DANA SCOTT

Existence and Description in Formal Logic

The problem of what to do with improper descriptive phrases has bothered logicians for a long time. There have been three major suggestions of how to treat descriptions usually associated with the names of Russell, Frege and Hilbert-Bernays. The author does not consider any of these approaches really satisfactory. In many ways Russell's idea is most attractive because of its simplicity. However, on second thought one is saddened to find that the Russellian method of elimination depends heavily on the scope of the elimination. Further, the semantical meaning of Russell's transformation is not all that clear; although it could be made quite precise. Frege's use of a null entity for the denotation of an improper description has of course an immediate semantical interpretation, but the arbitrary choice of a null entity in each domain is really not very natural. In many axiomatic theories, Euclidean geometry for example, the choice of a distinguished point is not possible or even very desirable. Bernays in [1] used Fregean descriptions with a kind of 'local' null entity carried along in the notation itself. This idea, though clear and workable, is not very elegant in the author's opinion.

It is curious that in ordinary mathematical practice having undefined function values, a situation close to using improper descriptions, does not seem to trouble people. A mathematician will often formulate conditionals of the form

if $f(x)$ exists for all $x < a$, then . . .

and will not give a moment's thought to the problem of the meaning of $f(a)$. More careful authors never use a description or a function value unless it has been previously proved that its value exists. This style led Hilbert-Bernays in [3] to the point of requiring such a proof before a formula containing the description can be considered as well formed. This suggestion is to be rejected on many grounds. As has often been pointed out, the class of well-formed formulas will hardly ever be recursive. Also the class of well-formed formulas will change upon the introduction of additional axioms. More serious is the fact that it is quite natural to employ descriptions *before* they have been proved to be proper. In axiomatic set theory in the discussion of recursive definitions, it is very tempting to give an explicit definition of the required function by means of a description and then prove a theorem of the form:

for all a in a set well-ordered by $<$,
if $f(x)$ *exists for all* $x < a$, *then* $f(a)$ *exists.*

It will then follow by transfinite induction that $f(a)$ exists for all a in the well-ordered set. Only a logician would have objections to this use of the 'exists'. It is the purpose of this paper to lay these objections to rest by presenting a formal theory of descriptions that corresponds quite faithfully to such natural modes of reasoning.

After the author had explicitly formulated his plan (December 1963), he discovered that around 1959–60 several other logicians had come to nearly the same idea: notably Hailperin and Leblanc in [2], Hintikka in [4] and [5], Rescher in [8] and Smiley in [10]. These papers have not received the attention they deserve; thus a complete exposition including a full discussion of the semantics required seems desirable. Further the author wishes to show how the idea can be applied to a theory like Quine's system of virtual classes in [7]. Quine, following Russell, employs contextual definitions which avoid giving an independent meaning to the virtual classes. The author will replace Quine's definitions by axioms and present a simple semantical interpretation for the theory. The paper will conclude with a model-theoretic discussion of eliminability of notions by contextual definitions. There is an interesting problem here that is left open.

The author is indebted to Professors Hintikka, D. Kaplan, Kreisel, Mostowski, Quine, A. Robinson, Suszko, Suppes, and Tarski who were kind enough to comment on earlier versions of this paper.

1. *Descriptions*

To simplify matters let us consider a first-order logic with just one non-logical constant: a binary predicate symbol R. The logical symbols are $\neg$ (for negation), $\rightarrow$ (for implication), $\forall$ (for universal quantification), $=$ (for equality) and I (for description). Note that I is an *inverted* capital I, which the author prefers to Russell's inverted iota. The other propositional connectives $\wedge$, $\vee$, $\leftrightarrow$, and the quantifier $\exists$ should be considered as introduced by definition, or better, the formulas involving them may be taken as abbreviations of formulas containing only the basic symbols. The individual variables are $v_0, v_1, \ldots, v_n, \ldots$. In the metalanguage x, y, z, w are metavariables ranging over the individual variables of the object language.

We define the notion of *term* and *formula* in the usual way:

(i) all variables are terms;
(ii) if α and β are terms, then $\alpha = \beta$ and $\alpha \mathsf{R} \beta$ are (atomic) formulas;
(iii) if Φ and Ψ are formulas and x is a variable, then $\neg \Phi$, $[\Phi \rightarrow \Psi]$ and $\forall x \, \Phi$ are formulas; while $\mathbf{I} \, x \, \Phi$ is a term.

The precise definition of *free* and *bound* variables need not concern us here, and it may be assumed as known.

To give a semantical interpretation of this language one first gives a structure $\langle A, R \rangle$, where A is a set (the domain of individuals) and R is a binary relation (the interpretation of the predicate symbol R). Then relative to the given structure one defines the *values* of formulas (they will be truth values) and terms (they will be objects) corresponding to the values given to the free variables. Before presenting this definition, it will be wise to consider some informal, motivating principles that have guided the choice of our precise formulation. Above all we wish to follow:

Principle 1. Bound individual variables should range only over the given domain of individuals

The author will not attempt to define what he means by 'range over' since he is sure everyone understands this statement. In case someone does not, he should wait to see the formal definition of value for the quantified formulas and for the descriptive phrases. The second principle is not so important, but the author wishes to include it with an eye to future applications:

Principle 2. The domain of individuals should be allowed to be empty

Finally, and very important for the basic idea of the paper, we have:

Principle 3. The values of terms and free variables need **not** *belong to the domain of individuals*

To see in a simple example the usefulness of Principle 3, consider the question raised by Mostowski in [6] in connection with the empty domain. Namely, it is 'clear' that the formula

$$x \mathsf{R} x \rightarrow x \mathsf{R} x$$

is valid in all domains including the empty one. Similarly the sentence

$$[x \mathsf{R} x \rightarrow x \mathsf{R} x] \rightarrow \exists y [y \mathsf{R} y \rightarrow y \mathsf{R} y]$$

is valid, because if x is given a value in the domain, then there is some value of y to satisfy the formula within the quantifier. On the other hand, the formula

$$\exists\, y[y \mathrel{\mathsf{R}} y \rightarrow y \mathrel{\mathsf{R}} y]$$

is *not* valid in the empty domain; hence, the valid formulas are not closed under the rule of modus ponens when the empty domain is included. The fallacy (or better inconvenience) here lies in allowing the second formula to be valid. The first formula is completely valid no matter what value we assign to x. The second formula will fail in the empty domain, however, if we recognize Principle 3. To have a valid formula we must modify the implication to read:

$$[x \mathrel{\mathsf{R}} x \rightarrow x \mathrel{\mathsf{R}} x] \wedge \exists\, y[x = y] \rightarrow \exists\, y[y \mathrel{\mathsf{R}} y \rightarrow y \mathrel{\mathsf{R}} y]$$

We can make this point better after the precise definition of value is given.

Before presenting the definition of value, we must still decide what to do with the improper descriptive phrases. Under the guidance of Principle 3 we are no longer required to give such terms values *within* the given domain. Indeed, it seems much better to give an improper description a value definitely *outside* of the domain, thereby emphasizing its impropriety. The way to do this is to assign to each domain A a null entity $\ast_A$ such that $\ast_A \notin A$. This is much easier than trying to make the null entity belong to A as Frege wished (especially when A is empty!). Assuming a reasonable set theory, we could let $\ast_A$ be the set of all sets belonging to A which are non-self-members. Thus $\ast_A \subseteq A$ but $\ast_A \notin A$. Assuming the so-called Axiom of Regularity we could even take $\ast_A = A$. The exact choice is quite irrelevant as long as we agree $\ast_A \notin A$. Now we are ready for the definition of value.

In the following we shall write $\mathfrak{A} = \langle A, R\rangle$ for short; while s will denote an *assignment* which is simply a function whose domain is the set of integers $N = \{0, 1, 2, \ldots\}$. For $i \in N$, $s(i)$, or simply s_i, is the value we wish to assign to the variable v_i. We define

$$s(i/a) = (s \sim \{\langle i, s_i\rangle\}) \cup \{\langle i, a\rangle\},$$

in other words $s(i/a)$ is like s except the ith value s_i has been replaced by a. We shall read

$$\models_{\mathfrak{A}} \Phi[s]$$

as: the (truth) value of the formula Φ is *true* for the assignment s relative to the structure $\mathfrak{A}$, or better, s *satisfies* Φ in $\mathfrak{A}$, or also, Φ is *true* at s in $\mathfrak{A}$. The symbol

$$\| \alpha\, [s] \|_{\mathfrak{A}}$$

is read: the (object) value of the term α for the assignment s relative to the structure $\mathfrak{A}$, or better, the value of α at s in $\mathfrak{A}$. (Maybe the use of 'in' is bad here, because in view of Principle 3 $\| \alpha [s] \|_{\mathfrak{A}} \in A$ need not be so.) The exact clauses of the recursive definition of these notions are as follows:

$$\| v_i [s] \|_{\mathfrak{A}} = s_i$$

$$\models_{\mathfrak{A}} \alpha \; \mathsf{R} \; \beta[s] \textit{ iff } \langle \| \alpha [s] \|_{\mathfrak{A}}, \| \beta [s] \|_{\mathfrak{A}} \rangle \in R$$
$$\models_{\mathfrak{A}} \alpha = \beta[s] \textit{ iff } \| \alpha [s] \|_{\mathfrak{A}} = \| \beta [s] \|_{\mathfrak{A}}$$
$$\models_{\mathfrak{A}} \neg \Phi[s] \textit{ iff not } \models_{\mathfrak{A}} \Phi[s]$$
$$\models_{\mathfrak{A}} [\Phi \rightarrow \Psi] [s] \textit{ iff if } \models_{\mathfrak{A}} \Phi[s], \textit{ then } \models_{\mathfrak{A}} \Psi[s]$$
$$\models_{\mathfrak{A}} \forall v_i \, \Phi[s] \textit{ iff for all } a \in A, \models_{\mathfrak{A}} \Phi[s(i/a)]$$

$$\| \mathbf{I} \, v_i \, \Phi[s] \|_{\mathfrak{A}} = \begin{cases} a & \textit{if } a \textit{ is the unique element of } A \textit{ such that } \models_{\mathfrak{A}} \Phi[s(i/a)]; \\ *_A & \textit{if there is no such element.} \end{cases}$$

We say that Φ is *valid* in $\mathfrak{A}$ and write $\models_{\mathfrak{A}} \Phi$ to mean that $\models_{\mathfrak{A}} \Phi[s]$ holds for all assignments s. We say that Φ is *universally valid* and write $\models \Phi$ to mean that $\models_{\mathfrak{A}} \Phi$ holds for all structures $\mathfrak{A}$.

The question now is to find an axiomatization of the universally valid formulas. Note first that these two rules are correct:

(MP) *If* $\models \Phi$ *and* $\models [\Phi \rightarrow \Psi]$, *then* $\models \Psi$.

(UG) *If* $\models [\Phi \rightarrow \Psi]$ *and* x *is not free in* Φ, *then* $\models [\Phi \rightarrow \forall x \Psi]$.

Next note that these schemata comprise only valid formulas:

(S0) Φ, *if* Φ *is a tautology,*

(S1) $\forall x [\Phi \rightarrow \Psi] \rightarrow [\forall x \Phi \rightarrow \forall x \Psi]$,

(S2) $\forall y \, \exists x [x = y]$,

(S3) $\alpha = \alpha$,

(S4) $\Phi(x/\alpha) \wedge \alpha = \beta \rightarrow \Phi(x/\beta)$,

where $\Phi(x/\alpha)$ is the result of substituting α for all the free occurrences of x in Φ rewriting bound variables if necessary.

In as much as $\exists x \Phi$ abbreviates $\neg \forall x \neg \Phi$, it is easy to see that the rule (UG) includes the rule

(EG) *If* $\models [\Psi \rightarrow \Phi]$ *and* x *is not free in* Φ, *then* $\models [\exists x \Psi \rightarrow \Phi]$.

Also using the schemata, especially (S1), we can show that

$$\models \forall x\, \Phi \wedge \exists x \Psi \rightarrow \exists x [\Phi \wedge \Psi]$$

and

$$\models \forall x [\Phi \rightarrow \Psi] \rightarrow [\exists x\, \Phi \rightarrow \exists x \Psi].$$

Using these together with (S4) we establish easily the validity of the schema

$$\text{(UI)} \qquad \forall x\, \Phi \wedge E\, x [x = \alpha] \rightarrow \Phi(x/\alpha),$$

where x is not free in the term α. This is the correct version of the law of universal instantiation which is valid not only when the domain is empty, but also when the values of terms are allowed to be outside the domain. Using (S2) we can also show

$$\models \forall y [\forall x\, \Phi \rightarrow \Phi(x/y)],$$

which some authors would take as an axiom but which is superfluous when principles of equality are available. On the other hand (S2) is practically a special case of this last schema. Replace Φ by the formula $\neg\, x = y$ obtaining

$$\models \forall y [\forall x [\neg\, x = y] \rightarrow \neg\, y = y],$$

from which we derive

$$\models \forall y [y = y \rightarrow \exists x [x = y]],$$

and then

$$\models [\forall y [y = y] \rightarrow \forall y \exists x [x = y]].$$

In view of (S3), we can now easily obtain (S2). So it is really just a matter of taste as to which schemata are chosen as the fundamental ones.

To understand better what is going on here, consider the meaning of

$$\exists x [x = \alpha]$$

under our semantical rules. When x is not free in α, then

$$\models_{\mathfrak{A}} \exists x [x = \alpha]\, [s]$$

holds if and only if $\| \alpha[s] \|_{\mathfrak{A}} \in A$. Let us call the elements of A the (properly) existing individuals (of the particular structure $\mathfrak{A}$). Then to say that $\exists x [x = \alpha]$ is true means that the value of α exists (properly). Is that not exactly what $\exists x [x = \alpha]$ ought to mean? Thus if $\forall x\, \Phi$ is true, it is not correct to conclude that $\Phi\, (x/\alpha)$ is true *unless* the value of α exists. Again, is that not quite reasonable? 'To be is to be the value of a bound variable', as Quine would say.

Turning now to the descriptive operator we have first of all this valid schema:

$$\text{(I 1)} \quad \forall\, y[y = \mathrm{I}\, x\, \Phi \leftrightarrow \forall\, x[x = y \leftrightarrow \Phi]],$$

where y is not free in Φ. In words: an existing individual is the value of a descriptive phrase if and only if it is indeed the unique individual satisfying the formula of the phrase. As a consequence of (I 1) we have at once:

$$\models \exists\, y[y = \mathrm{I}\, x\, \Phi] \leftrightarrow \exists\, y \forall\, x[x = y \leftrightarrow \Phi];$$

that is, proper phrases are the only ones whose values exist. What of improper phrases? According to our definition of value they are all given the same value $\ast_A$. Now the term $\mathrm{I}\, v_0[\neg\, v_0 = v_0]$ clearly is an improper descriptive phrase; call it $\ast$ for short. The rest of the definition of value for descriptive phrases can be expressed by the schema:

$$\text{(I 2)} \quad \neg\, \exists\, y[y = \mathrm{I}\, x\, \Phi] \rightarrow \ast = \mathrm{I}\, x\, \Phi.$$

The converse of this implication already follows from (I 1).

One important reason for insisting that improper descriptions all assume the same improper value is to have this highly useful law of extensionality:

$$\models \forall\, x[\Phi \leftrightarrow \Psi] \rightarrow \mathrm{I}\, x\, \Phi = \mathrm{I}\, x\, \Psi$$

This would not be valid if one wanted 'the golden mountain' and 'the round square' to have different values. While making unkind remarks about 'the golden mountain', Russell also rejected this law of extensionality, which this author considers an unfortunate choice. Of course, Russell was particularly interested in eliminating descriptions altogether, and we now must discuss that question.

Using (I 1) and (I 2) we can almost completely eliminate descriptions, because we have the schema of elimination:

$$\begin{aligned}\text{(IE)} \quad \Psi(y/\mathrm{I}\, x\, \Phi) \leftrightarrow\ & \exists\, y[\forall\, x[x = y \leftrightarrow \Phi] \wedge \Psi] \vee \\ & [\neg\, \exists\, y \forall\, x[x = y \leftrightarrow \Phi] \wedge \Psi(y/\ast)],\end{aligned}$$

where the variable y is not free in Φ. Several applications of (IE) will confine all occurrences of the descriptive operator to the following contexts:

$$v_i\, \mathsf{R}\, \ast,\ \ast\, \mathsf{R}\, v_i,\ \ast\, \mathsf{R}\, \ast,$$
$$v_i = \ast,\ \ast = v_i,\ \ast = \ast.$$

Now, assuming that a formula Θ has no free variables, the equality formulas can be eliminated, because the last one is true, and the first two are always false in contexts where the variable v_i is bound. To be able to eliminate descriptions completely we would have to add a new schema such as:

$$\text{(I 3)} \quad \ast = \alpha \vee \ast = \beta \to \neg\, \alpha \,\mathrm{R}\, \beta.$$

However, this schema is not valid with our present semantics. It *is* valid in those structures $\mathfrak{A} = \langle A, R \rangle$, where $R \subseteq A \times A$. This restriction that the relation of a structure should be confined to existing individuals is not at all desirable, as we shall see when we discuss Quine's virtual classes.

We could have validated (I 3) by choosing $\ast_A$ to lie outside the field of the relation R. Again this is not too desirable, because it is often felt that the valid formulas should be closed under substitution of formulas for predicate symbols. Clearly (I 3) becomes invalid when R is replaced by $=$. So for pure logic we reject (I 3). When giving axioms for a theory on the other hand, a schema like (I 3) might be very reasonable. Then in that theory complete elimination of descriptions from sentences would be possible.

In summary the author feels that it is fair to say that the theory of descriptions presented here combines the best features of Russell's and Frege's theories. With Frege, we preserve the laws of identity and the extensionality of the descriptive operator without giving improper descriptions an unintended proper designation. Assuming the very reasonable (I 3), we would be in complete agreement with Russell in non-equality atomic contexts, for from (I 3) we could derive:

$$\models \alpha \,\mathrm{R}\, \mathrm{I}\, x\, \Phi \leftrightarrow \exists\, y[\forall\, x[x = y \leftrightarrow \Phi] \wedge \alpha \,\mathrm{R}\, y].$$

This possibility is of course excluded by Frege.

We shall not pause here to give the proof that every valid formula can be derived from (S0)–(S4), (I 1)–(I 2) by the rules (MP) and (UG), because in Section 3 a more general completeness proof will be presented in full.

2. *Virtual Classes*

In his new book [7] Professor Quine makes thorough use of what he calls *virtual classes* to simplify the development and comparison of various systems of set theory. For example, the different kinds of existential assumptions about the real classes can be presented in a uniform manner in Quine's notation. More than that, with mild assumptions on real classes, the reduction of arithmetic to class theory can be very conveniently described in Quine's style; so the device has considerable appeal.

In the author's opinion, the only thing missing in Quine's presentation is a semantical analysis of the notion of virtual classes. No doubt Quine feels no need for such an analysis, since his class symbols are all eliminable by design. Virtual classes function mainly as an aid in condensing long formulas; the programme is successful owing to the transfer of standard set-theoretical notions from the real to the virtual. Nevertheless, semantical insights can be helpful in understanding a formal system; especially when one can check formulas without having to first eliminate the contextually defined notions. In the presentation to be given here, virtual classes will be treated axiomatically, Quine's contextual definitions will be proved as theorems, and the model theory for the system will naturally suggest itself along the lines of what we did for descriptions.

Our language will be much like the first-order language of Section 1, except we replace the predicate symbol R by the symbol $\in$ for membership. Further we drop the descriptive operator for the time being and use instead the operator of class abstraction; thus the terms are now either single variables or expressions of the form

$$\{x : \Phi\}$$

where x is a variable and Φ a formula. The construction of compound formulas proceeds as before.

As axioms and rules of inference for the theory we use (MP), (UG), and (S0)–(S4) as before, except the notions of terms and formulas must be understood in the new sense. In addition we employ three principles governing the behaviour of membership and abstraction:

(Q 1) $\forall y[y \in \alpha \leftrightarrow y \in \beta] \rightarrow \alpha = \beta$,

(Q 2) $\alpha \in \beta \rightarrow \exists y[y = \alpha]$,

(Q 3) $\forall y[y \in \{x : \Phi\} \leftrightarrow \exists x[x = y \wedge \Phi]]$,

where the variable y is not free in α, β, or Φ. The last schema could also have been written in the form

(Q 3′) $\forall y[y \in \{x : \Phi\} \leftrightarrow \Phi(x/y)]$

Combining (Q 2) and (Q 3) note that

(Q 3″) $\alpha \in \{x : \Phi\} \leftrightarrow \exists x[x = \alpha] \wedge \Phi(x/\alpha)$

is a consequence, where x is not free in the term α.

It is quite easy to see that every theorem provable in the present theory is provable in Quine's theory. Note that (S3) and (S4) are Quine's 6.4 and 6.6. Next (Q 1) follows from Quine's 2.7; (Q 2) follows from 6.9 and 6.12; and (Q 3) is given on p. 17 of [6]. Conversely, except for our allowing the empty domain of (real) individuals, all of Quine's theory can be deduced from ours. In particular, the contextual definitions 2.1, 2.7, and 5.5 are provable at once as biconditionals, and the lone axiom 4.1 is a special case of (S4). We are not concerned here with the additional axioms on the existence of real classes.

To discuss the models of Quine's theory, we first remark that having virtual class forces us to contemplate *many* improper individuals and not just *one* as was the case for descriptions. However, a virtual class is completely determined by its real members. So let us take as models structures of the form $\mathfrak{A} = \langle A, E\rangle$, where $E \subseteq A \times A$, and where the relation E is *extensional* in A. In other words the structure $\mathfrak{A}$ must satisfy the sentence

$$\forall x \forall x'[\forall y[y \in x \leftrightarrow y \in x'] \rightarrow x = x']$$

in the usual sense. The *elements* of A will correspond to the real classes; while the *subsets* of A will correspond to the virtual classes—well, not quite. We must identify the real classes with the virtual classes having the same members. This is best done by making the values of terms *always* be subsets of A. To get the correspondence between the elements of A and the subsets of A we define a function $\dot{E}$ on A such that for $a \in A$,

$$\dot{E}(a) = \{b \in A : \langle b, a\rangle \in E\}.$$

By virtue of the extensionality of E, this is a one-one correspondence between elements of A and certain subsets of A. Next in the definition of value we make these changes:

$$\models_{\mathfrak{A}} \alpha \in \beta[s] \textit{ iff for some } a \in \| \beta[s] \|_{\mathfrak{A}},\ \dot{E}(a) = \| \alpha[s] \|_{\mathfrak{A}}.$$

$$\models_{\mathfrak{A}} \forall v_i\, \Phi[s] \textit{ iff for all } a \in A,\ \models_{\mathfrak{A}} \Phi[s(i/\dot{E}(a))]$$

$$\| \{v_i : \Phi\} \|_{\mathfrak{A}} = \{a \in A : \models \Phi[s(i/\dot{E}(a))]\}$$

All the other clauses remain the same. In the definition of validity we make the restriction that assignments should have *only* subsets of A as values. Even if A is empty there is one subset of A; so this restriction does not cause any trouble.

Note that if $a, b \in A$ and $s_0 = \dot{E}(a)$ and $s_1 = \dot{E}(b)$, then

$$\models_{\mathfrak{A}} v_0 \in v_1[s] \textit{ iff } \langle a, b\rangle \in E,$$

and

$$\models_{\mathfrak{A}} v_0 = v_1[s] \textit{ iff } a = b.$$

Hence, if Φ is a formula without free variables and without the abstraction operator, then Φ is true in $\mathfrak{A}$ in the new sense if and only if it is true in the sense of Section 1.

It is easily checked that all the schemata (S0)–(S4) and (Q 1)–(Q 3) are valid in all extensional structures. The converse, that all formulas valid in all extensional structures are provable in the present theory, will follow from the completeness theorem of Section 3. Thus we have a full explanation of the model theory for Quine's system.

When he introduced the descriptive operator into his system, Quine, the modern day champion of contextual definition, abandoned the Russell approach in favour of Frege's idea. That seems a bit odd, does it not? The explanation is probably this: with Russell's elimination the formula $\alpha = \alpha$ is not always valid, whereas Quine wants this law of equality. Besides, it is a waste of effort to introduce new operators by contextual definition when an explicit definition is at hand. Quine chose this definition (more or less):

$$\mathrm{I}\, x\, \Phi = \{x : \exists\, y[\forall\, x[x = y \leftrightarrow \Phi] \wedge z \in y]\}.$$

Thus when $\neg \exists\, y\, \forall\, x[x = y \leftrightarrow \Phi]$ holds, $\mathrm{I}\, x\, \Phi$ denotes the empty class. Since Quine wants the empty class to be a real class, we see that the improper description is behaving in the Fregean manner.

Another definition was open to Quine, however, namely:

$$\mathrm{I}\, x\, \Phi = \{z : \exists\, y[\forall\, x[x = y \leftrightarrow \Phi] \wedge z \in y] \vee [\neg \exists\, y\, \bigwedge x\, [x = y \leftrightarrow \Phi] \wedge \neg z \in z]\}$$

If we let Δ be the term $\{z : \neg z \in z\}$, then by the argument of the Russell paradox we can prove that

$$\neg \exists\, y[y = \Delta].$$

Hence, with the revised definition we prove exactly (I 1) and (I 2) with $\ast$ replaced by Δ. Of course $\ast = \Delta$ is at once provable, so the former theory is recaptured. The author strongly feels that this path to descriptions is much more in harmony with the concept of virtual classes than is the version adopted by Quine.[1]

Let us see now what happens to the elimination of descriptions with the definition just proposed. In view of (Q 2) we have first:

$$\mathrm{I}\, x\, \Phi \in \beta \leftrightarrow \exists\, y[\forall\, x[x = y \leftrightarrow \Phi] \wedge y \in \beta],$$

where y is not free in Φ or β. That would please Russell. In the other argument place we have:

[1] See, however, Professor Quine's remarks quoted at the end of this section.

$$\alpha \in \mathbf{I}\, x\, \Phi \leftrightarrow \exists\, y[\forall\, x[x = y \leftrightarrow \Phi] \wedge \alpha \in y] \vee$$
$$[\neg \exists\, y\, \forall\, x[x = y \leftrightarrow \Phi] \wedge \exists\, y[y = \alpha \wedge \neg\, y \in y]].$$

That would probably confound followers of Russell or Frege; the author hopes it is not too displeasing to the Quine school, however.

Parallel to the revision of the definition of descriptions, the author would like to also suggest a revision of Quine's definition of *function value*. Let us assume along with Quine enough axioms to guarantee the existence of ordered pairs of real classes. In the author's notation, the definition of function value will read:

$$\phi(\xi) = \mathbf{I}\, y\, \exists\, x[x = \xi \wedge \langle y, x\rangle \in \phi],$$

where x and y are variables not free in ϕ and ξ. Quine gave as his definition:

$$\phi(\xi) = \mathbf{I}\, y[\langle y, \xi\rangle \in \phi].$$

This is 'defective' not only because the wrong kind of description was used, but because when ξ does not exist (i.e. $\neg \exists\, x[x = \xi]$), then

$$\langle y, \xi\rangle = \{\{y\},\ \{y, \xi\}\} = \{\{y\}\} = \langle y, y\rangle.$$

By chance there might be a unique y with $\langle y, y\rangle \in \phi$, and we are uncomfortable. Now Quine avoids this unpleasantness by restricting attention to the class

$$arg\, \phi = \{x : \exists\, z\, \forall\, y[y = z \leftrightarrow \langle y, x\rangle \in \phi]\}$$

Using the proposed *new* definition we can simplify this last equation to:

$$arg\, \phi = \{x : \exists\, y[y = \phi(x)]\}.$$

Further, there is no need to avoid unintended function values, for we can prove quite generally:

$$\exists\, y[y = \phi(\xi)] \leftrightarrow \xi \in arg\, \phi.$$

This last biconditional reads so well that it seems justification enough for the revised definition. (This discussion of function value is improved over an earlier version at the suggestion of David Kaplan.) The general principle to be applied to such questions is this: things should exist only when it is intended that they exist. It seems quite remarkable to the author that there is a flexible enough formalism that actually allows us to follow this principle.

In connection with these suggestions, Professor Quine wrote to the author on May 3, 1965, as follows:

'The redefinition of description and of function value that you propose sacrifices an advantage that I had gone out of my way for: the freedom to substitute descriptions for bound variables without regard to special existence premisses. This freedom, touched on in pp. 58, 68, and 107 of *Set Theory and Its Logic*, covers a lot, since so many notations are defined as function values and ultimately as descriptions. It even covers function values where function and argument are rendered by Greek letters, without presumption of existence; cf. p. 68. Also it covers arithmetical expressions containing Greek letters; cf. p. 107. Without this freedom the book would be appreciably more laboured. Perhaps you could devise alternative conventions, on your basis, that would work smoothly too; but then I'd want to see some trial runs for comparison.'

Professor Quine is quite justified in asking for trial runs for comparison, and the author will try to apply these comparisons in future publications. For the time being the reader is asked to consider the merits of the proposal on the grounds of 'naturalness' as indicated above. He should also imagine having to make all existential assumptions explicit, and ask himself whether unrestricted substitution is to be preferred over the gain of information obtained by using formulas with explicitly displayed assumptions.

3. *General operators*

The system that we shall treat here will be of the same type as the systems of Sections 1 and 2. The language will involve a binary predicate symbol R and a variable binding operator O of the same syntactical category as the operators of description and abstraction. Thus

$$\mathsf{O}\, x\, \Phi$$

is a term when x is a variable and Φ is a formula. As before we shall not assume that the values of terms are necessarily in the range of the individual variables. A convenient way to express this is to consider structures of the form

$$\mathfrak{A} = \langle A, A_*, R, O\rangle,$$

where A is a set (the domain of properly existing individuals), A_* is a *non-empty* superset of A (the domain of 'improper' individuals), R is a binary relation where $R \subseteq A_* \times A_*$, and O is a function defined on *subsets* of A taking values in A_*. The definition of value is now modified in this particular:

$$\| \mathsf{O}\, v_i\, \Phi[s] \|_{\mathfrak{A}} = O(\{a \in A : \models_{\mathfrak{A}} \Phi[s(i/a)]\})$$

Further, the definition of validity is changed so that $\models_{\mathfrak{A}} \Phi$ means that $\models_{\mathfrak{A}} \Phi[s]$ only for assignments s where the values of the function s are *all* in the set A_*. When the set A_* is explicitly mentioned it is not reasonable to allow the values of assignments to the free variables to be completely arbitrary.

If $\mathfrak{A} = \langle A, R\rangle$ is a structure in the sense of Section 1, then we can correlate with it a structure in the new sense, namely:

$$\mathfrak{A}_* = \langle A, A_*, R, I\rangle,$$

where

$$A_* = A \cup \{\ast_A\} \cup \text{field}\,(R),$$

and where I is defined on subsets $X \subseteq A$ so that:

$$I(X) = \begin{cases} a & \text{if } X = \{a\}, \\ \ast_A & \text{if } X \neq \{a\} \text{ for all } a \in A. \end{cases}$$

It is then easy to prove that for an assignment s with values in A_*, $\models_{\mathfrak{A}} \Phi[s]$ holds in the old sense if and only if $\models_{\mathfrak{A}*} \Phi[s]$ holds in the new sense, and that $\| \alpha[s] \|_{\mathfrak{A}} = \| \alpha[s] \|_{\mathfrak{A}*}$. Of course the symbol O should be replaced by $\mathbf{I}$ to make sense of this last statement.

If $\mathfrak{A} = \langle A, E\rangle$ is a structure in the sense of Section 2, then the correlated structure is

$$\mathfrak{A}_0 = \langle A_0, A_*, E_0, J\rangle,$$

where $A_0 = \{\dot{E}(a) : a \in A\}$, A_* is the set of all subsets of A, J is the identity function on A_*, and the relation E_0 is defined for $X, Y \in A_*$ so that:

$$XE_0Y \textit{ iff for some } a \in Y, \dot{E}(a) = X.$$

Again, for assignments with values in A_*, the old and new definitions of value agree completely. Therefore, the structures considered here do properly generalize those used in earlier examples.

Aside from (S0)–(S4) which are all valid in the present sense, we have also

(O1) $\quad [\forall\, x[\Phi \leftrightarrow \Psi] \to \mathsf{O}\, x\, \Phi = \mathsf{O}\, x\, \Psi]$,

and

(O2) $\quad \mathsf{O}\, x\, \Phi = \mathsf{O}\, y\, \Phi(x/y)$,

where the variable y is not free in Φ.

These kinds of schemata were not needed explicitly in Sections 1 and 2 because the required formulas were in each case deducible from the others given. In the general case they are the only schemata required that involves the operator $\mathbf{O}$ in a special way. We want now to show that a formula is universally valid if and only if it is deducible from (S0)–(S4), (O1), (O2) by the rules (MP) and (UG). Let $\vdash \Phi$ mean that the formula Φ is so deducible. All we need to prove is that if not $\vdash \Phi$, then there is a structure $\mathfrak{A}$ and an assignment s such that not $\models_{\mathfrak{A}} \Phi[s]$.

To this end let Φ_0 be a particular formula such that not $\vdash \Phi_0$. Let

$$\Psi_0, \Psi_1, \ldots, \Psi_n, \ldots$$

be a list containing *every* formula at least once such that Ψ_0 is $\neg \Phi_0$, and if Ψ_n is of the form

$$\neg \forall v_i \Phi$$

then Ψ_{n+1} is of the form

$$\exists v_{j+1}[v_{j+1} = v_j] \wedge \neg \Phi(v_i/v_j),$$

where v_j is the first variable not free in $\Psi_0, \Psi_1, \ldots, \Psi_n$. It is easy to show that such a sequence exists. We define by recursion the sequence of formulas

$$\Psi'_0, \Psi'_1, \ldots, \Psi'_n, \ldots$$

where Ψ'_n is $\neg \Psi_n$ or Ψ_n according as

$$\vdash [\Psi'_0 \wedge \ldots \wedge \Psi'_{n-1} \rightarrow \neg \Psi_n]$$

or not.

We let

$$M = \{\Psi'_n : n \in N\}.$$

Clearly $\Psi_0 \in M$ and the set M of formulas has these properties:

(i) *if* $\vdash \Phi$, *then* $\Phi \in M$,

(ii) $\neg \Phi \in M$ *iff* $\Phi \notin M$,

(iii) $[\Phi \rightarrow \Psi] \in M$ *iff* $\Phi \notin M$ *or* $\Psi \in M$,

(iv) $\forall v_i \Phi \in M$ *iff for all* j *if* $\exists v_{j+1}[v_{j+1} = v_j] \in M$, *then* $\Phi(v_i/v_j) \in M$.

So far the details of the proof are just as in any standard version as the completeness proof for first-order logic based on the method due to Henkin.

Let T be the set of all terms, and define an equivalence relation $\equiv$ on the set T by the condition that

$$\alpha \equiv \beta \text{ iff } [\alpha = \beta] \in M$$

The equivalence class of a term α is denoted by $\alpha/\equiv$. We define

$$A = \{v_j/\equiv \; : \; \exists \, v_{j+1}[v_{j+1} = v_j] \in M\},$$

and

$$A_* = \{\alpha/\equiv \; : \alpha \in T\}.$$

The relation $R \subseteq A_* \times A_*$ is defined by the equation

$$R = \{\langle \alpha/\equiv, \beta/\equiv \rangle : [\alpha \,\mathsf{R}\, \beta] \in M\},$$

and the operator O is defined for $X \subseteq A$ so that:

$$O(X) = \begin{cases} \mathsf{O}\, v_i \; \Phi/\equiv \textit{ if } X = \{v_j/\equiv \; \in A : \Phi(v_i/v_j) \in M\} \\ \quad v_0/\equiv \textit{ if there is no such formula } \Phi. \end{cases}$$

For the particular assignment s where $s_i = v_i/\equiv$, we wish to show that $\models_{\mathfrak{A}} \Phi_0[s]$ does not hold where $\mathfrak{A} = \langle A, A_*, R, O \rangle$ is the structure just defined. This cannot be done quite directly: one must prove by induction that if Φ is a formula, α is a term, and s is an assignment where $s_i = \alpha_i/\equiv$, then

$$\models_{\mathfrak{A}} \Phi[s] \textit{ iff } \Phi(v_0/\alpha_0, v_1/\alpha_1, \ldots, v_n/\alpha_n, \ldots) \in M,$$

and

$$\| \alpha[s] \|_{\mathfrak{A}} = \alpha(v_0/\alpha_0, v_1/\alpha_1, \ldots, v_n/\alpha_n, \ldots)/\equiv,$$

where the notation on the right-hand sides indicates simultaneous substitution of terms for free variables. Again this step is just like the corresponding step in the usual proofs, and conditions (i)–(iv) on M were explicitly chosen so that the argument would work out.

In case the additional axioms of Sections 1 or 2 were added, the structure $\mathfrak{A}$ just obtained could be modified directly to obtain the structure in the earlier sense that is required.

4. *Eliminability*

A *sentence* is a formula without free variables. A *theory* is a set of sentences containing all universally valid sentences and closed under the rule of *modus ponens*. The operator O is *eliminable* in a theory T, if for each formula Φ there is a formula Ψ not containing O such that

$$\forall\, v_0 \,\forall\, v_1 \ldots \forall\, v_{m-1}[\Phi \leftrightarrow \Psi]$$

belongs to T, where the free variables of Φ and Ψ are among v_0, $v_1, \ldots, v_{m-1}$.

The theory based on schemata (I 1)–(I 3) from Section 1 (where O replaces the symbol I) is a theory in which O is eliminable. Similarly for the theory based on (Q 1)–(Q 3) (where O replaces the abstraction operator). The purpose of this section is to give necessary and sufficient model-theoretic conditions for O to be eliminable in a theory T. The conditions found will be very close to those of Beth's Definability Theorem (cf. [9]).

If $\mathfrak{A} = \langle A, A_*, R, O\rangle$ and $\mathfrak{A}' = \langle A', A'_*, R', O'\rangle$ are two structures, we say that $\mathfrak{A}$ and $\mathfrak{A}'$ are *weakly isomorphic* if there is a one-one function mapping the set A on to the set A' such that for all $a, b \in A$,

$$\langle a, b\rangle \in R \textit{ iff } \langle f(a), f(b)\rangle \in R'.$$

When s is an assignment with values in A, we let $f((s))$ denote the assignment with values in A' such that

$$f((s))_i = f(s_i).$$

The condition of f to give a weak isomorphism can be equivalently stated as:

$$\models_{\mathfrak{A}} \Phi[s] \textit{ iff } \models_{\mathfrak{A}'} \Phi[f((s))]$$

for all assignments s with values in A and all formulas Φ not containing the operator O. We shall say that f gives a *strong isomorphism* if this last biconditional holds for *arbitrary* formulas Φ.

A *model* for a theory is of course a structure for which all sentences of the theory are true. We can now state the theorem on eliminability:

The operator O *is eliminable in a theory* T *if and only if whenever two models of* T *are weakly isomorphic by a certain one-one function, they are also strongly isomorphic by the same function.*

If O is eliminable in T, then it is clear that weak isomorphism implies strong isomorphism. The converse will be proved by applying Beth's theorem to a suitable first-order theory with many predicate symbols but without operators.

Let T be a theory for which weak isomorphism implies strong isomorphism. Introduce new predicate symbols S^Φ corresponding to each formula Φ in the original sense. The predicate S^Φ will be a m-place predicate, where m is the least integer such that the free variables of Φ are among $v_0, v_1, \ldots, v_{m-1}$. Consider the extension of T obtained by adjoining these sentences as axioms:

$$\forall v_0 \forall v_1 \ldots \forall v_{m-1}[\mathsf{S}^\Phi(v_0, v_1, \ldots, v_{m-1}) \leftrightarrow \Phi].$$

The theory T_0 is the set of sentences of the extension of T involving the predicates R and S^{Φ} but *not* O. It is obvious that O is eliminable in T if and only if *all* the S^{Φ} are definable in T_0 in terms of the predicate R in the ordinary sense of first-order definability.

According to Beth's theorem, to show that the S^{Φ} are definable in terms of R it is enough to show that two models

$$\mathfrak{A}_0 = \langle A, R_0, \ldots, S^{\Phi}, \ldots \rangle$$

and

$$\mathfrak{A}_0' = \langle A', R_0', \ldots, S'^{\Phi}, \ldots \rangle$$

of T_0, where $\langle A, R_0 \rangle$ and $\langle A', R_0' \rangle$ are isomorphic by a function f, are also isomorphic by the same function f. To prove this we will construct structures

$$\mathfrak{A} = \langle A, A_*, R, O \rangle$$
$$\mathfrak{A}' = \langle A', A_*', R', O' \rangle$$

such that for all assignments s with values in A and for all formulas Φ

$$\models_{\mathfrak{A}0} S^{\Phi}(v_0, v_1, \ldots, v_{m-1})\,[s] \textit{ iff } \models_{\mathfrak{A}} \Phi[s];$$

similarly for $\mathfrak{A}_0'$ and $\mathfrak{A}'$. Now by assumption $\langle A, R_0 \rangle$ and $\langle A', R_0' \rangle$ are isomorphic. Hence $\mathfrak{A}$ and $\mathfrak{A}_0'$ are weakly isomorphic; therefore strongly isomorphic. But this means that $\mathfrak{A}_0$ and $\mathfrak{A}_0'$ are isomorphic, all by the same function we started with. It will be enough to show how to construct $\mathfrak{A}$ from $\mathfrak{A}_0$; actually it will be easier to construct a structure $\bar{\mathfrak{A}}$ which is strongly isomorphic to the structure we want.

First let U be the set of all pairs $\langle \alpha, s \rangle$ where α is a term of the original language and s is an assignment with values in A. We define an equivalence relation $\equiv$ on the set U:

$$\langle \alpha, s \rangle \equiv \langle \beta, t \rangle \textit{ iff } \models_{\mathfrak{A}0} S^{\alpha=\beta'}(v_0, \ldots, v_{m-1}, v_m, \ldots, v_{m+n-1})\,[u],$$

where m is the least integer such that the free variables of α are among $v_0, \ldots, v_{m-1}$; n is the least integer such that the free variables of β are among $v_0, \ldots, v_{n-1}$; β' is the term $\beta(v_0/v_m, \ldots, v_{n-1}/v_{m+n-1})$; and u is the assignment where

$$u_i = \begin{cases} s_i \textit{ if } i < m, \\ t_{i-m} \textit{ if } i \geqslant m. \end{cases}$$

We let $\langle \alpha, s \rangle / \equiv$ be the equivalence class of $\langle \alpha, s \rangle$ in U and put

$$\bar{A}^* = \{\langle \alpha, s \rangle / \equiv \; : \langle \alpha, s \rangle \in U\}.$$

and

$$\bar{A} = \{\langle v_0, s \rangle / \equiv \; : \langle v_0, s \rangle \in U\}$$

We note that A and $\acute{A}$ are in a one-one correspondence by the function e such that for $a \in A$,

$$e(a) = \langle v_0, s\rangle/\equiv$$

for all assignments s with $s_0 = a$. The relation $\bar{R}$ on $\acute{A}_*$ is such that

$$\langle\langle\alpha, s\rangle/\equiv, \langle\beta, t\rangle/\equiv\rangle \in \bar{R} \textit{ iff}$$
$$\models_{\mathfrak{B}0} \mathsf{S}^{\alpha R \beta'}(v_0, \ldots, v_{m-1}, v_m, \ldots, v_{m+n-1})\,[u],$$

where m, n, β' and u are determined as before. Finally to define $\bar{O}$, we let r be a fixed assignment with values in A and for $X \subseteq \acute{A}$ we set

$$\bar{O}(X) = \begin{cases} \langle \mathbf{O} v_i\, \Phi, s\rangle/\equiv & \textit{if } X = \\ \qquad \{e(a) \in \acute{A} : \models_{\mathfrak{B}0} \mathsf{S}_{\Phi}\,(v_0, \ldots, v_{m-1})\,[s(i/a)]\} \\ \langle v_0, r\rangle/\equiv & \textit{if there is no such formula } \Phi \textit{ and assignment } s. \end{cases}$$

The desired properties of the structure

$$\bar{\mathfrak{B}} = \langle \acute{A}, \acute{A}_*, \bar{R}, \bar{O}\rangle$$

will be established by proving for all formulas Φ, all terms α, and all assignments s with values in A that

$$\models_{\mathfrak{B}} \Phi[e((s))] \textit{ iff} \models_{\mathfrak{B}0} \mathsf{S}^{\Phi}(v_0, \ldots, v_{m-1})\,[s],$$

and

$$\| \alpha[e((s))] \|_{\mathfrak{B}} = \langle \alpha, s\rangle/\equiv.$$

This result on eliminability is not very satisfactory. The operators of Sections 1 and 2 are eliminable in a much stronger sense: for example, the schema (IE) gives practically a wholesale way of eliminating the descriptive operator. Similar things may be said for Quine's abstraction operator. In other words to eliminate an occurrence of $\mathbf{I}\, x\, \Phi$ we need only examine the context in which this term is found; we do not have to make our elimination depend on any peculiarities of the formula Φ within the scope of the operator. The author has no idea what kind of model-theoretic conditions would correspond to this *uniform* eliminability that we always have when operators are introduced by contextual definitions. It seems like an interesting problem.

BIBLIOGRAPHY

1. P. Bernays and A. A. Fraenkel. *Axiomatic set theory*, Amsterdam (1958), 226 pp.
2. T. Hailperin and H. Leblanc. 'Non-designating singular terms.' *The philosophical review*, vol 68 (1959), pp. 239–243.

3. D. Hilbert and P. Bernays. *Grundlagen der Mathematik*, Bd. I (1934), Bd. II (1939), Berlin, 471 + 498 pp.
4. J. Hintikka. 'Existential presuppositions and existential commitments.' *The journal of philosophy*, vol. 56 (1959), pp. 125–137.
5. J. Hintikka. 'Towards a theory of definite descriptions.' *Analysis* (Oxford), vol. 19, no. 4 (1959), pp. 79–85.
6. A. Mostowski. 'On the rules of proof in the pure functional calculus of the first order.' *The Journal of Symbolic Logic*, vol. 16 (1951), pp. 107–111.
7. W. V. Quine. *Set theory and its logic*, Harvard (1963), xv + 359 pp.
8. N. Rescher. 'On the logic of existence and denotation.' *The philosophical review*, vol. 69 (1959), pp. 157–180.
9. A. Robinson. *Introduction to model theory and to the metamathematics of algebra*, Amsterdam (1963), ix + 284 pp.
10. T. Smiley. 'Sense without denotation.' *Analysis* (Oxford), vol. 20, no. 4 (1960), pp. 125–135.