=(\iota x)\Sigma_{CP}Sx \; \& \; (\exists y)(OMLy \; \& \; Kxy)$

By DESCRIPTIONS, it follows that according to Crime and Punishment, Raskolnikov is a student who killed an old moneylender, i.e.,

(16) $\Sigma_{CP}Sr \; \& \; (\exists y)(OMLy \; \& \; Kry)$

By Σ_{CP}-SUB, it follows that Raskolnikov exemplifies being a student who killed an old moneylender, i.e.,

(17) $\Sigma_{CP}[\lambda x \; Sx \; \& \; (\exists y)(OMLy \; \& \; Kyx)]r$

So there is a property which Raskolnikov exemplifies according to Crime and Punishment. By D_{31}, Raskolnikov is a character of that story. So by placing the story operator immediately after the iota operator in the description, we guarantee that the object denoted, if there is one, is a character of the story.

Finally, note that (15) is a true identity statement. From (14') and (15), it follows that according to Crime and Punishment, Porphyry arrested Raskolnikov, i.e.,

(18) $\Sigma_{CP}Apr$

The above results should establish at least a *prima facie* case for thinking that stories and characters are abstract objects. The groundwork has been laid for further investigations which might fill in more details.

§5. Modelling The Fountain of Youth

We now examine another class of English sentences which seem to be true *a priori*. They have the form "The F_1, F_2,..., F_n is G," where G is logically implied by one of the F_i and where there isn't an object which (uniquely) exemplifies F_1, F_2,..., F_n. Here are some examples:

(1) The fountain of youth is a fountain

(2) The set of all non-self-membered sets is a set

(3) The existent golden mountain is colored

For considerations similar to those in Chapter II, §2, we translate the English definite descriptions as A-object descriptions. Except in these cases, we translate "the F_1, F_2,..., F_n" as "$(\iota z)(G)(zG \equiv F_1 \Rightarrow G \vee F_2 \Rightarrow G \vee \ldots \vee F_n \Rightarrow G)$," where "$F \Rightarrow G$" means that necessarily, everything exemplifying F exemplifies G.

Consequently, the English descriptions in (1)-(3) are represented as following, using obvious abbreviations:

(a) $(\iota z)(G)(zG \equiv FY \Rightarrow G)$[26]

(b) $(\iota z)(G)(zG \equiv [\lambda x\ Sx \ \& \ (y)(y \varepsilon x \equiv Sy \ \& \ y \varepsilon y)] \Rightarrow G)$

(c) $(\iota z)(F)(zF \equiv E! \Rightarrow F \vee G \Rightarrow F \vee M \Rightarrow F)$

In the metalanguage, we signify this reading of the definite article as "the_A," and we assimilate the reading of the definite article proposed in Chapter II, §2 to this reading. Let's abbreviate (a)-(c) respectively as $(\iota z)\psi_1$ - $(\iota z)\psi_3$. By A-DESCRIPTIONS, it follows that any property satisfying the formula on the right of the biconditional in ψ is encoded by the object denoted by the entire description.

Take (a) for example. Since being a fountain is logically implied by being a fountain of youth, it follows that $(\iota z)\psi_1 F$, where "F" here denotes being a fountain. Our representations of (1)-(3) turn out to be theorems:

(1') $(\iota z)\psi_1 F$

(2') $(\iota z)\psi_2 S$

(3') $(\iota z)\psi_3 C$

This reading of the English definite article may prove to be useful for another group of problem cases. Consider (4):

(4) Necessarily, the teacher of Alexander is a teacher

If we symbolixe the description in (4) as "$(\iota x)Txa$," (4) would be false. That's because our definite descriptions are rigid designators, and even if we assume that there was a unique object which exemplified teaching Alexander and that that object was Aristotle, it seems wrong to suppose that Aristotle taught Alexander in every possible world. Many philosophers have then concluded that if we are to preserve the truth of (4), we need to translate the English description as a non-rigid description or eliminate the description altogether.

But this conclusion is not warranted. We can preserve the necessary truth which seems to be embedded in (4) another way. We can translate the English description as in (d):

(d) $(\iota z)(G)(zG \equiv [\lambda y\ Tya] \Rightarrow G)$

Abbreviating (d) as $(\iota z)\psi_4$, we then read (4) as:

(4') $\Box(\iota z)\psi_4 T$

It is true that necessarily the$_A$ teacher of Alexander encodes being a teacher. Again, we suppose the copula "is" to be ambiguous.

Philosophers since Russell have supposed that "scope distinctions" were the source of the trouble with the following arguments:

(4) Necessarily, the teacher of Aristotle is a teacher

(5) Aristotle is the teacher of Alexander

∴ (6) Necessarily, Aristotle is a teacher

(7) Necessarily, nine is greater than seven

(8) Nine is the number of planets

∴ (9) Necessarily, the number of planets is greater than seven

In these arguments, the premises appear to be true and the conclusions false. Russellians have explained the apparent invalidity by supposing that (4) and (9) exhibit *de re/de dicto* ambiguity. On the Russellian analysis, the definite descriptions get eliminated in terms of existential and uniqueness clauses, but in the presence of the modal operator, there is a way to eliminate the description so that these clauses appear before the operator (wide scope, *de re*) and a way to eliminate the description so that these clauses appear after the operator (narrow scope, *de dicto*). For example, two readings of (4) are:

(4") $(\exists x)(Txa\ \&\ (y)(Tya \rightarrow y{=}x)\ \&\ \Box Tx)$ (*re*)

(4"') $\Box(x)((\exists y)(Tya\ \&\ (u)(Tua \rightarrow u{=}y)\ \&\ y{=}x) \rightarrow Tx)$ (*dicto*)

(4) is false when read as (4") and true when read as (4"'). On the

other hand, (9) is false when read *de dicto* and true when read *de re*.

There is an alternative way of looking at the matter, however. We may agree that (4) and (9) are ambiguous, but we don't suppose that there is a problem of scope. That's because we don't eliminate the descriptions in terms of existential and uniqueness clauses. By using descriptions constructed with propositional formulas to represent a reading of the English descriptions, the question of scope doesn't arise. The arguments turn out to be perfectly valid. We translate "the teacher of Alexander" as "$(\iota x)Txa$" and "the number of planets" as "$(\iota x)Nx\{y|y \text{ is a planet}\}$." Let's abbreviate our translations as $(\iota x)\phi_5$ and $(\iota x)\phi_6$, respectively. Then we represent the above arguments as follows:

(4a) $\Box T(\iota x)\phi_5$

(5a) $a=(\iota x)\phi_5$

∴ (6a) $\Box Ta$

(7a) $\Box 9 > 7$

(8a) $9=(\iota x)\phi_6$

∴ (9a) $\Box(\iota x)\phi_6 > 7$

In both arguments, the conclusion follows by a simple application of =E. It's just that (4a) is false--the description rigidly denotes Aristotle. If we want to translate (4) as a truth, we have to translate it as (4') above. (4') and (5a) don't imply (6a).

(9a) is true--the description "$(\iota x)\phi_6$" rigidly denotes the number nine. By λ-conversion on (9a), we get that $[\lambda y\ \Box y>7](\iota x)\phi_6$, i.e., the number of planets exemplifies being necessarily greater than seven. However, (9a) doesn't do justice to an apparent falsehood which (9) seems to express. Such a falsehood might be (9b):

(9b) $\Box(\iota z)(F)(zF \equiv [\lambda x\ \phi_6] \Rightarrow F)[\lambda y\ y>7]$

(9b) asserts that necessarily the$_A$ number of planets encodes being greater than seven. This is false since the property of being greater than seven is not logically implied by the property of being a number which numbers the planets. Clearly, (7a) and (8a) don't imply (9b).[27]

Our explanation of these matters consists in translating English descriptions into descriptions of our formal language. We don't systematically eliminate the definite article of English in terms of existential and uniqueness clauses. Rather, we suppose that the word "the" in natural language is captured by the iota operator. But the different formulas behind the iota represent the different ways the English description may be functioning in a given sentence.

CHAPTER IV ENDNOTES

[1]See F.P. Ramsey, [1927].

[2]The material in this section was first developed during the writing of the first draft in Fall 1979.

[3]I'm indebted to Blake Barley for noting this simplification.

[4]I've presented the proof in great detail primarily so that the reader may become familiar with the inner workings of the non-standard modal system. Intuitively, the proof is quite simple. We can visualize the proof more easily by "thinking semantically," using the notion of a possible world as primitive: by definition, at some possible world, all and only the true propositions at that world are encoded in z_0. So "go" to an arbitrary such world and see how z_0 behaves there. It must encode, for an arbitrary proposition Q^0, either $[\lambda y\ Q^0]$ or $[\lambda y\ {\sim}Q^0]$, depending on whether Q^0 or ${\sim}Q^0$ is true at that world. Since A-objects rigidly encode their properties from world to world (LA9), z_0 encodes either $[\lambda y\ Q]$ or $[\lambda y\ {\sim}Q]$ back at the actual world.

Although this semantic visualization of the proof greatly simplifies matters, it has some curious disadvantages in the context of the present work. Were we to use this proof technique in the text, it would often be confusing as to whether we were using the word "world" as defined in D_{17} or in its primitive sense as used in Chapter III. If our A-object theoretical reconstruction of worlds is correct, this confusion wouldn't be significant. But for purposes of clear exposition, it seems inappropriate to use the word "world" one way in the statements of the theorems, and another way in the presentation of the proofs. A second disadvantage of this simplification is that it fosters the wrong impression. I don't want the reader to think as if certain A-objects can represent the worlds; I want the reader to think that worlds just are A-objects.

For these reasons, the exposition of the theorems and proofs in the text will be proof-theoretic. Those who prefer to think model-theoretically should be able to satisfy themselves that the theorems are provable, without attending so closely to the proofs provided in the text.

[5]The proof is in Appendix D.

[6]See L. Wittgenstein, [1921].

[7]See D. Lewis, [1968].

[8]The material in this section was first developed in my [1979e], written for an independent study on Leibniz under Robert Sleigh's supervision.

[9]The best attempt I know of to make this view precise in orthodox theory is Benson Mates [1968].

[10]Parsons was the first to attempt a precise modelling of monads in an object theoretical framework. See his [1978], [1980]. Parsons' results are proven as metatheorems, with the notion of possible world as primitive. Nevertheless, two of his metatheorems served as the inspiration and prototypes for the results which follow. Castañeda claims to have suggested similar results along these lines in [1974],

p. 24. The reader is encouraged to evaluate his suggestion.

[11]Compare Parsons [1978], p. 147, R_2, and [1980], Ch. VIII, Metatheorem 1, §3.

[12]Compare Leibniz [1686b], §9, and [1714], §56.

[13]Compare Parsons [1978], p. 147, R_3.

[14]Compare Parsons [1978], p. 147, R_4, and [1980], VIII, §3, Metatheorem 2.

[15]See Leibniz [1686a], pp. 44, 52, 57, 63, and 64. These references are to G.41, 47, 51, 56, and 57.

[16]B. Partee notes that nothing has been said to distinguish stories from essays. The intentions of the author may be relevant here.

[17]Suppose Story(z_0). ($\rightarrow$) Assume z_0P, for an arbitrary property P. Since z_0P, Vac(P). So for some proposition Q^0, $P=[\lambda y\ Q^0]$. So $z_0[\lambda y\ Q^0]$ & $P=[\lambda y\ Q^0]$. By EI, $(\exists F^0)(\Sigma_{z_0} F^0$ & $P=[\lambda y\ F^0])$; that is, ϕ^P_F. So by A-DESCRIPTIONS, $(\iota z')(z'F \equiv \phi)P$.

($\leftarrow$) Assume $(\iota z')(z'F \equiv \phi)P$. Then ϕ^P_F. By reversing the reasoning, z_0P.

Consequently, z_0 and $(\iota z')(F)(z'F \equiv \phi)$ encode exactly the same properties, so they are identical.

[18]Of course, it may be a matter for literary debate as to which propositions are true according to a given story. And the construction of principles which help us to decide the conditions under which a given proposition is true according to a story poses an interesting philosophical problem. That's because the sentences inscribed by the author in the manuscript (or uttered in a storytelling) are not the

only sentences which denote propositions true according to the story. By far, the majority of propositions true according to the story are not explicitly stated. Most are the result of an extrapolation process which facilitates communication between the author and his audience. The principles governing the extrapolation process are rather mysterious (see Parsons, [1980], Ch. VII). However, we need not concern ourselves with such mysteries, since the place to begin investigation is with the authorship relation--a relation we take as primitive. To the extent that this relation is unclear, so, too, will our proposal be. However, it should be said that this really reflects a genuine unclarity in our pretheoretical conceptions of the relevant stories.

[19] I believe that this is an important result. It seems to me that much of the potential fiction has for affecting us is bound up in our being able to project ourselves into unreal circumstances which involve objects with which we're already familiar.

[20] The Seven Per Cent Solution, a novel by Nicolas Meyer, is supposedly about the secret life of Sherlock Holmes. In this novel, Holmes is a heroin addict. The novel is meant to be consistent with the Conan Doyle novels.

[21] Compare Parsons' [1980], Ch. III, §2.

[22] By N-CHARACTERS and A-DESCRIPTIONS.

[23] Whenever the notion of relevant entailment ("$\underset{R}{\Rightarrow}$") is formulated precisely, we can replace Σ-SUB by the following axiom:

$$(\phi \underset{R}{\Rightarrow} \psi \ \&\ \Sigma_s\phi) \rightarrow \Sigma_s\psi$$

This latter axiom seems to be a principle intimately connected (governing?) the extrapolation process. That is, when we add a proposition to our "maximal account" of the novel, we want to add all the propositions relevantly entailed by this proposition. Σ-SUB should therefore be justified as a special instance of this more general principle, since it should be a consequence of the axioms for relevant entailment that $\phi \underset{R}{\Rightarrow} [\lambda x\ \phi^x_o]o$.

[24]D. Lewis poses the following rhetorical question in his [1978], p. 37: "Is there not some perfectly good sense in which Holmes, like Nixon, *is* [his emphasis] a real-life person of flesh and blood?" We agree that there is. The sense of "is" in question is "encodes."

[25]If we want to represent "Holmes is a famous fictional detective," we suppose that being famous ("F") is an extranuclear property, and that this is a property Holmes exemplifies. Consequently, we get: Fh & F-detective(h).

However, see Appendix E for a possible method of construing [λx Fx & F-detective(x)] as denoting an *abstract* property.

[26]If we were concerned primarily with the sentence "Fernando de Soto searched for the fountain of youth," we would suppose that "the fountain of youth" was intended to denote a character of a legend. We would then apply the analysis of the previous section. But our purposes in this section are somewhat different.

[27]There are other ways to get a false reading of (9) which use descriptions of A-objects:

(9c) $\Box(x)(WCor(x,\ (\iota z)(F)(zF \equiv [\lambda y\ \phi 6^y_x] \Rightarrow F)) \rightarrow x>7)$

(9d) $(w)([\lambda y\ y>7](\iota x)\Sigma_w \phi_6)$

CHAPTER V

THE TYPED THEORY OF ABSTRACT OBJECTS

The typed version of our theory commits us not only to abstract objects, but also to abstract properties, abstract relations, abstract properties of properties, abstract properties of relations, etc. We can use these entities to model impossible relations, like the symmetrical, non-symmetrical relation, and fictional relations, like simultaneity.[1] However, the primary motivation for developing the typed theory is to account for the data concerning the propositional attitudes.

The verbs of propositional attitude (e.g., believes, knows, desires, hopes, expects, discovers, etc.), often combine with the word "that" and an English sentence to produce logically problematic predicates like "believes that Cicero was a Roman," and "hopes that Kennedy is elected President." Frege noticed that terms (simple, complex names) inside these propositional attitude constructions exhibit rather strange behavior. In particular, Frege noticed that from the fact that someone believes that $\ldots\tau_1\ldots$, it doesn't follow that they believe that $\ldots\tau_2\ldots$, even when $\tau_1=\tau_2$ (where $\ldots\tau_1\ldots$ is any English sentence in which term τ_1 occurs, and $\ldots\tau_2\ldots$ is the result of replacing one occurrence of τ_1 with τ_2).[2] For example, each of the following triads of English sentences is consistent:

(1) S believes that Cicero was a Roman
(2) S doesn't believe that Tully was a Roman
(3) Cicero is Tully

(4) S believes that Socrates was the teacher of Plato
(5) S doesn't believe that the son of Phaenarete was the teacher of Plato
(6) Socrates is the son of Phaenarete

(7) S believes that x is French fire engine blue
(8) S doesn't believe that x is Crayola crayon blue
(9) French fire engine blue is Crayola crayon blue.[3]

It seems that the law of identity elimination (=E) doesn't preserve truth when applied to terms in propositional attitude contexts, and this constitutes the problem of "the logically deviant behavior of terms in intermediate contexts."

If in a given case, the law of identity elimination appears to fail, philosophers call the belief (context) *de dicto*, and distinguish it from a belief (context) *de re*, in which identity elimination preserves truth. When S's belief is *de re*, it does follow from the facts that S believes that $...\tau_1...$ and $\tau_1=\tau_2$, that S believes that $...\tau_2...$.

To account for this phenomenon of *de dicto* propositional attitudes, Frege theorized that there must be distinct entities, "senses," associated with the terms τ_1 and τ_2. These entities lend the term with which they're associated information, or cognitive value by serving somehow to <u>re</u>-present the object or relation denoted by the term. This "mode of presentation" embodied by the sense of the term stores information about the denotation of the term, assuming it to have one. And it is the sense of the term which the term denotes when it is situated in a *de dicto* context. Frege would argue that identity elimination is a perfectly good rule of inference; it is just that

English terms are ambiguous, and have different denotations when they're in and out of *de dicto* contexts. Identity elimination preserves truth when you substitute terms which have the same denotation.

Using the theory we have so far, we could construe the senses of English names and descriptions which denote objects as abstract objects. An association of abstract objects with English terms would allow us to picture how a given term had "information" or "cognitive" value. Abstract objects could "re-present" an object denoted by a term by encoding properties the object exemplified. They could serve to store information by encoding many such properties. Finally, they could serve as the denotation of the term when the term is located inside *de dicto* contexts.

Such an association between terms denoting objects and abstract objects is one of the most important features of the language developed in this chapter. We use this language to translate data similar to (1)-(6) in §1 of Chapter VI. However, (7)-(8)-(9) constitute an example of the *de dicto* phenomenon with respect to English names which denote relations. "French fire engine blue" and "Crayola crayon blue." are names of certain properties--properties which we could suppose to be identical. In order to account for the logically deviant behavior of these names, we associate with them abstract properties--properties which encode properties of properties. These abstract properties can lend property names their information value--they could store information about the properties denoted by such names by encoding properties of them. And these abstract properties can serve as the denotation of these names when the name is located in a *de dicto* context.

Similarly with English names which denote relations--we utilize abstract relations, relations which encode properties of relations, to serve as their sense. A completely general account of the senses of names of relations in the type hierarchy requires that we have abstract entities at each type which encode properties of the entities of that type. This is by far the most interesting application of the typed version of our theory.

In what follows, we shall use the word "object" in a new manner. The things which we have been calling "objects" will now be called "individuals." We shall now use the term "object" to discuss any kind of entity whatsoever--existing and abstract individuals, existing and abstract properties and relations, existing and abstract properties of properties (relations), etc. Thus, we call the developments in the next few pages "the typed theory of abstract objects," and we affectionately refer to it as "metaphysical hyperspace."

§1. The Language

In the usual manner, we recursively define the set of types.[4] For our purposes, we may think of them as a group of symbols which serve to categorize simultaneously the terms of the language and the entities they denote.

TYPES

1. "i" is a type (for individuals)
2. "p" is a type (for propositions)
3. If $t_1,\ldots,t_n$ are types, then $(t_1,\ldots,t_n)/p$ is a type (for relations)

The properties (and the expressions which named them) which we used in Chapters I-IV were of type i/p. The relations were of type (i,...,i)/p. But now we have an infinitely branching hierarchy.

A. Primitive terms. Officially, we use a_1^t, a_2^t, ... as names, and x_1^t, x_2^t, ... as variables for objects of each type t. These are the only primitive terms of the language. However, whenever a, b, c, ... and x, y, z, ... appear without typescripts, we assume they denote (range over) individuals (unless their first occurrence in a formula has a typescript and it is understood that the typescripts have been omitted for convenience from the later occurrences). Also, we use $P^{(t_1,\ldots,t_n)/p}$, $Q^{(t_1,\ldots,t_n)/p}$, ... and $F^{(t_1,\ldots,t_n)/p}$, $G^{(t_1,\ldots,t_n)/p}$, ... as names and variables for objects of relational types $(t_1,\ldots,t_n)/p$. And we use P^p, Q^p, ... and F^p, G^p, ... as names and variables for objects of type p. We indicate other abbreviations when the occasion arises.

It will be convenient to distinguish certain names for special purposes. We use $E!^{t/p}$ as the existence predicate for objects of type t. We use $Ex^{((t_1,\ldots,t_n)/p,\ t_1,\ldots,t_n)/p}$ as the explicit exemplification predicate, for all types $t_1,\ldots,t_n$. We use $B_1{}^{(i,p)/p}$, $B_2{}^{(i,p)/p}$, ... to translate the verbs of propositional attitudes. Finally, we use $R^{(t,t,i)/p}$ as the representation predicate--an object of type t represents another object of type t with respect to an individual of type i.

In addition to these terms, we utilize our usual list of grammatical symbols: connectives: ~, &; quantifier: $\exists$; lambda: λ;

iota: ι; box: $\Box$; parentheses and brackets: (,), [,]. We add to this list a one-place sentential operator: that-.

B. Formulas and terms. We simultaneously define (propositional) formula and term of type t. The definition has ten clauses and is rather complex--we sometimes insert extended comments and give examples between the clauses:

1. All primitive terms of type t are terms of type t
2. If τ is a term of type p, then τ is a (propositional) formula
3. Atomic exemplification: If ρ is a term of type $(t_1,\ldots,t_n)/p$ and $\tau_1,\ldots,\tau_n$ are terms of type $t_1,\ldots,t_n$, respectively, then $\rho\tau_1\ldots\tau_n$ is a (propositional) formula

We call ρ the initial term of the atomic exemplification formula. $\tau_1,\ldots,\tau_n$ are called the argument terms. We call any atomic exemplification formula which has the Ex predicate as the initial term an explicit exemplificational formula.

4. Atomic encoding: If ρ is a term of type t/p, and τ is a term of type t, $\tau\rho$ is a formula
5. Molecular, Quantified, and Modal: If ϕ and ψ are (propositional) formulas and α is a variable of any type, $(\sim\phi)$, $(\phi \;\&\; \psi)$, $(\exists\alpha)\phi$, and $(\Box\phi)$ are (propositional) formulas

In the usual manner, we define:

D_1 x^t is abstract ("$A!^{t/p}x$") $=_{dfn}$ $\Box{\sim}E!^{t/p}x$

We use z^t variables to range over the abstract objects of type t.

By inserting the parenthetical remarks in the definition thus far, it follows that ϕ is a propositional formula iff ϕ has no encoding subformulas. This constitutes an expansion of our old notion. We now allow formulas ϕ with quantifiers of any type to be propositional--in particular, formulas ϕ with quantifiers binding initial variables are propositional. For example, $(\exists F^{i/p})Fx^i$ is propositional. Formulas relevantly similar to it in earlier languages would not have been. However, $(\exists G^{i/p})(x^iG \& {\sim}Gx)$ and $(G)(xG \rightarrow Gx)$ are not propositional.

6. Complex terms of relational type: If ϕ is any propositional formula which has no quantifiers binding initial variables, and $\alpha_1,\ldots,\alpha_n$ are any variables with types $t_1,\ldots,t_n$, respectively, and none of the α_i's are initial terms in ϕ, then $[\lambda\alpha_1\ldots\alpha_n\ \phi]$ is a term of type $(t_1,\ldots,t_n)/p$

Given this clause, there are three important restrictions on the formulas ϕ which may appear behind the λ: (1) ϕ may not contain encoding subformulas, (2) ϕ may not contain any quantifiers binding initial variables, and (3) no variable bound by a λ may appear as an initial variable somewhere in ϕ. There is a way, however, to effectively eliminate the latter two restrictions. $[\lambda x\ (\exists F^{i/p})Fx]$ violates restriction (2). But we could suppose that it abbreviates $[\lambda x\ (\exists F^{i/p})ExFx]$, which will denote the first projection of the exemplification relation of type $(i/p,i)/p$.[6] Also, $[\lambda F^{i/p}\ Fb]$ violates restriction (3). But we could suppose that it abbreviated $[\lambda F^{i/p}\ ExFb]$, which will denote the property which results by plugging the exemplification relation of type $(i/p,i)/p$ in its second place by the denotation of b.

We can make this abbreviation procedure general and eliminate restrictions (2) and (3) in two steps. First we define the <u>explicit exemplificational form</u> ("ϕ^{Ex}") <u>of</u> a propositional formula ϕ as follows:

(i) If ϕ is any primitive term of type p, $\phi^{Ex} = \phi$

(ii) If $\phi = \rho\tau_1 \ldots \tau_n$, then if $\rho \neq Ex$, $\phi^{Ex} = Ex\rho\tau_1 \ldots \tau_n$ and if $\rho = Ex$, $\phi^{Ex} = \phi$

(iii) If $\phi = (\sim\psi)$, $(\psi \; \& \; \chi)$, $(\exists\alpha)\psi$, or $(\Box\psi)$, $\phi^{Ex} = (\sim\psi^{Ex})$, $(\psi^{Ex} \; \& \; \chi^{Ex})$, $(\exists\alpha)\psi^{Ex}$, and $(\Box\psi^{Ex})$, respectively

Secondly, we propose the following definition schema:

> Where ϕ is any propositional formula and $\alpha_1, \ldots, \alpha_n$ are variables with types $t_1, \ldots, t_n$, respectively, and either ϕ contains a quantifier binding an initial variable or one of the α_i's is an initial variable somewhere in ϕ, then $[\lambda\alpha_1 \ldots \alpha_n \; \phi]$ abbreviates $[\lambda\alpha_1 \ldots \alpha_n \; \phi^{Ex}]$

So if $\phi = (\exists G^{i/p})(Gx^i \; \& \; Gb)$, then $[\lambda x \; \phi]$ abbreviates $[\lambda x \; (\exists G^{i/p})$ $(ExGx \; \& \; ExGb)]$. The effect is that <u>any</u> propositional formula may appear after a λ. So $[\lambda x \; (\exists G^{i/p})(xG \; \& \; \sim Gx)]$ and $[\lambda x \; (G)(xG \rightarrow Gx)]$ are still ill-formed, and this prevents the known paradoxes.

7. Complex propositional terms: If ϕ is any propositional formula, ϕ and that-ϕ are terms of type p

8. Sense terms: If κ^t is any primitive name of type t, and σ is any primitive term of type i, $\underline{\kappa}^t_\sigma$ is a term of type t

Intuitively, what this does is give us a means for denoting the abstract object an individual associates with a given name as its sense. We suppose, with Frege, that the sense of a name varies from person to person (see VI, §1). For example, "Socrates" and "Frege" are names of type i, so "$\underline{\text{Socrates}}_{\text{Frege}}$" is a sense term of type i. It shall denote the abstract individual which serves as the sense of the name "Socrates" with respect to Frege. "French fire engine blue" is a name of type i/p. So "$\underline{\text{French fire engine blue}}_{\text{Frege}}$" is a sense term of type i/p and will denote the abstract property which serves to represent the property of being French fire engine blue to Frege. This abstract property encodes properties of type (i/p)/p, i.e., properties of i/p-properties.

9. Object descriptions: If ϕ is any formula with one free variable x of type t, then $(\iota x^{t})\phi$ is a term of type t

For example, where "T" denotes the (i,i)/p-relation of teaching, and "p" denotes Plato, $(\iota x)Txp$ reads "the teacher of Plato." Where "C" denotes the (i/p)/p-property of being a color, and "L" denotes the preference relation of type (i,i/p,i/p)/p, and "m" denotes Mary, $(\iota x^{i/p})(Cx \;\&\; \sim(\exists y)(Cy \;\&\; Lmyx))$ might read: the i/p-property x such that x is a color which Mary prefers to all others" (i.e., "Mary's favorite color").

10. Sense descriptions: If ϕ is any propositional formula with one free x variable of type t, $\underline{(\iota x^{t})\phi}$ is a term of type t

These sense descriptions will help us to model the senses of English definite descriptions. $\underline{(\iota x^t)\phi}$ shall end up denoting the abstract object of type t which encodes just the property $[\lambda x^t\ \phi\ \&\ (y)(\phi^y_x \rightarrow y=_E x)]$ ("being the unique ϕ"). For example, $\underline{(\iota x^i)Txp}$ shall denote the abstract individual which encodes just the property of being the teacher of Plato. When we concern ourselves specifically with the fact that the English description "the teacher of Plato" exhibits logically deviant behavior inside *de dicto* attitude contexts, we shall translate the English as we normally would into the standard type-theoretic language and then underline it. By doing so, we will have formed an expression which denotes the sense of the English description.

Finally, we say that τ is a <u>term</u> iff there is a type t such that τ is a term of type t.

<u>C. Definitions for identity.</u>

D_2 $\quad x^t =_E t^y =_{dfn} \Diamond E!^{t/p}x\ \&\ \Diamond E!^{t/p}y\ \&\ \Box(F^{t/p})(Fx \equiv Fy)$

D_3 $\quad x^t = y^t =_{dfn} x^t =_E y^t \vee (A!^{t/p}x\ \&\ A!^{t/p}y\ \&\ \Box(F^{t/p})(xF \equiv yF))$

These definitions work for objects of every type. We have now defined identity$_E$, instead of taking it as primitive. That's because $[\lambda xy\ x=_E y]$ is "well-formed"--it abbreviates a much longer, well-formed λ-expression. So we won't need an axiom governing identity$_E$.

When t=i, D_3 reduces to our old definition of general identity among objects. $[\lambda xy\ x=y]$ is still not a well-formed λ-expression, nor does it abbreviate a well-formed λ-expression. When t=i/p, D_3 says that properties $F^{i/p}$ and $G^{i/p}$ are identical just in case they are identical$_E$ i/p or they are both abstract and encode the same i/p/p-properties. Consequently, we no longer need a special definition of identity among property types. If we want to say $F^{t/p} =_E G^{t/p}$, we use D_2. If we want to say $F^{t/p} = G^{t/p}$, we use D_3. It will be provable that $F^{t/p} = G^{t/p} \equiv (x^t)(xF \equiv xG)$, i.e., t/p-properties F and G are identical iff they are encoded by the same t-objects.[7] So there are two conditions on the identity of t/p-properties--(i) they must both exemplify (or encode, if they're abstract) the same t/p/p-properties, and (ii) they must be encoded by the same t-objects.

Nor do we need to specially define identity among propositions and relational objects. $F^p = G^p$, $F^p =_E G^p$, $F^{(t_1,\ldots,t_n)/p} = G^{(t_1,\ldots,t_n)/p}$, and $F^{(t_1,\ldots,t_n)/p} =_E G^{(t_1,\ldots,t_n)/p}$ are already defined using D_2 and D_3.[8]

<u>§2. The Semantics</u>

<u>A. Interpretations</u>. An interpretation, I, of our type theoretic language is any octuple, $\langle W, w_0, D, ext_w, L, ext_A, sen, F\rangle$, which meets the conditions described in this subsection. The first member of I is a non-empty class, W, called the class of <u>possible worlds</u>. The second member of I, w_0, is a member of W and is called the <u>actual world</u>. The third member of I, D, is a non-empty class called the <u>domain of objects</u>. D is the union of a collection of non-empty, indexed, classes, i.e., $D = \bigcup_{t\in TYPE} D_t$. Each class in the collection, D_t, is called the <u>domain of objects of</u> type t. We call D_i the <u>domain of individuals</u>, D_p the <u>domain of propositions</u>, $D_{t/p}$ the <u>domain of properties of type t objects</u>, $D_{(t_1,\ldots,t_n)/p}$ the <u>domain of n-place</u>

relations among objects with types $t_1,\ldots,t_n$, respectively. We use "o^t" as a metalinguistic variable ranging over the objects in $\mathcal{D}_t$.

For convenience, we call the class of all objects with types not equal to i the class of higher order objects and we use "R" to denote this domain. So $R = \bigcup_{t \neq i} \mathcal{D}_t$. R is closed under all the logical functions specified in L, the fifth member of an interpretation. R may be subdivided into domains of relational types $R_{(t_1,\ldots,t_n)/p}$ and the domain of propositions R_p. We use "r^t" as a metalinguistic variable ranging over the higher order objects of type t.

We also let "A_t" denote the class of abstract objects of type t. $A_t = \{o^t \mid (w)(o^t \notin ext_w(F(E!^{t/p})))\}$, where ext_w and F are the fourth and eighth members of the interpretation, as defined below. We use "a^t" as metalinguistic variables ranging over the members of A_t.

The fourth member of I, ext_w, is a function defined on $R X W$ as follows:

(a) ext_w: $R_{(t_1,\ldots,t_n)/p} \, X \, W \rightarrow (\mathcal{D}_{t_1} X \mathcal{D}_{t_2} X \ldots X \mathcal{D}_{t_n})$

(b) ext_w: $R_p \, X \, W \rightarrow \{T,F\}$

Thus, the ext_w function distributes an exemplification extension at each world to all the higher order objects.

The fifth member of I, L, is a class of logical functions with members: $PLUG_j$, $PROJ_j$, $CONV_{j,k}$, $REFL_{j,k}$, $VAC_{j,t'}$, $CONJ$, NEG, and NEC. These functions are defined as follows:

(a) $PLUG_j$ is a function from $\bigcup_{1 \leq j \leq n} R_{(t_1,\ldots,t_j,\ldots,t_n)/p} \, X \, \mathcal{D}_{t_j}$

into $\bigcup_{1 \leq j \leq n} R_{(t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_n)/p}$ subject to the following conditions:

(1) for n>1, $ext_w(PLUG_j(r^{(t_1,\ldots,t_n)/p}, o^{t_j})) =$

$\{<o^{t_1},\ldots,o^{t_{j-1}},o^{t_{j+1}},\ldots,o^{t_n}> \mid$

$<o^{t_1},\ldots,o^{t_{j-1}},o^{t_j},o^{t_{j+1}},\ldots,o^{t_n}> \varepsilon$

$ext_w(r^{(t_1,\ldots,t_n)/p})\}$

(2) for n=1, $ext_w(PLUG_1(r^{t/p}, o^t)) =$

$$\begin{cases} \text{T iff } o^t \ \varepsilon\ ext_w(r^{t/p}) \\ \text{F otherwise} \end{cases}$$

(b) $PROJ_j$ is a function from $\bigcup_{1\leq j\leq n} R_{(t_1,\ldots,t_j,\ldots,t_n)/p}$ into $\bigcup_{1\leq j\leq n} R_{(t_1,\ldots,t_{j-1},t_{j+1},\ldots,t_n)/p}$ subject to the conditions:

(1) for n>1, $ext_w(PROJ_j(r^{(t_1,\ldots,t_n)/p})) =$

$\{<o^{t_1},\ldots,o^{t_{j-1}},o^{t_{j+1}},\ldots,o^{t_n}> \mid$

$(\exists o^{t_j})(<o^{t_1},\ldots,o^{t_{j-1}},o^{t_j},o^{t_{j+1}},\ldots,o^{t_n}> \varepsilon$

$ext_w(r^{(t_1,\ldots,t_n)/p}))\}$

(2) for n=1, $ext_w(PROJ_1(r^{t/p})) =$

$$\begin{cases} \text{T iff } (\exists o^t)(o^t \ \varepsilon\ ext_w(r^{t/p})) \\ \text{F otherwise} \end{cases}$$

(c) $CONV_{j,k}$ is a function from $\bigcup_{1\leq j<k\leq n} R_{(t_1,\ldots,t_j,\ldots,t_k,\ldots,t_n)/p}$

into $\bigcup_{1 \le j < k \le n} R_{(t_1,\ldots,t_{j-1},t_k,t_{j+1},\ldots,t_{k-1},t_j,t_{k+1},\ldots,t_n)/p}$

subject to the following condition:

$$ext_w(CONV_{j,k}(r^{(t_1,\ldots,t_n)/p})) =$$

$$\{<o^{t_1},\ldots,o^{t_{j-1}},o^{t_k},o^{t_{j+1}},\ldots,o^{t_{k-1}},o^{t_j},o^{t_{k+1}},\ldots,o^{t_n}> \mid$$

$$<o^{t_1},\ldots,o^{t_j},\ldots,o^{t_k},\ldots,o^{t_n}> \varepsilon\ ext_w(r^{(t_1,\ldots,t_n)/p})\}$$

(d) $REFL_{j,k}$ is a function from $\bigcup_{1 \le j < k \le n} R_{(t_1,\ldots,t_j,\ldots,t_k,\ldots,t_n)/p}$ into $\bigcup_{1 \le j < k \le n} R_{(t_1,\ldots,t_j,\ldots,t_{k-1},t_{k+1},\ldots,t_n)/p}$ subject to the following condition:

$$ext_w(REFL_{j,k}(r^{(t_1,\ldots,t_n)/p})) =$$

$$\{<o^{t_1},\ldots,o^{t_j},\ldots,o^{t_{k-1}},o^{t_{k+1}},\ldots,o^{t_n}> \mid$$

$$<o^{t_1},\ldots,o^{t_j},\ldots,o^{t_k},\ldots,o^{t_n}> \varepsilon\ ext_w(r^{(t_1,\ldots,t_n)/p})$$

and $o^{t_j}=o^{t_k}\}$

(e) $VAC_{j,t'}$ is a function from $(\bigcup_{1 \le j \le n+1} R_{(t_1,\ldots,t_j,\ldots,t_n)/p})$ $\cup\ R_p$ into $(\bigcup_{1 \le j \le n+1} R_{(t_1,\ldots,t_{j-1},t',t_j,t_{j+1},\ldots,t_n)/p})$ $\cup\ R_{t'/p}$ subject to the conditions:

(1) if $t=(t_1,\ldots,t_j,\ldots,t_n)/p$, then $1 \le j \le n+1$ and

$$ext_w(VAC_{j,t'}(r^t)) =$$

$\{<o^{t_1},\ldots,o^{t_{j-1}},o^{t'},o^{t_j},o^{t_{j+1}},\ldots,o^{t_n}> \mid$

$<o^{t_1},\ldots,o^{t_j},\ldots,o^{t_n}> \varepsilon\ ext_w(r^t)\}$

(2) if t=p, then j=1 and $ext_w(VAC_{1,t'}(r^p)) =$

$\{o^{t'} \mid ext_w(r^p) = T\}$

(f) *CONJ* is a function from *RXR* into *R* subject to the following conditions:

(1) if $t=(t_1,\ldots,t_n)/p$ and $t'=(t_1',\ldots,t_m')/p$, then $ext_w(CONJ(r^t,s^{t'})) =$

$\{<o^{t_1},\ldots,o^{t_n},o^{t_1'},\ldots,o^{t_m'}> \mid <o^{t_1},\ldots,o^{t_n}>$

$\varepsilon\ ext_w(r^t)$ and $<o^{t_1'},\ldots,o^{t_m'}> \varepsilon\ ext_w(s^{t'})\}$

(2) if $t=(t_1,\ldots,t_n)/p$ and $t'=p$, then

$ext_w(CONJ(r^t,s^p)) =$

$\{<o^{t_1},\ldots,o^{t_n}> \mid <o^{t_1},\ldots,o^{t_n}> \varepsilon\ ext_w(r^t)$

and $ext_w(s^p) = T\}$

(3) if t=p and $t'=(t_1,\ldots,t_m)/p$, then

$ext_w(CONJ(r^p,s^{t'})) =$

$\{<o^{t_1},\ldots,o^{t_m}> \mid ext_w(r^p) = T$ and

$<o^{t_1},\ldots,o^{t_m}> \varepsilon\ ext_w(s^{t'})\}$

(4) if t=p and t'=p, then $ext_w(CONJ(r^p,s^p)) =$

$$\begin{cases} T \text{ iff } ext_w(r^p) = T \text{ and } ext_w(s^p) = T \\ F \text{ otherwise} \end{cases}$$

(g) *NEG* is a function from *R* into *R* subject to the conditions:

(1) if $t=(t_1,\ldots,t_n)/p$, then $ext_w(NEG(r^t)) =$

$\{<o^{t_1},\ldots,o^{t_n}> \mid <o^{t_1},\ldots,o^{t_n}> \not\varepsilon\ ext_w(r^t)\}$

(2) if t=p, then $ext_w(NEG(r^p)) =$

$$\begin{cases} T \text{ iff } ext_w(r^p) = F \\ F \text{ otherwise} \end{cases}$$

(h) *NEC* is a function from *R* into *R* subject to the conditions:

(1) if $t=(t_1,\ldots,t_n)/p$, then $ext_w(NEC(r^t)) =$

$\{<o^{t_1},\ldots,o^{t_n}> \mid (w')(<o^{t_1},\ldots,o^{t_n}> \varepsilon\ ext_{w'}(r^t))\}$

(2) if t=p, then $ext_w(NEC(r^p)) =$

$$\begin{cases} T \text{ iff } (w')(ext_{w'}(r^p) = T) \\ F \text{ otherwise} \end{cases}$$

This completes the definitions of the logical functions. The sixth member of *I*, ext_A, is a function defined on $\bigcup_{t\varepsilon TYPE} R_{t/p}$. For a given type t, ext_A maps $R_{t/p}$ into D_t. ext_A assigns each higher order property of t-objects an <u>encoding</u> extension among these objects.

Let N_t be the set of primitive names of type t of our language. Then, the seventh member of I is the sense function, sen, which maps $D_i \times N_t$ into A_t (the set of abstract objects of type t). For convenience, we index the sen function to its first argument. Thus, for a given individual o, sen_o associates with a given name κ^t of type t an abstract object of type t. We call $sen_o(\kappa)$ the <u>sense</u> of κ <u>with respect to</u> o.[9] Intuitively, if "Socrates" is a name of type i, then sen_{Frege}("Socrates") is the abstract individual which serves as the sense of the name "Socrates" with respect to Frege. We shall assign this object to the sense term "$\underline{\text{Socrates}}_{\text{Frege}}$." And we shall make it a logical truth that $\underline{\text{Socrates}}_{\text{Frege}}$ represents Socrates to Frege.[10] We shall sometimes superscript the sense function to the type of the name upon which it is operating. For example, $sen^{i/p}_{Frege}$("French fire engine blue") is the abstract i/p-property which serves as the sense of "French fire engine blue" with respect to Frege.

The eighth member of I is a function, F, defined on the primitive names *and* on the closed sense terms of the language. For each name κ^t of type t, $F(\kappa^t) \in D_t$. For each closed sense term $\underline{\kappa}^t_{\sigma}$ of type t, $F(\underline{\kappa}^t_{\sigma}) = sen^t_{F(\sigma)}(\kappa^t)$. Recall that sense terms can have only primitive terms as subscripts. So the closed sense terms will have only primitive names as subscripts.[11]

In addition, we place the following two restrictions on F:

(1) $ext_w(F(\mathrm{Ex}^{((t_1,\ldots,t_n)/p,\ t_1,\ldots,t_n)/p})) =$
$\{\langle r^{(t_1,\ldots,t_n)/p}, o^{t_1},\ldots, o^{t_n}\rangle \mid$
$\langle o^{t_1},\ldots,o^{t_n}\rangle \in ext_w(r^{(t_1,\ldots,t_n)/p})\}$

So F must assign to the explicit exemplification predicate a relation with the "appropriate" extension.

(2) $ext_w(F(R^{(t,t,i)/p})) =$

$\{<a^t, o^t, o^i> \mid (\exists \underline{\kappa}_\sigma)(F(\underline{\kappa}_\sigma) = a^t$

$\& F(\kappa) = o^t \& F(\sigma) = o^i)\}$

Thus, "$R^{(t,t,i)/p}$" denotes any three place relation which objects a^t, o^t, and o^i bear to one another iff there is some closed sense term $\underline{\kappa}_\sigma$ such that a^t is the sense of κ with respect to $o^i(F(\sigma))$ and o^t is the denotation of κ. We say that a^t <u>represents</u> o^t <u>with respect to</u> o^i.

Finally, we call $ext_w(F(E!^{t/p}))$ the set of objects of type t which <u>exist at</u> w ("E^t_w"). We call $ext_{w_0}(F(E!^{t/p}))$ the set of <u>existing</u> objects of type t ("E^t"). And we call $\{o^t \mid (\exists w)(o^t \in ext_w(F(E!^{t/p})))\}$ the set of <u>possibly existing</u> objects of type t ("PE^t").

<u>B. Assignments and denotations</u>. For the most part, the definitions partitioning the λ-expressions are similar to those developed in Chapter III, §2, B. However, we need to type the added place in the definition of vacuous expansion. We also need to concern ourselves with argument variables (rather than the "object" variables of Chapter III) throughout these definitions.

If μ is an arbitrary λ-expression, $[\lambda\alpha_1 \ldots \alpha_n\ \phi]$, μ is defined as follows:

1. If $(\exists j)(1 \leq j \leq n$ and α_j doesn't occur free in ϕ and t' is the type of α_j and j is the least such number), then μ is the <u>j,t'-vacuous</u>

<u>expansion of</u> $[\lambda\alpha_1 \ldots \alpha_{j-1}\alpha_{j+1} \ldots \alpha_n\ \phi]$.

2. If μ is not a <u>j,t'-vacuous expansion</u>, then if $(\exists j)(1 \leq j \leq n$ and α_j is not the j^{th} free argument variable in ϕ and j is the least such number), then where α_k is the j^{th} free argument variable in ϕ, μ is the <u>j,k^{th}-conversion of</u> $[\lambda\alpha_1 \ldots \alpha_{j-1}\alpha_k\alpha_{j+1} \ldots \alpha_{k-1}\alpha_j\alpha_{k+1} \ldots \alpha_n\ \phi]$.
3. If μ is neither of the above, then
   (a) if $\phi = (\sim\psi)$, μ is the <u>negation of</u> $[\lambda\alpha_1 \ldots \alpha_n\ \psi]$
   (b) if $\phi = (\psi\ \&\ \chi)$, and ψ and χ have no free argument variables in common, then where $\alpha_1, \ldots, \alpha_m$ are the variables in ψ and $\alpha_{m+1}, \ldots, \alpha_n$ are the variables in χ, μ is the <u>disjoint conjunction of</u> $[\lambda\alpha_1 \ldots \alpha_m\ \psi]$ <u>and</u> $[\lambda\alpha_{m+1} \ldots \alpha_n\ \chi]$.
   (c) if $\phi = (\exists\beta)\psi$, and β is the j^{th} free argument variable in ψ, then μ is the <u>j^{th}-projection of</u> $[\lambda\alpha_1 \ldots \alpha_{j-1}\beta\alpha_{j+1} \ldots \alpha_n\ \psi]$.
   (d) if $\phi = (\Box\psi)$, then μ is the <u>necessitation of</u> $[\lambda\alpha_1 \ldots \alpha_n\ \psi]$.
4. If μ is none of the above, then if $(\exists j)(1 \leq j \leq n$ and α_j occurs free in more than one place in ϕ and j is the least such number), then where:
   (a) m is the number of free argument variables between the first and second occurrances of α_j,
   (b) ϕ' is the result of replacing the second occurrence of α_j with a new variable β (with the same type as α_j),
   (c) $k = j+m+1$, then

   μ is the <u>j,k^{th}-reflection of</u> $[\lambda\alpha_1 \ldots \alpha_{j+m}\beta\alpha_k \ldots \alpha_n\ \phi']$.
5. If μ is none of the above, then if τ is the leftmost argument term occurring in ϕ, then where
   (a) k is the number of free argument variables occurring before τ,

(b) ϕ' is the result of replacing the first occurrence of τ by a new variable β (with the same type as τ),

(c) j=k+1, then

μ is the j^{th}-plugging of $[\lambda\alpha_1\ldots\alpha_{k-1}\beta\alpha_j\ldots\alpha_n\ \phi']$ by τ.

6. If μ is none of the above, then

(a) ϕ is atomic

(b) $\alpha_1,\ldots,\alpha_n$ is the order in which these variables first occur in ϕ,

(c) $\mu = [\lambda\alpha_1\ldots\alpha_n\ \rho^n\alpha_1\ldots\alpha_n]$, for some term ρ^n, and

(d) μ is called elementary

I-assignments. If given an interpretation *I* of the language, an *I*-assignment will be any function, $\mathfrak{g}$, defined on the primitive variables of the language such that when α is a variable of type t, $\mathfrak{g}(\alpha) \in \mathcal{D}_t$.

Denotations. If given an interpretation *I* and an *I*-assignment $\mathfrak{g}$, we recursively define the denotation of term τ with respect to *I* and $\mathfrak{g}$ ("$d_{I,\mathfrak{g}}(\tau)$") as follows:

1. where κ is any primitive name, $d_{I,\mathfrak{g}}(\kappa) = F(\kappa)$
2. where $\underline{\kappa}_\sigma$ is any closed sense term, $d_{I,\mathfrak{g}}(\kappa) = F(\kappa)$
3. where α is any primitive variable, $d_{I,\mathfrak{g}}(\alpha) = \mathfrak{g}(\alpha)$
4. where $\underline{\kappa}^t_\sigma$ is any open sense term of type t,

   $d_{I,\mathfrak{g}}(\underline{\kappa}^t_\sigma) = sen_{d_{I,\mathfrak{g}}(\sigma)}(\kappa^t)$
5. where μ is an elementary λ-expression $[\lambda\alpha_1\ldots\alpha_n\ \rho^n\alpha_1\ldots\alpha_n]$,

   $d_{I,\mathfrak{g}}(\mu) = d_{I,\mathfrak{g}}(\rho^n)$
6. where μ is the j^{th}-plugging of ξ by τ,

   $d_{I,\mathfrak{g}}(\mu) = PLUG_j(d_{I,\mathfrak{g}}(\xi),\ d_{I,\mathfrak{g}}(\tau))$

7. where μ is the j^{th}-projection of ξ, $d_{I,\mathcal{f}}(\mu) = PROJ_j(d_{I,\mathcal{f}}(\xi))$

8. where μ is the j,k^{th}-conversion of ξ, $d_{I,\mathcal{f}}(\mu) = CONV_{j,k}(d_{I,\mathcal{f}}(\xi))$

9. where μ is the j,k^{th}-reflection of ξ, $d_{I,\mathcal{f}}(\mu) = REFL_{j,k}(d_{I,\mathcal{f}}(\xi))$

10. where μ is the j,t'-vacuous expansion of ξ,

$d_{I,\mathcal{f}}(\mu) = VAC_{j,t'}(d_{I,\mathcal{f}}(\xi))$

11. where μ is the conjunction of ξ and ζ,

$d_{I,\mathcal{f}}(\mu) = CONJ(d_{I,\mathcal{f}}(\xi), d_{I,\mathcal{f}}(\zeta))$

12. where μ is the negation of ξ, $d_{I,\mathcal{f}}(\mu) = NEG(d_{I,\mathcal{f}}(\xi))$

13. where μ is the necessitation of ξ, $d_{I,\mathcal{f}}(\mu) = NEC(d_{I,\mathcal{f}}(\xi))$

14. where μ is any propositional formula ϕ, $d_{I,\mathcal{f}}(\phi)$ is defined as follows:

(a) if ϕ is primitive term of type p, $d_{I,\mathcal{f}}(\phi)$ is already defined

(b) if $\phi = \rho^n\tau_1 \ldots \tau_n$, $d_{I,\mathcal{f}}(\phi) =$

$PLUG_1(PLUG_2(\ldots(PLUG_n(d_{I,\mathcal{f}}(\rho^n), d_{I,\mathcal{f}}(\tau_n)), \ldots),$
$d_{I,\mathcal{f}}(\tau_2)), d_{I,\mathcal{f}}(\tau_1))$

(c) if $\phi = (\sim\psi)$, $d_{I,\mathcal{f}}(\phi) = NEG(d_{I,\mathcal{f}}(\psi))$

(d) if $\phi = (\psi \,\&\, \chi)$, $d_{I,\mathcal{f}}(\phi) = CONJ(d_{I,\mathcal{f}}(\psi), d_{I,\mathcal{f}}(\chi))$

(e) if $\phi = (\exists\alpha^t)\psi$, $d_{I,\mathcal{f}}(\phi) =$

$$\begin{cases} PROJ_1(d_{I,\mathcal{f}}([\lambda\alpha^t\ \psi])) \text{ iff } \alpha^t \text{ is not an initial variable occurring in } \psi \\ PROJ_1(d_{I,\mathcal{f}}([\lambda\alpha^t\ \psi^{Ex}])), \text{ otherwise} \end{cases}$$

(f) if $\phi = (\Box\psi)$, $d_{I,\mathcal{f}}(\phi) = NEC(d_{I,\mathcal{f}}(\psi))$

15. where that-ϕ is any complex propositional term, $d_{I,\mathcal{f}}(\text{that-}\phi) = d_{I,\mathcal{f}}(\phi)$

16. where $(\iota x^t)\phi$ is any object description, $d_{I,\mathcal{f}}((\iota x^t)\phi) =$

$$\begin{cases} o^t \text{ iff } (\exists g')(g' =_{x^t} g \;\&\; g'(x^t)=o^t \;\&\; g' \text{ satisfies } \phi \text{ with respect to } \\ \quad w_0 \;\&\; (g'')(g'' =_{x^t} g \;\&\; g''(x^t)=o^{t\prime} \;\&\; g'' \text{ satisfies } \phi \text{ with respect to } \\ \quad w_0 \rightarrow o^{t\prime}=o^t)) \\ \text{undefined, otherwise} \end{cases}$$

17. where $\underline{(\iota x^t)\phi}$ is any sense description, $d_{I,g}(\underline{(\iota x^t)\phi}) =$ $d_{I,g}((\iota z^t)(F^{t/p})(zF \equiv F=[\lambda x\ \phi \;\&\; (y^t)(\phi^y_x \rightarrow y=_E x)]))$[12]

C. Satisfaction. Given an interpretation I and an I-assignment g, we may define g satisfies ϕ with respect to w as follows:

1. If ϕ is any primitive term of type p, g satisfies ϕ with respect to w iff $ext_w(d_{I,g}(\phi)) = T$
2. If $\phi = \rho^n\tau_1\ldots\tau_n$, g satisfies ϕ with respect to w iff $\langle d_{I,g}(\tau_1),\ldots,d_{I,g}(\tau_n)\rangle \in ext_w(d_{I,g}(\rho^n))$
3. If $\phi = o\rho^1$, g satisfies ϕ with respect to w iff $d_{I,g}(o) \in ext_A(d_{I,g}(\rho^1))$
4. If $\phi = (\sim\psi)$, g satisfies ϕ with respect to w iff g fails to satisfy ψ with respect to w
5. If $\phi = (\psi \;\&\; \chi)$, g satisfies ϕ with respect to w iff g satisfies both ψ and χ with respect to w
6. If $\phi = (\exists\alpha^t)\psi$, g satisfies ϕ with respect to w iff $(\exists g')(g'$ satisfies ψ with respect to $w)$
7. If $\phi = (\Box\psi)$, g satisfies ϕ with respect to w iff $(w')(g$ satisfies ϕ with respect to $w')$

D. Truth under an interpretation. ϕ is true under I iff every I-assignment g satisfies ϕ with respect to w_0.

§3. The Logic

A. The logical axioms. Axioms LA1-LA9, DESCRIPTIONS, and λ-EQUIVALENCE are transposed from the modal theory with typescripts where needed:

LA 1: $\phi \rightarrow (\psi \rightarrow \phi)$

LA 2: $(\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \chi))$

LA 3: $(\sim\phi \rightarrow \sim\psi) \rightarrow ((\sim\phi \rightarrow \psi) \rightarrow \phi)$

LA 4: $(\alpha)\phi \rightarrow ((\exists\beta)\beta{=}\tau \rightarrow \phi^{\tau}_{\alpha})$, provided τ is substitutable for α in ϕ[13]

LA 5: $(\alpha)(\phi \rightarrow \psi) \rightarrow (\phi \rightarrow (\alpha)\psi)$, provided α isn't free in ϕ

LA 6: $\Box\phi \rightarrow \phi$

LA 7: $\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$

LA 8: $\Diamond\phi \rightarrow \Box\Diamond\phi$

LA 9: $(x^{t})(F^{t/p})(\Diamond xF \rightarrow \Box xF)$

DESCRIPTIONS: where ψ is any atomic (or identity) formula, the universal closure of the following is an axiom:

$$\psi_{\alpha}^{(\iota x^{t})\phi} \equiv (\exists! y^{t})(\phi^{y}_{x} \;\&\; \psi^{y}_{\alpha}) \;\&\; (\exists! y^{t})\phi^{y}_{x}$$

λ-EQUIVALENCE: for any formula ϕ where $F^{(t_1,\ldots,t_n)/p}$ isn't free, the universal closure of the following is an axiom:

$$(x^{t_1})\ldots(x^{t_n})\Box([\lambda\alpha_1\ldots\alpha_n\ \phi]x^{t_1}\ldots x^{t_n} \equiv \phi^{x^{t_1},\ldots,x^{t_n}}_{\alpha_1\ ,\ldots,\alpha_n})$$

In addition to these logical axioms, there are four other logical truths validated by our semantic structure:

LA 10: $(F^{(t_1,\ldots,t_n)/p})(x^{t_1})\ldots(x^{t_n})(ExFx^{t_1}\ldots x^{t_n} \equiv Fx^{t_1}\ldots x^{t_n})$

LA10 is a consequence of our restrictions on the F function of any interpretation. It tells us that the explicit exemplification predicate works as it should.

LA 11: where ϕ is any propositional formula, the universal closure of the following is an axiom: that-ϕ = ϕ

LA 12: where $\underline{\kappa}^{t}_{\sigma}$ is any closed sense name, the following is an axiom:

$$A!^{t/p}\underline{\kappa}^{t}_{\sigma} \ \& \ R^{(t,t,i)/p}\underline{\kappa}^{t}_{\sigma}\kappa\sigma$$

Recall that we read the second conjunct of LA12 as: "the sense of the name κ with respect to σ represents κ to σ." Again, the paradigm example is: $\underline{\text{Socrates}}_{\text{Frege}}$ represents Socrates to Frege. LA12 is also a consequence of the restrictions we placed on the F function.

LA 13: where ϕ is any propositional formula, the universal closure of the following is an axiom:

$$\underline{(\iota x^{t})\phi} = (\iota z^{t})(F^{t/p})(zF \equiv F=[\lambda x^{t}\ \phi \ \& \ (y^{t})(\phi^{y}_{x} \rightarrow y=_{E}x)])^{14}$$

LA13 is a consequence of clause 17 in the definition of $\underline{\text{denotation}}_{I,\delta}$. Our sense descriptions denote Platonic Forms of type t (i.e., abstract objects of type t which encode a single t/p-property) which encode an individuating property (i.e., one which at most one object of type t can exemplify).

B. Rules of inference.

1. $\rightarrow$E: If $\vdash \phi\rightarrow\psi$ and $\vdash\phi$, then $\vdash\psi$
2. UI: If $\vdash\phi$, then $\vdash(\alpha^{t})\phi$, where t is any type
3. $\Box$I: If $\vdash\phi$, then $\vdash\Box\phi$, where ϕ has no descriptions

Unofficially, we use the usual derived rules of inference and proof techniques.

We call the metaphysical system which consists of the interpreted typed language (without the unusual complex terms or distinguished predicates), together with LA1-LA9, DESCRIPTIONS, λ-EQUIVALENCE, and the rules of inference, the typed object calculus. The addition of the unusual kinds of complex terms and distinguished predicates, together with their interpretation and logic (especially LA10-LA13), constitutes a special modification of the typed object calculus which has been designed specifically for the data about propositional attitudes.

§4. The Proper Axioms

We assert the following four axioms to be true *a priori*:

AXIOM 1 ("NO-CODER^t"): $(x^t)(\Diamond E!^{t/p}x \rightarrow \sim(\exists F^{t/p})xF)$

AXIOM 2 ("IDENTITY"): $(\alpha=\beta) \rightarrow (\phi(\alpha,\alpha) \rightarrow \phi(\alpha,\beta))$, where $\phi(\alpha,\beta)$ is the result of replacing some, but not necessarily all, free occurrences of α by β in $\phi(\alpha,\alpha)$, provided that β is substitutable for the occurrences of α which it replaces.

AXIOM 3 ("A-OBJECTS^t"): For any formula ϕ where x^t isn't free, the universal closure of the following is an axiom:

$$(\exists x^t)(A!^{t/p}x \;\&\; (F^{t/p})(xF \equiv \phi))$$

AXIOM 4 ("NECESSARY EXISTENCE"): For every type t, t≠i, the following is an axiom:

$$(x^t)(\Diamond E!^{t/p}x \rightarrow \Box E!^{t/p}x)$$

Since we've defined $=_E$ for all types t, we no longer need E^t-IDENTITY. NO-CODER^t, IDENTITY, and A-OBJECTS^t should be straightforward, given our familiarity with their counterparts in Chapters I and III. Note that abstract objects of type t might encode abstract t/p-properties, as well as (possibly) existing ones.[15] The $F^{t/p}$-quantifier in A-OBJECTS^t ranges over all t/p-properties.

We have added one extra axiom to the typed theory to preserve the intuition that higher order objects are not contingent beings. Since higher order objects either possibly exist or fail to possibly exist, it follows from NECESSARY EXISTENCE that either they necessarily exist or they necessarily fail to exist. Philosophers who do not share the intuition that higher order objects aren't contingent beings may not wish to embrace this axiom.

I believe that NO-CODER^t, IDENTITY, A-OBJECTS^t, and NECESSARY EXISTENCE are jointly consistent, though it may be a while before a proof is discovered.[16] We have taken steps to prevent the offending instances of property abstraction from being denoted. However, we have assumed that the presence of "higher order" quantifiers (that is, quantifiers binding initial variables) in λ-expressions is not the source of paradoxes. Consequently, we have allowed such λ-expressions to be well-formed (or at least abbreviate well-formed λ-expressions). We have also assumed that it is safe to have abstract objects of type t encode abstract properties of type t/p. I don't think that these moves reintroduce paradoxes, but they might. Should they do so, there are still ways to weaken the theory and preserve many of the applications which follow. There is a great deal of investigation which must be done before we can feel confident that this particular version of the theory is consistent.

As usual, we add abbreviations for the appropriate English gerunds to our primitive vocabulary. In addition, we add abbreviations for English proper names--names which aren't necessarily associated with works of fiction.[17] Finally, we use the distinguished constants $B_1^{(i,p)/p}$, $B_2^{(i,p)/p}$, ... to abbreviate the verbs of propositional attitudes such as believes, hopes, knows, expects, etc.

CHAPTER V ENDNOTES

[1]There may even be a way to model our metaphysical notions like Form, Monad, World, etc., as abstract properties. Recall that we can't guarantee that they are real properties because their definitions involve encoding formulas. So maybe they are abstract. See Appendix E for the attempt to model them as abstract properties.

[2]Frege [1892], pp. 56-78.

[3]I'd like to thank Mark Aronszajn for contributing this example. It was a result of a discussion with Mark that I discovered that the type theory could be used to model the senses of expressions denoting higher order objects.

[4]See Russell and Whitehead, [1910]. Also Church [1940], pp. 56-58.

[5]I'm indebted to Barbara Partee here, whose comments on the syntax of the type theory in the first draft helped me to be more careful in the final formulation.

[6]It should be clear what the type of the Ex predicate is. In what follows, we frequently omit its type.

[7]The proof is similar to the one described in Chapter I, §4, where we justified the definition of "$F^1=G^1$."

[8]We may need to add $F^p =_E G^p \equiv [\lambda y\ F^p]=[\lambda y\ G^p]$ if it is not derivable. The same goes for relation identity$_E$.

[9]If we wanted to follow Frege a little more closely, we would

define the set of senses of type t, S_t, as follows:

$$S_t =_{dfn} \{o|o \in A_t \;\&\; (w)((\exists o')(r^{t/p})(o \in ext_A(r) \rightarrow o' \in ext_w(r)) \rightarrow (\exists ! o')(r^{t/p})(o \in ext_A \rightarrow o' \in ext_w(r)))\}$$

Thus, the objects in S_t are the abstract objects of type t which have at most one weak correlate. This preserves Frege's intuition that senses *determine* at most one object. For example, any A-object of type t which encoded an "individuating" concept of type t/p would be in S_t, where

$$r^{t/p} \text{ is an individuating concept } =_{dfn} (w)((\exists o^t)(o \in ext_w(r)) \rightarrow (\exists ! o^t)(o \in ext_w(r)))$$

Should one decide that Frege's constraint on senses is essential, one would have to redefine sen so that it mapped $D_i \times N_t$ into S_t.

[10]Note that if κ^t is not in the vocabulary of o or if o is an abstract individual without representational capabilities, then we may suppose that $sen_o(\kappa^t)$ is the null object of type t.

[11]Should it become necessary, we could expand this device by allowing any complex term of type i to serve as subscripts on sense terms. This complication need not be developed for our purposes in Chapter VI.

[12]"z^t" ranges over abstract objects of type t. Consequently, the assignment to "$\underline{(\iota x^t)\phi}$" is the abstract object of type t which encodes just the t/p-property of being the ϕ.

[13]So τ must have the same type as α.

[14]Again, recall "z^t" ranges over abstract objects of type t.

[15]I believe that we can do this consistently. However, the

theory could be weakened so that A-objects of type t encode only (possibly) existing objects of type t/p. One might think that it is an encouraging sign that the semantics is all set for abstract t/p-properties to have encoding extensions that are non-empty.

[16]See footnote 8. We may need to add axioms for identity$_E$ among propositions and relations.

[17]We sharply distinguish between those English terms which simply lack denotation from those which denote non-existent or abstract objects (see Parsons [1979c]). If there are English proper names which simply lack a denotation, then we need to revise our specification that F map all the primitive names to a denotation. We are already covered by LA4, which handles terms which might lack a denotation.

CHAPTER VI

APPLICATIONS OF THE TYPED THEORY

§1. Modelling Frege's Senses (I)

> Frege's explanation, by way of ambiguity, of what appears to be the logically deviant behavior of terms in intermediate contexts is so theoretically satisfying that if we have not yet discovered or satisfactorily grasped the peculiar intermediate objects in question, then we should simply continue looking.
>
> David Kaplan[1]

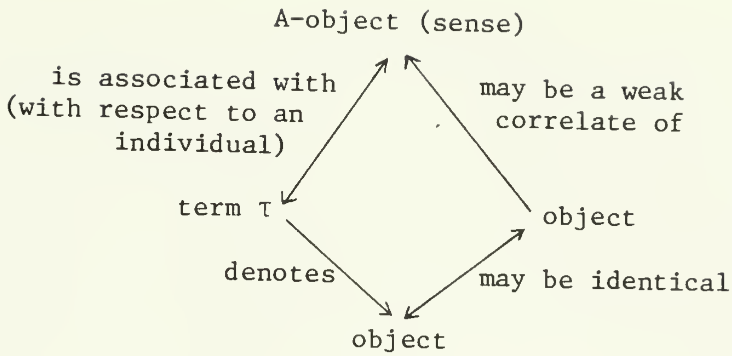
In this section, we translate and discuss the propositional attitude data which involve English names and definite descriptions that denote individuals. The data sentences are labelled (A)-(U), and because we are supposing with Frege that certain English terms occurring in them are ambiguous, there are several readings possible for each one. These readings are provided immediately after the particular datum is presented, and a discussion usually follows. In these discussions (in this section only), we revert to using the word "object" to refer to individuals and the word "property" to refer to properties of individuals (i.e., i/p-properties).

Also in these discussions, we shall modify somewhat the standard Fregean metalinguistic and metaphysical terminology. On the strict Fregean view, a term <u>expresses</u> its sense and <u>denotes</u> its denotation. And it is also said that the sense of a term <u>belongs to</u> the denotation of the term. Pictorially, these relationships are sometimes represented as follows:

$$
\begin{tikzcd}
 & \text{sense} \\
\text{term } \tau \arrow[ur, "\text{expresses}"] \arrow[dr, "\text{denotes}"'] & \\
 & \text{denotation} \arrow[uu, "\text{belongs to}"']
\end{tikzcd}
$$

Now we shall talk as if terms do denote their denotations (this is made precise by our definition of *denotation*$_{I,\delta}$, V, §2), but we shall not suppose that terms "express" their senses. Instead, we shall talk about the A-object which is *associated with* the term *with respect to* an individual. Sometimes, we shall say that the A-object *serves as* the sense of the term *with respect to* a given individual. We assume with Frege that the sense of a term (especially proper names) varies from person to person. The special sense terms (and their interpretation) that we added to our language in Chapter V help us to represent this phenomenon and help to make the above terminology precise.

For reasons which will soon become apparent, we shall not talk in terms of the metaphysical "belonging to" relationship between senses and denotations. Instead, we shall talk about the weak correlates which the A-objects that serve as senses may have. Should the A-object have one or more weak correlates, we do not suppose that any of these objects necessarily serves as the denotation of the term in question. Diagrammatically, we get:



A. S believes that Lauben is late

(.1) Bsthat-Lℓ (*de re*)

(.2) Bsthat-$\underline{\mathrm{L}\ell}_{s}$ (*de dicto*)

Suppose I feel ill one morning and resolve to stop in at the first physician's office I happen to pass on my way to work. I round a corner and see a sign on a door:[2]

DR. GUSTAV LAUBEN
General Practitioner
8:00 A.M. - 4:00 P.M.

At this point, I've now become part of a causal chain of events involving the name "Gustav Lauben."[3] I associate an A-object with this name--an A-object which serves as the sense of that name for me. We call that A-object "$\underline{\text{Lauben}}_{\text{Zalta}}$." $\underline{\text{Lauben}}_{\text{Zalta}}$ lends the name "Lauben" its cognitive significance or information value for me. It does so by encoding properties which serve to *re*-present to me the object that I suppose is denoted by the name. $\underline{\text{Lauben}}_{\text{Zalta}}$ may encode such properties as: being a (the) doctor whose office is at 15 High St., being a (the) doctor named "Gustav Lauben," being the doctor whose signpost this is, being a doctor who works from eight to four, etc.

Some other person, S, who first encounters the name "Lauben" under different circumstances will not associate with this name the A-object I've utilized. The A-object S utilizes will encode properties presented to S as being characteristic of the object named "Lauben." Thus, $\underline{\text{Lauben}}_S$ lends the name "Lauben" a cognitive significance for S which is distinct from the cognitive significance this name has for me.

On the theory we've developed in Chapter V, it is axiomatic that $\underline{\text{Lauben}}_S$ represents Lauben for S. This is one way of capturing Frege's principle that the sense of a term account for its information value. Frege also required, however, that the sense of a term determine at most one object, and that this object, should there be one, serve as the denotation of the term. It requires additional semantic complexity to capture these Fregean principles, and were we interested in a more strict modelling of Frege's ideas, we could modify our semantics.[4] However, we've chosen not to place these constraints on senses because: (a) the successful explanation of the data on which we've chosen to work doesn't seem to require that we have such constraints on senses, and (b) there are cases which suggest that Frege's principles are too strong.

Suppose that the week before I went to Lauben's office, the Medical Review Board stripped him of his license to practice, his medical school invalidated his degree, and he subsequently sold his office, never to return. It's just that no one bothered to take the sign down. In this situation, all of the properties which we have suggested might be encoded by $\underline{\text{Lauben}}_{\text{Zalta}}$ are not exemplified by Lauben. Lauben is not the weak correlate of $\underline{\text{Lauben}}_{\text{Zalta}}$; in fact, no object is.[5] Nevertheless, it seems reasonable to suppose that if I haven't learned about Lauben's recent calamity, $\underline{\text{Lauben}}_{\text{Zalta}}$, as described, still serves as the sense of "Lauben" and lends it cognitive significance.

To see this more clearly, suppose I knock at the office door and no one answers. I notice that it's just after 8:00 A.M. I believe that Lauben is late (our datum sentence). My belief is not *de re*, since I believe this without believing that the friend of Leo Peters is late (suppose Lauben is Peters' unique friend). So my belief is *de dicto*.

Now even though Lauben is not the weak correlate of $\underline{\text{Lauben}}_{\text{Zalta}}$, the latter A-object could still be instrumental in helping me to construct a proposition which serves as the "object" of my *de dicto* belief. We may suppose that the propositional object of my *de dicto* belief is $PLUG_1(\mathit{being\ late}, \underline{\mathit{Lauben}}_{\mathit{Zalta}})$. Had my belief been *de re*, the propositional object of my belief would have been $PLUG_1(\mathit{being\ late}, \mathit{Lauben})$. My *de dicto* belief will be true just in case this latter proposition is true.[6] Since Lauben no longer comes to his office, my *de dicto* belief is a false one.

Consequently we seem to be able to describe the important facts of this situation without having to suppose that $\underline{\text{Lauben}}_{\text{Zalta}}$ has a (unique) weak correlate, and without having to suppose that one of its weak correlates has to serve as the denotation of "Lauben." It is for this reason that we have chosen not to further complicate our semantics in order to present a more strict modelling of Frege's ideas.[7]

We have, however, validated another one of Frege's principles in the process--the A-object which serves as the sense of the term also serves as the denotation of the term inside *de dicto* belief contexts. To make this clearer, let's look at our two readings A.1 and A.2 in more detail. Our *de re* reading is $\text{Bsthat-L}\ell$ (A.1). From A.1, we can prove (1)-(4):

(1) $(\exists x)\text{Bsthat-L}x$

(2) $[\lambda x\ \text{Bxthat-L}\ell]s$

(3) $[\lambda x\ \text{Bsthat-L}x]\ell$

(4) $[\lambda xy\ \text{Bxthat-L}y]s\ell$

On the assumption that Lauben exists, A.1 also implies (5):

(5) $(\exists y)(E!y\ \&\ \text{Bsthat-L}y)$

We've symbolized the *de dicto* reading of A as $\text{Bsthat-L}\ell_{\underline{s}}$ (A.2). Thus, it is the sense of the name "Lauben" with respect to S which serves as the denotation of the name inside the *de dicto* belief context. From A.2, we may prove (6)-(10):

(6) $(\exists x)\text{Bsthat-L}x$

(7) $(\exists x)(A!x\ \&\ \text{Bsthat-L}x)$ (LA 12)

(8) $[\lambda x\ \text{Bxthat-L}\ell_{\underline{s}}]s$

(9) $[\lambda x\ \text{Bsthat-L}x]\ell_{\underline{s}}$

(10) $[\lambda xy\ \text{Bxthat-L}y]s\ell_{\underline{s}}$

Quantification into the belief context works <u>normally</u>, as does λ-conversion.

An examination of another case should help--here is one inspired by Quine's work.[8]

B. Ralph believes that Cicero was a Roman

(.1) Brthat-Rc (*de re*)

(.2) Brthat-R$\underline{c}_r$ (*de dicto*)

C. Ralph doesn't believe that Tully was a Roman

(.1) ~Brthat-Rt (*de re*)

(.2) ~Brthat-R$\underline{t}_r$ (*de dicto*)

D. Cicero is Tully

(.1) c=t

(.2) $c=_E t$

The triad B-C-D (and other triads like it) constitutes a paradigm case where the English proper name exhibits logically deviant behavior. From B.1 and D.1 (or D.2), it follows that Brthat-Rt. Identity elimination works normally. But from B.2 and D.1, nothing follows. And there is no reason to think that B.2, C.2, and D are jointly inconsistent. From B.2 and C.1, we can conclude both that $\underline{c}_r \neq t$ and that $R\underline{c}_r \neq Rt$. From B.2 and C.2, we can conclude both that $\underline{c}_r \neq \underline{t}_r$ and that $R\underline{c}_r \neq R\underline{t}_r$. Thus, we follow Frege in thinking that it is the ambiguity of the English proper name inside *de dicto* contexts which accounts for its logically deviant behavior.

We now precisely define the conditions under which someone has a true belief. Let us define the <u>erasure of</u> a formula ϕ ("ϕ^*") as the formula which results by deleting all the underlines and subscripts from terms occurring in ϕ. So where $\phi = R\underline{t}_r$, $\phi^* = Rt$. We now define:

D_4 S truly believes that ϕ ("TBsthat-ϕ") $=_{dfn}$ Bsthat-ϕ & ϕ^*

So from B.1 or B.2 and the supposition that Cicero was a Roman, it follows that Ralph has a true belief. From A.1 or A.2 and the supposition that Lauben isn't late, it follows that S doesn't have a true belief.

Note that in the case of *de dicto* readings, S can truly believe that ϕ even when ϕ is false. In B.2, the propositional object of Ralph's belief is a false proposition--by our AUXILIARY HYPOTHESIS, no A-object exemplifies the (nuclear) property of being a Roman. This false proposition, however, is just a neutral object which helps Ralph to represent $PLUG_1$(*being a Roman, Cicero*). Consequently, we must abandon a certain principle some philosophers hold about true belief. The principle that S truly believes that ϕ iff S believes that ϕ and ϕ is true must be given up, not just because it is inconsistent with our treatment of belief, but also because doing so allows us to construe the logic of propositional attitude contexts as another application of the philosophical logic of encoding properties. The usefulness of abandoning the old principle is a good reason for doing so.

E. S believes that Lauben was mugged

(.1) Bsthat-$M\ell$ (*re*)

(.2) Bsthat-$M\underline{\ell}_s$ (*dicto*)

F. S' believes that Lauben was mugged

(.1) Bs'that-$M\ell$ (*re*)

(.2) Bs'that-$M\underline{\ell}_{s'}$ (*dicto*)

Suppose I go to a party in the evening of the day I knocked on Lauben's door. Suppose also that Lauben is in good standing in the medical community, but that he just didn't go to work that day. Leo Peters (Lauben is his unique friend, and I'm unaware of their relationship) is there and I overhear him say "Dr. Lauben was mugged last night." The proposition I grasp when I hear this utterance is $PLUG_1$(*being mugged*, $\underline{\textit{Lauben}}_{\textit{Zalta}}$). My belief is *de dicto*, because I believe that Lauben was mugged without believing that the friend of Leo Peters was mugged. "$\underline{\textit{Lauben}}_{\textit{Zalta}}$" may or may not be the (semantic) name of the A-object I associated with "Lauben" this morning. As I went through the day, I might have been involved in another context in which the name was used. The new information I gather might get "encoded" by associating some distinct A-object which encodes all the old and new properties I now use to *re*-present Lauben to me via the name "Lauben." For now, however, let's suppose that this name retains its earlier cognitive significance.

Now is there any reason to believe that the proposition I grasped when Peters uttered his sentence was the same proposition that Peters was entertaining? Suppose his belief were *de dicto*. It seems like there would be "more intimate" properties encoded by $\underline{\text{Lauben}}_{\text{Peters}}$ than are encoded by $\underline{\text{Lauben}}_{\text{Zalta}}$. We've supposed that Lauben is Peters' friend, and there might be a very complex A-object which Peters associates with "Lauben." Although the fact that $\underline{\text{Lauben}}_{\text{Peters}} \neq \underline{\text{Lauben}}_{\text{Zalta}}$ is not a guarantee, it seems likely that the propositional object of Peters' *de dicto* belief may differ from the object of my *de dicto* belief.

Despite the fact these propositions may differ, there may still be good reason for thinking communication has taken place. A full discussion of how the communicative process operated in this situation would take us too far afield. We would have to discuss the intentions of the speaker to refer to Lauben, determine whether the speaker succeeded in referring to Lauben, and these might involve a discussion of the presuppositions of the context of the utterance. Even if we had a reasonable understanding of these features of the communicative process, it now seems in order to consider two further features. And they are, the degree to which $\underline{\text{Lauben}}_{\text{Peters}}$ and $\underline{\text{Lauben}}_{\text{Zalta}}$ are "similar" A-objects and the kind of correspondence there is between the properties these A-objects encode and the properties Lauben exemplifies.

In the ideal case, $\underline{\text{Lauben}}_{S}$ and $\underline{\text{Lauben}}_{S'}$ will be identical (or one will be embedded in the other) and Lauben will be the unique weak correlate of both of them. At the other extreme, $\underline{\text{Lauben}}_{S}$ and $\underline{\text{Lauben}}_{S'}$ will have no properties in common and Lauben will be the unique weak correlate of neither of them. Communication takes place to a greater or lesser degree depending on whether the former or the latter of these two extremes is more closely approximated. So despite the fact that "Lauben was mugged" might be used by S to express one proposition and used by S' to construct another proposition, communication between S and S' takes place to a greater degree if both $\underline{\text{Lauben}}_{S}$ and $\underline{\text{Lauben}}_{S'}$ encode, for the most part, properties which Lauben exemplifies. In the cases where $\underline{\text{Lauben}}_{S}$ and $\underline{\text{Lauben}}_{S'}$ have little in common (with Lauben), communication is rather crude and not straightforward. Yet even in these latter cases, it is important to note that the language is holding everything together (as we might expect for *de dicto* beliefs). $\underline{\text{Lauben}}_{S}$ and $\underline{\text{Lauben}}_{S'}$ would have in common the fact that they are both associated with the name "Lauben." If Lauben was mugged, S and S' have true beliefs.

G. S doesn't believe that the friend of Leo Peters was mugged

(.1) ~Bsthat-M(ιx)Fxp (*re*)

(.2) ~Bsthat-M<u>(ιx)Fxp</u> (*dicto*)

Recall that we established that my belief that Lauben was mugged was *de dicto* by the fact that I didn't also believe that the friend of Leo Peters was mugged. But the English definite description exhibits logically deviant behavior inside belief contexts as well. On the *de re* reading of G, the proposition that I fail to believe is $PLUG_1$(*being mugged, the friend of Leo Peters*), i.e., $PLUG_1$(*being mugged, Lauben*). On the *de dicto* reading of G, the proposition I fail to believe is $PLUG_1$(*being mugged,* <u>*the friend of Leo Peters*</u>). <u>The friend of Leo Peters</u> is the abstract object which encodes just the property of being the friend of Leo Peters (by LA 13). That is,

$$\underline{(\iota x)Fxp} = (\iota z)(G)(zG \equiv G=[\lambda x\ Fxp\ \&\ (y)(Fyp \rightarrow y=_E x)])$$

<u>The friend of Leo Peters</u> serves as the sense of "the friend of Leo Peters." It lends the English description cognitive significance and information value. It also has at most one weak correlate, and in this case, its unique weak correlate happens to be the denotation of the description. Finally, <u>the friend of Leo Peters</u> serves as the denotation of the description when the description is inside a *de dicto* context.

H. Ralph believes that the man in the brown hat is a spy

(.1) Brthat-S$(\iota x)\phi_1$ (*re*)

(.2) Brthat-S$\underline{(\iota x)\phi_1}$ (*dicto*)

I. Ralph doesn't believe that the mayor of the town is a spy

(.1) ~Brthat-S$(\iota x)\phi_2$ (*re*)

(.2) ~Brthat-S$\underline{(\iota x)\phi_2}$ (*dicto*)

J. Ralph believes that the mayor of the town isn't a spy

(.1) Brthat-$\sim S(\iota x)\phi_2$ (*re*)

(.2) Brthat-$\sim S\underline{(\iota x)\phi_2}$ (*dicto*)

K. Ortcutt is both the man in the brown hat and the mayor of the town

(.1) $o = (\iota x)\phi_1$ & $o = (\iota x)\phi_2$

(.2) $o =_E (\iota x)\phi_1$ & $o =_E (\iota x)\phi_2$

If Ralph's belief in H is *de re*, the object of his belief is $PLUG_1$(*being a spy, the man in the brown hat*), i.e., $PLUG_1$(*being a spy, Ortcutt*). Given H.1 and K, we may conclude (11):

(11) Brthat-So

Given I.1 and K, we may conclude (12):

(12) ~Brthat-So

So H.1 and I.1 are inconsistent. Ortcutt himself is the constituent of the propositional object of the *de re* belief--the descriptions inside the relevant belief ascriptions "contribute" their denotation to the proposition. From H.1 and DESCRIPTIONS, we also get (13):

(13) $(\exists ! y)(\phi 1^{y}_{x}$ & Brthat-Sy)

However, let's suppose Ralph's belief is *de dicto*. The object of his belief is $PLUG_1$(*being a spy, <u>the man in the brown hat</u>*). From H.2 and K, nothing follows. From H.2 and I.2, it follows that <u>the man in the brown hat</u> $\neq$ <u>the mayor of the town</u>. If the mayor of the town is a spy, then it follows from H.2 that Ralph has a true belief. (14) also follows from H.2, given LA 13 and DESCRIPTIONS:

(14) $(\exists ! x)(A!x$ & $(F)(xF \equiv F=[\lambda x\ \phi_1$ & $(y)(\phi 1^{y}_{x} \rightarrow y=_E x)])$ & Brthat-Sx)

Besides these, we have the usual consequences of H.2 based on existential introduction and λ-conversion:

(15) $(\exists x)\text{Brthat-S}x$

(16) $[\lambda x\ \text{Bxthat-S}\underline{(\iota x)\phi_1}]r$

(17) $[\lambda x\ \text{Brthat-S}x]\underline{(\iota x)\phi_1}$

(18) $[\lambda xy\ \text{Bxthat-S}y]r\underline{(\iota x)\phi_1}$

Note that H.1 and J.1 ascribe contradictory beliefs to Ralph. Given K, H.1 implies (11) and J.1 implies (19):

(19) $\text{Brthat-}{\sim}\text{So}$

From (11) and (19) we get (20):

(20) $(\exists F^0)(\text{Brthat-}F^0\ \&\ \text{Brthat-}{\sim}F^0)$

However, (20) doesn't imply that Ralph believes a contradiction.

If H.2 and J.2 correctly describe Ralph's state of mind, then we cannot prove that Ralph has inconsistent beliefs. From H.2 and J.2, we cannot deduce that $\underline{(\iota x)\phi_1} \neq \underline{(\iota x)\phi_2}$, but we can prove this from the plausible assumption that the property of being the man in the brown hat is distinct from the property of being the mayor of the town. Since $\underline{(\iota x)\phi_1} \neq \underline{(\iota x)\phi_2}$, no substitutions into H.2 and J.2 would lead us to think that Ralph has inconsistent beliefs. However, from J.1 or J.2 and the fact that Ortcutt is a spy, we can prove that Ralph has a false belief, where,

S falsely believes that ϕ ("FBsthat-ϕ") $=_{dfn}$

Bsthat-ϕ & ~TBsthat-ϕ

L. Ralph believes that the shortest spy is a spy

(.1) $\text{Brthat-S}(\iota x)\phi_3$ (*re*)

(.2) $\text{Brthat-S}\underline{(\iota x)\phi_3}$ (*dicto*)

If L.1 expresses what Ralph believes, then his belief would be of interest to the FBI. (21) follows from L.1:

(21) $[\lambda x\ Brthat\text{-}Sx](\iota x)\phi_3$

If Bond is the shortest spy, then (22) follows from L.1, and (23) follows from (21) or (22):

(22) Brthat-Sb

(23) $[\lambda x\ Brthat\text{-}Sx]b$

If we assume that Bond exists and that an existence claim is built into ϕ_3, then we can generalize on (21)-(23) to get:

(24) $(\exists x)(E!x\ \&\ Brthat\text{-}Sx)$

(25) $(\exists y)(E!y\ \&\ [\lambda x\ Brthat\text{-}Sx]y)$

All of this results because the propositional object of Ralph's belief has an existing object, namely Bond, as a constituent.

None of these results follow if L.2 expresses what Ralph believes. There is no way to use "exportation" on L.2 to produce (24) or (25). We can only reap the "standard" inferences from L.2 based on existential and lambda introduction:

(26) $(\exists x)Brthat\text{-}Sx$

(27) $[\lambda x\ Brthat\text{-}Sx]\underline{(\iota x)\phi_3}$

(28) $(\exists y)(A!y\ \&\ [\lambda x\ Brthat\text{-}Sx]y)$

I take it that the FBI wouldn't be interested by the fact that Ralph, like most everyone, uses an abstract object to represent whoever it is that is the shortest spy (in the absence of a *de re* belief).

M. Ralph believes someone is a spy

(.1) $(\exists x)Brthat\text{-}Sx$

(.2) $Brthat\text{-}(\exists x)Sx$

M.1 and M.2 disambiguate M. M.1 is similar to (24) and we might prefer to use the latter to read M properly. M.2 relates Ralph to the proposition $PROJ_1$(*spyhood*). No legitimate exportation on M.2 will get us to M.1.

N. S believes that Newton met Leibniz

(.1) Bsthat-Mnℓ (*re*)

(.2) Bsthat-M$\underline{n}_s\ell$ (*dicto/re*)

(.3) Bsthat-Mn$\underline{\ell}_s$ (*re/dicto*)

(.4) Bsthat-M$\underline{n}_s\underline{\ell}_s$ (*dicto*)

In order to determine which of the readings of N is the correct one, we have to examine data triads to discover how the names "Newton" and "Leibniz" are functioning.

O. Frege believes that Hesperus is Hesperus

P. Frege doesn't believe that Phosphorus is Hesperus

Q. Phosphorus is Hesperus

There are various ways to represent the triad O-P-Q. The preferred representation is as follows:

O.' Bfthat-$\underline{h}_f =_E h$

P.' ~Bfthat-$\underline{p}_f =_E h$

Q.' $p =_E h$

Suppose Frege as a young man is being taught the names of the stars. Early one evening, his teacher points out Venus and says "That's Hesperus--it's the first visible star of the evening." Frege becomes, at that moment, part of an historical, causal chain of events connecting him with the name "Hesperus." So we suppose that Frege associates an abstract object sense with the name. $\underline{\text{Hesperus}}_{\text{Frege}}$ may encode: being the star to which my teacher is pointing, being the first visible star of the evening, being named "Hesperus," being situated in position p in the western sky at 5:30 P.M. Thursday, December 7, 1860, etc. "Hesperus" would have different cognitive value for someone who learned the name in different circumstances.

Now suppose Frege's teacher points out Venus to Frege early the next morning and says "That's Phosphorus--it's the last star visible in the morning." The young Frege will associate some new, distinct A-object with "Phosphorus." That's because the features of the learning situation are radically different. The object pointed out to him is in a position of the sky that appears unrelated to the position of the object pointed out the evening before. The names introduced are distinct. There is no reason for Frege to believe that the object pointed out to him now is identical with the object pointed out to him the evening before.

So if Frege's teacher doesn't tell him that Phosphorus is Hesperus, Frege could believe that Hesperus is Hesperus without believing that Phosphorus is Hesperus. Although there are various ways to represent this data as a consistent triad, we have chosen the reading on which Frege believes $PLUG_1(PLUG_2(\mathit{Identity}_E, \mathit{Hesperus}), \underline{\mathit{Hesperus}}_{\mathit{Frege}})$ and fails to believe $PLUG_1(PLUG_2(\mathit{Identity}_E, \mathit{Hesperus}), \underline{\mathit{Phosphorus}}_{\mathit{Frege}})$.

R. John hopes that the strongest man in the world, whoever he is, beats up the man who just insulted him.

Preferred reading:

(R') Hjthat-B$\underline{(\iota x)\phi_4}(\iota x)\phi_5$ (*dicto/re*)

On the preferred reading of R, we interpret the first definite description as occupying a *de dicto* position and suppose that it contributes its sense to the proposition which is the object of John's hope.

S. Mary believes that the wife of Tully is the wife of Tully

T. Mary doesn't believe that the wife of Cicero is the wife of Tully

U. The wife of Cicero is the wife of Tully

Preferred representation:[9]

S.' $Bm\text{that-}\underline{(\iota x)Wxt}_m =_E (\iota x)Wxt$ (*dicto*)

T.' $\sim Bm\text{that-}\underline{(\iota x)Wxc}_m =_E (\iota x)Wxt$ (*dicto*)

U.' $(\iota x)Wxc = (\iota x)Wxt$

S-T-U is an interesting triad since it requires that we use the senses of the names "Tully" and "Cicero" with respect to Mary to construct the senses of the English descriptions "the wife of Tully" and "the wife of Cicero." That's because <u>the wife of Tully</u> and <u>the wife of Cicero</u> are identical, and so (S") and (T") are inconsistent:

S." $Bm\text{that-}\underline{(\iota x)Wxt} =_E (\iota x)Wxt$

T." $\sim Bm\text{that-}\underline{(\iota x)Wxc} =_E (\iota x)Wxt$

<u>The wife of Tully</u> and <u>the wife of Cicero</u> are identical because Cicero is Tully, and so being the wife of Tully just is being the wife of Cicero. Since these properties are identical, the object which encodes just being the wife of Tully is identical with the object which encodes just being the wife of Cicero. So we can't use (S") and (T") to help us understand how S-T-U is consistent because the proposition that <u>the wife of Tully</u> is identical$_E$ with the wife of Tully is identical with the proposition that <u>the wife of Cicero</u> is identical$_E$ with the wife of Tully.

So we *must* use the senses of "Tully" and "Cicero" with respect to Mary in order to suppose S-T-U is consistent. Thus, <u>the wife of Tully</u>$_m$ is a constituent of the propositional object of Mary's belief in

(S'). Though the wife of Tully$_m$ could have at most one weak correlate, it fails to have any. By the AUXILIARY HYPOTHESIS, A-objects fail to exemplify the (nuclear) property of having a wife.[10]

Our definition of true belief still works fine:

(29) $TBmthat\text{-}\underline{(\iota x)Wxt}_m =_E (\iota x)Wxt \equiv Bmthat\text{-}\underline{(\iota x)Wxt}_m =_E (\iota x)Wxt$ & $(\iota x)Wxt =_E (\iota x)Wxt$

Given (S') and given that there is a unique wife of Tully, it follows that Mary has a true belief.[11] If the negation of (T') represented Mary's state of mind, she would still have a true belief. If Mary believes that the wife of Cicero was not the wife of Tully, and this was correctly represented as $Bmthat\text{-}\underline{(\iota x)Wxc}_m \neq_E (\iota x)Wxt$, then she would have a false belief.

We may conclude this section with a few general remarks about our treatment of definite descriptions. G-L and R-U give us evidence for thinking that English descriptions have both a sense and a denotation. The sense of the definite description lends it cognitive value--a value to beings with representational capacities in that it enables them to recognize (or understand what it might be like to recognize) objects which have (never) been presented to them. If we are interested solely in describing the cognitive value of a given English description, we always have available to us a *sense*-description of our formal language. English descriptions don't come with their property terms marked as to whether the property denoted is exemplified or encoded by the object being described. And it might be that some other description of our language is better suited in having the intuitively right denotation of the English description. For example, we might prefer to use a description which contains an encoding subformula to translate "the student who killed an old moneylender," where this English description is meant to refer to Raskolnikov of Crime and Punishment (Chapter IV, §4). But insofar as we are interested purely in the phenomenon of the apparent deviant behavior of this English description inside *de dicto* contexts, it might just be that the student who killed an old moneylender serves well enough as its cognitive value. This depends on whether there is conclusive data which shows that the question of getting the denotation right and the question of explaining apparent deviant behavior are not independent.[12,13]

§2. Modelling Frege's Senses (II)

We now consider the propositional attitude data triads which involve English terms that denote higher order objects.

A. John believes that Woodie is a woodchuck

(.1) Bjthat-Ww (*de re*)

(.2) Bjthat-$\underline{W}_j$w (*de dicto*)[14]

B. John doesn't believe that Woodie is a groundhog

(.1) ~Bjthat-Gw (*de re*)

(.2) ~Bjthat-$\underline{G}_j$w (*de dicto*)

C. Being a woodchuck just is being a groundhog

(.1) G=W

(.2) $G=_{Ei/p}W$

A.1 and B.1 are inconsistent, given C.1 (C.2). A.1 asserts that John believes the proposition $PLUG_1$(*being a woodchuck, Woodie*), whereas B.1 asserts that John doesn't believe $PLUG_1$(*being a groundhog,* *Woodie*). But since C is a true identity statement which asserts that the properties of being a woodchuck and being a groundhog are identical, it's provable that these propositions are identical. So either A.1 or B.1 must be false.

A.2 and B.2 can both be true together, however. A.2 asserts that John believes $PLUG_1(\underline{\textit{being a woodchuck}}_{John}, \textit{Woodie})$. $\underline{\text{Being a woodchuck}}_{\text{John}}$ is the abstract property of individuals (i/p-property) which serves as the sense of the name "being a woodchuck" with respect to John. The $sen^{i/p}_{John}$ function of our semantics (Chapter V, §2, A) assigns "being a woodchuck" a member of $A_{i/p}$ (that is, a member of the abstract objects of type i/p). So "$\underline{\text{being a woodchuck}}_{\text{John}}$" denotes $sen^{i/p}_{John}$ ("being a woodchuck").

$\underline{\text{Being a woodchuck}}_{\text{John}}$ can be plugged up with any individual--our *PLUG* function is defined so that it operates on all i/p-properties. Consequently, $PLUG_1(\underline{\textit{being a woodchuck}}_{John}, \textit{Woodie})$ is a type p object and can serve as the propositional object of someone's belief.

B.2 asserts that John doesn't believe $PLUG_1(\underline{\textit{being a groundhog}}_{John}, \textit{Woodie})$. $\underline{\text{Being a groundhog}}_{\text{John}}$ is the abstract i/p-property which serves as the sense of "being a groundhog" with respect to John. If A.2 and B.2 are true, it follows both that $PLUG_1(\underline{\textit{being a woodchuck}}_{John}, \textit{Woodie}) \neq PLUG_1(\underline{\textit{being a groundhog}}_{John}, \textit{Woodie})$ and that $\underline{\textit{being a woodchuck}}_{John} \neq \underline{\textit{being a groundhog}}_{John}$. So as Frege predicted, the senses of the property terms flanking the identity sign in C are distinct.

This seems right--"woodchuck" and "groundhog" probably entered John's vocabulary under different circumstances. Maybe on one occasion he saw and was told he was seeing a woodchuck. He encoded properties of the property of being a woodchuck into an A-object which represented the property of being a woodchuck to him. And maybe on another occasion, someone described woodchucks to him improperly, in the process saying only that he was describing an animal called a "groundhog." John would not have known that in fact these properties are the same. Sentence C would be informative to him.

D. John believes that the chair in front of the class is Crayola crayon blue

(.1) $Bj\text{that-}CCB(\iota x^{i})\phi_{1}$ (*re*)

(.2) $Bj\text{that-}\underline{CCB}_{j}(\iota x^{i})\phi_{1}$ (*dicto*)

E. John doesn't believe that the chair in front of the class is French fire engine blue

(.1) $\sim Bj\text{that-}FFB(\iota x^{i})\phi_{1}$ (*re*)

(.2) $\sim Bj\text{that-}\underline{FFB}_{j}(\iota x^{i})\phi_{1}$ (*dicto*)

F. Crayola crayon blue just is French fire engine blue

(.1) $FFB=CCB$

(.2) $FFB=_{E}i/_{P}CCB$

D-E-F is analyzed analogously with A-B-C. We may suppose "French fire engine blue" and "Crayola crayon blue" to be names of the same shade of blue. So F is an informative identity statement about properties. As a boy, John may have become directly acquainted with this property. But the label on his Crayola crayons just read "blue," and he has never seen that shade of blue labeled "French fire engine."

G. John believes that Bill has the property of being a student

(.1) $Bj\text{that-}HasSb$ (*re*)

(.2) $Bj\text{that-}\underline{Has}_{j}Sb$ (*dicto*)

H. John doesn't believe that Bill exemplifies the property of being a student

(.1) ~Bjthat-ExSb (*re*)

(.2) ~Bjthat-$\underline{\text{Ex}}_j$Sb (*dicto*)

I. Having a property just is exemplifying a property

(.1) Ex=Has

(.2) Ex $=_E$(i,p,i)/p Has

G-H-I might describe a student beginning in philosophy, unaware of the technical sense philosophers have for the word "exemplifies." "Has" and "exemplifies" both denote relations of type (i/p,i)/p. We suppose that there are abstract objects of this type which serve as the senses of these names with respect to John. G.2 asserts that John believes $PLUG_1(PLUG_2(\underline{having}_{John}, Bill), being\ a\ student)$; whereas H.2 asserts that John doesn't believe $PLUG_1(PLUG_2(\underline{exemplifying}_{John}, Bill), being\ a\ student)$.

To handle our next triad, J-K-L, we need to add some functional notations to our language:

where ρ is a term of type $(t_1,t_2)/p$ and τ is a term of type t_1, then ρ(τ) is a term of type t_2.

Let us interpret this notation as follows:

$$d_{I,f}(\rho(\tau)) = d_{I,f}((\iota\alpha^{t_2})\rho\tau\alpha)$$

So ρ(τ) is the object which is such that τ bears ρ to it. Using this interpreted notation, we might construe adverbs as names of relations of type (i/p,i/p)/p, i.e., relations which relate two i/p-properties. For example, "slowly" might denote a relation between the property of walking and the property of walking slowly. In the language, "slowly" combines with "walk" to form "slowly(walk)," which denotes the property of walking slowly.

This gives us a means of representing J-K-L consistently.

J. John believes Bill walked bravely ....

(.1) Bjthat-B(W)b (*re*)

(.2) Bjthat-$\underline{B}_j$(W)b (*dicto*)

K. John doesn't believe that Bill walked courageously ....

(.1) ~Bjthat-C(W)b (*re*)

(.2) ~Bjthat-$\underline{C}_j$(W)b (*dicto*)

L. Walking bravely ... just is walking courageously ...

(.1) B(W) = C(W)

(.2) B(W) $=_E$ C(W)

Examples like G-L should demonstrate that our analysis for *de dicto* belief is generalizable throughout the types. This treatment of beliefs about higher order objects suggests a solution to the "paradox" of analysis. Central to this puzzle are data triads similar to the ones we've been discussing. Here's an example:

M. It's trivial that the concept brother is identical with the concept brother

N. It's not trivial that the concept male sibling is identical with the concept brother

O. The concept brother is identical with the concept male sibling

Although there are various ways to state the puzzle precisely, all we need to say is that the puzzle involves the question of how an identity statement like (O) can be (an) informative (analysis). Philosophers who believe that property terms denote sets and express properties, and who hold that "brother" and "male sibling" express the same property are left with no means of accounting for the informative nature of the identity statement formed by flanking an identity sign with the property *denoting* terms "the concept brother" and "the concept male sibling." What is to serve as the senses of these expressions?

We suppose here that property analyses are sentences which say that two properties are identical. We simply extend Frege's view of their informative character by supposing that there are distinct abstract properties which serve as the senses of the above property denoting expressions. In order to represent M-N-O correctly, we need to note that strictly speaking, triviality is person relative--what is trivial for one person may not be trivial for another. We assume that triviality is a relation between persons and propositions. Consequently, we introduce "T" to be a name of type (i,p)/p and we read "Txthat-ϕ" as: it is trivial for x that ϕ. This forges an analogy with the other propositional attitudes. Terms which follow the "it is trivial for x" prefix behave like terms in propositional attitude contexts--sometimes they denote their senses.

Clearly, if we're restricting ourselves to a discussion of a particular individual S, then (P')-(Q')-(R') would be the proper way to capture the triad P-Q-R:

P. It is trivial for S that the concept brother is identical with the concept brother

Q. It is not trivial for S that the concept male sibling is identical with the concept male sibling

R. The concept brother is identical with the concept male sibling

P.' $Tsthat\text{-}\underline{B}_s =_E B$

Q.' $\sim Tsthat\text{-}\underline{MS}_s =_E B$

R.' $MS=B$

The proposition asserted to be trivial by P is *PLUG*$_1$(*PLUG*$_2$(*identity*$_E$, *being a brother*), *being a brother*$_s$). Its triviality derives from the logical truth that being a brother$_s$ represents being a brother to S. The proposition that's not trivial according to (Q') is *PLUG*$_1$(*PLUG*$_2$(*identity*$_E$, *being a brother*), *being a male sibling*$_s$).

(P')-(Q')-(R') may be a good account of P-Q-R, but the original triad was M-N-O. How are we to represent it? Well, since the English prefix "it is trivial" as it occurs in (M) is not relativized to a particular individual, it seems that (M) asserts that it is trivial for everyone that the concept brother is identical with the concept brother. The relevant reading of (N) seems to be: everyone is such that it is not trivial for them that the concept male sibling is identical with the concept brother. If we recall that we have allowed primitive variables of type i to serve as subscripts for sense terms, then (M')-(N')-(O') seems to be the correct way to translate the data triad:

M.' $(x)Txthat\text{-}\underline{B}_x =_E B$

N.' $(x)\sim Txthat\text{-}\underline{MS}_x =_E B$

O.' $MS=B$

An even closer representation of M-N-O would be one which uses the iota-operator to capture the English definite article. Let us represent "the concept brother" as $(\iota y^{i/p})(y=B)$ and "the concept male sibling" as $(\iota y^{i/p})(y=MS)$. A consistent reading of M-N-O would be:

M." $(x)Txthat\text{-}(\iota y^{i/p})(y=\underline{B}_x) =_E (\iota y^{i/p})(y=B)$

N." $(x){\sim}Txthat\text{-}(\iota y^{i/p})(y=\underline{MS}_x) =_E (\iota y^{i/p})(y=B)$

O." $(\iota y^{i/p})(y=MS) = (\iota y^{i/p})(y=B)$

§3. Modelling Impossible and Fictional Relations

A. Impossible relations. Sometimes we think about "impossible" individuals. These are not individuals which are such that some contradiction is true. Rather, these are individuals like the_A round square, which encode incompatible properties. We can also think about "impossible" relations--the symmetrical non-symmetrical relation is one. Here is an *a priori* truth about this relation:

(1) The symmetrical non-symmetrical relation is symmetrical

We can't analyze "the symmetrical non-symmetrical relation" as a description involving exemplification formulas since it would fail to denote. There are no higher order objects which exemplify both being symmetrical and being non-symmetrical. (1) would be turned into a falsehood.

However, we may read the description as "the object which encodes symmetricality and non-symmetricality," and then generate the *a priori* truth that this object encodes symmetricality. Let us suppose that the properties in question are of type ((i,i)/p)/p.[15] We then have the following instance of A-OBJECTS:

$$(\exists x^{(i,i)/p})(A!^{((i,i)/p)/p}x \;\&\; (F^{((i,i)/p)/p})(xF \equiv F{=}S \vee F{=}\bar{S}))$$

This axiom, plus the definition of identity among (i,i)/p-objects, justifies our talking about *the* abstract relation which encodes being symmetrical and being non-symmetrical. As in Chapter II, we define:

the_A symmetrical non-symmetrical relation $=_{dfn}$

$$(\iota x^{(i,i)/p})(A!x \;\&\; (F)(xF \equiv F{=}S \vee F{=}\bar{S})),$$

dropping the obvious typescripts. It is then provable that the_A symmetrical non-symmetrical relation encodes symmetricality. This theorem represents (1).

The analysis proposed for data about the_A round square seems, therefore, to generalize within type theory to the data about the_A symmetrical non-symmetrical relation.

<u>B. Fictional relations</u>. When we see a sentence like "Einstein discovered that there is no such thing as simultaneity," how are we to understand it? Did Einstein discover that no two events ever exemplify the relation of simultaneity? Or did he discover that the simultaneity relation doesn't exist?[16] I'm not sure how to decide the issue, but the latter reading seems a legitimate option. We might therefore suppose that simultaneity is a fictional relation, and for our present purposes, we could suppose that it is native to the Newtonian (science) fiction. So let us identify it in a way analogous to our earlier work.

Let us suppose that events are special kinds of propositions, and that they are type p objects. A relation among two events would be of type (p,p)/p. Newtonian mechanics presupposes that simultaneity is a (possibly) existing relation of type (p,p)/p. However, in our view, it must be an abstract relation.[17] Let "s" be a name of type (p,p)/p which denotes this relation. We analyze (2) as (2)':

(2) Einstein discovered that simultaneity doesn't exist

(2)' $\text{Dethat-}{\sim}E!^{((p,p)/p)/p}s$

But which abstract relation does "s" denote? Well, we note that by typing the definitions of character ("Char(x^t,s)") and native ("Native(x^t,s)"), we may construct a typed N-CHARACTERS axiom such as:[18]

$$(x^t)(s)(\text{Native}(x,s) \rightarrow x=(\iota z^t)(F^{t/p})(zF \equiv \Sigma_s Fx))$$

Let "n" denote the story of Newtonian mechanics. We then get the following instance of N-CHARACTERS:

$$\text{Native}(s,n) \rightarrow s=(\iota z)(zF \equiv \Sigma_n Fs),$$

where "s" denotes simultaneity and is not a restricted variable. This may prove to be an interesting way to identify other fictional relations of disproven scientific theories.

Could someone write a story about non-scientific, fictional relations? Could we dream about non-existent relations? If these are genuine possibilities, we will have further data for the application of our type theory.

CHAPTER VI ENDNOTES

[1]D. Kaplan, [1968].

[2]I've borrowed this name, and a few others, from Frege's late essay [1918]. The discussion here is not meant to be understood in the context of that work, however.

[3]See K. Donnellan [1974], pp. 3-32, and [1972], p. 377. Also see S. Kripke [1972], p. 302.

[4]Placing such constraints adds the following complexities. As in footnote 9, Chapter V, §2, we would first define the set of senses of type t, S_t, as the abstract objects of type t which have at most one weak correlate. Then we would have to define, for each object o^t in $\mathcal{D}_t$, the set of senses of which o is the unique weak correlate ("$S_t(o)$"). Then we would require that the $sen^t_{o'}$ function assign to a given name κ an object drawn from $S_t(F(\kappa))$, i.e., an object drawn from the set of senses of which $F(\kappa)$ is the unique weak correlate. The assigned object would serve as the sense of κ with respect to o'. So where $F(\sigma) = o'$, $F(\underline{\kappa_\sigma}) = sen^t_{o'}(\kappa)$. The object determines $F(\kappa)$.

Although this succeeds in modelling Frege's ideas, I doubt that language works this way.

[5]We could even imagine a situation in which some other object besides Lauben was the weak correlate of $\underline{\text{Lauben}}_{\text{Zalta}}$.

[6]It follows from the fact that I believe the former and not the latter that these propositions are distinct. This *doesn't* follow from

the fact that Lauben $\neq$ $\underline{\text{Lauben}}_{\text{Zalta}}$. Recall the footnote in Chapter III, §4 where we showed that some propositions with distinct constituents might be identical.

[7]Cases in Kripke [1972] and Donnellan [1972], [1974] would be relevant here. However, it is slightly tricky to transpose their arguments designed to refute the Russellian view that names are disguised descriptions into arguments designed to refute the Fregean view that senses determine a unique object as the referent of the term.

[8]Quine [1956].

[9]Note that the English "is the wife of Tully" could be represented either as "$[\lambda x\ Wxt\ \&\ (y)(Wyt \rightarrow y{=}_E x)]$" or as "$=(\iota x)Wxt$."

[10]That is, any weak correlate of $\underline{\text{the wife of } \underline{\text{Tully}}_m}$ would be the wife of an abstract object. By the AUXILIARY HYPOTHESIS, this never happens.

[11]Mary's belief here is trivial because $\underline{\text{the wife of } \underline{\text{Tully}}_m}$ represents the wife of Tully to Mary. However, you might think that the propositional object of Mary's belief when she believes that the wife of Tully is the wife of Tully is an *a priori* truth. But consider the *a priori* truth that <u>the wife of Tully</u> encodes the property of being the wife of Tully. In our formal language, we express this as:

$$\underline{(\iota x)Wxt}\ [\lambda x\ Wxt\ \&\ (y)(Wyt \rightarrow y{=}_E x)]$$

Being an encoding formula, this sentence doesn't denote a proposition. We could, however, develop a new logical function $ENPLUG$ ("encoding plug"), which maps a property r and an object o to the proposition, $ENPLUG(r,o)$, which is such that $ext_w(ENPLUG(r,o))$ = T iff $o \in ext_A(r)$. This would give us propositions for atomic encoding formulas to denote propositions which, if true, would be *a priori*. Then we could represent the object of Mary's belief (above) as an *a priori* proposition.

We could also represent "Meinong believed that the round square is round" as a relation between Meinong and the proposition *ENPLUG(being round, the*$_A$ *round square)*. And we could represent "S believes that Holmes is clever" as a relation between S and the proposition *ENPLUG(being clever, Holmes)*. There are other possibilities here.

[12]There could be a problem here. Is the following triad conclusive data showing that these questions aren't independent?

(i) S believes that Dostoyevsky wrote about the student who killed an old moneylender (according to Crime and Punishment)

(ii) S doesn't believe Dostoyevsky wrote about the student who was arrested by Porphyry (in Crime and Punishment)

(iii) The student who killed an old moneylender (in Crime and Punishment) is the student arrested by Porphyry (in Crime and Punishment)

The following representation gets the denotation of the descriptions correct, but doesn't account for the (apparent?) consistency of the triad:

(i)' $\text{Bsthat-Wd}(\iota x)\Sigma_{CP}(Sx \;\&\; (\exists y)(OMLy \;\&\; Kxy))$

(ii)' $\sim\text{Bsthat-Wd}(\iota x)\Sigma_{CP}(Sx \;\&\; Apx)$

(iii)' $(\iota x)\Sigma_{CP}(Sx \;\&\; (\exists y)(OMLy \;\&\; Kxy)) = (\iota x)\Sigma_{CP}(Sx \;\&\; Apx)$

But we can't underline these descriptions to produce sense-descriptions. For an attempt to solve this problem, see Appendix E.

Also, Parsons has suggested that there is another kind of sentence which causes trouble: "Some biblical prophets are real, some are unreal, and some I'm unsure about."

To handle this sentence, let "B" denote the Bible, let "K" abbreviate the verb "knows," and let "P" denote being a prophet. Then consider:

$(\exists x)(\mathrm{Char}(x,B)\ \&\ \Sigma_B Px\ \&\ E!x)\ \&\ (\exists x)(\mathrm{Char}(x,B)\ \&\ \Sigma_B Px\ \&\ A!x)\ \&$

$(\exists x)(\mathrm{Char}(x,B)\ \&\ \Sigma_B Px\ \&\ {\sim}\mathrm{Ksthat}\text{-}E!x\ \&\ {\sim}\mathrm{Ksthat}\text{-}A!x)$

[13]In Appendix F, the reader will find a presentation and discussion of an objection developed by E. Gettier and T. Parsons. The objection is an attempt to produce two pieces of data which seem to have incompatible translations.

[14]Of course, there are other *de dicto* readings:

$B_j\text{that-}W\underline{w}_j$

$B_j\text{that-}\underline{W}_j\underline{w}_j$

The same goes for sentence (B) and many of the other sentences which follow. We are now presenting the preferred readings.

[15]There is typical ambiguity here. It should be unobjectionable. See Parsons [1979a].

[16]Some philosophers may prefer to say that Einstein discovered that there were an infinite number of simultaneity relations--one for each frame of reference. So we have to restate our datum sentence as "Einstein discovered that there is no such thing as absolute simultaneity."

[17]Recall that to say it is abstract is *not* to say that necessarily it fails to have a correlate.

[18]In this axiom, "s" is a restricted variable ranging over stories.

B I B L I O G R A P H Y

Bealer, George

[1981] Quality and Concept, London: Oxford University Press (forthcoming), 1981.

Castañeda, Hector-Neri

[1974] "Thinking and the Structure of the World," *Philosophia* 4 (1974), pp. 3-40.

Church, Alonzo

[1940] "A Formulation of the Simple Theory of Types," *Journal of Symbolic Logic* 5 (1940), pp. 56-68.

[1941] Calculi of λ-Conversion, Princeton: Princeton University Press, 1941.

Donnellan, Keith

[1972] "Proper Names and Identifying Descriptions," *Synthése* 40 (1972), pp. 356-380.

[1974] "Speaking of Nothing," *The Philosophical Review* 83 (1974), pp. 3-32.

Findlay, J. N.

[1933] Meinong's Theory of Objects and Values, London: Oxford University Press, 1963.

Frege, Gottlob

[1892] "On Sense and Reference," translated in Geach and Black (eds.), Translations of Frege's Philosophical Writings, Oxford: Blackwell, 1970, pp. 56-78.

[1918] "The Thought: A Logical Inquiry," translated by A. M. and M. Quinton, *Mind* 65 (1956), pp. 289-311.

Kaplan, David

[1968] "Quantifying In," *Synthése* 19 (1968), pp. 178-214.

Kripke, Saul

[1963] "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica* 16 (1963), pp. 83-94.

[1972] "Naming and Necessity," *Synthése* 40 (1972), pp. 253-354.

Leibniz, Gottfried Wilhelm

[1686a] The Leibniz-Arnaud Correspondence, H. T. Mason (ed.), Manchester: Manchester University Press, 1967.

[1686b] "The Discourse on Metaphysics," reprinted in Leibniz Philosophical Papers and Letters, L. Loemaker (ed.), Dordrecht: D. Reidel, 1969, pp. 303-330.

[1714] The Monadology, reprinted in Leibniz Philosophical Writings, London: J. M. Dent, 1973, pp. 179-195.

Lemmon, E. J.

[1965] Beginning Logic, London: Thomas Nelson, 1965.

Lewis, David

[1968] "Counterpart Theory and Quantified Modal Logic," *Journal of Philosophy* 65 (1968), pp. 113-126.

[1978] "Truth in Fiction," *American Philosophical Quarterly* 15 (1978), pp. 37-46.

Mally, Ernst

[1912] Gegenstandstheoretische Grundlagen der Logik und Logistik, Leipzig: Barth, 1912.

Mates, Benson

[1968] "Leibniz on Possible Worlds," Logic, Methodology and Philosophy of Science, Rootselaar and Stahl (eds.), Amsterdam: North Holland, 1968.

Meinong, Alexius

[1904] "Über Gegenstandtheorie," translated by Levi, Terrell, and Chisholm, Realism and the Background of Phenomenology, Glencoe: The Free Press, 1960, pp. 76-117.

[1915] Über Möglichkeit und Wahrscheinlichkeit, Leipzig: Barth, 1915.

Mendelson, Eliot

[1964] An Introduction to Mathematical Logic, New York: D. Van Nostrand, 1964.

Parsons, Terence

[1974] "Prolegomenon to Meinongian Semantics," *Journal of Philosophy* 71 (1974), pp. 561-580.

[1975] "A Meinongian Analysis of Fictional Objects," *Grazer Philosophische Studien* 1 (1975), pp. 73-86.

[1978] "Nuclear and Extranuclear Properties, Meinong, and Leibniz," *Nous* 12 (1968), pp. 127-151.

[1979a] "Type Theory and Ordinary Language," Linguistics, Philosophy, and Montague Grammar, S. Davis and M. Mithun (eds.), Austin: University of Texas Press, 1979.

[1979b] "The Methodology of Nonexistence," *Journal of Philosophy* 76 (1979), pp. 649-661.

[1979c] "Referring to Nonexistent Objects," *Theory and Decision* 11 (1979), pp. 95-110.

[1980] Nonexistent Objects, New Haven: Yale University Press, 1980.

Prior, Arthur N.

[1956] "Modality and Quantification in S_5," *Journal of Symbolic Logic* 21 (1956), pp. 60-62.

Quine, W. V. O.

[1956] "Quantifiers and Propositional Attitudes," The Ways of Paradox, New York: Random House, 1966, pp. 183-194.

[1960] "Variables Explained Away," Selected Logical Papers, New York: Random House, 1966, pp. 227-235.

Ramsey, F. P.

[1927] "Facts and Propositions," The Foundations of Mathematics, New York: Humanities Press, 1950, pp. 138-155.

Rapaport, William

[1976] Intentionality and the Structure of Existence, Ann Arbor: University Microfilms, 1976 (#77-1930).

[1978] "Meinongian Theories and a Russellian Paradox," *Nous* 12 (1978), pp. 153-180.

Rogers, Robert

[1971] Mathematical Logic and Formalized Theories, Amsterdam: North Holland, 1971.

Routley, Richard

[1979] Exploring Meinong's Jungle and Beyond, Departmental Monograph #3, Philosophy Department, Research School of Social Sciences, Australian National University, Canberra.

Russell, Bertrand

[1910] Principia Mathematica (with A. N. Whitehead), Cambridge: Cambridge University Press, 1910.

[1918] "The Philosophy of Logical Atomism," reprinted in Logic and Knowledge, Robert Marsh (ed.), New York: Macmillan, 1972, pp. 175-282.

Schönfinkel, Moses

[1924] "On the Building Blocks of Mathematical Logic," From Frege to Gödel, Heijenoort (ed.), Cambridge: Harvard University Press, 1967.

Strang, Colin

[1971] "Plato and the Third Man," Plato: A Collection of Critical Essays, Vlastos (ed.), New York: Doubleday, 1971, pp. 184-200.

Tarski, Alfred

[1944] "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* (1944), pp. 431-575.

Vlastos, Gregory

[1954] "The Third Man Argument," *The Philosophical Review* 63 (1954).

Wittgenstein, Ludwig

[1921] Tractatus Logico-Philosophicus, London: Routledge and Kegan Paul, 1961.

Zalta, Edward N.

[1979a] "Alternative Meinongian Semantics," unpublished.

[1979b] "An Alternative Theory of Nonexistent Objects," with Alan McMichael, *Journal of Philosophical Logic* 9 (1980), pp.

[1979c] "Plato and Predication," unpublished.

[1979d] "A Platonic Theory of Forms and an Application to The Sophist," unpublished.

[1979e] "Monads and Meinong," unpublished.

APPENDIX A

CLARK'S PADADOX

1.	$(\exists x)(A!x \,\&\, (F)(xF \equiv F=[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)]))$	
	A-OBJECTS, ϕ: $F=[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)]$	
2.	$A!a_0 \,\&\, (F)(a_0F \equiv F=[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)])$	EE, 1
3.	$(F)(a_0F \equiv F=[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)])$	&E, 2
4.	$a_0[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)] \equiv [\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)] = [\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)]$	UE, 3
5.	$[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)] = [\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)]$	=I
6.	$a_0[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)]$	≡E, 4, 5
7.	$[\lambda x\ (F)(xF \rightarrow Fx)]a_0$	Assumption
8.	$(F)(a_0F \rightarrow Fa_0)$	λ-E, 7
9.	$a_0[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)] \rightarrow [\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)]a_0$	UE, 8
10.	$[\lambda x\ (\exists F)(xF \,\&\, {\sim}Fx)]a_0$	→E 6, 9
11.	$(\exists F)(a_0F \,\&\, {\sim}Fa_0)$	λ-E, 10
12.	${\sim}(F)(a_0F \rightarrow Fa_0)$	QN, 11
13.	$[\lambda x\ (F)(xF \rightarrow Fx)]a_0 \equiv (F)(a_0F \rightarrow Fa_0)$	λ-EQUIV
14.	${\sim}[\lambda x\ (F)(xF \rightarrow Fx)]a_0$	≡E 12, 13
15.	${\sim}[\lambda x\ (F)(xF \rightarrow Fx)]a_0$	IP 7-14
16.	${\sim}(F)(a_0F \rightarrow Fa_0)$	≡E 13, 15
17.	$(\exists F)(a_0F \,\&\, {\sim}Fa_0)$	QN, 16
18.	$a_0R \,\&\, {\sim}Ra_0$	EE, 17

19.	a_0R	&E, 18
20.	$a_0R \equiv R=[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]$	UE, 3
21.	$R=[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]$	$\equiv$E 19, 20
22.	${\sim}Ra_0$	&E, 18
23.	${\sim}[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]a_0$	=E 21, 22
24.	$[\lambda x\ (\exists F)(xF\ \&\ {\sim}Fx)]a_0$	λI, 17

Contradiction, lines 23, 24

APPENDIX B

McMICHAEL'S PARADOX

1.	$(\exists x)(A!x \mathbin{\&} (F)(xF \equiv (\exists u)(F=[\lambda y\ y=u] \mathbin{\&} {\sim}uF)))$	A-OBJECTS
2.	$A!a_1 \mathbin{\&} (F)(a_1F \equiv (\exists u)(F=[\lambda y\ y=u] \mathbin{\&} {\sim}uF))$	EE, 1
3.	$a_1[\lambda y\ y=a_1]$	Assumption
4.	$a_1[\lambda y\ y=a_1] \equiv (\exists u)([\lambda y\ y=a_1]=[\lambda y\ y=u] \mathbin{\&} {\sim}u[\lambda y\ y=a_1])$	UE, 2
5.	$(\exists u)([\lambda y\ y=a_1]=[\lambda y\ y=u] \mathbin{\&} {\sim}u[\lambda y\ y=a_1])$	$\equiv$E, 3, 4
6.	$[\lambda y\ y=a_1]=[\lambda y\ y=a_2] \mathbin{\&} {\sim}a_2[\lambda y\ y=a_1]$	EE, 5
7.	$a_1=a_1$	=I
8.	$[\lambda y\ y=a_1]a_1$	λI, 7
9.	$[\lambda y\ y=a_2]a_1$	=E, 6, 8
10.	$a_1=a_2$	λE, 9
11.	${\sim}a_1[\lambda y\ y=a_1]$	=E, 6, 10
12.	${\sim}a_1[\lambda y\ y=a_1]$	IP, 3-11
13.	${\sim}(\exists u)([\lambda y\ y=a_1]=[\lambda y\ y=u] \mathbin{\&} {\sim}u[\lambda y\ y=a_1])$	$\equiv$E, 4, 12
14.	$(u)([\lambda y\ y=a_1]=[\lambda y\ y=u] \rightarrow u[\lambda y\ y=a_1])$	QN, 13
15.	$[\lambda y\ y=a_1]=[\lambda y\ y=a_1] \rightarrow a_1[\lambda y\ y=a_1]$	UE, 14
16.	$[\lambda y\ y=a_1]=[\lambda y\ y=a_1]$	=I
17.	$a_1[\lambda y\ y=a_1]$	$\rightarrow$E, 15, 16

Contradiction, 12, 17

APPENDIX C

INCORPORATING NON-RIGID DESIGNATORS

Had we decided to have non-rigid definite descriptions in our language, we would have needed to redefine the notions of <u>denotation</u>$_{\mathcal{I},f}$ and satisfaction. On the definition of denotation presented below, every term receives a denotation <u>with respect to</u> a possible world:

<u>Denotation</u>. Given an interpretation $\mathcal{I}$ and an $\mathcal{I}$-assignment f, we recursively define the <u>denotation of</u> term τ <u>at</u> world w <u>with respect to</u> $\mathcal{I}$ <u>and</u> f as follows:

1. where κ is any primitive name, $d_{\mathcal{I},f}(\kappa,w) = F_{\mathcal{I}}(\kappa)$
2. where α is any primitive variable, $d_{\mathcal{I},f}(\alpha,w) = f(\alpha)$
3. where $(\iota x)\phi$ is any non-rigid description, $d_{\mathcal{I},f}((\iota x)\phi,w) =$

$$\begin{cases} o \text{ iff } (\exists f')(f' \underset{x}{=} f \;\&\; f'(x)=o \;\&\; f' \text{ satisfies } \phi \text{ with respect} \\ \quad \text{to } w \;\&\; (f'')(f'' \underset{x}{=} f \;\&\; f''(x)=o' \;\&\; f'' \text{ satisfies } \phi \text{ with} \\ \quad \text{respect to } w \rightarrow o'=o)) \\ \text{undefined, otherwise,} \end{cases}$$

   where satisfaction is defined as in this appendix, below.

4. where $[\lambda\nu_1\ldots\nu_n\ \rho^n\nu_1\ldots\nu_n]$ is any elementary λ-expression, $d_{\mathcal{I},f}([\lambda\nu_1\ldots\nu_n\ \rho^n\nu_1\ldots\nu_n],w) = d_{\mathcal{I},f}(\rho^n,w)$
5. where μ is the i^{th}-plugging of ξ by o, $d_{\mathcal{I},f}(\mu,w) = PLUG_i(d_{\mathcal{I},f}(\xi,w),\ d_{\mathcal{I},f}(o,w))$
6. where μ is the i^{th}-projection of ξ, $d_{\mathcal{I},f}(\mu,w) = PROJ_i(d_{\mathcal{I},f}(\xi,w))$

7. where μ is the i,j^{th}-conversion of ξ, $d_{I,f}(\mu,w) =$ $CONV_{i,j}(d_{I,f}(\xi,w))$

8. where μ is the i,j^{th}-reflection of ξ, $d_{I,f}(\mu,w) =$ $REFL_{i,j}(d_{I,f}(\xi,w))$

9. where μ is the i^{th}-vacuous expansion of ξ, $d_{I,f}(\mu,w) =$ $VAC_i(d_{I,f}(\xi,w))$

10. where μ is the conjunction of ξ and ζ, $d_{I,f}(\mu,w) =$ $CONJ(d_{I,f}(\xi,w),\ d_{I,f}(\zeta,w))$

11. where μ is the negation of ξ, $d_{I,f}(\mu,w) =$ $NEG(d_{I,f}(\xi,w))$

12. where μ is the necessitation of ξ, $d_{I,f}(\mu,w) =$ $NEC(d_{I,f}(\xi,w))$

13. where ϕ is any propositional term, $d_{I,f}(\phi,w)$ is defined as follows:

    (a) if ϕ is a primitive propositional term, $d_{I,f}(\phi,w)$ is already defined

    (b) if $\phi = \rho^n o_1 \ldots o_n$, $d_{I,f}(\phi,w) = PLUG_1(PLUG_2(\ldots(PLUG_n(d_{I,f}(\rho^n,w),\ d_{I,f}(o_n,w)),\ldots),d_{I,f}(o_2,w)),d_{I,f}(o_1,w))$

    (c) if $\phi = (\sim\psi)$, $d_{I,f}(\phi,w) = NEG(d_{I,f}(\psi,w))$

    (d) if $\phi = (\psi \& \chi)$, $d_{I,f}(\phi,w) = CONJ(d_{I,f}(\psi,w),\ d_{I,f}(\chi,w))$

    (e) if $\phi = (\exists\nu)\psi$, $d_{I,f}(\phi,\ ) = PROJ_1(d_{I,f}([\lambda\nu\ \psi],w))$

    (f) if $\phi = \Box\psi$, $d_{I,f}(\phi,w) = NEC(d_{I,f}(\psi,w))$

Satisfaction. Given an interpretation I and an I-assignment f, we define f satisfies ϕ with respect to world w as follows:

1. where ϕ is any primitive propositional term, $\mathfrak{f}$ satisfies ϕ with respect to w iff $ext_w(d_{I,\mathfrak{f}}(\phi,w)) = T$
2. where $\phi = \rho^n o_1 \ldots o_n$, $\mathfrak{f}$ satisfies ϕ with respect to w iff $\langle d_{I,\mathfrak{f}}(o_1,w),\ldots,d_{I,\mathfrak{f}}(o_n,w)\rangle \varepsilon\ ext_w(d_{I,\mathfrak{f}}(\rho^n,w))$
3. where $\phi = o\rho$, $\mathfrak{f}$ satisfies ϕ with respect to w iff $d_{I,\mathfrak{f}}(o,w)\ \varepsilon\ ext_A(d_{I,\mathfrak{f}}(\rho,w))$
4. where $\phi = (\sim\psi)$, $\mathfrak{f}$ satisfies ϕ with respect to w iff $\mathfrak{f}$ fails to satisfy ϕ with respect to w
5. where $\phi = (\psi\&\chi)$, $\mathfrak{f}$ satisfies ϕ with respect to w iff $\mathfrak{f}$ satisfies both ψ and χ with respect to w
6. where $\phi = (\exists\alpha)\psi$, $\mathfrak{f}$ satisfies ϕ with respect to w iff $(\exists\mathfrak{f}')(\mathfrak{f}'$ satisfies ψ with respect to $w)$
7. where $\phi = (\Box\psi)$, $\mathfrak{f}$ satisfies ϕ with respect to w iff $(w')(\mathfrak{f}$ satisfies ψ with respect to $w')$

APPENDIX D

A PROOF

We want to show that $D_{16} \Rightarrow D_{17}$. So suppose z_0 is a D_{16}-world. So every property z_0 encodes is vacuous. Consequently, all we have to show is that $\Diamond(F^0)(\Sigma_{z_0} F^0 \equiv F^0)$. We're given that $\Diamond(F^0)(\Sigma_{z_0} F \rightarrow F^0)$. So let $\phi = \ulcorner(F^0)(\Sigma_{z_0} F^0 \rightarrow F^0)\urcorner$. We try to show that $\phi \rightarrow \psi$, where $\psi = \ulcorner(F^0)(\Sigma_{z_0} F^0 \equiv F^0)\urcorner$. Consequently, all we need to prove is that $Q^0 \rightarrow \Sigma_{z_0} Q^0$, for an arbitrary proposition Q^0. So assume Q^0, i.e., $\sim\sim Q^0$. By instantiating $\sim Q^0$ into ϕ, we get $\Sigma_{z_0} \sim Q^0 \rightarrow \sim Q^0$. So $\sim\Sigma_{z_0} \sim Q^0$. Since z_0 is a D_{16}-world, it is maximal. So $\Sigma_{z_0} Q^0$. So $\phi \rightarrow \psi$, and hence $\Box(\phi \rightarrow \psi)$. Since $\Diamond\phi$ is true by the definition of z_0, it follows by our S_5-theorem that $\Diamond\psi$.

APPENDIX E

MODELLING NOTIONS

Throughout this work, we've talked about notions. Syntactic notions like term, occurrence, the erasure of a formula ϕ, etc., aren't all that mysterious--they seem to be *bona fide* relations among linguistic objects. However, there is a group of metaphysical notions which are rather puzzling. These are all the notions constructed out of the primitive notion of encoding. These notions fall into two groups: defined notions (such as correlation, Form, Monad, World, complete, maximal, etc.) and paradoxical notions (such as exemplifying a property that is not also encoded, exemplifying every property that is encoded, and identity). The former group of notions may not be relations (since the formulas ϕ which would "express" them violate restrictions on λ-formation and RELATIONS) and it's provable that the latter group could not be relations--if they were, some contradiction would be true. The reason these notions are puzzling is because as ontologists, we like to avoid uncategorizable entities ("ontological danglers"). So if these notions aren't relations, what are they? Do they have independent ontological status? Or is talk about "notions" just a convenient reification, disguising metalinguistic talk about objects which satisfy definitions?

Even if the answer to the last question is yes, it might be worthwhile to look for a reification procedure whereby we do find some appropriate object in our ontology to code up, or go proxy for our notions. An easy, though risky way to do this would be to add an axiom which asserts that there is a *primitive* encoding relation and a relation behind every one of our *defined* notions (and leave the paradoxical notions to dangle). Until we have an easy way of confirming the consistency of the theories which result, such a procedure seems suspicious and unsystematic.

There may be a better way, however; one which allows us to find proxies for the primitive notion of encoding, the defined notions *and* the contradictory notions. The trick is to build the notional formula in question into the defining formula for an abstract relation. Consider the following two instances of A-OBJECTS:

(1) $(\exists z^{(i,i/p)/p})(F^{((i,i/p)/p)/p})(zF \equiv (\exists x^{i})(\exists H^{i/p})(xH \;\&$
$F=[\lambda G^{(i,i/p)/p} GxH]))$

(2) $(\exists z^{(i,i)/p})(F^{((i,i)/p)/p})(zF \equiv (\exists x^{i})(\exists y^{i})((H^{i/p})(yH \equiv Hx) \;\&$
$F=[\lambda G^{(i,i)/p} Gxy]))$

(1) says there is an abstract (i,i/p)/p-relation (between individuals and properties of individuals), z, which encodes a property of such relations, F, iff F is the property of: being the relation which relates an object x^{i} with a property $H^{i/p}$ it encodes. (2) says that there is an abstract relation among individuals, z, which encodes a property of such relations, F, iff F is the property of: being the relation which relates an object y^{i} with its correlate x^{i}. These two abstract relations are unique. It seems reasonable to suppose that they could represent the primitive notion of encoding and the defined notion of correlation, respectively.

Note that by abbreviating a crucial subformula, $(H^{i/p})(yH \equiv Hx)$, in (2), we could rewrite (2) as:

$$(2)' \quad (\exists z)(F)(zF \equiv (\exists x)(\exists y)(Cor(x,y)\ \&\ F=[\lambda G\ Gxy]))$$

(3) and (4) are further examples:

$$(3) \quad (\exists z^{i/p})(F^{(i/p)/p})(zF \equiv (\exists x^{i})(World(x^{i})\ \&\ F=[\lambda G^{i/p}\ Gx]))$$

$$(4) \quad (\exists z^{i/p})(F^{(i/p)/p})(zF \equiv (\exists x^{i})((\exists H^{i/p})(xH\ \&\ {\sim}Hx)\ \&\ F=[\lambda G^{i/p}\ Gx]))$$

These abstract properties could serve to represent the notions of being a world and being a non-self-correlate, respectively. Necessarily, the latter fails to have a weak correlate--if it did, some contradiction would be true.

It is important to see that this modelling of a contradictory notion doesn't reintroduce Clark's Paradox. Let us use $[\lambda x\ (\exists H)(xH\ \&\ {\sim}Hx)]$ to denote the abstract property that (4) gives us. That is, we're using what previously had been an ill-formed λ-expression to name a unique abstract object. But we cannot allow this λ-expression to be used in instances of λ-EQUIVALENCE. That's because we've permitted abstract objects of type t to encode abstract properties of type t/p--so we know that there would be an abstract object of type i which encodes $[\lambda x\ (\exists H)(xH\ \&\ {\sim}Hx)]$ (the abstract property given by (4)). This would be the first move in Clark's Paradox. The second would be to suppose it exemplified $[\lambda x\ (F)(xF \rightarrow Fx)]$ (an abstract property constructed using an axiom like (4)). But we can stop the paradox from developing to completion by not allowing λ-conversion in the (formerly ill-formed) λ-expressions we've just used to name our abstract properties. Metaphysically speaking, this means that we are not supposing that the abstract properties we've chosen to represent the notions of being a non-self-correlate and being a weak correlate of one's self have in their respective exemplification extensions just the objects which are non-self-correlates or which are weak-self-correlates. And in general, we do not suppose that the abstract relations which represent our defined and contradictory notions have in their exemplification extensions all and only the n-tuples of objects which satisfy the defining formulas of these relational notions. This prevents the paradoxes from being reintroduced.

A question then immediately arises. If the proxy abstract relations don't have the "appropriate" exemplification extensions, why is our modelling procedure worthwhile? Well, the answer is that it is useful. If we generalize our procedure, we can handle data which we couldn't handle before. First, here's our generalization:

where ϕ is any non-propositional formula, and $\alpha_1,\ldots,\alpha_n$ are any variables of types $t_1,\ldots,t_n$, respectively, then:

$$[\lambda\alpha_1\ldots\alpha_n\ \phi] =_{abbr} (\iota z^{(t_1,\ldots,t_n)/p})(F^{((t_1,\ldots,t_n)/p)/p})$$
$$(zF \equiv (\exists\alpha_1)\ldots(\exists\alpha_n)(\phi\ \&\ F=[\lambda G^{(t_1,\ldots,t_n)/p}\, G\alpha_1\ldots\alpha_n]))$$

This gives us a way to easily name the abstract object which goes proxy for the defined or contradictory notion "expressed" by ϕ. This procedure allows us to represent the following data triad:

A. S believes that Dostoyevsky wrote about the student who killed an old moneylender according to <u>Crime and Punishment</u>

B. S doesn't believe that Dostoyevsky wrote about Raskolnikov

C. Raskolnikov is the student who killed the old moneylender according to <u>Crime and Punishment</u>

To get the denotation for the English definite description correct, it must be symbolized as $(\iota x)\Sigma_{\underline{CP}}(Sx \;\&\; (\exists y)(OMLy \;\&\; Kxy))$. This is the way we did things in Chapter IV, §4. But we can't underline our representing description and turn it into a sense description because it's not constructed out of a propositional formula. "$\underline{(\iota x)\Sigma_{CP}\phi}$" is not a well-defined sense description. So we didn't have a definite description in our formal language which had the right denotation and which, when underlined, represented the sense of the English definite description.

But now we can do this. We can suppose that "$\underline{(\iota x)\Sigma_{CP}\phi}$" was defined as follows:

$$\underline{(\iota x)\Sigma_{CP}\phi} =_{dfn} (\iota z)(F)(zF \equiv F{=}[\lambda x\, \Sigma_{CP}\phi \;\&\; (y)(\Sigma_{CP}\phi^{y}_{x} \rightarrow y{=}_{E}x)])$$

The λ-expression used in this definition abbreviates a definite description of an abstract property. $\underline{(\iota x)\Sigma_{CP}\phi}$ is the abstract object which encodes just this abstract property. Consequently, we may represent A-B-C as follows:

(A') Bsthat-Wd$\underline{(\iota s)\Sigma_{CP}\phi}$ (*dicto*)

(B') ~Bsthat-Wd$\underline{r}_{s}$ (*dicto*)

(C') $r{=}(\iota x)\Sigma_{CP}\phi$

This should give the reader a good idea how to handle the data in footnote 12, Chapter VI, §1. But what about the following data:

D. S believes that the person who killed an old moneylender according to Crime and Punishment was a student

E. S doesn't believe that Raskolnikov was a student

F. Raskolnikov is the person who killed an old moneylender according to Crime and Punishment

To handle data like this, we would need to incorporate the new logical function described in footnote 11, Chapter VI, §1. *ENPLUG*(r,o) is the proposition that o encodes r. It would be denoted by "$\tau\rho$," where τ denotes o and ρ denotes r. We then get:

(D') $\mathrm{Bsthat}\text{-}\underline{(\iota x)\Sigma_{CP}(Px \ \& \ (\exists y)(OMLy \ \& \ Kxy))}S$

(E') $\sim\mathrm{Bsthat}\text{-}\underline{r}_s S$

(F') $r=(\iota x)\Sigma_{CP}(Px \ \& \ (\exists y)(OMLy \ \& \ Kxy))$

But there is still a problem about making this more general. Consider G and H:

G. S believes that Porphyry arrested the student who killed an old moneylender.

H. S doesn't believe that Porphyry arrested Raskolnikov

To represent G and H correctly, and in a way which suggests a completely general treatment, we recall that "$\Sigma_{CP}\phi$" abbreviates "$CP[\lambda y\ \phi]$." So can we represent G and H as:

(G') $\mathrm{Bsthat}\text{-}CP[\lambda y\ Ap\underline{(\iota x)\Sigma_{CP}(Sx \ \& \ (\exists u)(OMLu \ \& \ Kxu))}]$

(H') $\sim\mathrm{Bsthat}\text{-}CP[\lambda y\ Ap\underline{r}_s]$

The general solution is to use *ENPLUG* on the vacuous property encoded by <u>Crime and Punishment</u> to get the proposition that <u>Crime and Punishment</u> encodes the vacuous property. This vacuous property is denoted by a λ-expression, $[\lambda y\ Ap(\iota x)\Sigma_{CP}(Sx \ \& \ (\exists u)(OMLu \ \& \ Kxu))]$; but in trying to capture the sense of the English description involved, we have to underline the translating description, thereby denoting an abstract object which encodes an abstract property.

APPENDIX F

AN OBJECTION AND A REPLY

The Gettier-Parsons objection is that the theory, plus our techniques for applying it, seems to commit us to a pair of sentences which cannot both be true. Consider the following case of John, an I.R.S. agent who dabbles in philosophy.

Suppose John is assigned to watch a lottery and determine whether it has been fixed. The draw has not yet taken place, but the winner, whoever it is, will earn \$100,000. Suppose also that John, having accepted the theory of abstract objects and the AUXILIARY HYPOTHESIS, thinks both that the property of earning \$100,000 is nuclear and that no abstract objects exemplify this property.

Given such a situation, the following two English sentences, (V) and (W), may both be true:

(V) John doesn't believe that the winner of the lottery, whoever it is, will fail to earn \$100,000

(W) John believes that the abstract object which encodes just the property of being the winner fails to exemplify earning \$100,000

Now let "j" denote John, "W" denote the property of winning the lottery, and "E" denote the property of earning \$100,000. The techniques for applying the theory developed in Chapter VI, §1, apparently commit us to the following two translations:

(V') $\sim Bj\text{that-}\sim E\underline{(\iota x)Wx}$

(W') Bjthat--$\sim E(\iota z)(F)(zF \equiv F=[\lambda y\ Wy\ \&\ (u)(Wu \rightarrow u=_E y)])$

However, (V') and (W') cannot both be true since it is a principle of the theory (LA13) that:

$\underline{(\iota x)Wx} = (\iota z)(F)(zF \equiv F=[\lambda y\ Wy\ \&\ (u)(Wu \rightarrow u=_E y)])$

A simple application of =E shows that (V') and (W') are incompatible.

So, put bluntly, the objection is that (V) and (W) seem to be compatible pieces of data which receive incompatible translations. To reply to this objection, we either deny that both (V) and (W) are pieces of data or deny that both (V') and (W') correctly translate (V) and (W), respectively. In the first part of the following reply, we argue that either (W) is not a datum or (W') does not correctly translate (W).

To do this, we must recall our discussion in the Introduction (where we tried to give the reader some idea of what the data was) and appeal to the reader's own intuitions about what constitutes data. Clearly, beliefs about specific A-objects, and beliefs involving the technical notions of being abstract and encoding a property could not have been part of the data we were trying to explain. Though we may have had beliefs about Sherlock Holmes or the round square, these beliefs were pretheoretical and any ascription of such beliefs would be couched in non-technical English.

Consequently, if we want to think of (W) as a datum, we have to suppose that the words "abstract" and "encodes" as they occur in (W) have their *non-technical* senses. This supposition makes it easier to argue that (W') is not the proper translation of (W).

We may summarize our first response in the following argument:

Data sentences do not contain technical terms. So if "abstract" and "encodes" in (W) are technical terms, (W) is not a datum. If "abstract" and "encodes" in (W) are non-technical terms, (W') is not the proper translation of (W). Therefore, either (W) is not a datum or (W') is not the proper translation of (W).

Although the above argument may defuse the objection as stated, some questions immediately arise: Do we acquire new beliefs about abstract objects during the exposition of the theory? Do such beliefs constitute new data to be explained by the theory? If one thinks that the answer to these two questions is "yes," then one must presuppose that the word "believes" as it occurs in the technical English sentences like (W) has the same sense as it does when it occurs in our original, non-technical data. This may require an argument, since it is not clear that when we say that we acquire new "beliefs" about abstract objects as a result of the presentation of the theory, we mean the same thing that we do when we ascribe "beliefs" to others in natural language.

If one has good justification for thinking that "believes" is used univocally, and that (W) is a new datum to be explained, then finding a consistent representation of (V) and (W) remains a puzzle. If this constitutes the only puzzle the theory cannot explain, then it must be admitted that the outstanding, recalcitrant data of the Russellian paradigm has been successfully explained. It is only natural that along with the more finely grained view of things that accompanies a paradigm shift, new puzzles not characterizable from within the old paradigm should arise. Let us, then, close with a few brief comments about this new puzzle.

There are two suggestions which might prove fruitful in future investigations. The first is to suggest that (W') may, nevertheless, not be the correct translation of (W). Maybe we should avail ourselves of the technique developed in Appendix E for representing the senses of non-propositional descriptions and translate (W) as (W"):

(W") $Bj\text{that-}{\sim}E\underline{(\iota z)(F)(zF \equiv F=[\lambda y\ Wy\ \&\ (u)(Wu \rightarrow u=_E y)])}$

By doing this, we suppose that it is the sense of the English description in (W) which is a constituent of the propositional object of belief. We've identified the sense of such a description in Appendix E as an abstract object which encodes an abstract property.

An alternative suggestion is to suppose (W') is the correct translation of (W), but that (V') is not the correct translation of (V). Such a suggestion would require that we give up the specific treatment proposed in Chapter VI. But we could replace it with an alternative translation procedure which incorporates abstract objects. For example, maybe (V) is to be translated as (V"), where "R" here denotes our representation relation:

(V") $(\exists z)(Rz,(\iota x)Wx,j\ \&\ {\sim}Bj\text{that-}{\sim}Ez)$

So (V) will be true just in case some abstract object represents the winner for John (maybe this abstract object just is <u>the winner</u>) and John fails to believe that it fails to earn \$100,000. We would leave our treatment of the *de re* cases just as they are, but propose a general analysis along these lines for the *de dicto* cases. So there still may be of involving abstract objects in the analysis of belief.