Part III

Unification

7

Extensionality and Meaning

Up to this point in my investigation of properties, relations, and propositions I have taken a free and easy approach toward three issues: the relation between extensional and intensional logic, the relation between the semantic theory of truth and the theory of meaning, and the relation between the two traditional conceptions of PRPs, conceptions 1 and 2. First, I have formulated intensional logic as if it were an autonomous subject above and beyond extensional logic. Secondly, I have characterized the semantics for intensional language by means of a Tarski-style theory of truth, and in so doing I have proceeded as if the theory of truth were an autonomous subject independent of the theory of meaning. Finally, I have developed the two traditional conceptions of PRPs side by side as if each one were fully correct and autonomous from the other. Taking these steps considerably simplified the investigation of several highly complex problems. But the three issues must be confronted directly before a fully satisfactory treatment of PRPs can be attained.

My goal will be unification: a construction of intensional logic within an extensional logic; a derivation of the theory of truth from a unified semantic theory based on a single meaning relation, and finally, a synthesis of the two traditional conceptions of PRPs within a theory of qualities and concepts.

In the classical tradition of Frege and Russell, the first two issues—the relation between extensional and intensional logic and the relation between the theory of truth and the theory of meaning—are treated in tandem. I will follow the same practice in this chapter. In the following chapter, which is more metaphysical than linguistic in character, I will discuss the relation between the two traditional conceptions of PRPs. The resulting synthesis will then be applied to a number of outstanding problems in modern philosophy. These applications comprise the final two chapters.

37. The Thesis of Extensionality

Why is there intensionality in language? Why, in addition to extensional language, for which Leibniz's law and the substitutivity of equivalent formulas are valid, is there any language for which these laws are invalid?

In chapter 1 we took a preliminary look at these questions. When one looks more closely, one may take either of two fundamentally different attitudes toward this question. According to the first, it is simply assumed that there really are violations of the laws of extensional logic. We are to take it as a fact of life that the extensional and the intensional are two primitively different kinds of linguistic forms having different kinds of logic. According to this attitude, one should not attempt to explain away intensionality in language. Rather, one should simply codify both kinds of logic, extensional and intensional, more or less as they appear to the naive eye. This attitude toward intensionality in language underlies the now dominant movement in modal logic originated by C. I. Lewis and carried on by Carnap in *Meaning and Necessity*[1] and, more recently, by Hintikka, Kripke, and their school. This attitude is also held by Montague in his 'Universal Grammar' and 'The Proper Treatment of Quantification in English'. And, of course, this is the attitude originally held by Russell in the first edition of *Principia Mathematica*, where we find the first comprehensive formal intensional logic.[2]

According to the second attitude toward intensionality in language, there are no genuine violations of the laws of extensional logic; all *prima facie* intensional phenomena are surface phenomena that can be explained away. All language, and all logic, is at bottom extensional. This attitude is, of course, the one held by Frege in 'Über Sinn und Bedeutung'. There Frege outlined a theory, subsequently formalized by Church, by means of which all *prima facie* intensional phenomena can be explained away. According to this theory, all *prima facie* deviations from extensional logic are produced by the fact that in the problematic contexts certain expressions do not name what they usually name; instead they name the intensional entities that they usually express. Thus, for Frege and Church, there is no genuinely intensional language; when *prima facie* intensional language is properly analysed, it turns out to be extensional language concerning intensional entities.

The attitude that intensionality in language is only an apparent phenomenon is also evident in Carnap's thesis of extensionality, which he advanced and defended in *The Logical Structure of the World* and *The Logical Syntax of Language* (only to abandon it in *Meaning and Necessity*). The thesis was the subject of lively debate from Wittgenstein's *Tractatus* to Carnap's *The Logical Syntax of Language*. Since then, however, it has been largely neglected, no doubt as a result of the prevalence of the first attitude, i.e., the one Carnap took in *Meaning and Necessity*. In *The Logical Syntax of Language* Carnap formulates the thesis of extensionality as follows:

> *a universal language of science may be extensional*; or, more exactly: for every given intensional language S_1, an extensional language S_2 may be constructed such that S_1 can be translated into S_2. (p. 245)

The truth of the thesis of extensionality would immediately suggest an account of intensionality in language. An especially perspicuous statement of this account may be given when one appeals to the theory of transformational grammar. If, as the thesis of extensionality insures, every intensional sentence can be translated into an extensional sentence, then every intensional sentence can be treated as the result of applying a meaning-preserving transformation to the original extensional sentence. The deep structure of language would be fully extensional, and intensionality in language would be a mere surface phenomenon. Such an account of intensionality in language would be quite general; indeed, Frege's account could be viewed as a special case of it.

In logical theory there are two competing general methodologies, one liberal and one conservative, and the conflict between the two attitudes toward intensionality in language may be viewed as an instance of this methodological conflict. According to the liberal methodology, when one attempts to extend the scope of logical theory to uncharted territory, one should feel free to adopt a pragmatic approach. Specifically, one should not feel constrained to formulate the laws of logic in such a way that they exhibit maximum generality. All that is required is that the new theory should work as a self-contained whole. According to the conservative methodology, by contrast, one is constrained to formulate the laws of logic in such a way that they do exhibit maximum generality: if a law is valid for a natural fragment of the language under consideration, then it should, if at all possible, be valid for the entire language. It is not enough that a logical theory should work as a self-contained whole.[3] There is in addition a demand for generality.

This methodological conflict shows up in numerous places in logical theory. For example, it is seen in the two classical theories of definite descriptions, Frege's and Russell's. On Frege's theory, definite descriptions are the result of applying an unanalysed name-forming operator (e.g., 'the α such that') to open-sentences (e.g., '$A\alpha$').[4] Sentences containing definite descriptions are then treated as if the definite descriptions were full-fledged names. However, in view of the phenomenon of vacuous definite descriptions, Frege's theory has the effect of overturning certain logical laws that held prior to the introduction of definite descriptions, namely, existential generalization and the law of the excluded middle. To the extent that Frege is willing thus to put limits on the syntactic domain in which these laws are valid, he may be viewed as a proponent of the liberal methodology, which does not demand maximum generality. In contrast to Frege's theory of definite descriptions, Russell's theory does not have the effect of overturning any logical laws that held prior to the introduction of definite descriptions. For on Russell's theory, sentences containing definite descriptions arise from the application of a meaning-preserving transformation to sentences that, ultimately, contain no definite descriptions. Hence, all logical laws that held prior to the introduction of definite descriptions—including the law of the excluded middle and existential generalization—hold for the enlarged language which contains definite descriptions. Thus, with regard to definite descriptions Russell's theory tends to maximize generality in the laws of logic. Inasmuch as this was his goal, Russell may be viewed as a proponent of the conservative methodology, which demands maximum generality.

Our present concern is the treatment of intensionality in language. The Lewis-style modal logician begins his study with a standard extensional language. To this language he adjoins new primitive operators $\Box$ and $\Diamond$, thereby obtaining an enlarged class of well-formed formulas. Since $\Box$ and $\Diamond$ are taken as primitive, the logical syntax of each new formula is identified with its surface syntax. As a result, the new formulas must be treated as genuine violators of the laws of extensional logic, i.e., laws that hold for the language to which the modal operators are adjoined. However, this loss of generality makes no difference to the Lewis-style modal logician. All that matters to him is that the logic for the enlarged language can be characterized in terms of his logical syntax. In this, the Lewis-style modal logician may be viewed as a proponent of the liberal methodology. By contrast, on the treatment of intensionality associated with the thesis of extensionality, all *prima facie* intensional expressions arise from the application of meaning-preserving transformations to expressions that ultimately conform to the laws of extensional logic. Hence, the laws that hold in the original language also hold in the enlarged language, generally. Thus, this treatment of intensionality promotes the maximum generally demanded by the conservative methodology.

This methodological conflict over the treatment of intensionality also shows up in the theories held by Russell and Frege. This time, however, Russell, not Frege, is the one representing the liberal methodology, for he posits primitive intensional forms as well as primitive extensional forms. And Frege, not Russell, is the one representing the conservative methodology, for he advocates a fully extensional logic for *prima facie* intensional language.[5] Other things being equal, methodological vacillations like these ought to be avoided.

It cannot be denied that the liberal methodology has led to a number of valuable advances in logic. Indeed, whereas the conservative approach often bogs down in the face of demanding constraints, the liberal methodology is free and easy. And inasmuch as it frees logical structure to reflect surface grammatical forms, it removes one of the barriers to a familiar and natural logical syntax. These reasons justify the liberal methodology at least as a short-term strategy. However, as a long-term strategy, only the conservative methodology guarantees the kind of maximally general theory that is the highest ideal in science. So it too should be explored.

Similarities among intensional expressions in natural language lead us to the thesis, defended in §§6–9, that there is a single intensional abstraction operation—represented by the bracket notation $[\ ]_\alpha$—that underlies all apparent intensionality in language. Now suppose that following the liberal methodology one treats this intensional abstraction operation as if it were primitive. (This is the strategy followed in chapter 2.) One would then arrive at an intensional logic that is highly successful at least on its own terms. To stop here would be to adopt the liberal methodology. Can this approach to intensional logic be brought into line with the conservative methodology, which calls for a logic that at bottom is fully extensional? That is, in accordance with the thesis of extensionality, is there a way to treat sentences containing intensional abstracts as the outcome of applying a meaning-preserving transformation to sentences that ultimately are extensional? Can intensional abstracts be defined, directly or contextually, within extensional language?

According to §8, intensional abstracts are semantically correlated, not with linguistic entities, but instead with intensional entities. In view of this, it would be unwise to base an extensional definition of intensional abstraction upon any of the nominalistic approaches (such as the one that Carnap offered in his original defense of the thesis of extensionality). It would also be unwise to pursue Frege's theory, for his approach to intensionality runs into troubles on such matters as Davidson's finite learnability requirement (desideratum 13; see §8) and quantifying-in (desideratum 5; see §11).[6]

Despite troubles in Frege's approach, one can draw inspiration from his informal doctrine that all *prima facie* intensional language is no more than extensional language about intensional entities. According to §8, an intensional abstract is semantically correlated with the intensional entity that the abstracted formula expresses. The key to giving an extensional definition of intensional abstraction, then, is to find a way to give extensional descriptions of the intensional entities semantically correlated with intensional abstracts. The algebraic theory of intensional entities is the crucial ingredient. On the resulting analysis, an intensional abstract would be treated as a transformation from a structural description of an intensional entity; further, this structural description would be stated in terms of the fundamental algebraic logical operations (conjunction, negation, existential generalization, etc.), and finally, the syntactic structure of the intensional abstract would stand in an easy one-one correlation to the syntactic structure of the structural description.

Some illustrations will make plain how this works. For example, 'the proposition that there exists an x such that Fx' is transformed from the (structural) definite description 'the proposition that is the existential generalization of the property F-ness'; 'the property of being an x such that Fx and Gx' is transformed from the definite description 'the property that is the conjunction of the properties F-ness and G-ness'; and so on for more and more complex intensional abstracts. (Both here and below I use the terms 'F-ness' and 'G-ness' for heuristic purposes only. Primitive property and relation names (e.g., 'red', 'love', etc.) would take over their role in the final analysis.)

Notice in these examples that when the predicate '*F*' occurs within an intensional abstract, it does not actually occur as a predicate. Instead, it occurs in effect as a name, for it is correlated via the transformation to an occurrence of the name '*F*-ness'. And this occurrence of the name '*F*-ness' names the very intensional entity that is the meaning of the predicate '*F*'. Thus, the transformation has the effect of giving us extensional occurrences of names that name the meanings of the predicates to which they are correlated. Much the same thing goes for *prima facie* occurrences of formulas within intensional abstracts. When a formula seems to occur within an intensional abstract, what one really has is an expression that is correlated via the transformation to an extensional occurrence of a definite description of the intensional entity that is the meaning of the formula. Hence, in general, the transformation has the effect of giving us extensional occurrences of singular terms that denote the meanings of the predicates and formulas to which they are correlated. This, of course, is reminiscent of certain aspects of the higher-order theories both of Frege and of the Russell of the first edition of *Principia Mathematica*.

With these preliminary remarks in mind I am now ready to outline how one could give a comprehensive account of intensionality in language. Consider first the apparent violation of Leibniz's law produced by co-denoting names (or indexicals). If names (indexicals) have no description content, then, as I will argue in §39, these apparent violations of Leibniz's law are pragmatic, not semantic, phenomena akin to those responsible for Mates' puzzle (see §18), and they can be given pragmatic explanations in terms of Grice's theory of conversational implicature (see details, §39). On the other hand, if names (indexicals) do have descriptive content, these apparent violations of Leibniz's law are special cases of the apparent violations produced by definite descriptions. However, the latter can be explained away by means of Russell's theory of descriptions.[7] According to this account, the apparent intensionality of a given occurrence of a definite description (e.g., the occurrence of 'the author of *Waverley*' in 'George IV wished to know whether Scott was the author of *Waverley*') is blamed on the apparent intensionality of occurrences of constituent descriptive predicates or formulas (e.g., 'x is an author of *Waverley*'). Thus, each apparent instance of intensionality produced by a singular term either is explained away pragmatically or is reduced to the apparent intensionality of certain occurrences of predicates of formulas. However, according to the proposed extensional analysis of intensional abstraction, every time we seem to have an occurrence of a predicate or formula that violates the principle of substitutivity of equivalents, the guilty occurrences are not genuine occurrences of predicates or formulas at all, and so they do not constitute genuine violations of this substitutivity principle. Given this extensional analysis of intensional abstraction, the substitutivity principle that is relevant to these linguistic contexts is again Leibniz's law. For according to the analysis, each problematic occurrence of a predicate or a formula is actually an occurrence of an expression correlated via the transformation to an occurrence of a singular term. Now, according to the analysis, this occurrence of a singular term denotes the meaning of the predicate or formula to which it is correlated syntactically. Consequently, the apparent intensional occurrence of a predicate or a formula could lead to a genuine violation of extensional logic only if there were genuine violations of the principle of the substitutivity of synonymous predicates and formulas, i.e., genuine substitutivity failures for predicates or formulas that have the same meaning. However, violations of this substitutivity principle would be instances either of the paradox of analysis or of Mates' puzzle and, hence, could be handled by means developed elsewhere (the paradox of analysis, §20; Mates' puzzle, §39) independently of the present issue concerning the thesis of extensionality.[8] And so by this chain of analyses one could eliminate all apparent violations of the principles of extensional logic. Logic and language would be at bottom extensional. Apparent intensional language would be extensional language concerning intensional entities.

In the remainder of this section I will spell out the details of this extensional analysis of intensional abstraction. In the next section I will take up the topic of meaning, which has figured informally throughout our discussion of the thesis of extensionality and indeed throughout all the preceding chapters. In the final section of this chapter I will, as promised, take up names, indexicals, and Mates' puzzle.

I begin by constructing a first-order extensional language L.*

* Some readers might wish to skip over this technical material.

The primitive symbols of L are the following:

Operators:	$\&, \neg, \exists$
Predicates:	$\mathscr{Conj}^3, \mathscr{Neg}^2, \mathscr{Exist}^2, \mathscr{Exp}^2,$ $\mathscr{Inv}^2, \mathscr{Conv}^2, \mathscr{Ref}^2, \mathscr{Pred}^3$ $F^1_1, F^1_2, F^1_3, \ldots, F^q_p$
Names:[9]	$\bar{F}^1_1, \bar{F}^1_2, \bar{F}^1_3, \ldots, \bar{F}^q_p$
Variables:	$x, y, z, \ldots$
Punctuation:	(,).

The formulas of L are built up in the usual way. As in L_ω, the predicate F^2_1 is singled out as a distinguished logical predicate and is to be rewritten as $=$. The predicates $\mathscr{Conj}^3$, $\mathscr{Neg}^2$, $\mathscr{Exist}^2$, $\mathscr{Exp}^2$, $\mathscr{Inv}^2$, $\mathscr{Conv}^2$, $\mathscr{Ref}^2$, $\mathscr{Pred}^3$ are also singled out as distinguished logical predicates. The semantics for L is done in the usual Tarskian extensional manner with certain added conditions (set forth in a moment) which insure the proper interpretation of these additional logical predicates. Since these added conditions only narrow the class of models for L, every L-formula that is true in all Tarskian models will be true in all models in the narrower class. Thus, any L-formula provable in standard first-order quantifier logic with identity will be valid on the intended semantics for L; this is true in particular for all instances of Leibniz's law and the substitutivity of equivalent formulas. In this precise sense, then, L is a fully extensional language. Notice also that L satisfies Davidson's finite learnability requirement since it has a finite number of primitive constants. The narrowed class of interpretations of L may be obtained in a short-cut way. Consider an arbitrary Tarskian model $\langle \mathscr{D}, \mathscr{R} \rangle$ for L. ($\mathscr{D}$ is the universe of discourse, and $\mathscr{R}$ is an extensional interpretation of the primitive predicates of L relative to $\mathscr{D}$.) Then, the model $\langle \mathscr{D}, \mathscr{R} \rangle$ is *admissible* if and only if it is associated with a standard model $\langle \mathscr{M}, \mathscr{I} \rangle$ for L_ω such that the following rules hold:[10]

$$
\begin{array}{lll}
\mathscr{D} = \mathscr{D}_{\mathscr{M}} & \mathscr{R}(\mathscr{Conj}^3) = \text{Conj} & \mathscr{R}(\mathscr{Inv}^2) = \text{Inv} \\
\mathscr{R}(F^j_i) = \mathscr{G}(\mathscr{I}(F^j_i)) & \mathscr{R}(\mathscr{Neg}^2) = \text{Neg} & \mathscr{R}(\mathscr{Conv}^2) = \text{Conv} \\
\mathscr{R}(\bar{F}^j_i) = \mathscr{I}(F^j_i) & \mathscr{R}(\mathscr{Exist}^2) = \text{Exist} & \mathscr{R}(\mathscr{Ref}^2) = \text{Ref} \\
 & \mathscr{R}(\mathscr{Exp}^2) = \text{Exp} & \mathscr{R}(\mathscr{Pred}^3) = \bigcup_{k \geq 0} \text{Pred}_k.
\end{array}
$$

An admissible model $\langle \mathscr{D}, \mathscr{R} \rangle$ for L is type 1 (type 2) if and only if it is associated in this way with an L_ω-model $\langle \mathscr{M}, \mathscr{I} \rangle$ in which the model structure $\mathscr{M}$ is type 1 (type 2). Relative to the admissible models for L one may give a standard Tarskian definition of truth for L. Finally, relative to the notion of truth for L, the appropriate notions of validity may then be defined: a formula of L is valid_1 (valid_2) if and only if it is true in all admissible type 1 (type 2) models for L.

Using L's distinguished logical predicates we may contextually define certain functional constants 'conj', 'neg', 'exist', 'exp', 'inv', 'conv', 'ref', 'pred_0', 'pred_1', . . . , 'pred_k', ... (The fact that $\mathscr{R}(\mathscr{P}red^3)$ partitions into the predication operations Pred_0, Pred_1, . . . is what makes the functional constants 'pred_k' definable in L: $\text{pred}_k(x, y) = z$ iff_{df} $\mathscr{P}red^3(x, y, z)$ and z's degree is greater than the degree of x by $k-1$.) As long as L is interpreted in the admissible ways, these contextual definitions will always pick out the intended fundamental logical operations Conj, Neg, ... Now let $[A]_\alpha$ be any intensional abstract of L_ω. Recall the inductive definition of the denotation function given in §14. Let the definition of $\text{D}_{\mathscr{I}\mathscr{A}\mathscr{M}}(`[A]_\alpha')$ be written out fully so that no occurrences of $\text{D}_{\mathscr{I}\mathscr{A}\mathscr{M}}$ remain in the definiens. The resulting expression θ consists of (1) quotation names of predicates and variables of L_ω, (2) 'Conj', 'Neg', 'Exist', 'Exp', 'Inv', 'Conv', 'Ref', 'Pred_0', 'Pred_1', . . ., and (3) '$\mathscr{I}$', '$\mathscr{A}$', commas, and parentheses. Make the following changes in θ: (1) replace each quotation name of the predicate F_i^j with the associated name $\bar{F}_i^j$ and each quotation name of a variable with the variable itself; (2) replace 'Conj', 'Neg', . . . with the associated functional constant 'conj', 'neg', . . . which was contextually defined in L; (3) delete '$\mathscr{I}$' and '$\mathscr{A}$' and all associated occurrences of parentheses. The resulting complex expression θ^* is one of the function-*cum*-argument terms that were defined contextually in L. The L_ω-term $[A]_\alpha$ is then defined in L as follows: $[A]_\alpha =_{\text{df}} \theta^*$.

Some elementary examples should help to illustrate how this definition works. Consider the L_ω-term $[F_1^1 x]^x$. The result of expanding the definition of $\text{D}_{\mathscr{I}\mathscr{A}\mathscr{M}}(`[F_1^1 x]^x')$ is $\text{Pred}_0(\mathscr{I}(`F_1^1'), \mathscr{A}(`x'))$. Then after the steps (1)–(3) above, the result is $\text{pred}_0(\bar{F}_1^1, x)$. So, one gets $[F_1^1 x]^x =_{\text{df}} \text{pred}_0(\bar{F}_1^1, x)$, i.e., the proposition that $F_1^1 x$ $=_{\text{df}}$ the proposition that is the absolute predication of F_1^1-ness of x. For a second example, consider the L_ω-term $[(\exists x)(F_1^1 x \;\&\; F_2^1 x)]$. The definition of $\text{D}_{\mathscr{I}\mathscr{A}\mathscr{M}}(`[(\exists x)(F_1^1 x \;\&\; F_2^1 x)]')$ is $\text{Exist}(\text{Conj}(\mathscr{I}(`F_1^1'), \mathscr{I}(`F_2^1')))$, which, after application of the steps (1)–(3), yields $\text{exist}(\text{conj}(\bar{F}_1^1, \bar{F}_2^1))$. So, $[(\exists x)(F_1^1 x \;\&\; F_2^1 x)] =_{\text{df}} \text{exist}(\text{conj}(\bar{F}_1^1, \bar{F}_2^1))$, i.e., the proposition that something is both F_1^1 and $F_2^1 =_{\text{df}}$ the proposition that is the existential generalization of the conjunction of F_1^1-ness and F_2^1-ness.

Given the intended semantics for L, the adequacy of the proposed extensional definition of intensional abstraction is confirmed by the following little theorem, which has a straightforward inductive proof:

> Let $\langle \mathscr{M}, \mathscr{I} \rangle$ be an L_ω-model; $\langle \mathscr{D}, \mathscr{R} \rangle$, the associated admissible L-model; $\mathscr{A}$, an assignment; $[A]_\alpha$, an L_ω-term, and θ^*, the translation of $[A]_\alpha$ into L. Then, $[A]_\alpha$ denotes, relative to $\langle \mathscr{M}, \mathscr{I} \rangle$ and $\mathscr{A}$, the same entity as θ^* denotes, relative to $\langle \mathscr{D}, \mathscr{R} \rangle$ and $\mathscr{A}$.[11]

What then of the truth of the thesis of extensionality? This thesis is of course a philosophical thesis, as is, for example, Church's thesis on effective computability. As such, it cannot be proven or disproven. But the prospect of its truth has looked dim to most people recently since even its technical feasibility has seemed out of reach. What I have just shown is that there are no technical barriers to its truth. Its truth turns instead on an ongoing methodological conflict in logical theory.

In closing, I should like to emphasize that the intensional ontology of PRPs is, ironically, what makes possible the defense of the thesis of extensionality. The moral is that those who wish to be extensionalists in logic may be so, but only if they are intensionalists in ontology. Short of artificial limitations on the natural domain of logic, this conclusion seems unavoidable.

38. Semantics

Semantics is the theory concerning the fundamental relations between words and things. Up to now I have been making free and uncritical use of a style of semantics that was developed by Tarski. In Tarskian semantics one defines what it takes for a sentence in a language to be true relative to a model. This puts one in a position to define what it takes for a sentence in a language to be valid. Since validity is what interests logicians, Tarskian semantics often proves quite useful in logic. Despite this, Tarskian semantics neglects meaning, as if truth in language were autonomous. This seems wrong, and it is time to address this issue.

Under the leadership of Donald Davidson,[12] many philosophers and linguists propose to identify the theory of meaning with the theory of truth—or, alternatively, to eliminate entirely the theory of meaning in favor of the theory of truth. The central problem with this approach is that it seems unable to characterize basic facts about what words mean. For example, it seems unable to give a satisfactory explanation of how '$(\forall x)(Ax \equiv Bx)$' can be both true and different in meaning from '$(\forall x)(Ax \equiv Ax)$'. For '$(\forall x)(Ax \equiv Bx)$', if true, has the same truth conditions as '$(\forall x)(Ax \equiv Ax)$'. Learning the meaning of '$(\forall x)(Ax \equiv Bx)$'—or learning how to use '$(\forall x)(Ax \equiv Bx)$'—requires more than learning that this sentence is true if and only if $(\forall x)(Ax \equiv Ax)$.[13]

According to commonsense semantics, the theory of truth is not autonomous from the theory of meaning, unlike what Tarskian semantics would suggest. Instead, the theory of truth is a derived theory obtained from the theory of meaning:

'A' is a true sentence $\mathit{iff}_{\mathrm{df}}$ 'A' expresses a true proposition.

'A' is true if and only if what 'A' means is true. (What it takes for a proposition to be true is an antecedent question to be settled by the theory of PRPs; see §45.) Thus, once commonsense semantics is properly developed, the Tarskian theory of truth, though useful in mathematical logic, becomes inessential to the semantics for natural language.

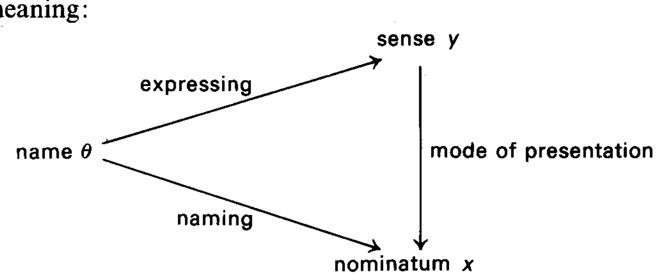
Commonsense semantics takes *meaning*, *naming*, and *expressing* to be the fundamental relations between words and things. (Referring is another relation between words and things that commonsense semantics deems important. I will take up the relation of referring later in this section.) True enough, commonsense semantics has never been given an adequate rigorous formulation. Aside from this, though, none of its critics has been able to make good his claim that the basic commonsense concepts of meaning, naming, and expressing are unsuited to their charge.[14] Because of this, and because the commonsense theory is simpler and more elegant than its competitors, it is the theory one ought to try to develop.

One should first get clear about how meaning, naming, and expressing are related to each other. Since other semantic theories may be viewed as variations on the two classical theories—Frege's and Russell's[15]—my discussion will center on the two classical theories. I will argue that Frege's theory has several faults which are avoided by Russell's theory and that Frege's theory has no advantages over Russell's.

According to Frege, every primitive symbol that is not a variable, and every well-formed string of symbols that does not contain free variables, is a name. Furthermore, according to Frege, there exist *two* fundamentally different meaning relations: expressing and naming. Every name expresses something. This kind of meaning is called a *sense* (*Sinn*). And, at least in an ideal language, every name names something. This kind of meaning is called a *nominatum* (*Bedeutung*).[16] The relations of expressing and naming satisfy the following further general principle:

> θ names x if and only if there is a y such that θ expresses y and y is a mode of presentation of x.

Here Frege takes *mode of presentation* (*Art des Gegebenseins*) to be an unanalysed extralinguistic relation akin to the relation of representation posited by representationalists. (See §42 for more on representationalism.) Thus, Frege gives us the following picture of meaning:



Frege's theory is beset with many difficulties. To begin, it offends our semantic common sense. Verbs such as 'repeats', 'runs', 'chews': is it credible that these name (or refer to) something, as Frege's theory requires? Is it any more credible that sentences, like 'The cat is on the mat', name (or refer to) something? Common sense grants that verbs and sentences do *express* something; but they have no other kind of meaning than this.[17] It is also common sense that names name something (unless they are vacuous). Frege's theory holds that, in addition, a name expresses something that determines what the name names. Although intuition is less clear here, there are provoking arguments (such as those in Kripke's

'Naming and Necessity') that names express no such thing. These offences to common sense only begin the difficulties for Frege's theory. Others arise when the theory is used in his construction of a semantics for *prima facie* intensional languages. Frege's infinite sense-hierarchy renders the resulting syntax forbiddingly complex. And the theory runs into perhaps insuperable difficulties in connection with quantifying-in and with Davidson's finite learnability requirement.

This is where Russell's theory of meaning comes in, for it is in the clear on the above counts. According to the theory, there is only one meaning relation, one fundamental relation between words and things. There is, however, a fundamental *syntactic* distinction between names, on the one hand, and verbs and sentences, on the other. This syntactic distinction is then used to define the subsidiary semantic relations of naming and expressing. An expression θ names x if and only if θ is a name that means x. ("When you said 'John', whom did you mean?") And an expression θ expresses x if and only if θ is a verb or an open or closed sentence that means x. Naming is simply meaning restricted to names, and expressing is simply meaning restricted to verbs and sentences. So Russell gives us the following simple picture:

$$\text{name, verb, or sentence } \theta \xrightarrow{\text{meaning}} x.$$

Whence, if one wishes, one can derive:

$$\text{name } \theta \xrightarrow{\text{naming}} x$$

$$\text{verb or sentence } \theta \xrightarrow{\text{expressing}} x.$$

At the same time, Russell's theory provides a semantics for intensional language. Whether one treats intensional abstraction as undefined (as in the intensional language L_ω) or as defined (as in the extensional language L of §37), the semantics will be a standard Russellian semantics. Either way, the approach runs into no difficulties over quantifying-in or over Davidson's finite learnability requirement.

So far so good. But is Russell's theory as adequate as Frege's? After all, Frege was not led to his theory only by architechtonic concerns. There are two arguments standardly given in behalf of

Frege's theory of meaning. The first is that the theory is an essential ingredient in the extensional analysis of intensional language. However, I have already given just such an analysis without Frege's infinite sense-hierarchy, so the first argument fails. The second argument is that only Frege's theory is able to explain satisfactorily certain elementary meaning phenomena epitomized by Frege's famous '$a = a$'/'$a = b$' puzzle, to which I now turn.

In 'Funktion und Begriff' and 'Über Sinn und Bedeutung' Frege poses a question that may be put as follows: how can '$a = b$' be true yet different in meaning from '$a = a$'? Frege believes that an adequate answer to this question requires invoking his two-kinds-of-meaning semantics. According to this semantics, although sense determines nominatum, nominatum does not determine sense. Thus 'a' and 'b' can differ in sense even when they have the same nominatum. For this reason, '$a = a$' and '$a = b$' also can differ in sense even when they have the same nominatum (i.e., truth value).

Russell believes that Frege's question can be adequately answered without positing two kinds of meaning. Sticking to his own one-kind-of-meaning semantics, Russell answers Frege's question by means of a two-part *syntactic* theory. First, he holds that, if '$a = b$' is true but different in meaning from '$a = a$', then 'a' or 'b' is an overt or covert definite description or extensional abstract. Secondly, Russell holds that definite descriptions and extensional abstracts are incomplete symbols[18]—definite descriptions being analysed with the theory from 'On Denoting' and extensional abstracts, with the theory from *Principia Mathematica*.[19] The effect of this procedure is to shift the weight of the explanation away from expressions that would seem to be subjects; Russell instead places the weight of the explanation on underlying predicates and formulas. The strategy succeeds because the only kind of meaning that a predicate or a formula can in any natural sense be said to have is the intensional entity that it expresses. The fact that such intensional entities can be equivalent without being identical is what makes it possible for the predicates and formulas to be equivalent without being identical in meaning. This, in turn, is what makes it possible for the initial '$a = a$'/'$a = b$' sentences to be true without being identical in meaning. (In what follows I will focus on definite descriptions since, if Russell's answer works for them, then it also works for extensional abstracts.)

Over the years there has been a wide range of doubts about the adequacy of Russell's answer to Frege's question. Let me begin by considering the doubts aimed at the second part of Russell's theory. These doubts fall into two main kinds.

First, they arise in connection with the paradox of analysis. (For example, if it is possible for someone to believe that G(the F) and yet not believe that $(\exists v)((Fu \equiv_u v = u) \& Gv)$, would this show that Russell's analysis of definite descriptions is in error?) These doubts, however, arise for analyses in general and are not a special problem for Russell's analysis of definite descriptions. Any adequate resolution of the paradox of analysis should handle doubts about the special case of Russell's analysis. These doubts can be allayed, for example, by an adaptation of the resolution offered in §20.

Secondly, there are doubts, prompted by Strawson's 'On Referring' and Donnellan's 'Reference and Definite Descriptions', concerning Russell's view that definite descriptions are semantically incomplete symbols. Yet on this matter there is a forceful defense of Russell based on the methods developed by Paul Grice in 'Logic and Conversation' and 'Definite Descriptions in Russell and in the Vernacular'. On the picture that emerges, although definite descriptions usually do have a reference, referring, unlike naming, is a *pragmatic* relation not a *semantic* relation. Thus, definite descriptions, while pragmatically complete symbols (they typically refer in conversational contexts), are semantically incomplete: their being co-referential in conversational context does not make them alike in any kind of genuine semantic meaning.[20]

I now come to the doubts aimed at the first part of Russell's answer to Frege's question, i.e., Russell's theory that, if '$a = b$' is true yet different in meaning from '$a = a$', then 'a' or 'b' is an overt or covert definite description (or extensional abstract). Two main doubts arise here. The first springs from the Mill-Kripke doctrine that ordinary names have no descriptive content (see Kripke, 'Naming and Necessity'). But Fregean semantics, not Russellian semantics, is this doubt's proper target. Indeed, the Mill-Kripke doctrine is a straightforward consequence of Russellian semantics when that theory is conjoined with the syntactic theory that ordinary names are genuine names. (An analogous doubt concerns indexicals. Throughout this section I omit discussion of indexicals since what I say about Russell on ordinary names applies analogously to Russell on indexicals. Indexicals will be explicitly discussed in the following section.)

Finally, there is the doubt raised by Alonzo Church,[21] the most prominent American advocate of Frege's theory. The thrust of Church's doubt seems to be this. Let it be granted that any isolated instance of Frege's puzzle can be explained in a Russellian framework by bringing to the surface occurrences of covert definite descriptions. Nonetheless, such a procedure must use new constants, among which are often new predicates. Nothing in Russell's theory guarantees that such new constants themselves will not be responsible for further instances of Frege's puzzle. Now although such further instances of Frege's puzzle might, in turn, be explained by still further use of Russell's methods, this process obviously should not go on forever. But nothing in Russell's theory insures that it will not. Frege's approach, in contrast, runs into no comparable difficulty.

Let me take a moment to show that Church's doubt is unfounded. The point of Russell's theory (i.e., that overt or covert definite descriptions are responsible for all instances of Frege's puzzle) is that it permits him to shift the weight of the explanation onto the predicates and formulas embedded in those descriptions. Church believes that in higher-order languages (such as those fashioned after the language of *Principia Mathematica*) these predicates and formulas could generate new instances of Frege's puzzle. The reason is that in higher-order languages predicates and formulas are linguistic subjects (see §23), making strings such as '$F = F$', '$F = G$', '$A = A$', and '$A = B$' well-formed. It does not follow from this, of course, that Church's doubt applies to first-order languages with Russellian semantics. Though Russell's procedure shifts the weight onto predicates and formulas, this can generate no such instances of Frege's puzzle since predicates and formulas are just not linguistic subjects in a first-order setting. So Church's doubt that Russell's theory can handle all instances of Frege's '$a = b$'/'$a = a$' puzzle all but evaporates when one properly distinguishes Russell's theory of meaning from the combined theory consisting of Russell's theory of meaning and the *Principia Mathematica* theory that logical syntax is at bottom higher-order. Indeed, I argued (§§10, 22–6) that the higher-order approach is defective on several counts; logical syntax is at bottom first-order.

All that could now keep Church's doubt alive is the worry that there are other kinds of *prima facie* linguistic subjects (besides spurious higher-order "names") that render Russell's theory of meaning less adequate than Frege's. But what could these conceivably be? Concerning intensional abstracts,[22] extensional abstracts, definite descriptions, and ordinary functional constants,[23] it is clear that the Russellian theory is at least as adequate as Frege's. So the question comes down to the issue of ordinary names (and indexicals). But it is easy to show that in its treatment of these expressions, Russell's theory is again at least as adequate. There are two relevant theories on the content of ordinary names, Frege's and Mill's. According to Frege's theory, associated with each ordinary name is a descriptive content that serves to determine its nominatum; according to Mill's theory, ordinary names lack such a content. I will consider each of these theories in turn.

Suppose, with Frege, that each ordinary name 'a' has an associated descriptive content. Then, according to Russell's syntactic theory, 'a' is not a genuine name; genuine names do not express anything. So according to Russell, 'a' should be treated as a disguised definite description: $a =_{\mathrm{df}} (\iota v_i)F^1_k(v_i)$, where the predicate F^1_k is interpreted so as to express the property Fregeans would associate with 'a'. (Conventions governing scope and the introduction of the new variable v_i are naturally in force.) This familiar maneuver enables the Russellian to handle ordinary names if they do have associated descriptive content. On the other hand, suppose, with Mill, that they do not. In this case, ordinary names may be simply treated as special undefined singular terms rather akin to variables with fixed assignments. Now since on this theory ordinary names 'a' and 'b' have no descriptive content, it is not possible for '$a = b$' to be true yet different in semantic meaning from '$a = a$'. So no genuine Fregean puzzles could arise.[24] Thus, whether or not ordinary names have associated descriptive contents, a Russellian semantics for them is as adequate as a Fregean semantics once the proper first-order logical syntax is identified. Russell's theory of meaning, therefore, is at least as adequate as Frege's. This shows that Church's doubt about Russell's theory of meaning is unfounded.

Since the Russellian account of the '$a = a$'/'$a = b$' puzzles is at least as adequate as Frege's and since Russell's theory of meaning is superior to Frege's theory on the several other counts reviewed above, Russell's is the better theory. If Frege's theory on the descriptive content of names is right, then the Russellian will treat all names as abbreviations for contextually defined definite descriptions. If one does this in a first-order extensional setting, the resulting language will contain no names. Using the algebraic apparatus of §§13–14, one can define the Russellian meaning function $\mathrm{M}_{\mathscr{I}\mathscr{M}}$ for this language simply as follows:

$$\mathrm{M}_{\mathscr{I}\mathscr{M}}(F_i^j) = \mathscr{I}(F_i^j)$$
$$\mathrm{M}_{\mathscr{I}\mathscr{M}}(A) = \mathrm{D}_{\mathscr{I}\mathscr{A}\mathscr{M}}([A]_{v_1 \ldots v_j})$$

where $\mathscr{M}$ is an algebraic model structure, $\mathscr{I}$ is an interpretation, $\mathscr{A}$ is an assignment, and $v_1, \ldots, v_j$ are the free variables of A in order of their first free occurrences.[25] One may then use $\mathrm{M}_{\mathscr{I}\mathscr{M}}$ to define the Fregean sense function $\mathrm{S}_{\mathscr{I}\mathscr{M}}$ and the Fregean reference function $\mathrm{R}_{\mathscr{I}\mathscr{M}}$:

$$\mathrm{S}_{\mathscr{I}\mathscr{M}}(\theta) = \mathrm{M}_{\mathscr{I}\mathscr{M}}(\theta)$$
$$\mathrm{R}_{\mathscr{I}\mathscr{M}}(\theta) = \mathscr{G}(\mathrm{M}_{\mathscr{I}\mathscr{M}}(\theta))$$

where θ is any predicate or formula.[26] If, on the other hand, Mill's theory of names is correct, then one is to use an analogous procedure except that names are given a meaning in much the same way that one might give variables fixed assignments. In either case, the definability of the Fregean sense and reference functions in terms of the Russellian meaning function shows that Fregean semantics provides no semantic information not already provided by the essentially simpler Russellian semantics.

What makes this Russellian semantics viable is the intensional ontology of PRPs. It is natural to wonder, then, which of the two traditional conceptions of PRPs pertains to the semantics for natural language. Are the meanings of natural language formulas PRPs of conception 1 or conception 2? The work of Paul Grice makes this question easy to answer.[27] Grice is able to define what it is for a speaker to mean a given propostion by performing an intentional action. The definition is given in terms of the intentions with which the speaker performs the action, including in particular his intentions to get his hearers to believe the proposition on the basis of a certain preferred inference route. Using the intentionalist analysis as a first step, Grice is then able to analyse how a sentence or a word comes to mean what it does for a community of speakers. That is, Grice is able to analyse how through their intentional activity a community of speakers comes to invoke an abstract semantical relation (such as the relation $\mathrm{M}_{\mathscr{I}\mathscr{M}}$ that I just characterized) as the meaning relation for the language they actually speak. In order for a given pure abstract semantics to become the semantics for an actual natural language, the speakers of that language must stand in certain complex intentional relations to that pure abstract semantics and, thence, to the meanings isolated by it. Intentionality therefore is the link-up between pure abstract semantics and the semantics for a spoken natural language. Now given this intentionalist analysis, the type of propositions that come to be the meanings of sentences in natural languages must of necessity be the type that are typically intended, believed, remembered, etc. And the type of propositions that typically serve this function are precisely conception 2 propositions. Therefore, it is conception 2 PRPs that pertain to the semantics for natural language.

When I asserted at the beginning of this section that commonsense semantics had never been given an adequate rigorous formulation, I had two things in mind. First, previous attempts to rigorously formalize commonsense semantics have mistakenly formalized Frege's semantics, not Russell's. Yet Russell's is the commonsense theory. Secondly, previous attempts to formalize commonsense semantics have utilized the possible-worlds technique, which is based on conception 1, not conception 2. Yet conception 2 is the one relevant to the commonsense semantics for natural language. These deficiencies have now been remedied.

39. Pragmatics

I now shift to four problems in the logic for intentional matters: (1) *prima facie* substitutivity failures involving co-denoting names, (2) *prima facie* substitutivity failures involving co-denoting indexicals, (3) Mates' puzzle (which concerns *prima facie* substitutivity failures involving synonymous predicates and formulas),[28] and (4) Geach's problem of intentional identity (which concerns *prima facie* quantification over non-actual possibilia).[29] Recently, there has been much provocative investigation of these issues;[30] to attempt definitive solutions of them at this point would be premature. I aim only to explore some candidate solutions. My purpose in doing so is to convey the explanatory power of the theory of PRPs and, in turn, to suggest that this theory provides a general framework within which these four problems can be solved eventually. If I am right, solving these problems is best viewed as a matter of fine-tuning among the applications of the theory, fine-tuning that does not threaten the theory's underlying Platonistic character.

Let us consider names and indexicals first. I have shown that Russell's theory of meaning is compatible with both classical theories on the content of ordinary names (and indexicals), Frege's theory and Mill's theory. It is not the job of the theory of PRPs to decide the Frege/Mill controversy; all that matters is that the theory be adaptable to whichever doctrine is correct. Assuring oneself of its adaptability is relatively straightforward if Frege's doctrine is correct; for then the substitutivity problems involving such singular terms either submit to traditional solutions (Frege's or Russell's) or collapse into instances of Mates' puzzle, which can be dealt with in the manner suggested later in this section. If, on the other hand, Mill's theory is right, then the substitutivity problems involving names (and indexicals) would call for a non-traditional approach. In a moment I will propose such an approach. If it is successful, then the adequacy of the framework of PRPs is guaranteed, no matter which theory on the content of ordinary names (and indexicals) is correct.

Suppose for the sake of discussion that Mill's theory is correct. Let us compare our three substitutivity problems—Mates' puzzle and the two involving primitive singular terms. Recall from §18 that, unlike the paradox of analysis, whose source is ignorance of definitions of concepts, Mates' puzzle originates in some form of linguistic (or historical or social) ignorance. For example, consider someone x who knows what the verb 'chew' expresses but is ignorant that the verb 'masticates' expresses the same thing. In this situation it might be natural to affirm

(1) x does not know that whatever masticates chews.

while denying

(2) x does not know that whatever chews chews.

And this is so despite the fact that the literal Russellian meanings of sentences (1) and (2) are the same.[31] Now according to Mill's theory, ordinary names have no descriptive content; they only name. If this is right, observe how similar Mates' problem is to the substitutivity problem for co-denoting ordinary common names in intentional sentences. The types of ignorance responsible for these problems appear analogous. E.g., ignorance that whatever masticates chews would seem quite on a par with ignorance that pot is marijuana, that consumption is tuberculosis, that lorries are trucks, that pumas are cougars, that filbert is hazelnut, etc. And if this is so for the substitutivity problem for ordinary common names ('pot', 'consumption', 'lorry', 'puma', etc.), what reason could there be for thinking that the substitutivity problem for ordinary proper names is different? It would seem that one has no choice but to treat ignorance that Scott is Sir Walter, that Tully is Cicero, etc. on a par with ignorance that whatever masticates chews. (Of course, one must be careful to distinguish genuine names from descriptions masquerading as names. For example, 'H_2O' is no name but a description short for something like 'the compound whose molecules bind together 2 hydrogen and 1 oxygen atom'.[32]) In fact, if both Mill's doctrine and Russellian semantics are right, one can all but prove that the substitutivity problem for all ordinary names—proper as well as common—is on a par with Mates' substitutivity puzzle. The argument goes as follows. The only difference between the two problems lies in the fact that the former problem involves ordinary names whereas the latter problem involves verbs. However, given Russell's theory of meaning, there is only one kind of meaning, and genuine names stand in this same meaning relation to their meanings as do verbs to theirs. Further, according to Mill's theory, ordinary common and proper names are genuine names. It follows that ordinary names stand in the very same meaning relation to their meanings as do verbs to theirs. Hence, there can be no relevant *semantical* difference between ordinary names and verbs. Consequently, there can be no relevant semantical difference between the substitutivity problem for ordinary names and Mates' puzzle. The only difference between the two problems therefore is *syntactical*: one concerns ordinary names and the other concerns verbs. Indeed, the two problems are really species of the same general problem: the problem of the substitutivity of synonyms.

Much the same thing holds for the substitutivity problem involving demonstratives. For example, suppose that a given speaker utters 'this' while pointing directly to a certain object in plain view and 'that' while pointing through some complex optical apparatus to what turns out to be the same object. In this situation it might be natural for this speaker to deny

(3) I believe that this = that.

while affirming

(4) I believe that this = this.

Yet from a semantical point of view demonstratives have elusive descriptive content. If none can be found, it would seem that, like the substitutivity problem for Millian names, the substitutivity problem for demonstratives is just a syntactical variant of Mates' puzzle and, hence, that all three substitutivity problems are species of the same general problem. If this is so, the three problems call for a unified solution.

I will now sketch a metaphysical theory of belief which should bring one closer to a solution. (This theory easily generalizes to the other problematic intentional relations.) Consider a normal conversational context in which it would be appropriate to utter 'x believes that A'. Typically, the believer x must satisfy *two* conditions. First, he must be what I call *cognitively committed* to the proposition that is the meaning of the embedded sentence A; he need not be acquainted (in the traditional epistemological sense[33]) with this proposition, however. Secondly, x must be convinced of a proposition (often, but not always, a different one) with which he is acquainted. (I call this a *conviction in acquaintance* or $conviction_{acq}$ for short.) It is in virtue of this conviction in acquaintance that x is cognitively committed to the proposition literally expressed by A. Now a person is cognitively committed to all those propositions of which he is convinced in acquaintance, but the converse does not hold. This is crucial. In daily social intercourse, when we rely on another believer for information about the world, we focus on those of his cognitive commitments of which he is not convinced in acquaintance. For these cognitive commitments deal directly with the objects in the world and ignore the individual modes of epistemic access to those objects, which usually are of no special interest and which also are difficult for us to discover. However, when we wish to explain a believer's actions, we ultimately look to his convictions in acquaintance. For these are what figure in his deliberations about what to do; he typically is not even immediately aware of many of his associated cognitive commitments.[34]

Let us apply this metaphysical scheme to some problematic cases. For example, suppose that I have severe amnesia and that I sincerely utter the sentence 'I believe that I am not George Bealer'. In this situation I would be cognitively committed to the proposition literally expressed in the context by the embedded sentence 'I am not George Bealer'; i.e., I would be cognitively committed to the necessarily false proposition $[x \neq x]^x$, where I am x. But I would be convinced in acquaintance of a different proposition, perhaps $[x \neq$ "George Bealer", as he is called$]^x$, where I am x. Or even better, simply $[x \neq$ "George Bealer"$]^x$, where the quotation marks are "scare quotes". (Without taking a position on the proper analysis of scare quotes, one may be confident that it involves some form of metalinguistic allusion.) It is in virtue of such a conviction$_{acq}$ that I would be cognitively committed to the necessary falsehood $[x \neq x]^x$. Of course, I would not be convinced$_{acq}$ of $[x \neq x]^x$; that would take gross irrationality whereas I only suffer from amnesia.[35]

Or consider the example given in Kripke's 'A Puzzle About Belief' (pp. 254 ff). A Frenchman Pierre, who has only seen photos of London, sincerely utters the sentence 'Londres est jolie'. Later Pierre moves to London. After learning English, he sincerely utters the sentence 'London is not pretty', but he does so without knowing that 'Londres' and 'London' name the same city. Indeed, he still would sincerely utter 'Londres est jolie'. Using the above metaphysical scheme one would say that Pierre is cognitively committed to both the proposition that London is pretty and its negation, i.e., to both $[u \Delta v]^{uv}$ and $[u \not\Delta v]^{uv}$, where $u =$ London and $v =$ prettiness. He is cognitively committed to the first proposition in virtue of being convinced in acquaintance of a further proposition—e.g., the proposition that "Londres", as it is called, is pretty. And he is cognitively committed to the negation of the first proposition in virtue of being convinced in acquaintance of still another proposition—e.g., the proposition that "London", as it is called, is not pretty. (As before, I am using scare quotes.) Pierre's logical acumen is not under suspicion, however, for logical acumen gets tested only against those cognitive commitments that a person is immediately aware he has, such as those that are convictions$_{acq}$. And the two propositions of which Pierre is convinced in acquaintance are logically independent.

Before going further, I should meet a possible worry.[36] Suppose in the above story that Pierre is a rather primitive fellow who has never articulated any of the metalinguistic concepts belonging to linguistic theory. This would show, so the worry supposes, that Pierre could not be convinced in acquaintance of the propositions ["Londres", as it is called, is pretty], ["London", as it is called, is not pretty], or anything like that, for such propositions appeal to metalinguistic concepts. This line of argument, however, overlooks the paradox of analysis. Consider an analogy. Suppose that Pierre has never articulated a geometric theory and that he is ignorant of how to define (analyse) what a circle is. Still, if Pierre is convinced$_{\mathrm{acq}}$ that there are circles, then one can truly say 'Pierre is convinced$_{\mathrm{acq}}$ that there are loci of points in the same plane equidistant from a common point'. And this is so even though Pierre might be brought up short by the utterance. Using the apparatus developed for resolving the paradox of analysis (see §20), one can easily explain what is going on. There are *two* different propositions denoted by the 'that'-clause 'that there are loci of points...':

$[(\exists y)y \text{ is a locus of points} \ldots]$
$[(\exists y)\underline{y \text{ is a locus of points} \ldots}]$.

Pierre has no conviction$_{\mathrm{acq}}$ concerning the former proposition, for it involves the analysed concept of circularity. However, he does have a conviction$_{\mathrm{acq}}$ concerning the latter proposition; this proposition (which is just the proposition $[(\exists y)\underline{y \text{ is a circle}}]$) does not involve an articulated definition (analysis) of what a circle is. Now on analogy, the fact that Pierre lacks an articulated linguistic theory provides no evidence whatsoever that he lacks convictions$_{\mathrm{acq}}$ concerning metalinguistic propositions. For Pierre might simply be ignorant of how to define (analyse) the metalinguistic propositions with which he is acquainted. (This sort of ignorance is surely pervasive.) Of course, care must always be taken in the formal statement of unanalysed convictions$_{\mathrm{acq}}$. In Pierre's case, for example, we might want to represent his convictions$_{\mathrm{acq}}$ with something like the following:

["Londres" as it is called is pretty],
["London" as it is called is not pretty].[37]

With this worry allayed, let us now consider how intentional verbs, e.g., 'believe', behave in natural language. There are two idealized positions on this question. The first is that 'believe' literally expresses a concept that applies only to what one is convinced of in acquaintance.[38] The second is that it expresses a concept that applies only to one's cognitive commitments. Now I am rather persuaded that in everyday speech each of these concepts operates at least pragmatically, if not semantically. Indeed, I would not be surprised if in ordinary language 'believe' expresses both concepts, or even some composite of them. At the same time, I am persuaded that with suitable maneuvering either of these two idealized positions can be made to fit all the linguistic data, and that this can be done within the general framework provided by the theory of PRPs.[39] I need not decide here which position is best. For illustrative purposes, though, I will sketch a version of the second position.

On this version 'believes' literally expresses a concept that applies just to what one is cognitively committed to. Yet when a speaker sincerely utters 'x believes that A' in conversation, he typically does two things. First, he asserts that x is cognitively committed to $[A]$—i.e., to the proposition that is the meaning of the sentence A. Secondly, he presupposes that there is some conviction in acquaintance that is the vehicle of x's cognitive commitment to $[A]$. (Such presuppositions arise through mechanisms of the sort isolated by Paul Grice in his conversational pragmatics; see his 'Logic and Conversation' and 'Definite Descriptions in Russell and in the Vernacular'.) In most contexts the identity of this conviction in acquaintance is irrelevant, and the speaker leaves it indefinite. But in some contexts its identity becomes of interest, and the speaker intends to indicate at least roughly what it is. In these contexts the utterance of 'x believes that A' carries a conversational implicature that x has a conviction in acquaintance which falls within the indicated range.

Consider an example to see how this works. If I sincerely utter 'x believes that most pot is grown in Colombia', I assert that x is cognitively committed to the proposition that most pot is grown in Colombia. I also presuppose that x has this cognitive commitment in virtue of some conviction in acquaintance. But just which one, since it is of little importance, I leave indefinite. If I sincerely utter 'x believes that most marijuana is grown in Colombia', my assertion is the same as before, and so is my presupposition; the substitution of 'marijuana' for 'pot' changes neither of these. Next suppose that I sincerely utter 'x believes that pot = pot'. Here, I assert that x is cognitively committed to the proposition that pot = pot, and I presuppose that x has this cognitive commitment in virtue of some conviction in acquaintance. As in the other cases the identity of the conviction$_{\text{acq}}$ is of no relevance, so I leave it indefinite. But suppose I sincerely utter 'x believes that pot = marijuana'. Although what I assert about x's cognitive commitment remains unchanged, the conversational pragmatics becomes different. Since 'pot' and 'marijuana' are synonymous, utterances of 'x believes that pot = pot' and 'x believes that pot = marijuana' make the same assertion about x's cognitive commitment. But since the former sentence provides such a simple way to make this assertion, an utterance of the latter sentence signals that there is some special reason for not using the simpler form. Therefore, such an utterance must conversationally implicate something beyond what the sentence expresses semantically. The ripest candidate for this conversational implicature would be something that concerns the conviction in acquaintance underlying x's cognitive commitment. The conversationally salient feature of the sentence is its lexical complexity. This suggests that the implicature concerning x's conviction in acquaintance has something to do with the lexical items themselves. A thing's being called by a certain name is an obvious mode of epistemic access to that thing, so the conversational implicature would often be simply that x's conviction in acquaintance is some metalinguistic proposition such as the proposition that "pot" = "marijuana" (scare quotes again) or something like that. (Of course, depending on the context, non-linguistic modes of epistemic access are often conversationally more relevant than linguistic modes, and the implicature is affected accordingly.) So it is that in this case the substitution of 'marijuana' for 'pot' sharply affects the conversational pragmatics. Utterances of 'x believes that pot = pot' and 'x believes that pot = marijuana' conversationally say quite different things, and this explains why their truth values can differ.

It is no coincidence that in the last case the substitution of 'marijuana' for 'pot' sharply affects conversational pragmatics. True, substitution of co-denoting names—and co-denoting demonstratives and synonymous predicates—usually does not have this sort of pragmatic effect, for in most conversational contexts the interest in an utterance of a belief sentence lies in the literally expressed cognitive commitment. However, in some situations contextual signals shift interest to the conviction in acquaintance underlying the literally expressed cognitive commitment. Here the particular way in which the cognitive commitment is expressed becomes relevant, for it provides contextual cues about the believer's epistemic access to his cognitive commitment and, hence, about what his conviction in acquaintance is. Substitution can thus affect whether an utterance of a belief sentence concerns only the literally expressed cognitive commitment or whether it concerns in addition the underlying conviction in acquaintance, and it can also affect the identity of such conversationally implicated convictions in acquaintance. Substitutions seem invalid in exactly those cases in which they have one of these pragmatic effects, and these are precisely the cases of *prima facie* substitutivity failures I set out in this section to explain. If this is correct, then the three types of substitutivity puzzles—and their explanations—belong to conversational pragmatics, not semantics.[40]

Now the point of the above exercise has not been to give a definitive solution to these problems. Rather, it has been to provide evidence that their solution can be carried out within the general framework provided by the theory of PRPs. I will now attempt a similar exercise for our last puzzle in the logic for intentional matters, namely, Geach's problem of intentional identity.

Geach tells the following little story. A reporter visits a region where there is a rumor that a witch is on the loose. Although the reporter does not himself believe in witches, he makes the following report about the beliefs of two locals:

> Hob believes that the witch blighted the sheep, and
> Nob believes that she killed the cow.

Geach asks us to assume that there is something true in what the reporter has said; the problem is to characterize what it is. (We need not assume that the uttered sentence is literally true; I doubt that it is.) The problem is one of *intentional* identity because the reporter's statement would seem to imply that Hob and Nob have a belief about the same witch even though no witch exists except, as it were, in the minds of Hob and Nob. But what on earth does this mean?

Someone might try to solve this problem by augmenting the ontology of PRPs with non-actual possibilia, non-existent subsistents, or intentional inexistents. Doing so might permit one to represent the reporter's statement as being about some non-actual, non-existent, or inexistent witch. Though nothing in the theory of PRPs rules it out formally, this strategy taxes the principle of ontological economy, and it does violence to commonsense realism, a view that, if possible, one ought to hold on to.[41]

Where else might one look for a solution to Geach's problem? Evidently the only alternative is to analyse the reporter's statement in such a way that the problematic beliefs of Hob and Nob are propositions that involve descriptive witch-concepts that bear some suitable relation to one another. The simplest example of such a descriptivist analysis is this:

> For some descriptive concept *w*, Hob believes that the *w* witch blighted the sheep and Nob believes that the *w* witch killed the cow.

However, Geach argues against this analysis. His argument is that, though Hob and Nob might conceivably share a descriptive witch-concept, the truth of the reporter's statement does not require that they do.

One might try to meet Geach's argument by formulating a more sophisticated descriptivist analysis. It would be preferable, though, to preserve the form of the initial simple analysis while also doing justice to Geach's intuition that Hob and Nob need not share any descriptive witch-concept. The theory of belief sketched earlier in this section permits this. It is entirely possible that there is no descriptive concept *w* such that Hob is convinced in acquaintance that the *w* witch blighted the sheep and Nob is convinced in acquaintance that the *w* witch killed the cow. This seems to be the basis of Geach's intuition. However, this in no way prohibits Hob and Nob from having other convictions in acquaintance, ones that would give them precisely the sort of cognitive commitments that would validate the simple descriptivist analysis. If Hob and Nob were to have such convictions in acquaintance, Geach's problem about them would be solved.

To make this solution plausible, I will describe a situation in which Hob and Nob would have the relevant sort of convictions in acquaintance and cognitive commitments. First, Hob is convinced in acquaintance that the witch who is at the root of the reference tree to which he is presently a party blighted the sheep. This conviction in acquaintance gives Hob a cognitive commitment to the proposition that the witch who is at the root of reference tree *R* blighted the sheep. Secondly, Nob is convinced in acquaintance that the witch who is at the root of the reference tree to which he is presently a party killed the cow. This gives Nob a cognitive commitment to the proposition that the witch who is at the root of reference tree *S* killed the cow. Finally, reference tree *R* (i.e., the one to which Hob is a party) and reference tree *S* (i.e., the one to which