J Philos Logic (2012) 41:505–517
DOI 10.1007/s10992-011-9184-4

If Logic, Definitions and the Vicious Circle Principle

Jaakko Hintikka

Received: 10 August 2010 / Accepted: 14 March 2011 / Published online: 19 April 2011
© Springer Science+Business Media B.V. 2011

Abstract In a definition $(\forall x)((x \in r) \leftrightarrow \mathrm{D}[x])$ of the set r, the definiens D[x] must not depend on the definiendum r. This implies that all quantifiers in D[x] are independent of r and of $(\forall x)$. This cannot be implemented in the traditional first-order logic, but can be expressed in IF logic. Violations of such independence requirements are what created the typical paradoxes of set theory. Poincaré's Vicious Circle Principle was intended to bar such violations. Russell nevertheless misunderstood the principle; for him a set a can depend on another set b only if $(b \in a)$ or $(b \subseteq a)$. Likewise, the truth of an ordinary first-order sentence with the Gödel number of r is undefinable in Tarki's sense because the quantifiers of the definiens depend unavoidably on r.

Keywords (In)dependence · IF logic · Definitions · Vicious circle principle · Truth-definition

1 Correcting Frege's Mistake

Gottlob Frege is usually credited with having constructed (discovered and codified) the central part of our common logic, the logic of quantifiers (plus propositional connectives). If so, he did not finish the job. He left at least ones serious flaw in his construct. This flaw was not a random mistake. It is due to Frege's failure to understand fully the nature of the most important ingredients of his edifice, the (existential and universal) quantifiers. This mistake did not only affect Frege's own work. Subsequent logicians have until recently repeated his mistake. This paper will

J. Hintikka (✉)
Department of Philosophy, Boston University,
745 Commonwealth ave., Boston, MA 02215-1401, USA
e-mail: hintikka@bu.edu

J. Hintikka
Collegium for Advanced Studies, University of Helsinki,
Fabianinkatu 24, P.O. Box 4, 00014 Helsinki, Finland

show through case studies how logicians' continued failure to recognize the mistake has led to serious gaps in the contemporary logical theory and misguided the development of both logical theory and set theory.

What was the aspect of the meaning of quantifiers that Frege did not appreciate? What is the semantical job description of quantifiers, anyway? The answer that one is likely to receive from any logician or philosopher is that they range over a class of values. An existentially quantified sentence is like a disjunction of its substitution-instances and a universally quantified sentence is like the analogous conjunction.

This "ranging over" is obviously part of what quantifiers do. But is it their whole job? Frege thought so, and expressed his thought in the form of his interpretation of quantifiers as higher-order predicates expressing the nonemptiness and the exceptionlessness of lower-order predicates. The same view is sometimes regimented into the so-called substitutional interpretation of quantifiers, an interpretation endorsed in so many words for instance by Saul Kripke. The same idea is built into Tarski-type truth-definitions and even into most of the theories of so-called generalized quantifiers.

Yet the "ranging over" idea covers only a part of the semantical function of quantifiers. They have another, more intriguing task. By the formal dependence of a quantifier (Q_2y) on another one, say (Q_1x), we can express the actual ("material") dependence of the variable y on the variable x. Such dependence relations are vital to science and highly important in everyday life. In logical terms, they are expressed by the Skolem functions of the given sentence. The list of arguments of a Skolem function corresponding, say, to a quantifier ($\exists x$) shows what other variables x (and what constants) depends on in that context. Quantifiers can, and arguably should, be considered as being little more than deputies for the corresponding Skolem functions. Purely logically, the role of these dependence relations manifests itself among other ways in the importance of quantifier ordering, for example in the so-called epsilon-delta definitions of sundry mathematical concepts. It is revealing that Frege never (as far as I know) paid any serious attention to such definitions even though they could have offered him a most impressive example of nontrivial uses of the logic of his own *Begriffsschrift.* At the same time, ironically, if Frege had undertaken this prominent application of his logic in full awareness of Weierstrass's work, he would have been forced to realize its inadequacy. For Weierstrass had defined and was using, over and above the usual concepts of convergence, continuity, differentiability etc., the corresponding uniformity concepts (uniform convergence, uniform continuity etc.). These are important in analysis (see e.g. Bann in Grattan-Guinness [1], sec. 3.12) but cannot be defined in Frege's logic or for that matter in our usual first-order logic.

It is obvious that the questions of dependence and ordering among quantifiers are independent (sic) of their ranges. We could for instance have a many-sorted quantification theory with different variables ranging over different classes of values and yet raise all the same questions of order and dependence as you raise about one-sorted quantifiers. Even if you are a segregationist and use different quantifiers for men and women, there still is a difference in meaning between "some man loves every woman" and "every woman is loved by some man".

But how do the formal relations of dependence and independence of (expressions of) quantifiers manifest themselves in the usual symbolic logic? In the kind of logic

Frege founded, the only way of expressing the independence of a quantifier (Q_2y) or (Q_1x) is for it to occur in the formal scope of (Q_1x), usually expressed by a pair of brackets

$$(Q_1x)(\text{——}\ (Q_2y)(\text{———})\text{———}) \tag{1.1}$$

But it is easily seen that not all patterns of dependence and independence between quantifiers can be expressed in this way. (The relation of scope inclusion has properties that dependence relations need not have.) This failure is a formal symptom of Frege's mistake. It is corrected in what is known as independence-friendly (IF) logic. There the independence of (Q_2y) of a quantifier (Q_1x) within whose formal scope it occurs is expressed by writing it (Q_2y/Q_1x). Likewise a quantifier like (Qz/b) is independent of b. In the game-theoretical semantics, where formal dependence is construed as informational dependence, this means that the choice of the value of z is made in ignorance of which individual b is. If needed, the dependence of (Q_3z) on (Q_1x) can be indicated by writing it $(Q_3z//Q_1x)$. It is easy to see how we can dispense with the single slash/if we can use the double one//, and vice versa. In principle, we could even forego new symbols altogether by liberating the use of parentheses.

In terms of the Skolem function f corresponding to a quantifier (Q_2y) in a context, the dependence of another one (Q_1x) is shown by the presence of x among the arguments of f, and independence by its absence. This shows that we must also take into account a possible dependence of (Q_2y) on a constant b. Such dependence is again expressed by the presence of b among the arguments of f. (For a full explanation, see below.) The awkwardness of the usual first-order logic is illustrated by the fact that, since Skolem functions are not notationally expressed, this distinction is not expressed in any simple way.

A semantics for IF logic is obtained from the usual game-theoretical semantics by modelling dependence relations between variables as informational dependencies between the moves prompted by their respective quantifiers. For instance, a variable y bound to a quantifier (Q_2y) depends on a variable x bound to (Q_1x) iff the move prompted by (Q_2y) is made by a player who knows what happened at the move prompted by (Q_1x). Ignorance of that move can be expressed by (Q_2y/Q_1x), no matter what the formal scopes of (Q_1x) and (Q_2y) are.

Thus game-theoretical semantics assigns a meaning to any slashed or unslashed quantifier. In what is generally known as IF logic, this possibility is availed of only in the case of the (in)dependence of existential force quantifiers on universal force quantifiers. However, no further explanations are need in the case of other kinds of (in)dependence relations between quantifiers. Such extensions of the usual IF logic are explored in Hintikka [6] and in Hintikka and Symons [7].

Likewise a move prompted by a quantifier (Q_ox) can be made in a semantical game in ignorance of what "b" refers to. This makes a difference. For instance,

$$(\exists y/b)(\exists u/a)F[a, y, b, u] \tag{1.2}$$

differs from

$$(\exists y)(\exists u)F[a, y, b, u] \tag{1.3}$$

in the same way as

$$(\forall x)(\forall z)(\exists y/\forall z)(\exists u/\forall x)F[x,y,z,u] \tag{1.4}$$

differs from

$$(\forall x)(\forall z)(\exists y)(\exists u)F[x,y,z,u]. \tag{1.5}$$

The Skolem functions of $(\exists y)$ in the four cases can be $f(a)$, $f(a,b)$, $f(x)$, $f(x,z)$, and correspondingly for $(\exists u)$. It is easy to find interpretations of F which result in different truth-values for (1.2) and (1.3).

Skolem functions are game-theoretically considered choice functions guiding the players' selections of suitable values of variables. Hence, complete arrays of Skolem functions for a IF sentence S codify strategies for the verifier in the correlated semantical game $G(S)$ with S. In IF logic in the usual narrow sense we need to heed only the verifier's strategies. The falsifier can use any strategy. Hence in this case no guidance in the form of choice functions need to be considered.

In general the independence of quantifiers of constant can be expressed in terms of quantifiers only. Thus

$$(\exists x/b)F[x,b] \tag{1.6}$$

is equivalent to

$$(\exists y)((y=b)\&(\exists x/\exists y)F[x,y]). \tag{1.7}$$

Likewise

$$(\forall x/b)F[x,b] \tag{1.8}$$

is equivalent to

$$(\forall y)((y=b)\supset(\forall x/\forall y)F[x,y]). \tag{1.9}$$

This illustrates the need (possibility) of going beyond IF logic in narrow sense by considering dependencies between similar quantifiers (existential on existential ones and universal on universal ones).

Similar remarks apply to propositional connectives. Their semantics, too, is determined by their role in semantical games. They, too, can be dependent or independent of each other and/or of quantifiers, and vice versa. In practice, the most important consideration is to distinguish between the strong (dual) negation ~and the contradictory negation. Since the law of excluded middle does not hold in IF logic for ~, we must allow a sentence (except for identities) to be neither true nor false.

For other aspects of IF logic, see e.g. Hintikka [5] as well as Sandu et al. (forthcoming). It is not an alternative to the traditional (Fregean) first-order logic or a special development of this traditional logic, but an improved version of it. Its details are not relevant here. The continued use of the received first-order logic rather than IF logic perpetuates Frege's mistake.

IF logic, suitably extended, is a much richer logic than the received first-order logic or its godfather, Frege's *Begriffsschrift*. For instance, it is IF logic that is needed for the definition of uniformity concepts. (The logic Weierstrass was using in his mathematical practice was therefore IF logic rather than our usual first-order logic.)

2 Theory of Definition

Does quantifier independence matter? One area where it obviously does is the theory of definition. One of its cornerstones is indeed the noncircularity requirement that says that the definiens must not depend in any way on the definiendum. It is virtually obvious that this requirement cannot be implemented by means of traditional first-order logic but requires IF logic or sometimes the equivalent.

We can take as a simple example an attempted definition of an individual b by means of first order logic. Then it must be required that in an explicit definition of the form

$$(\forall x)((x = b) \leftrightarrow D[x]) \tag{2.1}$$

the definiens $D[x]$ must be independent of the definendum b. For this purpose it must be further required that no expression for b may occur in $D[x]$. Usually, this is taken to be not only a necessary but a sufficient condition of noncircularity.

In a definition like (2.1) the equivalence sign $\leftrightarrow$ must express the identity of truth values (true, false or indefinite) lest instances of *tertium non datur* are smuggled in. Hence $(A \leftrightarrow B)$ will express the same as $(A \,\&\, B) \vee (\sim A \,\&\, \sim B) \vee (\neg A \,\&\, \neg\sim A \,\&\, \neg B \,\&\, \neg\sim B)$. (Similar remarks apply to the use of $\leftrightarrow$ below.)

However, it is perfectly obvious that this notational independence is not enough to make the definens $D[x]$ independent in all relevant respects of the definiendum b. For one thing, all the quantifiers in $D[x]$ must be independent of b. In the usual first-order treatment of definition, not only is this requirement left unheeded. The requirement cannot always be implemented without IF logic or equivalent.

This example shows that the usual first-order treatments of the theory of definitions are seriously incomplete.

This prompts the question of how so gross a defect has not been pointed out (or at least not pointedly enough). A possible (but lame) excuse might be perhaps that quantifiers have tacitly been taken to be independent of constants. However, this does not help very much, for then the newly defined individuals do not obey the usual logical laws. For instance, from (2.1) one cannot infer

$$(\exists y)(\forall x)((x = y) \leftrightarrow D[x]) \tag{2.2}$$

For even if the quantifiers in $D[x]$ can be taken to be independent of b in (2.1), in (2.2) they cannot according to the usual received notation be taken to be independent of $(\exists y)$.

This sort of independence and independence requirement will be said to be of the first kind. It is not only the only kind of independence that has to be required of definitions. In (2.1) not only must the quantifiers in $D[x]$ be independent of b. Clearly, $(\forall x)$ must also be independent of b. This will be called independence of the second kind.

This kind of independence must be required because otherwise the attempted definition would be blatantly circular. We would be trying to define (as it were an unknown individual) b by specifying what it itself must be like. The need of requiring the second kind of independence also is seen from the usual truth-definitions for first-order logic. They show that one of the substitution-values of the definiens on which its truth-value depends is $D[b]$

The proper form of first-order definitions of the individual b is therefore, not (2.1) but

$$(\forall x/b)((x = b) \leftrightarrow D^*[x]) \tag{2.3}$$

where $D^*[x]$ is like $D[x]$ except that all quantifiers are independent of b.

There is still a third kind of independence relevant to definability. It has to be required that in (2.3) all the quantifiers must also be independent of $(\forall x)$. This is the subtlest of the three kinds of independence. Formally, this requirement is almost obvious. From (2.3) it is seen that in order for the definition to be applicable, it must be true that

$$(\forall x/b)((x = b) \supset D^*[b]) \tag{2.4}$$

Now if $D^*[x]$ there were any quantifiers that depend on $(\forall x)$, by the substitutivity of identicals they must in $D^*[b]$ depend on b. This would violate the noncircularity requirement.

The third independence requirement can be made intuitive by reference to game-theoretical semantics. There the Skolem functions of a sentence S tell us how to find the "witness individuals" that show the truth of S. The arguments of these Skolem functions are the individuals that need to be known at the time of an application of the function. Now if in $D^*[x]$ some of the Skolem functions depend on the value of x being known, then the witness individuals showing the truth of $D^*[b]$ depend on the definiendum individual b's being known. But this would be blatantly circular.

This third independence requirement has wider consequences. The first two requirements affect only definitions of individuals. The third one applies also to definitions of predicates and functions. For instance, according to the third requirement in a definition of the form

$$(\forall x)\,(A(x) \supset D^*[x]) \tag{2.4}$$

all the quantifiers in the definiens $D^*[x]$ must be independent of $(\forall x)$.

None of these different independence requirements can be satisfied without IF logic (or equivalent). IF logic is therefore indispensable for a satisfactory logical theory of definition. This is especially important in set theory. There sets are treated as individuals (values of first-order quantifiers), wherefore all three kinds of independence requirements apply.

The three independence requirements are different. Yet they have a common root in the idea of independence of the kind that can only be expressed in IF logic. Together they show that the usual expositions of definitions and definability in first-order logic are seriously defective. Its defects are automatically removed by using IF logic instead of the received first-order logic.

All these remarks pertain to attempted definitions in terms of first-order logic. The notion of independence nevertheless applies to quantifiers (and other logical notions) in any language. Hence the independence requirements formulated in this section are fully relevant also to definitions formulated in natural language or in any other notation. Needless to say, an application of these principles may be complicated by the intricacies of natural language.

3 Defining Truth

These observations are not only relevant to a generic theory of definition. They have application to the definability of different philosophically interesting concepts.

One of the most interesting such concepts is the notion of truth. For seventy-some years, philosophers have been laboring in a theoretically important sense under the dogma that truth is not definable for a formal language in the same language. This is supposed to have been proved strictly by Alfred Tarski. In reality, all that Tarski strictly proved is that truth for a first-order theory using the traditional first-order logic cannot be defined in that same theory.

Tarski also claimed that the concept of truth cannot be used coherently in an actual "colloquial" language. This claim will be considered below at the end of this section.

But this undefinability is puzzling even as a technical result. It prompts questions that Tarski does not answer. What Tarski is dealing with are interpreted first-order languages that are self-applied, perhaps by means of a technique like Gödel numbering. A paradigmatic case in point is the language of elementary arithmetic. Defining truth for such a language means defining a number-theoretical predicate $T(x)$ that applies to the Gödel number $g(S)$ of a sentence S if and only if S is true. Tarski expressed this idea by means of what he calls the T-schema. It says that a truth predicate T applies to a Gödel number $g(S)$ of sentence S if and only if S, formally expressed

$$(T) \qquad T(g(S)) \leftrightarrow S$$

Tarski requires that a satisfactory truth definition must satisfy all the particular instances of (T). So why cannot we simply generalize universally in (T) and turn it into a truth definition? Tarski never tells us, and subsequent logicians and philosophers have not done much better.

What makes this silence intriguing is that there does not seem to be any particular problem in constructing a truth definition for instance for the language of elementary arithmetic by converting (T) into a proper definition. In an elementary arithmetic S is a numerical sentence that can be constructed arithmetically from its Gödel number $g(S)$. Hence S can be thought of as being of the form $S[h(g(S))]$ where $h(x)$ is an arithmetical function. Hence it might seem as if we could turn (T) into a definition, e.g.

$$(\forall x)((x = g(S)) \supset (T(x) \leftrightarrow S[h(x)])) \tag{3.1}$$

Here $h(x)$ is like the Skolem function corresponding to a quantifier. Having x as one of its arguments violates our third independence requirement. In other words, the quantifiers that could do the job of $h(x)$ would have to depend on $(\forall x)$.

This violation of the third independence requirement is the true reason why the T-schema (T) cannot be converted into a truth definition and more generally why a truth definition for a first-order language using the traditional first-order logic cannot be formulated in the same language. This impossibility can therefore be said to be but a consequence of Frege's mistake. This reason has for instance nothing directly to do with paradoxes like the paradox of the liar.

The other side of the same conceptual coin is that as soon as the third independence requirement can be satisfied, truth becomes definable in the sense

Tarski and others have been dealing with. This can happen by means of second-order logic (cf. below). It can also happen on the first-order level simply by using IF logic.

The third requirement can normally be satisfied as soon as the language in question is rich enough to express independence. There is no reason to think that this cannot be done in ordinary language. Hence there is, Tarski notwithstanding, no reason to think that we cannot use the concept of truth completely consistently in ordinary discourse (or in philosophical discourse) and if needed to define it.

4 Set-Theoretical Paradoxes and the Notion of Independence

The notions of dependence and independence have also played an important albeit largely tacit role in the history of set theory. The best known paradoxes of set theory can be seen as consequences of Frege's mistake. Take, for instance, Russell's paradox [3]. It arises from an attempt to define the set r of all sets that are not members of themselves. Russell seems to have thought initially that the existence of such a set is a truth of "common sense". The definition of r would in the logic Russell used be of the form

$$(\forall x)((x \in r) \leftrightarrow \neg(x \in x)) \tag{4.1}$$

A contradiction nevertheless results by substituting r for x.

Now (4.1) violates the second kind of independence requirement. (Strictly speaking, it violates an obvious extension of this requirement from definitions of individuals to definitions of sets.) A paradox-free form of the definition of r would be

$$(\forall x/r)((x \in r) \leftrightarrow \sim(x \in x)) \tag{4.2}$$

This sentence is not false in all models. It is not false in a model in which ($r \epsilon r$) is neither true nor false, as it can be in IF logic.

It was not clear at first to Russell what is wrong with (4.1). On the surface there are other prima facie explanations of the paradox. One can purely formally escape the contradiction by exempting r from the range of the variable x and thereby barring the substitution of r for x in (4.1). The problem with this kind of suggestion is that it does not seem to be in keeping with the usual semantics of first-order logic.

Russell thought that there is some kind of circularity in (4.1). But how that diagnostic idea can be spelled out is not clear, at least it was not initially clear to Russell.

Other paradoxes and other threats of contradiction emerged in the same period in the history of set theory. Some of the most striking of them can likewise be traced back in the light of hindsight to violations of sundry independence requirements. Russell and others tried to find the greatest common denominator in them, in order to be able to eliminate the problems. This can be seen as having been made difficult (again in the light of much later insights reported above) by the difference between different independence requirements. For instance, in (4.1) there are no quantifiers in the definiens, and hence the first and third independence requirements do not come into play. This makes Russell's paradox less than a fully instructive test case in the study of set-theoretical paradoxes. Furthermore, the fact that in ordinary language the exclusive reading of quantifiers is usually presupposed tends to make (4.1) look

even more natural, and also encourages a solution that turns on excluding r from the range of the quantifier $(\forall x)$.

Another paradox that played a role in the history of set theory is Richard's paradox. Poincaré [10] explains it as follows:

> Let us consider all the decimal numbers that can be defined with the help of a finite number of words. These decimal numbers form an aggregate [set] E, and it is easy to see that this [set] is denumerable—that is to say, it is possible to *number* the decimal numbers of this [set] from one to [countable] infinity.

Poincaré then points out how by reference to E we can define a sequence of natural numbers that is the decimal expansion of a number N not in E. N is thus definable and not definable.

Contrary to what some writers have claimed, it does not matter that the definition that Poincaré is apparently talking about is formulated in ordinary language (in "words," as Poincaré puts it.) Essentially the same argument can be carried out in a formalized set theory that is self-applied so that we can in it discuss in it its own syntax and semantics, for instance by means of some variant of Gödel numbering.

What is wrong with Richard's argument? Poincaré presents what he calls "the true solution" [10, pp.189-190, 9, pp.307-308 and 11] paradoxically crediting it to Richard himself. The purported solution (slightly paraphrased) goes as follows (cf. the statement of the paradox above): The set E is to be taken as the set of all numbers that can be defined by a finite number of words *without introducing the notion of the set E itself.* Otherwise the definition of E would contain a vicious circle, for we cannot define E by means of the set E itself. He adds that the same solution works for other paradoxes.

Poincaré is here ruling out from all definitions some kind of circularity that he calls a vicious circle. This is supposed to disarm Richard's paradox and presumably others as well. Poincaré's principle came to be known as the Vicious Circle principle (VCP) and came to be considered as a key to the solution of the paradoxes of set theory, especially by Russell.

5 Vicious Circle Principle

But what precisely is this principle? Definitions violating it are also called impredicative. It is still to-day unclear in the literature what predicativity is. An acknowledged expert on this subject, Solomon Feferman, writes about predicativity as follows

> While this term suggests that there is a single idea involved, what history will show is that there are a number of ideas of predicativity which may lead to different analyses. ([2], p. 590)

What is suggested here is that, Feferman notwithstanding, there is a single idea involved in the sundry manifestations of impredicativity, viz. illicit variable dependence, in other words, a violation of one of our independence requirements. Admittedly, this idea surfaces in different variants in different set-theoretical paradoxes, like the different independence requirements in sec. 2.

It has also been a difficult idea for logicians to grasp. As Poincaré's contemporaries were quick to point out, there is no circularity in the definitions he was discussing according to the current concepts of definability. It is in fact apparently hard to fault either the definition of E or the definition of N (given E). For instance, there need not be in the definition of E any term that introduces E itself into it.

But what has been emphasized in this paper is that an object can be tacitly "introduced" into a definiens in another way, viz by quantifiers that depend on it. When the definendum is "introduced" into a definiens in this way, we have a violation of the first independence requirement. The word "introduce" is in fact strikingly apt here. The meaning of quantifiers is determined by their role in semantical games. In these games, new individuals are introduced as substitution-values of variables that are bound to the quantifier that prompts the introductory move. If quantifier is not independent of a constant r, a move in a semantical game can literally introduce r into a play of the game

It seems to me overwhelmingly natural to take Poincaré's own interpretation of the VCP to amount precisely to a violation of the first independence requirement. This reading of VCP will be called the Poincaré interpretation. Unfortunately, he did not try to formulate his point by using a precise logical term, largely because such terms were not routinely available to Poincaré. A critic might say that this failure to spell out his own point belies his claim of the fruitlessness of symbolic logic. On the other hand, Poincaré's rejection of the Frege-type notation in which not enough attention is paid to relations of quantifier dependence and independence probably helped him to see the conceptual situation more accurately.

Strictly understood and applied, VCP can only deal with those paradoxes that turn on a violation of the first independence requirement. As we saw, not all the others have this origin. For instance, Russell's paradox is due to a violation of the second independence requirement.

What happened was that in effect, Russell and others tried to use their version of VCP to resolve the paradoxes other than Richard's and put set theory and logic to a safe basis. In this process, VCP was understood in ways different from Poincaré's but that it nevertheless was assigned the key role in the resolution of what Russell called "vicious circle fallacies".This difference in the way Poincaré and Russell understood the VCP is not belied by the fact that Russell verbally accepted Poincaré's principle in his [12]. The difference between their respective interpretations is shown clearly among other things by the fact that Poincaré's remarks concern the definition of the set E of definable numbers, whereas Russell blames the paradox on the definition of N (see [13], p.61). However, this explanation does not resolve other paradoxes, at least not directly.

The analysis of the independence requirements or definitions in section 2 above enables us to put developments into a perspective. Richard's paradox is solved by Poincaré's version of VCP because it is due to a violation of the first independence requirement. In the condition of membership in E, the definition of its members must contain only quantifiers independent of E. The definition of N does not satisfy this requirement.

But Russell, like Frege, did not understand the idea that a quantifier could depend informationally on a set. He understood only two other ways in which a set S_2 can

depend on another set S_1: Either S_1 is a member of S_2 or it is a subset of S_2. Accordingly, when a set S_o defined by a set-valued quantifier (Qx), S_o must not be a possible value or a subset of x. As Russell and Whitehead put it ([13], p. 37): "Whatever involves *all* of a collection, must not be one of the collection." In a more modern terminology, Russell's version of VCP says that a set can only be defined by means of quantifiers if it not itself a (potential) value of those quantifiers. This will be called Russell's interpretation of VCP. It seems to have been how most others, too, interpreted VCP. It calls for several comments.

One kind of evidence for the prevalence of Russell's interpretation are the criticisms that were leveled against VCP. On Russell's interpretation, critics pointed out that VCP seems to rule out far too much. Definitions of mathematical objects in terms of quantifiers one of whose potential values they are not uncommon in mathematics, as Peano (1973, p. 214) and Zermelo (van Heijenoort 1967, pp 190-191) were quick to point out.

What VCP does on Russell's interpretation is to force him to adopt his theory of types, simple as well as ramified. For according to it, an object cannot be defined by means of another entity of a higher order. In this theory, the universe of discourse as well as its quantifiers is stratified into types and orders. A set of a given type can only have members of a lower type, and a quantifier of order n could draw its values from objects of a lower order, at most $(n\text{-}1)$. Thus Russell's interpretation of VCP led him into his ramified type theory.

This theory cannot be deemed to have been successful. Russell's introduction of his Axiom of Reducibility is a tacit acknowledgement of the validity of Peano's and Zermelo's criticism of Russell's interpretation VCP.

Russell's interpretation is misguided also in another respect. A typical violation of VCP is the dependence of a quantifier in the definiens on the definiendum: The definendum occurs as one of the arguments of the Skolem function introduced by the quantifier. If instead the Skolem function is introduced by a higher order quantifier (function quantifier), the arguments of the function introduced can be chosen freely, and hence do not have to include the forbidden definendum.

Hence Russell was diametrically wrong. His version of VCP forbids definitions of entities in terms of higher-type quantifiers. Such definitions can be quite innocent. For instance, truth (the set of true sentences) can be defined by reference to the existence of Skolem functions, while in reality the problems arise from use of same-type quantifiers.

Looking back at the line of thought carried out in this paper, it can be said that Richard's paradox exemplifies the first independence requirement, Russell's paradox the second one and the Tarski-type undefinability truth the third requirement.

In the perspective on the history of set theory suggested by these observations, the entire tradition of type theory (especially the ramified type theory) came about because Russell misunderstood the import of Poincaré's VCP. And this misunderstanding is but a variant of Frege's mistake of neglecting the dependence-expressing function of quantifiers which can thus be blamed for the main paradoxes of set theory.

The future of set theory is going to depend on the use of methods that do justice to this function, not on attempts to repair the details of Frege's and Russell's particular ideas

6 Truth and First-Order Set Theories

These results have implications that go beyond the problematic centered around the VCP and the developments in set theory it prompted.

Among other things, they also indicate the reason why it is easy to define truth for a first-order theory by means of second-order quantifiers. The argument of such variables gives us a way of satisfying the third independence requirement by explicitly listing the first-order variables they depend on (or are independent of).

Indeed, it is easy to define truth for an interpreted first-order language by means of the corresponding second-order logic. It is also easily seen that the resulting definition belongs to the $\sum_1^1$ fragment of second-order logic. For instance, the truth definition for *S* can say that there exists a full set of Skolem functions for S. This applies also to an IF first-order language. But the $\sum_1^1$ fragment of second-order logic can be translated into the corresponding first-order IF language, which therefore allows for truth definitions for itself and for the traditional first-order languages. Tarski's undefinability result holds for traditional first-order theories only because the traditional first-order logic does not allow the representation of all the relevant independence relations between quantifiers. In brief, it holds only in virtue of Frege's mistake.

But this poses a striking problem. If truth can be defined for a language by means of higher-order quantifiers, should it not be definable in an axiomatic set theory? Truth was definable for first-order languages in second-order language because the values of higher-order function variables, have each a definite set of arguments. This includes Skolem functions, which means that no extra arguments can be smuggled in by formal rules interpretation so as to violate the third independence requirement.

This question has a sharp edge. There does not seem for instance to be any reason why a second-order definition of truth in terms of the existence of Skolem functions should not be expressible in the language of set theory using traditional logic, without resorting to IF logic. And yet Tarski's impossibility theorem shows that one cannot define truth for a first-order set theory just because it is first-order. We do in fact have a real problem here. The key to understanding what is involved is to recognize that the paradox comes about if the variables of axiomatized set theory are taken to be actual higher-order entities, sets, functions etc., in the normal model-theoretical sense of collections of members of the universe of discourse, classes of ordered pairs and a unique second member, etc. If we could insist that the language of a first-order set theory be interpreted in this way, then there would exist a valid truth definition. Since this is impossible, first-order axiomatic set theories cannot be so interpreted. Not to mince my words, these "set theories" are not about sets. The statements they use cannot consistently be interpreted to be about sets in the literal, realistic sense. Since the models of first-order theories are structures of particulars, such a theory simply is not a theory of sets.

The possible implications of this insight for mathematical practice cannot be adequately discussed here. Perhaps, it is nevertheless not premature to take this theoretical consideration as an illustration of the mess into which Frege's mistake has led set theorists.

Using set theory as a test case nevertheless helps to answer an overarching historical question. It might appear that when everything is said but not yet done, it

would have been almost trivially easy to avoid the whole problematic discussed in this paper simply by recognizing Frege's mistake and slightly liberating our logical notion. That would have meant using IF first-order logic rather than the Frege-Russell one, but it was noted in sec.1 how mathematicians had practically done so already. (Didn't Russell's Tripos examination cover anything that involved uniformity concepts?)

Even though a switch from the traditional Frege-Russell first-order logic thus seems from a logical point of view to be a small step, for a set theorist who wants to actually apply her or his theory, it would loom very large. The most eye-catching characteristic of IF logic is the failure of *tertium non datur*. This failure means that we have to accommodate unsharp concepts, including unsharp sets. But this would have meant modifying the very concept of set. This concept had been characterized by Cantor precisely in terms of a sharp membership relation. A set S is something such that one can in principle tell about any object x whether it belongs to S or not. Even if there are fuzzy sets, there surely must in any case be sharp sets that we can theorize about, or so it easily seems. It is indeed a remarkable result that we apparently cannot build a satisfactory theory of sharp sets without introducing into it unsharp sets, too. This prompts other questions. Is the conceptual situation the same with other categories? Can we develop a satisfactory theory of functions in which variables do not always have a sharp value?

Perhaps the need for unsharp sets makes it psychologically and historically understandable that logicians and mathematicians did not know how to implement Poincaré's insight.

References

1. Bann, R. (1980). Developments in the foundations of mathematics 1870-1910. In I. GrattanGuinness (Ed.), *From calculus to set theory* (pp. 220–255). Princeton: Princeton U.P.
2. Feferman, S. (2005). Predicativity. In S. Shapiro (Ed.), *Oxford handbook of philosophy of mathematics and logic* (pp. 590–624). New York: Oxford U.P.
3. Garciadego, A. R. (1992). *Bertrand Russell and the origins of set-theoretical paradoxes*. Basel: Birkhäuser.
4. Heijenoort, Jean van, 1967, *From Frege to Gödel. A Source Book in Mathematical Logic*. Harvard University Press, Cambridge, MA.
5. Hintikka, J. (1996). *Principles of mathematics revisited*. Cambridge: Cambridge U.P.
6. Hintikka, J. (2009). A proof of nominalism: An exercise in successful reduction in logic. In A. Hieke & H. Leitgeb (Eds.), *Reduction—abstraction—analysis* (pp. 1–14). Heusenstam: Ontos Verlag.
7. Hintikka, J., & Symons, J. (forthcoming). Game-theoretical semantics as a basis ofgeneral logic.
8. Peano, G. (1973). In H. Kennedy (Ed.), *Selected Works*. Toronto: University of Toronto Press.
9. Poincaré, H. (1905-06) Les mathématiques et la logique Part II, *Revue métaphysique etmorale*. 13: pp. 815–835 and 14: 17–34, 294–317.
10. Poincaré, H. (1908) *Science et Méthode,* E. Flammarion, Paris. In English as *Science andMethod*, Dover, New York, 1952.
11. Poincaré, H. (1909). Reflexions sur les deux notes precedents. *Acta Mathematica, 32*, 195–200.
12. Russell, B. (1906). Les paradoxes de la logique. *Revue de métaphysique et de morale, 14*, 627–650.
13. Russell, B., & Whitehead, A. N. (1910). *Principia mathematica* (Vol. 1). Cambridge: Cambridge U.P.
14. Sandu, G., Mann, A. L., & Sevenster, M. (forthcoming). *Independence-friendlyLogic: A Game-theoretical Approach.* Cambridge U.P., Cambridge.