QUINE'S INDIVIDUALS

DANA SCOTT

University of California, Berkeley, California, U.S.A.

Professor Quine has suggested that with regard to theories of membership it would be possible to allow for the existence of *non-classes* or *individuals* by interpreting the formula '$x \in y$' as synonymous with '$x = y$' in the case that y is an individual. This suggestion is by no means a commitment, for as he stresses in [7, p. 123] the situation never arises in the formal development where it is known that any object is actually an individual in this sense. The fact that the situation did not arise in the course of one particular development is not a conclusive argument, however. By some bizarre chain of deductions, might we not prove from Quine's axioms for class membership that individuals do in fact exist? It seems clear from his remarks that Professor Quine does not imagine such a deduction is possible, nor would anyone else believe this who understands the suggestion and feels confident that the axioms are consistent. Nothing supports belief like proof, and it will be the purpose of this note to demonstrate that if Quine's axioms are consistent, then they remain so upon the adjunction either of the sentence

$$(*) \qquad `\exists y \forall x[x \in y \leftrightarrow x = y]'$$

or of its negation. Take your choice.

For the sake of simplicity, we shall not discuss the system of [7] but rather the system of [6], usually called *New Foundations*, or NF for short. As Wang in [8] has shown, every model for NF can be extended to a model of ML, the system of [7]; hence, a consistency and independence proof for (*) relative to NF easily yields one for ML. Of course, Wang's proof is not a finitary argument, because it makes use of the semantical notion of *definability* within a model. The proof given here for NF is strictly finitary, and no doubt a similar argument can be applied to ML.

For our purposes, we shall imagine NF formulated in first-order logic with *identity* and with the *descriptive operator*. The only non-logical constant is the binary predicate symbol '$\in$'. The logical constants are those in the following list

$$\wedge, \vee, \neg, \rightarrow, \leftrightarrow, =, \forall, \exists, \mathbf{T},$$

where the last symbol denotes the descriptive operator. The variables are those in the list

$$x, y, z, w, x', y', z', w', x'', \cdots, \text{etc.}$$

The note published here is not the paper read by the author at the Congress. After the presentation of the original paper, Professor E. Specker pointed out that he had published similar and overlapping material from his *Habilitationsschrift* of 1951. There was very little new to recommend the author's remarks, and so he decided to substitute another article.

Brackets '[' and ']' are used (and misused) with the (binary) propositional connectives in the usual way, but neither they nor any other additional punctuation is required in such contexts as $\ulcorner\neg\Phi\urcorner$, '$x \in y$', '$x = y$', $\ulcorner\exists x\Phi\urcorner$, $\ulcorner Tx\Phi\urcorner$, where Φ is a formula. It is assumed as clear what is meant by the statements 'Φ is a formula', 'ϕ is a term', "'x' is free in the formula Φ', and so on. As indicated, capital Greek letters (except 'Λ') will range over formulas, while the lower case Greek letters range over terms. If ϕ is a term and Φ a formula, then $\Phi(\phi)$ is the result of substituting ϕ for all free occurrences of 'x' in Φ', where Φ' is the first alphabetic variant of Φ (in some suitable ordering of all formulas) which has no bound variables identical with any free variables of ϕ. Similarly for $\phi(\psi)$. We shall write '$\Phi(y)$' rather than the more correct 'Φ('y')' to indicate substitution of particular variables for 'x'.

The theory of the descriptive operator as developed in [3] is to be applied here. From this point of view the term '$\mathbf{T}x\, x = x$' denotes an object identical with the denotations of all improper descriptive phrases. That neither the axioms of logic nor the axioms of NF indicate more definitely the nature of this privileged object is totally irrelevant. However, we must discuss how the notion of a *stratified* formula as defined in [7, pp. 157 f.] is to be carried over to this extended theory. In the first place, to avoid tiresome circumlocution, we assume that the notion is so defined that if a formula is stratified, then so are all of its alphabetic variants. Thus to check a given formula for stratification, rewrite all its bound variables so that no letter occurs twice with the operators '$\forall$', '$\exists$', '$\mathbf{T}$'. Then in assigning numerals to terms, be sure to assign $\ulcorner\mathbf{T}x\Phi\urcorner$ the same numeral as assigned to 'x', and make certain that both sides of an equation receive the same numeral. Otherwise proceed exactly as described by Quine. For example, '$x = y \wedge x \in y$' is not stratified, while the strange formula '$\mathbf{T}x\, x = x \in \mathbf{T}y\, y = y$' is stratified.

We may take as the axioms of NF the following:

(I) '$\forall x \forall y[\forall z[z \in x \leftrightarrow z \in y] \to x = y]$';

(II) $\ulcorner\exists y \forall x[x \in y \leftrightarrow \Phi]\urcorner$,

where Φ is a stratified formula in which 'y' is not free. (I) is the *axiom of extensionality*, while (the closures of) the formulas mentioned in (II) are the *axioms of class existence*.

A term τ will be called *level* if its only free variable is 'x' and if the formula $\ulcorner x = \tau(x)\urcorner$ is stratified. If we imagine for a moment that τ defines a "function" which maps the object x to the object denoted by τ, then the condition of being *level* means that the function values are of the same "level" or "type" as the arguments. In any case, whatever we have in the back of our minds, it turns out that level terms are very useful for giving reinterpretations of the membership relation. To be specific, for each formula Φ, let Φ^τ denote the result of simultaneously replacing [1] all portions of Φ of the form

[1]This replacement might better be described as a substitution of the formula $\ulcorner x \in \tau(y)\urcorner$ for the atomic formula '$x \in y$' in the sense of the $\check{S}$ substitution of Church [2, pp. 192 f.].

$\ulcorner\phi \in \psi\urcorner$, where ϕ, ψ are terms, by the formula $\ulcorner\phi \in \tau(\psi)\urcorner$. If τ is a level term and Φ is stratified, then it is very easy to see that Φ^τ is also stratified. The importance of this process is indicated by the following result.

METATHEOREM.[2] *Suppose τ is a level term for which there is a proof in NF of the sentence*

$$\ulcorner\forall y \exists z \forall x[y = \tau(x) \leftrightarrow x = z]\urcorner.$$

Then if Ψ is provable in NF, so is Ψ^τ.

PROOF. It is obvious that we have only to check the cases where Ψ is an axiom of NF. Take the case where Ψ is axiom (I). The sentence to be shown provable is

$$\ulcorner\forall x \forall y[\forall z[z \in \tau(x) \leftrightarrow z \in \tau(y)] \rightarrow x = y]\urcorner.$$

In view of axiom (I) this is equivalent to the sentence

$$\ulcorner\forall x \forall y[\tau(x) = \tau(y) \rightarrow x = y]\urcorner,$$

which is a direct consequence of the assumption on τ.

Next consider the case where Ψ is one of the axioms of (II). Here the formula to be shown provable is (the closure of) a formula of the form

$$\ulcorner\exists y \forall x[x \in \tau(y) \leftrightarrow \Phi^\tau]\urcorner,$$

where 'y' is not free in the formulas Φ and Φ^τ. From the assumption on τ it follows that the sentence

$$\ulcorner\forall z \exists y\; z = \tau(y)\urcorner$$

is provable. Hence, it is sufficient to verify that

$$\ulcorner\exists y \forall x[x \in y \leftrightarrow \Phi^\tau]\urcorner$$

is provable. But since Φ^τ is stratified, (the closure of) this formula is an axiom of (II), and hence is indeed provable.

Let us call a level term τ which satisfied the hypothesis of the Metatheorem a *permutation*. An immediate corollary of the above result is that if τ is a permutation such that Ψ^τ is provable in NF, and if NF is consistent, then the *sentence Ψ is consistent with the axioms of* NF. Clearly the exhibition of τ together with a proof of Ψ^τ yields a finitary consistency proof of Ψ relative to the axioms of NF. This is the method that will be applied to the sentence (*) and to its negation.

For the moment let Ψ be the formula (*) and let τ be a permutation. Notice that Ψ^τ is provably equivalent to the sentence

---

[2]That a similar result holds for other theories of class membership was first brought to the author's attention by the paper [4] of Reiger and reported on in the review [5]. Subsequently Specker pointed out that the idea is really contained in Bernays' article [1, see pp. 83 f.], but there it is applied only to models of the axioms, and the finitary character of the method is not stressed. Still, it seems proper to credit the idea to Bernays; hence, the remarks in the review [5] are unfair or at least misleading. The application to NF does not seem to have been suggested before, however.

(**) $$\ulcorner \exists y\, \tau(y) = \iota(y) \urcorner,$$

where $\iota(y)$ is the term denoting the *unit set* of y, or in other words ι is the term '$\mathrm{T}z\forall w[w \in z \leftrightarrow w = x]$'. Our first question, then, is whether we can find a permutation for which (**) is provable. The answer is yes, for take τ to be the term

$$\begin{aligned}&\ulcorner \mathrm{T}z[[x = \Lambda \wedge z = \iota(\Lambda)] \vee \\ &[x = \iota(\Lambda) \wedge z = \Lambda] \vee \\ &[\neg x = \Lambda \wedge \neg x = \iota(\Lambda) \wedge z = x]]\urcorner,\end{aligned}$$

where Λ is the term denoting the *empty set*, i.e. the term '$\mathrm{T}z\forall w \neg w \in z$'. In less formal terms, τ is the permutation that "interchanges" the sets Λ and $\iota(\Lambda)$.[3] That τ is level and satisfies all the conditions is obvious.

The second question is whether there exists a different term τ for which the negation of (**) is provable. This construction requires somewhat more thought, since care is always necessary to produce a level term.

To facilitate the definition of the new τ, two auxiliary terms ρ and σ are useful; they are respectively the following:

$$\begin{aligned}&\ulcorner \mathrm{T}z[[x = \Lambda \wedge z = \iota(\Lambda)] \vee \\ &[\neg x = \Lambda \wedge \neg \Lambda \in x \wedge z = \Lambda] \vee \\ &[\Lambda \in x \wedge \neg \iota(\iota(\Lambda)) \in x \wedge z = \iota(\iota(\Lambda))] \vee \\ &[\Lambda \in x \wedge \iota(\iota(\Lambda)) \in x \wedge z = \iota(\Lambda)]]\urcorner, \\ &\ulcorner \mathrm{T}z\exists w[x = \iota(w) \wedge \forall z'[z' \in z \leftrightarrow [z' = w \vee z' = \rho(w)]]]\urcorner\end{aligned}$$

Notice first that both ρ and σ are level terms, but of course neither is a permutation. The term σ will only be applied to unit sets, and the fact that the descriptive phrase is sometimes improper will have no influence on the argument. The provability of these sentences involving ρ and σ is required next:

(i) $\ulcorner \forall w\, \neg w = \rho(w)\urcorner$;

(ii) $\ulcorner \forall w \forall z'[z' \in \sigma(\iota(w)) \leftrightarrow [z' = w \vee z' = \rho(w)]]\urcorner$;

(iii) $\ulcorner \forall w \forall w'\, \neg \sigma(\iota(w)) = \iota(w')\urcorner$;

(iv) $\ulcorner \forall w\, \neg \sigma(\iota(w)) = w\urcorner$;

(v) $\ulcorner \forall w \forall w'[\sigma(\iota(w)) = \sigma(\iota(w')) \rightarrow w = w']\urcorner$.

The proof of (i) is clear by inspection of the definition of ρ. Sentence (ii) follows easily from the definition of σ by taking into account the fact that ρ is level and by applying a suitable instance of the axiom schema (II). Sentence (iii) is now a direct consequence of (i) and (ii) and the characteristic property of unit sets. To prove (iv), speaking informally, assume that $\sigma(\iota(w)) = w$. By (ii) we have $\forall z'[z' \in w \leftrightarrow [z' = w \vee z' = \rho(w)]]$; hence, $\neg w = \Lambda$. If $\neg \Lambda \in w$, then $\rho(w) = \Lambda$, and so $\Lambda \in w$; therefore, $\Lambda \in w$ and $\rho(w) = \Lambda$. On the other hand it is clear from the definition of ρ that

[3]For the case of the system of [1], this permutation was suggested by Bernays [1, p. 83] for an analogous purpose. Reiger also uses this permutation in [4].

$[\rho(w) = \Lambda \rightarrow \neg\Lambda \in w]$; thus a contradiction is reached. Finally to establish (v), assume that $\sigma(\iota(w)) = \sigma(\iota(w'))$ and $\neg w = w'$. From (ii) it follows at once that $w = \rho(w')$ and $w' = \rho(w)$. From the definition of ρ we see that there are only three possibilities for w, namely: Λ, $\iota(\Lambda)$, and $\iota(\iota(\Lambda))$. The corresponding values of $\rho(w)$, that is to say w', are by definition $\iota(\Lambda)$, $\iota(\iota(\Lambda))$, and Λ. In none of these three cases is the equality $w = \rho(w')$ correct. The desired conclusion now follows.

The intuitive content of the sentences (iii) and (v) above is that the class of objects of the form $\iota(w)$ and the class of objects of the form $\sigma(\iota(w))$ are two disjoint classes that are in a one-one correspondence by means of the mapping determined by σ. Whenever such a situation exists, the correspondence can always be extended to a permutation of all objects (which is actually an "involution", as we shall see). This remark brings us to the formal definition of the term τ:

$$\begin{aligned}&\ulcorner \mathbf{T}z[\exists w[x = \iota(w) \wedge z = \sigma(\iota(w))] \vee \\ &\exists w[x = \sigma(\iota(w)) \wedge z = \iota(w)] \vee \\ &[\neg\exists w[x = \iota(w) \vee x = \sigma(\iota(w))] \wedge z = x]]\urcorner.\end{aligned}$$

It is worthwhile to note that not only is τ a level term, but also the descriptive phrase is always used properly in this particular context.

From the foregoing provable sentences about σ, it is easy to deduce these two sentences:

$$\ulcorner \forall x\, \tau(\tau(x)) = x\urcorner;$$
$$\ulcorner \forall w\, \tau(\iota(w)) = \sigma(\iota(w))\urcorner.$$

The first implies that τ is a permutation; while the two together show us that if $\tau(y) = \iota(y)$, then $\tau(\tau(y)) = y = \tau(\iota(y)) = \sigma(\iota(y))$, which is impossible for any y by virtue of (iv) above. In other words, the negation of the sentence (**) is actually provable for this permutation τ. The proof of the relative consistency and independence of (*) is thus complete.

REFERENCES

[1] BERNAYS, P. A system of axiomatic set theory, Part VII. *Journal of Symbolic Logic*, Vol. 19 (1954), pp. 81–96.
[2] CHURCH, A. *Introduction to Mathematical Logic*. Princeton (1956), ix+378 pp.
[3] KALISH, D., and R. M. MONTAGUE. Remarks on descriptions and natural deduction. *Archiv für Mathematische Logik und Grundlagenforschung*, Vol. 3 (1957), pp. 50–73.
[4] REIGER, L. A contribution to Gödel's axiomatic set theory, I. *Czechoslovak Mathematical Journal*, Vol. 7 (82) (1957), pp. 323–357.
[5] SCOTT, D. Review of [4]. *Journal of Symbolic Logic*, Vol. 23 (1958), pp. 216 f.
[6] QUINE, W. V. New Foundations for Mathematical Logic. *From a logical point of view*, Harvard (1953), pp. 80–101.
[7] QUINE, W. V. *Mathematical Logic* (Revised Edition), Harvard (1952), xii+346 pp.
[8] WANG, H. A formal system of logic. *Journal of Symbolic Logic*, Vol. 15 (1950), pp. 25–32.