THE REVIEW OF SYMBOLIC LOGIC
Volume 9, Number 3, September 2016

VON NEUMANN'S CONSISTENCY PROOF

LUCA BELLOTTI
Department CFS (Philosophy), University of Pisa

Abstract. We consider the consistency proof for a weak fragment of arithmetic published by von Neumann in 1927. This proof is rather neglected in the literature on the history of consistency proofs in the Hilbert school. We explain von Neumann's proof and argue that it fills a gap between Hilbert's consistency proofs for the so-called elementary calculus of free variables with a successor and a predecessor function and Ackermann's consistency proof for second-order primitive recursive arithmetic. In particular, von Neumann's proof is the first rigorous proof of the consistency of an axiomatization of the first-order theory of a successor function.

§1. Introduction. In 1925, the twenty-one-year-old John von Neumann (then still Neumann János) wrote and submitted to the *Mathematische Zeitschrift* a long paper, *Zur Hilbertschen Beweistheorie*. The paper would have been published two years later (Von Neumann 1927; it is reprinted in Von Neumann 1961, 256–302). The paper contains a consistency proof for a fragment of first-order arithmetic by a variant of Hilbert's Substitution Method.

An in-depth account of this work of von Neumann's is lacking in the current literature on consistency proofs and the ε-Substitution Method. In this paper we try to fill this gap. We shall consider the following questions. What is the system of which von Neumann proves consistency? How does it relate to the systems treated by Hilbert and by Ackermann? What are the distinctive features of von Neumann's Substitution Method? In which meta-theory can von Neumann's proof be carried out?

Von Neumann's consistency proof is for a system containing full predicate logic and axioms for a successor function. This fills an apparent gap in the commonly accepted historical account of the development of consistency proofs and the Substitution Method. Von Neumann's paper also contains a clear and insightful presentation of Hilbert's proof-theoretic approach (*Ansatz*) and a detailed critique of the most important consistency proof produced until then in the Hilbert school, namely Ackermann's proof (Ackermann 1925). Moreover, von Neumann gives a very precise (although slightly peculiar) definition of the general notion of a formal system (using—probably for the first time in the literature—axiom schemas), a rigorous delimitation of the specific formal system treated, together with many interesting side remarks (e.g., on decidability, on choice principles, on definitions by recursion, etc.). We herein concentrate strictly on the proof, and we do not attempt to address all the other themes and aspects of this very rich work of von Neumann's.[1]

Received: May 15, 2014.

[1] This holds, in particular, for the interesting topic of a possible comparison between Von Neumann's prudential attitude on the *Entscheidungsproblem* shown in this paper (1927, 11 ff.), and Bernays'—and especially Hilbert's—optimism on the problem in those years: this would

doi:10.1017/S1755020316000198

The present paper is organized as follows. In Section 2 we review the literature on von Neumann's proof. In Section 3 we attempt to locate von Neumann's proof in the commonly accepted historical scheme of the development of consistency proofs in the Hilbert school. In Section 4 we present the details of the formal system of which von Neumann proves consistency. In Section 5 we give a detailed description of von Neumann's proof.

§2. Review of the literature. A lot of scattered references to von Neumann's paper exist in the literature. One is even (in a footnote) on the first page of Gödel's incompleteness paper. There is an early critical comment on von Neumann's symbolism in Leśniewski (1929), to which von Neumann replied in (1931); there is a brief account of his result, without details, in Heyting (1955, 52–54), and a somewhat cryptic one in Ulam's obituary of von Neumann (1958, 13–14). Ackermann briefly reviewed the work in the *Jahrbuch über die Fortschritte der Mathematik*.

Concerning the more recent historical literature one can observe that a satisfactory account of von Neumann's proof is missing. One reason is probably that, as Zach writes in (2003, 226),

> Von Neumann's paper of 1927, the only other major contribution to proof theory in the 1920s, does not entirely fit into the tradition of the Hilbert school, and we have no evidence of the extent of Hilbert's involvement in its writing.

Nevertheless, as we shall argue, von Neumann's proof is very close to the methods of the Hilbert school and in fact provides the first rigorous application of—essentially—the Substitution Method to the full predicate calculus extended by axioms for a successor function.

The merits of von Neumann's work are not totally overlooked in the literature. On the one hand, early recognition by Hilbert, Ackermann and Bernays of the importance of von Neumann's work is documented. In his address to the International Congress of Mathematicians in 1928 (Hilbert 1928, 1929), Hilbert writes that "the consistency proof of the ε-axiom for the natural numbers has been accomplished by the works of Ackermann and von Neumann". After reading von Neumann's paper, Ackermann acknowledged the problems in his own proof of ε-reduction and reworked his own proof using von Neumann's notion of *Grundtypus*.[2] On the other hand, partly because of the common misconceptions of that time about exactly which system had been proved to be consistent, partly because of the nonhistorical preoccupations of the working mathematicians involved in the endeavor of consistency proofs, partly also because of the fact that von Neumann was sort of an outsider to the Hilbert school, it is fair to say that no precise assessment of the merits of von Neumann's work and of its exact relationship to the work of Hilbert and Ackermann can be found in the early literature.

---

require a systematic study of the relevant texts (mainly some of the material collected in Hilbert 2013) which is beyond the scope of the present work.

[2] The following two excerpts from Mancosu–Zach–Badesa (2009) and Zach (2003) summarize the situation: "Ackermann continued to work on the proof, amending and correcting the ε-substitution procedure even for first-order ε-terms. These corrections used ideas of von Neumann (1927), which was completed already in 1925" (2009, 398). "In 1927, Ackermann developed a second proof of ε-substitution, using some of von Neumann's ideas (in particular, the notion of an ε-type, *Grundtyp*). The proof is unfortunately not preserved in its entirety, but references to it can be found in the correspondence" (2003, 242).

The recent historical literature also recognizes the importance of von Neumann's work but leaves it as an open problem to figure out what exactly are the reasons of this importance. The most thorough account of von Neumann's proof in the recent literature are the following two excerpts from Mancosu–Zach–Badesa (2009, 398 and fn. 89):

> Von Neumann (1927) used a different terminology from Ackermann, and the precise connection between Ackermann's and von Neumann's proofs is not clear. Von Neumann's system does not include the induction axiom explicitly, since induction can be proved once a suitable second-order apparatus is available. Hence, the consistency proof for the first-order fragment of his theory does not include induction, whereas Ackermann's system has an induction axiom in the form of the second ε-axiom, and his substitution procedure takes into account critical formulas of this second kind. Another significant feature of von Neumann's proof is the precision with which it is executed: von Neumann gives numerical bounds for the number of steps required until a solving substitution is found.
>
> Von Neumann (1927) is remarkable for a few other reasons. Not only is the consistency proof carried out with more precision than those of Ackermann, but so is the formulation of the underlying logical system. For instance, the set of well-formed formulas is given a clear inductive definition, application of a function to an argument is treated as an operation, and substitution is precisely defined. The notion of axiom system is defined in very general terms, by a rule which generates axioms (additionally, von Neumann remarks that the rules used in practice are such that it is decidable whether a given formula is an axiom). Some of these features von Neumann owes to König (1914).

In the present paper we intend to locate more precisely this work of von Neumann's within the historical scheme of the development of consistency proofs and of the Substitution Method proposed in Mancosu–Zach–Badesa (2009), to which we adhere.

Part of the difficulty in isolating the merits of von Neumann's proof comes from the confusion about the strength of the systems for which consistency had been proved. This confusion was very common at the time. In particular, the role of induction was somewhat underestimated. The simultaneous development of logic and mathematics that Hilbert promoted took forms that are not familiar nowadays. We will argue that von Neumann's work witnesses a very insightful understanding of the limits of the Substitution Method. Indeed the consistency proof does in full rigor as much as was possible at the time without using transfinite induction and without substantially new ideas (such as those introduced by Gentzen).

§3. Early consistency proofs. We briefly recall the stages of the development of consistency proofs in the first years of the 1920s. We adhere to the accounts presented in Zach (2003) and Mancosu–Zach–Badesa (2009).

We are particularly interested in Stage II and III of Hilbert's program as identified there. Stage II is a consistency proof for the so-called *elementary calculus of free variables* with successor and predecessor. This axiom system consists of axioms for propositional logic, identity axioms, and two axioms for a successor and a predecessor function: $a + 1 \neq 0$ and $\delta(a + 1) = a$ (for the full list of axioms see Zach 2003). The rules of inference are Modus Ponens and substitution for individual variables and formula variables.

Note that this system does not include quantifiers. As reported by Zach (2003), Hilbert's 1921/22 lecture notes contain a consistency proof for this system using a standard induction on the length of proofs. Stage III is a consistency proof for a system expanding the elementary calculus described above with definitions by primitive recursion and an induction rule. This system is also presented in Hilbert's lecture notes from 1921/22. The transition from this system to Ackermann's formulation of the ε-calculus is nicely described in Zach (2003). From our perspective it is interesting to observe that Hilbert's original formulation uses the τ-operator. This is a term-forming operator that applies to formulas (and derivatively to functions) and picks a counterexample. The τ-operator can be used to define quantifiers. If an axiom imposing that the counterexample is minimal is added then the τ-operator can also be used to derive the induction principle and the least number principle. Hilbert and Bernays later changed the notation—but not the meaning—of the τ-operator and used the ε-operator. Ackermann switched to the dual and now common meaning of the ε-operator (i.e., finding a witness).

It is important to observe in this context that Hilbert's (1923) and Hilbert–Bernays' (1923, 1923a) only contain a sketch of the consistency proof for the system with the substitution axioms. In particular, only the case in which the proof contains exactly one first-order ε-term $\varepsilon_a A(a)$ and the corresponding critical formula is treated explicitly. Zach (2003) identifies the following challenges left open after Hilbert's 1922/23 treatment.

1. Extend to cover more than one ε-term in the proof.
2. Take care of nested ε-terms.
3. Extend to second-order ε-terms.

These three points are those addressed by Ackermann in his 1924 dissertation (published in 1925). It is generally acknowledged that Ackermann only in part achieved his goals. In particular, while the treatment of point (3) above was partly successful, though to the limited extent of a consistency proof for a version of second-order PRA (essentially inaugurating the use of transfinite induction in consistency proofs), the treatment of the Substitution Method in its full generality (i.e., points (1) and (2) above) is known to be rather defective. Let us quote Zach (2003, 241) on this point.

> A preliminary assessment can, however, already be made on the basis of the outline of the substitution process above. Modulo some needed clarification in the definitions, the process is well-defined and terminates at least for proofs containing only least-number axioms (critical formulas corresponding to axiom (4)) of rank 1. The proof that the procedure terminates (paragraph 9 of Ackermann 1925) is opaque, especially in comparison to the proof by transfinite induction for primitive recursive arithmetic. The definition of a Substitution Method for second-order ε-terms is insufficient, and in hindsight it is clear that a correct termination proof for this part could not have been given with the methods available.

Ackermann himself was aware of the defects of this part of his dissertation and reworked the proof in the subsequent years.

According to the previous account of the history of consistency proofs between 1921 and 1924, it seems fair to say that there is a significant gap between the elementary calculus of free variables with successor and predecessor of Hilbert's 1921/22 lecture notes (defined above) and Ackermann's system of 1924.

In this perspective von Neumann's 1927 paper gains a special interest. The paper contains a fully satisfactory syntactic consistency proof for a system stronger than the system treated by Hilbert: von Neumann's system consists of full first-order predicate calculus plus axioms for a successor function. The proof is completely rigorous and does not share the opacity of Ackermann's (1925) treatment of the ε-Substitution Method. On the other hand, von Neumann's proof works for a system that is significantly weaker than the target system of Ackermann's (1925) and indeed weaker than the system used by Hilbert and Bernays in Stage III (1923, 1923a). The latter system includes definitions by primitive recursion and the second ε-axiom. But Hilbert only gives a sketch of the ε-reduction procedure for a very special case. Ackermann's (1925) system is even stronger than Hilbert's one in that it includes second-order definitions by primitive recursion. But, as previously remarked, it is commonly acknowledged that Ackermann's proof is defective (except for the subsystem corresponding to second-order PRA) and indeed Ackermann reworked his proof using ideas of von Neumann's.

It is true that the weak system whose consistency is proved by von Neumann can be shown to be consistent much more easily with the methods of Herbrand (1931) or Gentzen (1934), but this holds only because one then has their general strong results at disposal (results which of course were unknown in 1925).[3] The main interest of von Neumann's proof lies in the method adopted by him, a method which is (as will be evident after the detailed description of the proof) quite original in the context of the Hilbert school in the mid-Twenties, and whose possibilities of further development and application (by adopting stronger background assumptions) have not been explored yet.

The bottom-line seems to be as follows. Von Neumann's proof is the first fully rigorous syntactic proof of the consistency of a system of first-order arithmetic with quantifiers and a successor function. The other main merit of von Neumann's paper is the introduction of the notion of *Grundtypus*, which proved to be the right notion to rework and generalize Ackermann's 1924 proof, as later done by Ackermann himself, a work culminating in his (post-Gentzen) consistency proof for full Peano Arithmetic in 1940 (Ackermann 1940). The latter merit is already acknowledged in both the early (Hilbert–Bernays 1939, 122) and recent literature (Zach 2003, Mancosu–Zach–Badesa 2009).

§4. The theory. We here describe the formal system whose consistency von Neumann proves. Our description is, as much as possible, in current notations and respects current conventions. We warn the reader that in many cases these are different from von Neumann's conventions.

In particular, Von Neumann's terminology diverges from ours. We shall adhere to some extent to it, while in other cases we use current terminology. Von Neumann does not distinguish between propositions and terms and uses 'formula' (*Formel*) for both. We shall use the term *expression* instead, to avoid confusion to the contemporary reader. We shall also use 'formula', 'sentence', 'term' in their current usage. A τ*-term* is a term beginning with the τ*-operator* (see below).

The logical basis of the system is a version of first-order predicate calculus with identity, with negation and conditional as primitive connectives, and both quantifiers as primitives. The language has identity as the only binary relation, a single monadic predicate

[3] In fact, Herbrand proved in his thesis (Ch. 4) the consistency of this weak system also by a simple elimination of quantifiers, *without* his theorem.

Z (mnemonic for *Zahl*), a single constant 0, and a single unary function symbol $'$ (successor). A term-forming operator τ acting on formulas is present in the language (its action will be presently explained).

The axioms can be divided into four groups. Von Neumann gives axiom schemas (instead of axioms and substitution rules).

(I) Propositional axioms

Let A, B, C be closed formulas.

(1) $A \to (B \to A)$
(2) $(A \to (A \to B)) \to (A \to B)$
(3) $(A \to (B \to C)) \to (B \to (A \to C))$
(4) $(A \to B) \to ((B \to C) \to (A \to C))$
(5) $A \to (\neg A \to B)$
(6) $(A \to B) \to ((\neg A \to B) \to B)$

(II) Identity

Let C be a formula with at most one free variable; let a, b be closed terms.

(1) $a = a$
(2) $(a = b) \to (C(a) \to C(b))$

(III) Arithmetic

Let a, b be closed terms.

(1) $Z(0)$
(2) $Z(a) \to Z(a')$
(3) $\neg(a' = 0)$
(4) $(a' = b') \to (a = b)$

(IV) Quantifiers and τ-operator

Let A be a formula with at most the variable x free; let b be a closed term.

(1) $\forall x A \to A(b)$
(2) $A(b) \to \exists x A$
(3) $A(\tau^x A) \to \forall x A$
(4) $\exists x A \to A(\tau^x \neg A)$

The axioms in Groups (I)–(IV) are all the axioms of the formal system whose consistency is proved in the paper. The only deduction rule of the system is Modus Ponens (the axioms of Group IV make rules on quantifiers superfluous). Only closed formulas occur in proofs.

An important difference with respect to our ordinary formal languages is the presence, in the logical basis, of Hilbert's τ-operator. This operator was introduced by Hilbert in his 1922/23 lectures. The *τ-operator* applies to a formula A with respect to a variable x, possibly occurring free in the formula, and generates a term $\tau^x A$. When x is the only free variable in A, this term intuitively stands for a counterexample to the validity of A, if one exists. If there is at least a counterexample to the predicate expressed by a given formula with one free variable, then the term consisting of the τ-operator applied to that formula denotes any one of these counterexamples; if there are no counterexamples, it denotes (conventionally) an arbitrary fixed object (e.g., 0). In other words, the τ-operator chooses an example of the 'worst case' for a given predicate (the operator is nonextensional, since the choice is relative to the formula expressing the predicate): if even the example of the worst case satisfies the predicate, then everything does; dually, if there is something satisfying the predicate, then also the example of the worst case of the negation of the predicate must satisfy the predicate (otherwise, everything would satisfy the negation of the predicate). The point of axioms (3) and (4) is just to express this. To recall Hilbert's memorable example: *If Aristides the Just is corrupt, then anyone is corrupt*.

In Hilbert's treatment of 1922/23 axioms were included to force the τ-operator to give *the least* counterexample, if one exists. Von Neumann leaves this minimality axiom out. Thus, the axioms of Group IV are fully in the tradition of the Hilbert school of that time, except for the following points.

1. von Neumann retains Hilbert's τ-operator and does not adopt the dual ε-operator, as done instead by Ackermann in 1924.
2. von Neumann retains the quantifiers along with the τ-operator, and the respective axioms are formulated accordingly.
3. von Neumann does not include the minimality axiom for the τ-operator, which was included in Hilbert's treatment of 1922/23.

The first two points are merely conventional and cosmetic. The third point is substantial: that the minimality axiom, which can be used to derive the induction axiom, is left out of the system is a witness of von Neumann's remarkable insight.

Von Neumann also discusses two further groups of axioms. The first (Group V) includes a single axiom schema, which is a form of the full impredicative second-order Comprehension Schema (suitably adapted to the language and idiosyncratically formulated—we shall not give it here, to avoid notational complications). This schema is explicitly introduced by the author since it is necessary to allow the formal development of a system sufficient for classical analysis (first of all, after the introduction of a suitable new predicate, it allows the proof of the full second-order version of mathematical induction). But, of course, the consistency proof does not encompass the system including this schema, as von Neumann explicitly remarks. The second (Group VI) includes three axiom schemas, rigorously expressing for this system some basic rules that allow the usage of explicit (nominal) definitions; in our current logical practice these (notationally rather cumbersome) axiom schemas are not on the same language-level as the ones in the other groups. The consistency proof regards a system that at present we would formulate with the axiom schemas in Groups (I)–(IV) only.

It is commonly stated (see, e.g., van Heijenoort 1967, 489) that both von Neumann and Ackermann proved in fact the consistency of a subsystem of arithmetic with quantifier-free induction. In the case of von Neumann, this can be considered correct, recalling that one can prove that if a subsystem of arithmetic without induction is consistent, it remains consistent when a quantifier-free induction rule is added (this was known in the Hilbert school at the time). The case of Ackermann is more complicated, and we cannot discuss it here; we refer the reader to Zach (2003). As we have seen, Ackermann set out for a stronger system (including induction and definitions by second-order primitive recursion) but his 1924 proof is defective in many respects (see discussion above) so that it is hard to assess exactly which system is proved consistent. Ackermann's proof is criticized in detail by von Neumann in the last part of his paper. There (in the very last paragraph of the paper)[4] he also seems to imply the false statement that both he and Ackermann could even prove the consistency of a predicative subsystem of analysis (roughly what nowadays is known as ACA_0, a conservative extension of Peano Arithmetic PA). But it is well-known that the limitations of the consistency proofs obtained were not at all clear at the end of the Twenties (see Zach 2003 for discussion).

§5. Overview of the proof. We give a reasonably detailed exposition of von Neumann's proof. This serves two purposes. The first is to offer to the modern reader a presentation of von Neumann's proof stripped-off of some of the oddities of the notation and terminology of his times and of his own, although we chose to respect two momentous choices of his: to use τ's instead of ε's and to retain quantifiers (otherwise the whole proof should be fully reformulated, in fact completely altered, as will be evident after its exposition). The second is to give the reader enough ground to judge whether our proposed historical assessment is justified.

5.1. The proof strategy: part 1. Von Neumann starts with the consistency problem and shows how to reduce its solution, step by step, to the solution of gradually simpler and more specific problems, and ultimately to the satisfaction of purely combinatorial conditions, whose fulfilment is finally proved, thus demonstrating consistency. In the course of these reductions von Neumann reformulates the basics of the Substitution Method and, more importantly, introduces the key notion of a *Grundtypus*. We shall presently outline this sequence of reductions.

First of all, though, we briefly recall the main idea of the Substitution Method. The basic goal of the procedure is (roughly) to find suitable substitutions of terms in place of the variables acted upon by the operators which are to be eliminated (instances of the ε-operator or–here–of the τ-operator, which in their turn would allow an eliminative definition of quantifiers–although the latter are retained in the formal language here), in such a way that all the formulas so obtained (from given initial sets of formulas to be evaluated) satisfy certain requirements that will ultimately ensure the consistency of the formal system. Thus, Von Neumann's procedure is akin to the classical procedure of ε-substitution (see Hilbert–Bernays 1939, Tait 1965, Moser 2006, etc.), provided one recalls that the ε-operator chooses an example, while the τ-operator chooses a counterexample. His proof is carried out by means of a suitable enrichment of the theory whose consistency is to be proved, by finding witnesses for the τ-terms, chosen in such a way that the relevant formulas (axioms) are made true. Although the basic idea is similar, this marks a difference[5] with respect to the (sort of 'priority') arguments developed by Ackermann, which are the basis of the classical Substitution Method, officially presented by Bernays in the second volume of the *Grundlagen* (1939): this is based on successive replacements of terms, usually with 'backtracking', until a final solving substitution is found after a finite

[4] "The formalism thereby [by Ackermann's proof, after the necessary limitations to its scope] recognized as consistent makes it possible only the construction of a mathematics which corresponds to the semi-intuitionistic mathematics of the critics of set theory and analysis before Brouwer" (1927, 46), e.g., he adds, Russell's and Weyl's predicative systems.

[5] Whose precise extent cannot be discussed here, since this would require too much space and a technical competence on these methods in general that we do not have. These are also the reasons why we do not venture into a precise comparison with current, state-of-the-art uses of the Substitution Method, such as those in (e.g.) Buchholz–Mints–Tupailo 1996, Moser 2006, Moser–Zach 2006, etc.

number of steps (in the original formulations, termination must be proved by finitistic means).[6]

Step 1. The starting point is the fact that if a theory has a valuation then it is consistent. A *valuation*[7] here is a decidable partition (T, F) of the set of all the closed expressions in the language of the theory such that (a) all closed instances of axioms of the theory are in T, (b) $A \to B$ is in T iff $A \in F$ or $B \in T$, and (c) $A \in T$ iff $\neg A \in F$. If the theory has a valuation then it is consistent, since if there were a proof of a contradiction then the finitely many axioms occurring in the proof would be in T; the valuation would have to put in T both a formula and its negation (since we would have proved both); but this contradicts the last property in the definition of valuation.

Von Neumann first defines a valuation (T, F) that works (without modifications throughout the proof) for the first three groups of axioms. This is achieved as follows. Two special constants, W_T and W_F are introduced: they denote the truth values, True and False, respectively. We have to define a partition (T, F) of the set of *all* closed expressions. For those expressions which are not formulas, the partition is necessarily quite artificial; von Neumann decides (without consequences for the consistency proof) to define the partition for *all* the expressions (even the open ones). First he simply puts all variables and all constants except W_F into T, and W_F into F. Then we have the nontrivial clauses. A sentence $A \to B$ belongs to T iff either A belongs to F or B belongs to T (note that with the first three groups of axioms there is no problem in deciding which case holds). A sentence $\neg A$ belongs to T iff A belongs to F. A sentence $a = b$ belongs to T iff a and b, as expressions, are syntactically identical. A sentence $Z(a)$ belongs to T iff a is a numeral (in the usual sense, since we have 0 and successor in the language). All atomic sentences that do not fall under the previous cases belong (by default, without effect on the proof) to T. If a is an expression which has already been put into T or into F, then $\forall x(a)$, $\exists x(a)$ and $\tau^x(a)$ belong to T (this is again a degenerate case, with vacuous operators).

Step 2. The next move is to reduce the consistency problem to that of finding valuations *for each finite subset S of axioms of the target theory*. To this aim, von Neumann first introduces the notion of partial valuation. A *partial valuation* ('*Teilwertung*') with respect to a finite set S of axioms is simply a valuation relativized to that set of axioms. Thus, instead of referring to all the axioms, condition (a) (in the above definition of valuation) is limited to the axioms in the set S: all the elements of S receive the value True. In a partial valuation the assignment of truth values depends in general on the chosen set of axioms S, while in a valuation it does not. In general, the presence of quantifiers makes it necessary to find a partial valuation.

The problem is now reduced to finding a partial valuation for every finite set S of (closed instances of) axioms of the formal system. The point is that if for every finite set of axioms

[6] Ackermann developed his methods on the basis of Hilbert's original intuitions, and employed them, as we said above, both in Ackermann 1925 and (with corrections and refinements) 1940: we again refer the reader to Zach 2003 and Mancosu–Zach–Badesa 2009 for clear exposition and discussion (the analogy with priority arguments was suggested by an anonymous referee). By the way, a method which is to some extent similar to Von Neumann's one (more than to Ackermann's one) was also employed, in a different context, by Shoenfield 1967, Section 4.3; a detailed comparison, which again would require a long detour, is out of the scope of the present paper.

[7] *Wertung*: this notion is traced back by von Neumann to Julius König.

S of a given formal system there is a partial valuation with respect to S, then the formal system is consistent: as above, if there were a proof of a contradiction, then the finitely many axioms occurring in the proof would constitute such a set S; the partial valuation with respect to S would have to verify both a formula and its negation, but this contradicts the last property in the definition of partial valuation.[8]

Step 3. The next move is to reduce the problem of showing that a finite set of sentences has a partial valuation to that of defining a so-called reduction rule. A *reduction rule* or *procedure* is a rule that associates to each closed formula or term another formula or term, its *reduct*, that has the property that all the τ-operators and the quantifiers occurring in the original formula or term have been eliminated. The resulting formula or term is called *reduced* (see Def. 5.1 below). Thus, a reduction procedure is a method to eliminate all the operators (τ-operators and quantifiers) from closed expressions, finding suitable substitutions of terms for the variables bound by the operators which are to be eliminated.

Instead of directly defining a partial valuation (T_S, F_S) for each finite subset S of the axioms (as a partition of all closed expressions), one can define more simply for each S a reduction rule P_S, depending on S, and set

$$A \in T_S \Leftrightarrow P_S(A) \in T,$$

where $P_S(A)$ is the reduct of the expression A by means of the procedure P_S, and T is from the above fixed partition (T, F) of all expressions (see Step 1; since this is a partition, the same condition holds for false expressions). The reduction rule P_S depends on the finite set S of axioms, in the sense that different reduction rules may give different reducts for the same expression, suitable for different sets S.[9]

Von Neumann then imposes the minimal conditions on a reduction rule that guarantee that the resulting partition is a partial valuation. The following notion of *reduced expression* plays a central role here.

DEFINITION 5.1 (Reduced Closed Expression). *A closed expression is reduced if no closed subexpression (inclusive of the expression itself) begins with a quantifier or with a τ-operator.*

Thus, reduced closed expressions are, by definition, closed expressions without quantifiers or τ-operators.

To make sure that the partition (T_S, F_S) induced by a reduction rule P_S is a partial valuation, it is sufficient (by definition of partial valuation) to ensure that the reduction rule commutes with the connectives (the reduct of a negation is the negation of the reduct, etc.), with identity, and with the basic operations and predicates of the language (the reduct of a successor is the successor of the reduct, etc.), that the reduct of 0 is 0, and that if a and b

[8] This use of the notion of partial valuation apparently marks a first slight methodological divergence from the consistency proofs of the Hilbert school. In the latter the finiteness condition is derived from focusing on a purported formal proof of a contradiction. It seems that Von Neumann reasons here in terms of finiteness of proofs *in general* instead, and does not focus his procedure on a particular proof-figure. But this difference should not be overemphasized, and we do not insist on it: we have just seen that the relevant application of the notion of partial valuation to the consistency problem concerns just the purported proof of a contradiction.

[9] This is usual in substitution methods in general. Here, the choice of specific witnesses for the τ-terms will depend on the given set of axioms in which the τ-terms occur. But it will be possible to fully appreciate this point only below, when we inductively define a reduction procedure.

have the same reduct, then $C(a)$ and $C(b)$ (where C is any formula with at most one free variable) have the same reduct. For this purpose, it is sufficient to stipulate that any reduced formula is its own reduct, that any reduct is reduced, and that if a^* is the reduct of a, then $B(a)$ and $B(a^*)$ (where B is any formula with at most one free variable) have the same reduct. Thus, von Neumann singles out the following two conditions (the first is twofold) on any reduction procedure P, where RED denotes the set of reduced expressions.

(α) $a \in RED \Rightarrow P(a) = a$; for all a, $P(a) \in RED$.
(β) $P(B[x/a]) = P(B[x/P(a)])$

A further condition (III′b in von Neumann's paper) is necessary and will be the hardest to guarantee. This condition is simply the following.

> (Condition III′b) All elements of S that are instances of quantifiers axioms or of τ-axioms (Group IV) belong to T_S.

Step 4. Von Neumann makes one more simplifying step by singling out a special set of expressions and showing that it is enough to define a reduction rule on it. This is the set of *directly reducible expressions*, defined as follows.

DEFINITION 5.2 (Directly Reducible Expression). *A closed expression is directly reducible if it begins with a quantifier or a τ-operator but no proper closed subexpression of it begins with a quantifier or with a τ-operator.*

In particular, a τ-term is directly reducible if it has no proper subterm without free variables which is a τ-term, and the formula on which the τ-operator acts has no closed subformula in which quantifiers occur. A closed formula is directly reducible if it begins with a quantifier, it has no proper closed subformula in which quantifiers occur, and no τ-terms without free variables occur in it. Von Neumann makes the fundamental observation that in a reduction rule it is always sufficient to deal with directly reducible terms and formulas, since the other terms and formulas can then be reduced by sequences of reductions of directly reducible ones. In this way conditions (α) and (β) can always be ensured.

Step 5. The problem is now reduced to that of finding, for each finite set of sentences S, a rule P_S assigning reduced expressions to directly reducible expressions and satisfying Condition (III′b).

In general, in order to obtain a partial valuation for all the closed expressions in the language with respect to the finite set S of axioms, it is certainly sufficient to make true the reducts of all the (finitely many) instances of the axioms (of all the groups) in which the closed formulas and terms in the set S occur. For the other formulas it is then enough to assign the values according to the truth tables of the connectives.

Since the instances of axioms of Groups (I)–(III) can always be trivially made true (von Neumann shows this on p. 24 of his paper; see above, Step 1), the point is to make true the reduced formulas of all the instances of the axioms of Group IV (expressing the relations between quantifiers and τ-terms) in which the closed formulas and closed terms from the set S occur. These axioms are conditionals: thus (in view of Steps 1–3 above) we just have to give the suitable values to the reducts of their antecedents and consequents. Von Neumann hence formulates the following condition, which we call Condition (*).

> (Condition *) Let A be any formula in which at most x occurs free and let b be any closed expression such that A, x, b belong to S. Then the

P_S-reducts of all the instances of axioms of Group IV belong to T. These instances are the following.

$$\forall x A \to A(b),\, A(b) \to \exists x A,\, A(\tau^x A) \to \forall x A, \exists x A \to A(\tau^x \neg A).$$

A word of clarification is needed on what von Neumann means when he says here that the combination A, x, b *belongs* to S. This means that A, b *occur* in S as sub-expressions of elements of S (von Neumann first uses *gehörende* and later *vorkommende*, pp. 27 ff.), with at most x occurring free in A.

Step 6. The next crucial move is to restate the property that a reduction rule satisfies Condition (*) in terms of types. The notion of type is commonly recognized (e.g., by Bernays in Hilbert–Bernays 1939, 122) as the main contribution of von Neumann to the development of substitution methods for consistency proofs. From the definitions below, though, it will be clear that von Neumann's notion of type (and the more specialized notion of ground-type) is only akin to the current notion of ε-type or ε-matrix.[10] So we carefully stick to his terminology below, explaining it. This holds also as regards nesting and subordination of types (see the definition of subtype, Def. 5.4 below), and the use of dots instead of variables (a choice which, however, will have no consequence whatever for the proof).

Recall that an expression is a *proper subexpression* of another expression if it occurs in it and does not coincide with it; a subexpression is *free* in another if any variable occurring free in it is free (in every occurrence) also in the expression in which the subexpression occurs; a free proper subexpression is *maximal* if it is not a proper subexpression of any free proper subexpression.

DEFINITION 5.3 (Type; Substituent). *Let α be an expression. The type of α, $type(\alpha)$, is the formal expression obtained by removing from α all the maximal proper free subexpressions and replacing each one by a dot. The sequence of the removed subexpressions, from left to right, is the sequence of substituents of α.*

DEFINITION 5.4 (Subtype). *If α is an expression and β is a proper subexpression of α not free in α, then the type of β is a proper subtype of the type of α.*

We occasionally denote by $\sqsubset$ the subtype relation. Note that the notion of subtype is independent of the expressions considered, and depends only on their types. We associate a set of types to any finite set S of sentences (axioms) as follows.

DEFINITION 5.5 (Ground-types of S). *We define a set of types, called Ground-types of S, $GT(S)$, inductively.*

1. *If A, x belong to S, then $type(\tau^x A), type(\forall x A), type(\tau^x \neg A), type(\exists x A)$ are in $GT(S)$.*
2. *If $t \in GT(S)$ and $t' \sqsubset t$, then $t' \in GT(S)$.*

[10] But not identical with it: with respect to notation, here we have dots instead of fresh variables; more important, Von Neumann's notion of type (even restricted to terms) involves in its definition expressions in general, not only terms. The differences, however, are immaterial for all practical purposes in this kind of consistency proofs (apart, of course, from the use of τ instead of ε). For the notion of ε-matrix see Tait (1965), Moser (2006), etc. Wang (1963, 364) uses 'ε-category'. It is true that the word 'type' is overloaded in logic, but with this calque from German we want to be deliberately faithful to the terminology of the classics of the Hilbert school.

3. *If* $type(\tau^x A) \in GT(S)$ *then* $type(\forall x A) \in GT(S)$.
4. *If* $type(\forall x A) \in GT(S)$ *then* $type(\tau^x A) \in GT(S)$.
5. *If* $A = \neg A'$ *and* $type(\tau^x A) \in GT(S)$ *then* $type(\exists x A') \in GT(S)$.
6. *If* $type(\exists x A) \in GT(S)$, *then* $type(\tau^x \neg A), type(\forall x \neg A) \in GT(S)$.

Let us then divide the types so obtained into *groups*, putting in one and the same group, given a formula A (belonging to S), the types $\tau^x A$ and $\forall x A$, when the formula A has not the form $\neg A'$; and putting in one and the same group the types $\tau^x A$, $\forall x A$ and $\exists x A'$, when the formula A has such form. Let us enumerate the (finite) set of all groups in a standard way, ensuring that the groups with smaller index in the enumeration have as elements proper subtypes of the types in the groups with greater index.[11] The main induction will be precisely on the number of groups (see below): this is sufficient, since the types in the same group are related enough to be treated together.

The *Ground-expressions* ('*Grundformeln*') for the set S are all the substituents (in the sense of Def. 5.3) in the types of the terms and formulas in S, and in addition all the closed expressions occurring in the sentences of S. We call *Ground-constants* ($GC(S)$) and *Ground-operations* ($GO(S)$) of S the constants and operations occurring in the Ground-expressions for S (recall that, in general, in our language as constants we have only 0 and the truth values, and as operations we have only the connectives, for formulas, and successor, identity and the predicate Z, for terms). We always put the truth values W_T, W_F, among the ground constants. We omit reference to S below, when there is no danger of confusion.

The next key-notion is that of degree, which is defined with respect to a given reduction rule P. The *P-degree*, deg_P, measures, in general, the complexity of a *reduced* expression with respect to a reduction rule P.

DEFINITION 5.6 (P-degree). *Let P be a reduction rule.*

1. *If* $k \in GC$ *then* $deg_P(k) = 0$.
2. *If* $O \in GO$ *is an n-ary operation, and* $a_1, \ldots, a_n \in RED$ *such that* $\max(deg_P(a_i)) = d$, *then* $deg_P(O(a_1, \ldots, a_n)) = d + 1$.
3. *If* $t \in GT$ *with n dots, and* $a_1, \ldots, a_n \in RED$ *such that* $\max(deg_P(a_i)) = d$, *then* $deg_P(P(t[a_1, \ldots, a_n])) = d + 1$.

Note that point (2) accounts both for logical (propositional) complexity and for nesting of relations. In point (3), $t[a_1, \ldots, a_n]$ is the result of the substitution of the components of the n-tuple $(a_1, \ldots, a_n)$, in their order, in the n empty places of the type t.

Two remarks are essential for the rest of the proof. The first is that, *independently* of the chosen rule P but *depending* on S, we can determine a number $\bar{p}$ strictly larger than the P-degree of the P-reducts of Ground-expressions, for any reduction rule P. Von Neumann remarks that it is sufficient to take the successor of the maximum number of occurrences of operations and abstractions (i.e., τ-operators or quantifiers) in a Ground-expression from S. Note then that the number $\bar{p}$ *does depend* on the set of Ground-expressions, which is univocally determined by the choice of the system S of axioms. The second remark is

[11] This can be done, e.g., by defining (on the basis of the definitions of type and subtype) a suitable equivalence relation on expressions and then defining a suitable ordering on the equivalence classes, inducing the desired enumeration of groups. This can be achieved in various standard ways, all interchangeable for the present purpose. Since no more details on this will be needed below, we omit the tedious definitions.

that we can bound the number of closed expressions of P-degree $\leq d$ *independently of* P (but, again, depending on S), for any d. Let $\varphi(d)$ be an upper bound on that number.[12] Note that the function φ only depends on the number of Ground-constants, the number of Ground-operations, the number of groups, and the maximum number of argument places in the operations and of empty places in the types.

We now have to define, given S, a map $P_S : DR \to RED$ that also satisfies Condition (*), where DR denotes the set of directly reducible expressions. For notational brevity, sometimes we shall write (here and below, whenever this is not a source of confusion) $red_P()$ instead of $P_S()$, omitting reference to the given S, or even $red()$, without mentioning the given P. The general condition that must be satisfied for this purpose, denoted (T) by von Neumann (on page 30 of his paper), is the following.

(Condition T)

A type t has the form $\tau^x A$, $\forall x A$ (or $\exists x A'$ only in case $A = \neg A'$). Let it have n empty places (marked by dots). Then we have to satisfy the following, for each type under consideration.

- For all n-tuples $(a_1, \ldots, a_n)$ of reduced closed expressions of P-degree $< \bar{p}$ consider $t[a_1, \ldots, a_n]$, the result of the substitution of the components of the n-tuple, in their order, in the n empty places of the type. This has the form $\tau^x R$, $\forall x R$, or $\exists x R'$.
- If there exists a reduced closed expression b of P-degree $\leq \bar{p}$ such that[13] $red_P(R[x/b]) \in F$ then
  - $red_P(\tau^x R) :=$ one such b,
  - $red_P(\forall x R) := W_F$,
  - $red_P(\exists x R') := W_T$,
- Else
  - $red_P(\tau^x R) := 0$,
  - $red_P(\forall x R) := W_T$,
  - $red_P(\exists x R') := W_F$,

Condition (T) is sufficient for our purpose: if the condition is satisfied, then the truth-values of the reduced formulas (occurring as antecedents and consequents) will be such that they will make true all the conditionals (instances of the axioms on quantifiers) that should be made true according to Condition (*). Indeed, among the types mentioned in the condition there are all the ones occurring in the initial set S (by definition of Ground-type), and the reducts of the substituents and of the b's, as reducts of Ground-expressions, all have degree less than $\bar{p}$ (by definition of $\bar{p}$): hence, the condition encompasses all the cases that have to be considered. In Condition (*) we want to take care of all the quantifier axioms arising from choices A, b, x occurring in S, and we want to ensure that P maps all these axioms to T. The reducts under P of the antecedents and consequents of the axioms are exactly what we obtain in Condition (T). Consider for example the first

[12] We defer the discussion of the calculation of this bound to Section 5.4 below.

[13] Here, as von Neumann remarks, it would be sufficient to put '$<$' instead of '$\leq$', since (as we explain in the next paragraph) all the relevant reducts have degree strictly less than $\bar{p}$. Nevertheless, it turns out that by choosing '$\leq$' the correctness proof (given below) is made a bit simpler.

axiom: $\forall x A \rightarrow A[x/b]$. The reducts we need to take into consideration are $P(\forall x A)$ and $P(A[x/b])$. The reduct by P is defined either directly (if the formula is directly reducible) or by a sequence of steps acting on directly reducible formulas within the formula. Each step replaces a directly reducible formula by a reduced formula. Thus, the reduct by P of $\forall x A$ can be obtained as follows: from $\forall x A$ go to $type(\forall x A)$, then replace the dots by the P-reducts of the substituents of $\forall x A$. Now consider $A[x/b]$. The reduct by P is again obtained by a sequence of steps each one replacing a directly reducible subformula of the formula. Also, we know (see Step 3 above) that P satisfies $P(A[x/b]) = P(A[x/P(b)])$. Therefore the P-reduct of $A[x/b]$ is again obtained as one of the formulas considered under Condition (T).

In conclusion, given a finite set S of formulas, and thus given a finite set of Ground-types, divided into groups, and finite sets of Ground-constants and Ground-operations, respecting all the constraints given above, the goal is to find a corresponding reduction rule P such that Condition (T) is satisfied.

Step 7. The final move is to abstract from the unnecessary details and to reformulate Condition (T) by dropping all the unnecessary dependencies. Von Neumann observes that S in Condition (T) determines the groups of Ground-types, the sets of Ground-operations and Ground-constants, and the quantities $\bar{s}, \bar{c}, \bar{o}, \bar{n}, \bar{p}$ (where: $\bar{s}$ is the number of groups of Ground-types, $\bar{c}$ is the number of Ground-constants, $\bar{o}$ is the number of Ground-operations, $\bar{n}$ is the maximum number of argument places in the Ground-operations and of empty places in the types, and $\bar{p}$ is the fixed upper bound on the degree of the reducts of the Ground-expressions under any reduction rule). Condition (T) is then replaced by Condition (T′), which abstracts from S as much as possible. In Condition (T′) we deal with an arbitrary set (system) of types, which we declare as the Ground-types, an arbitrary set of operations, which we declare as the Ground-operations, and an arbitrary set of constants, containing W_R and W_F, which we declare as the Ground-constants. These choices uniquely determine the quantities $\bar{s}, \bar{c}, \bar{o}, \bar{n}, \bar{p}$. Then we ask that there exist a reduction rule P such that Condition (T) is satisfied with respect to P and to the chosen systems of types, operations, and constants, and corresponding (uniquely determined) quantities $\bar{s}, \bar{c}, \bar{o}, \bar{n}, \bar{p}$. It is here crucial that, as we observed above, the quantity $\bar{p}$ is independent of the reduction rule P. Our aim is now to prove that for any such choice there exists a reduction rule P that satisfies Condition (T) with respect to that choice.

5.2. The proof strategy: part 2. After Steps 1–7 are completed, it remains to prove that, for every choice of S, there exists a reduction rule P_S satisfying Condition (T). Von Neumann stresses the importance of finding P_S from S uniformly. In his terms, we must find a rule to associate a P_S to any given S. The needed uniformity is achieved by using an inductive proof strategy. By induction on the number $\bar{s}$ of groups of types (the base case, $\bar{s} = 0$, being trivial) von Neumann proves that there exists an adequate reduction rule for any system of types with $\bar{s}$ groups. We present the details of the induction step, when $\bar{s}$ has the form $\sigma + 1$.

The goal is to define the P-reducts of *all directly reducible closed expressions* whose type t is in one of the groups $g_1, \ldots, g_\sigma, g_{\sigma+1}$ in such a way that Condition (T) is satisfied by P with respect to $\bar{p}$ and the other parameters indicated above. The induction hypothesis is that a reduction rule P^* exists that satisfies Condition (T) for all the types in the groups up to and including g_σ with respect to the same parameters except—crucially—for the parameter $\bar{p}$, which is replaced by $\bar{p}^*$. We will see below how $\bar{p}^*$ has to be chosen in order for the argument to go through.

The approach taken by von Neumann will ensure that the P-reduct of all directly reducible closed expressions whose type is in $g_{\sigma+1}$ and whose substituents have P-degree $< \bar{p}$ will be defined. For the directly reducible closed expressions whose type is in $g_{\sigma+1}$ whose substituents do not have P-degree $< \bar{p}$, which are irrelevant, an arbitrary choice is made. On expressions whose type is not in $g_{\sigma+1}$ P will coincide with P^*. Recall that if a Ground-type in the group g_i is a proper subtype of another in the group g_j, then $i < j$.

Von Neumann's strategy is to resolve the first quantification in Condition (T) (which read: for all n-tuples $(a_1, \ldots, a_n)$ of reduced closed expressions of P-degree $< \bar{p}$) by *stratifying*, i.e., by proving successively the property for all n-tuples of P-degree ≤ 0, then for all n-tuples of P-degree ≤ 1, etc., up to for all n-tuples of P-degree $\leq \bar{p} - 1$. The definition of P will be obtained by *bootstrapping*.

The stratification and bootstrapping procedure goes as follows.[14] For succinctness we only deal with types of the form $\tau^x A$ (the treatment of the other types being derivative).

First we deal with the reduced closed expressions of P-degree ≤ 0. These by definition are the Ground-constants, whatever P may be. The set of all n-tuples of Ground-constants can be enumerated, let

$$C_1, \ldots, C_{k_0}$$

for some k_0 be such an enumeration. By insertion in the dots of the relevant type of the form $\tau^x A$ the C_is give rise to directly reducible expressions

$$\tau^x R_1, \ldots, \tau^x R_{k_0}.$$

We now have to assign a P-reduct to each of these. Suppose for the sake of exposition that this has been done.

We are then in a position to assign P-reducts to all the instantiations of the same type by expressions of P-degree ≤ 1, despite the fact that P is not completely defined yet. Indeed, the reduced closed expressions of P-degree ≤ 1 can only have one of the following three forms: (i) expressions of P-degree ≤ 0, (ii) P-reducts of the expressions obtained by filling the dots of types in $g_{\sigma+1}$ by expressions of P-degree 0, (iii) P-reducts of the expressions obtained by filling the dots of types in $g_1, \ldots, g_\sigma$ by expressions of P-degree 0. We know all these expressions: those arising from (i) are just the Ground-constants, those arising from (ii) are the reducts we just defined in treating $C_1, \ldots, C_{k_0}$, and those arising from (iii) we know since P and P^* coincide on types not in $g_{\sigma+1}$. We can thus enumerate all n-tuples of such expressions (different from $C_1, \ldots, C_{k_0}$), let them be

$$C_{k_0+1}, \ldots, C_{k_1},$$

for some k_1. By insertion in the dots of the relevant type the new C_is give rise to directly reducible expressions

$$\tau^x R_{k_0+1}, \ldots, \tau^x R_{k_1}.$$

We now have to assign a P-reduct to each of these.

The procedure is iterated along the same lines. The bootstrapping works since we can enumerate all n-tuples of closed expressions of P-degree $\leq i + 1$ after having defined the P-reducts of all instantiations of the type $\tau^x A$ by n-tuples of closed expressions of P-degree $\leq i$. Globally, we are assigning, in order, P-reducts of instantiations of the type $\tau^x A$ by the n-tuples

$$C_1, \ldots, C_{k_0}, C_{k_0+1}, \ldots, C_{k_1}, \ldots, C_{k_{\bar{p}-2}+1}, \ldots, C_{k_{\bar{p}-1}},$$

[14] Hardly surprisingly nowadays, but one should not forget that von Neumann writes in 1925.

where the n-tuples $C_{k_i+1}, \dots, C_{k_{i+1}}$ are the not-yet-considered n-tuples of expressions of P-degree $\leq i$. The enumeration $C_{k_i+1}, \dots, C_{k_{i+1}}$ is only possible after the P-reduct of all instantiations by C_j with $j < k_i + 1$ have been defined. This set-up obviously imposes the constraint that the procedure used to assign P-reducts to instantiations of the type $\tau^x A$ by the C_is can be repeated $k_{\bar{p}-1}$ times.

Now let us consider how the actual assignment of P-reducts to instantiations of the type $\tau^x A$ by the C_is is performed. As outlined, this is done group by group (of n-tuples), first treating the instantiations by $C_1, \dots, C_{k_0}$ and only then treating the instantiations by $C_{k_0+1}, \dots, C_{k_1}$, for the bootstrapping procedure to work. Within each group $C_{k_i+1}, \dots, C_{k_{i+1}}$, the instantiations are dealt with in the order of enumeration (which is arbitrary).

Here we are bound to use the induction hypothesis. Hence the P-reduct of a directly reducible expression $\tau^x R_i$ is chosen among the P^*-reducts of expressions $R_i[x/b]$ obtained by substituting x in R_i with some closed term b. The induction hypothesis is that P^* satisfies Condition (T) with respect to parameter $\bar{p}^*$ for expressions whose type is in $g_1, \dots, g_\sigma$. This ensures that the P^*-reduct of such an expression is well-defined as long as the P^*-degree of the term b is $\leq \bar{p}^*$. For the moment let us stipulate that when defining the P-reduct of $\tau^x R_i$ (for $i \in \{1, \dots, k_{\bar{p}-1}\}$) the bound on the P^*-degree of the candidates b is B_i. The induction hypothesis ensures that P^* satisfies Condition (T) as long as $B_i \leq \bar{p}^*$. Moreover, we require that $B_i \leq B_{i+1}$, since once a candidate b is accepted as the P-reduct of $\tau^x R_i$, it is fully entitled to be taken as a component of the new n-tuples instantiating the relevant type.[15] We will see below (in the next subsection) which further constraints are imposed in order to prove the correctness of the reduction rule P we are defining. In any case, the above-described procedure completely defines a rule P once the parameters $B_1, \dots, B_{k_{\bar{p}-1}}$ are fixed.

Let us sum up the constraints imposed on the parameters $\bar{p}^*$ and B_is by the above construction. For all i such that $1 \leq i \leq k_{\bar{p}-1}$,

1. $B_i \leq \bar{p}^*$.
2. $B_i \leq B_{i+1}$.

Since the values of the B_is are nondecreasing (by Constraint 2) we can express them as follows.

$$B_{i+1} = \max(\gamma_1, \dots, \gamma_i) + N_{i+1},$$

where for all i such that $1 \leq i \leq k_{\bar{p}-1}$, $N_i \geq 0$ and γ_i is the P^*-degree of the P-reduct of $\tau^x R_i$ as defined in the procedure. Further constraints on the N_is will arise in the proof of correctness of the definition of P. Since fixing the parameters $B_1, \dots, B_{k_{\bar{p}-1}}$ fixes the rule (as we have just seen), it follows that the above-described procedure completely defines a rule P once the parameters $N_1, \dots, N_{k_{\bar{p}-1}}$ are specified.

For the rest of the argument the following remarks will be crucial. Suppose we want to bound the P^*-degree of a reduced closed expression as a function of its P-degree. Such an expression, call it α, is obtained from Ground-constants by $< \bar{p}$ (resp. $\leq \bar{p}$)[16] repeated applications of Ground-operations or applications of types in $g_1, \dots, g_{\sigma+1}$

[15] Strictly speaking, to require $B_i \leq B_{i+1}$ for *all* i is more than we need, in view of the grouping of n-tuples just described. We put this constraint since in any case it will be essential in order to prove the correctness of the procedure, as we shall see in detail in the next subsection.

[16] The distinction is relevant in Condition (T), cf. fn. 12 above.

and successive P-reduction. We can distinguish two cases. Case 1: the last step in the P-degree history of α is an application of a type in $g_{\sigma+1}$ and successive P-reduction. Let this type be $\tau^x A$. Then the P-reduct has, by construction, P^*-degree equal to one of $\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}$. Case 2: not Case 1. Then the P-degree history of α can be subdivided as follows: a sequence of steps ending in the last application of a type in $g_{\sigma+1}$, followed by $< \bar{p}$ (resp. $\leq \bar{p}$) many applications of Ground-operations or types in $g_1, \ldots, g_\sigma$. In this case the P-reduct has, again by construction, P^*-degree less than (resp. less than or equal to) $\max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) + \bar{p}$. Thus, in general (resp. with '$<$' in place of '$\leq$'):

$$deg_P(\alpha) \leq \bar{p} \quad \Rightarrow \quad deg_{P^*}(\alpha) \leq \max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) + \bar{p}.$$

This bounding of the P^*-degree of a reduced closed expression on the basis of its P-degree is essential in the correctness proof below. It is the main constraint arising from the chosen definition procedure for P. Essentially, the bound says that reduced expressions of P-degree $\leq \bar{p}$ can have (as we have just shown by the above reasoning by cases) at most the sum of the P^*-degree of the reduct of the instantiation of the relevant type by one of the C_is ($i \in \{1, \ldots, k_{\bar{p}-1}\}$) plus $\bar{p}$.

After this abstract account, before proving the correctness of the whole construction (in the next subsection), it is better to give for the sake of clarity a step-by-step description of the way the induction works, following von Neumann. The induction step is of course the crucial point in this proof, and the place in which the specific, original features of von Neumann's approach turn out: hence it is useful to try to see in full detail what is really going on.

We have seen that the induction hypothesis is that we can always find a reduction procedure in the case $\bar{s} = \sigma$; we have to prove that we can always find a reduction procedure in the case $\bar{s} = \sigma + 1$. By induction hypothesis, as we have seen, we may suppose that we already have a procedure P^*, satisfying Condition (T) for all the groups up to and including g_σ, for the same constants and operations, and for the same parameters, except that in place of $\bar{p}$, we take a $\bar{p}^*$, which must be (for reasons that will be explained in the correctness proof in the next subsection) equal to $\bar{p}(2^{(\varphi(\bar{p}-1))^{\bar{n}}})$, where $\varphi(p)$ is the function (mentioned above, Step 6) giving the maximum number of reduced formulas having degree at most p. Note that the induction hypothesis allows this because of its generality.

For the types which do not belong to $g_{\sigma+1}$, we define the procedure P as identical with P^*. Thus, we have to define the procedure P only for the terms and formulas whose types belong to $g_{\sigma+1}$. We know that if a ground type in the group g_i is a proper subtype of another in the group g_j, then $i < j$. For clarity, we shall concentrate on terms and the τ-operator; the truth-functional evaluation of formulas can be obtained immediately.

Let us suppose that in the group of types $g_{\sigma+1}$ occurs, as the only τ-type, the type of the term $\tau^x A$, where A is a formula having at most x free, and the type has n empty places.[17]

Let us take all the n-tuples of closed terms whose P-degree is 0, where P is the reduction procedure we are defining. We have seen that we can do this, although P is not yet defined, since we already know that, in general, only the ground constants have degree 0. We enumerate these n-tuples in the following way (e.g., ordering them lexicographically, but this is irrelevant): $C_1, \ldots, C_{k_0}$.

[17] Of course, 'τ-type' abbreviates 'type of a τ-term', and we denote the type (as above, without loss of generality) with the corresponding term.

Now we make all possible substitutions of these n-tuples in the type considered, saturating each time the n empty places of the type, in order, with the n components of the n-tuple.

The first substitution in the type, by means of the first n-tuple C_1, will yield a certain τ-term without free variables, directly reducible, that we shall denote $\tau^x R_1$.

Now we take all the closed terms b whose P^*-degree has at most a certain value N_1, which must be (for reasons that again will emerge only in the correctness proof in the next subsection) equal to $\bar{p}(2^{(\varphi(\bar{p}-1))^{\bar{n}}-1})$, and we substitute them in turn in place of the variable x in the formula R_1, resulting from the considered τ-term by removing the initial τ-operator. Among these closed terms we search for a counterexample for the formula R_1, and we shall use precisely this counterexample to define the P-reduct of the τ-term. Given a certain b, let the formula resulting from its substitution in R_1 in place of x be $R_1[x/b]$, a formula whose reduct with respect to the procedure P^* we can compute, by induction hypothesis. This reduced formula will have the value True or the value False. If, for at least one b, the respective reduced formula has the value False, then the reduct with respect to P of the considered τ-term is defined as any one of these falsifying b's; otherwise, the reduct of the τ-term is defined as the constant 0. For the corresponding quantified formulas, in order to respect Condition (T), in the first case the reduct of the universal formula is accordingly defined as the value False and the reduct of the (possibly also occurring) existential formula as the value True; vice versa in the second case. The P-reduct of the τ-term, so defined, has a certain P^*-degree, known beforehand; we denote this degree by γ_1; by construction, γ_1 is at most N_1.

Let us proceed to the second substitution. We substitute C_2 in the type, obtaining now a τ-term $\tau^x R_2$. We take all closed terms b whose P^*-degree is at most $\gamma_1 + N_2$, where $N_2 = \bar{p}(2^{(\varphi(\bar{p}-1))^{\bar{n}}-2})$, and we substitute them in turn in place of x in R_2, thus obtaining certain formulas, that we denote in general by $R_2[x/b]$. Now we reduce all these formulas by P^* (we can do this by induction hypothesis), and we look at the resulting truth values. If at least once we have the value False, as reduct of the τ-term we take any one of the falsifying terms, otherwise we take 0, and we give the corresponding truth values to the quantified formulas, everything as above. We denote the P^*-degree of the P-reduced term so obtained by γ_2; clearly γ_2 is at most $\gamma_1 + N_2$.

The third substitution is treated in the same way, and we proceed like this up to and including the k_0-th substitution, each time considering the suitable degrees. The limitation on the P^*-degree of the terms b to be substituted is, in general, the following (it will be justified in the correctness proof below): when we substitute in the type the n-tuple C_k, then we consider terms b of P^*-degree at most $\mathrm{Max}(\gamma_1, \ldots, \gamma_{k-1}) + N_k$, where $N_k = \bar{p}(2^{(\varphi(\bar{p}-1))^{\bar{n}}-k})$.

Now we have to consider all the n-tuples of closed terms having P-degree at most 1, different from those previously considered. Although the procedure P is only partially defined, we know that we can do this, because we already knew all the terms of degree 0, and now (after the first k_0 steps just described) we also know the P-reducts of the terms resulting from the substitution in the considered type of the n-tuples of terms of degree 0; these reduced terms are the only ones, among the relevant terms of P-degree 1, that had not been determined yet. Let us denote $C_{k_0+1}, \ldots, C_{k_1}$ these further n-tuples of closed terms of P-degree at most 1. We now proceed with all the substitutions, as above, determining the P-reduced terms and the respective degrees.

In the same way we proceed with the n-tuples of closed terms of P-degree at most 2 (which can be determined after all the preceding steps), then with the n-tuples of terms of

P-degree at most 3, and so on, up to and including the case of the n-tuples of P-degree at most $\bar{p}-1$. The last n-tuple to be substituted will be $C_{k_{\bar{p}-1}}$, where $k_{\bar{p}-1} \leq (\varphi(\bar{p}-1))^{\bar{n}}$ by definition of φ, since $n \leq \bar{n}$.

For the record, the induction step can be formulated, in full generality, as follows.

Fix a value of p, $0 \leq p \leq \bar{p}-1$. Let all the n-tuples of closed terms of P-degree less than p be determined, and let also be determined the reducts of all the directly reducible terms obtained by substituting those n-tuples in the τ-type of the group $g_{\sigma+1}$. Let these n-tuples be $C_1, \ldots, C_{k_{p-1}}$, and let the respective directly reducible terms be $\tau^x R_\nu$, with ν ranging from 1 to k_{p-1}. Let γ_ν be the P^*-degree of the P-reduct of $\tau^x R_\nu$. We then know all the n-tuples, different from the ones already given, of closed terms of P-degree at most p (i.e., of P-degree less than $p+1$). Let these n-tuples be $C_{k_{p-1}+1}, \ldots, C_{k_p}$, and let the respective directly reducible terms be $\tau^x R_\nu$, with ν ranging from $k_{p-1}+1$ to k_p. The reducts of these terms $\tau^x R_\nu$, with ν ranging from $k_{p-1}+1$ to k_p, are inductively defined as follows. Fix a value of ν, $k_{p-1}+1 \leq \nu \leq k_p$. For all $\mu < \nu$ let everything be already determined, and let γ_μ be the P^*-degree of the reduct of $\tau^x R_\mu$, with μ ranging from 1 to $\nu-1$. Take all the closed terms b of P^*-degree at most $\mathrm{Max}(\gamma_1, \ldots, \gamma_{\nu-1}) + N_\nu$, where $N_\nu = \bar{p}(2^{(\varphi(\bar{p}-1))^{\bar{n}}-\nu})$. Determine the P^*-reducts of the closed formulas obtained by substituting x with these b's in the fixed formula R_ν. If among these P^*-reducts there is at least a formula with value False, then the P-reduct of $\tau^x R_\nu$ is defined as any one of the b's that, substituted to x in the formula R_ν, yield one of these closed formulas whose P^*-reduct is the value False; otherwise, it is defined as the constant 0. The reducts of the corresponding quantified formulas are defined accordingly: in the first case the reduct of the universal formula is the value False and the reduct of the (possibly also occurring) existential formula is the value True; vice versa in the second case. In any case, γ_ν denotes the P^*-degree of the P-reduct of $\tau^x R_\nu$. This concludes the induction step.

5.3. Correctness of the construction. It remains to prove that, for some choice of the parameters $\bar{p}^*, N_1, \ldots, N_{k_{\bar{p}-1}}$, the above-defined reduction P satisfies Condition (T). The choice has to respect the constraints arising from the definition of P and further constraints that will become apparent from how the argument for correctness can be framed, based on that definition.

Von Neumann shows that the following choice (given without explanation in the previous subsection) works for the purpose. The reasons for the choice will be clear only *after* the correctness proof.

$$\bar{p}^* = \bar{p}2^{\varphi(\bar{p}-1)^{\bar{n}}}$$

$$N_1 = \bar{p}2^{\varphi(\bar{p}-1)^{\bar{n}}-1}$$

$$N_{i+1} = \bar{p}2^{\varphi(\bar{p}-1)^{\bar{n}}-(i+1)}.$$

Recall that the function $\varphi(d)$ occurring in the exponents (defined above, Step 6) bounds the number of closed expressions of degree at most d (independently of P). Thus, $\varphi(\bar{p}-1)^{\bar{n}} \geq k_{\bar{p}-1}$ is the greatest possible number of n-tuples considered in the construction (since $n \leq \bar{n}$).

We present the correctness proof for P in such a way as to heuristically justify this choice. The proof is divided into two cases.

CASE 1. Consider a type in $g_1, \ldots, g_\sigma$. The point of this case is the following. P and P^* coincide on these types. Yet Condition (T) for P has a double universal quantification on (i) all n-tuples of expressions of P-degree $< \bar{p}$ and (ii) all expressions of P-degree $\leq \bar{p}$.

The P-degree of an expression depends on the reduction P in a nonlocal way. In particular, the P-degree of an expression can depend on the behavior of P on the types in $g_{\sigma+1}$, on which P^* is not defined.

It is enough to show that, for any reduced closed expression α,

$$deg_P(\alpha) \leq \bar{p} \Rightarrow deg_{P^*}(\alpha) \leq \bar{p}^*$$

and similarly with $<$. Then the conclusion holds because by induction hypothesis P^* satisfies Condition (T) with respect to parameter $\bar{p}^*$.

To carry out the argument we now—crucially—need to bound the P^*-degree of any reduced closed expression as a function of its P-degree. As observed above this bound is the following (for both $<$ and $\leq$):

$$\max(\gamma_1, \dots, \gamma_{k_{\bar{p}-1}}) + \bar{p}.$$

We then need to show that this quantity is $\leq \bar{p}^*$. To prove this crucial inequality we only need to observe that

$$\gamma_1 \leq N_1$$

$$\gamma_{i+1} \leq \max(\gamma_1, \dots, \gamma_i) + N_{i+1}.$$

Thus, by choice of the N_is, using a simple property of sums of exponentials,

$$\max(\gamma_1, \dots, \gamma_{k_{\bar{p}-1}}) \leq \bar{p} \sum_{j=1}^{k_{\bar{p}-1}} 2^{\varphi(\bar{p}-1)^{\bar{n}}-j} \leq \bar{p}\left(2^{\varphi(\bar{p}-1)^{\bar{n}}} - 1\right).$$

Note that $k_{\bar{p}-1}$ is the number of steps the construction needs to be performed, and since $k_{\bar{p}-1} \leq \varphi(\bar{p}-1)^{\bar{n}}$, terms in the above inequalities are well-defined.

CASE 2. Consider a type in $g_{\sigma+1}$. The goal is to prove that P satisfies Condition (T) for expressions whose type is in $g_{\sigma+1}$.

Condition (T) says that, for all directly reducible expressions $\tau^x R$, $\forall x R$, $\exists x R'$ arising from the types in $g_{\sigma+1}$ by filling the dots in the type with n-tuples of reduced closed expressions of P-degree $< \bar{p}$ (this is done by stratification in the construction), the following must hold.

$$(\forall b | deg_P(b) \leq \bar{p})(P(R[x/b]) \in T) \Rightarrow (P(\tau^x R) = 0),$$

$$(\exists b | deg_P(b) \leq \bar{p})(P(R[x/b]) \in F) \Rightarrow (P(\tau^x R) = \text{ one such } b).$$

Now, the types in $g_{\sigma+1}$ give rise to $\tau^x R_i$, $\forall x R_i$, $\exists x R'_i$ where i is the stratification index. The P-reduct is assigned it the way outlined in the previous section, which can be summarized in the following definition.

(Definition (i))

$$(\forall b | deg_{P^*}(b) \leq \max(\gamma_1, \dots, \gamma_{i-1}) + N_i)(P^*(R_i[x/b]) \in T) \Rightarrow P(\tau^x R_i) = 0.$$

$$(\exists b | deg_{P^*}(b) \leq \max(\gamma_1, \dots, \gamma_{i-1}) + N_i)(P^*(R_i[x/b]) \in F) \Rightarrow P(\tau^x R_i) = \text{one such } b.$$

Thus by Definition (i) $P(\tau^x(R_i))$ is either 0 or a b such that $deg_{P^*}(b) \leq \max(\gamma_1, \dots, \gamma_{i-1}) + N_i$ and such that $P^*(R_i[x/b]) \in F$.

Now von Neumann observes that the following condition is sufficient to ensure that P satisfies Condition (T) for the relevant types. The condition reads as follows, where $red()$ indicates the reduct by P or P^* (which coincide in the case at hand).

$$(**) \quad red(R_i[x/P(\tau^x R_i)]) \in T \Leftrightarrow (\forall b | deg_P(b) \leq \bar{p})(red(R_i[x/b]) \in T).$$

We reason as follows for proving that this condition (**) implies that P satisfies Condition (T).

Assume the condition. Suppose Condition (T) fails. First suppose it fails because the first implication in the definition of Condition (T) fails. That is,

$$(\exists b)(deg_P(b) \leq \bar{p} \wedge P(R_i[x/b]) \in F)$$

but

$$P(\tau^x R_i) \text{ is not one such } b.$$

Therefore $red(R_i[x/P(\tau^x R_i)]) \in T$, and by condition (**) it must be the case that $(\forall b | deg_P(b) \leq \bar{p})(red(R_i[x/b]) \in T$. But this contradicts our hypothesis.

Now suppose that Condition (T) fails because the second implication in the definition of Condition (T) fails. That is,

$$(\forall b | deg_P(b) \leq \bar{p})(P(R_i[x/b]) \in T),$$

but

$$P(\tau^x R_i) \neq 0.$$

Then, by Definition (i), $P(\tau^x R_i)$ is a b such that $deg_{P^*}(b) \leq \max(\gamma_1, \ldots, \gamma_{i-1}) + N_i$, and $P^*(R_i[x/b]) \in F$. Therefore, by condition (**) (right to left), there exists b such that $deg_P(b) \leq \bar{p}$ and $red(R_i[x/b]) \in F$. This contradicts the hypothesis.

We now prove that the condition (**) holds. We only prove the left-to-right direction, the other being trivial. Let us reason on the contrapositive form of the implication. Suppose that $\exists b$ such that $deg_P(b) \leq \bar{p}$ and $red(R_i[x/b]) \in F$. If we are able to show that $P(\tau^x R_i)$ is one such b, then $red(R_i[x/P(\tau^x R_i)]) \in F$ will follow.

By Definition (i) the contrapositive holds if the condition $deg_P(b) \leq \bar{p}$ is replaced by $deg_{P^*}(b) \leq \max(\gamma_1, \ldots, \gamma_{i-1}) + N_i$. Since we know that in general

$$deg_P(b) \leq \bar{p} \Rightarrow deg_{P^*}(b) \leq \max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) + \bar{p},$$

we are left with showing that

$$\max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) + \bar{p} \leq \max(\gamma_1, \ldots, \gamma_{i-1}) + N_i.$$

Now observe that, in general, it holds that

$$\gamma_{i+1} \leq \max(\gamma_1, \ldots, \gamma_i) + N_{i+1}.$$

Then by the choice of N_i and a simple property of sums of exponentials (as above), we have that

$$\max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) \leq \max(\gamma_1, \ldots, \gamma_i) + \bar{p}(2^{\varphi(\bar{p}-1)^{\bar{n}}-i} - 1).$$

It remains to show that, *in the case at hand*,

$$\max(\gamma_1, \ldots, \gamma_{i-1}) = \max(\gamma_1, \ldots, \gamma_i).$$

Assume $red(R_i[x/P(\tau^x R_i)]) \in T$ (we can do this, since this is the left-hand side of the equivalence (**) whose forward direction we are proving). By Definition (i), this can happen only if $P(\tau^x R_i) = 0$. Therefore the degree γ_i is 0 and $\max(\gamma_1, \ldots, \gamma_{i-1}) = \max(\gamma_1, \ldots, \gamma_i)$.

This concludes the correctness proof for P.

We wish to stress how the choice of the N_is is dictated by the last part of the argument we have just given.

Let us define a new function, $g(i)$, as a function yielding for every possible step i of the inductive construction, $1 \leq i \leq M = (\varphi(\bar{p}-1))^{\bar{n}}$, the number that must be added to the previous maximum P^*-degree (of the closed terms taken as counterexamples) in order to obtain the maximum P^*-degree for that step, as a function of the number, that here we shall denote by i, of tuples of closed expressions considered till then (one at the first step, two at the second, etc.; at the first step we simply take the maximum P^*-degree for that step). This is the same number that above was denoted by N_i. We have to univocally determine this function g, such that $g(i) = N_i$, $1 \leq i \leq M$.

Now, in order to carry out the correctness proof we have just given, we have seen that at the crucial step in the last part of the proof we need the following, for all i such that $1 \leq i \leq M$:

$$\max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) \leq \max(\gamma_1, \ldots, \gamma_i) + \bar{p}(2^{\varphi(\bar{p}-1)^{\bar{n}}-i} - 1),$$

which is equivalent to

$$\max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) + \bar{p} \leq \max(\gamma_1, \ldots, \gamma_i) + N_i.$$

By construction (see above), the maximum of the degrees $\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}$ is at most the number obtained by adding the maximum of the degrees considered up to a certain point (let them be $\gamma_1, \ldots, \gamma_i$, for a certain $i \leq k_{\bar{p}-1}$) to the sum of the entire possible course of values, from that point on, of the numbers that must be added to the previous maximum P^*-degree in order to obtain the current maximum P^*-degree, viz. the numbers N_i, for all $i \leq M$. I.e., we have:

$$\max(\gamma_1, \ldots, \gamma_{k_{\bar{p}-1}}) \leq \max(\gamma_1, \ldots, \gamma_i) + N_{i+1} + \cdots + N_M.$$

This formula holds for $1 \leq i \leq k_{\bar{p}-1}$ (if $k_{\bar{p}-1} = M$, in the special case $i = M$ the second member is simply $\max(\gamma_1, \ldots, \gamma_M)$, so that the formula is trivially verified). Then, it is sufficient to have, for all i such that $1 \leq i \leq M$:

$$\max(\gamma_1, \ldots, \gamma_i) + N_{i+1} + \cdots + N_M \leq \max(\gamma_1, \ldots, \gamma_i) + N_i - \bar{p}.$$

Hence, the function $g(i)$ we are looking for (which gives, by definition, the value of N_i) must be such that, for every $i < M$, the sum total of the course of its values for $i+1 \leq k \leq M$, with addition of $\bar{p}$, is the value of the function for the argument i (we take equality since we look for the minimal solution). I.e., we must have, *for all* $i < M$:

$$g(i) = g(i+1) + g(i+2) + \cdots + g(M) + \bar{p}.$$

This functional equation has (in the given integer interval) as unique solution (by a simple property of the geometric progression of common ratio 2) the function:

$$g(i) = \bar{p}(2^{M-i}).$$

The function $g(i)$ is thus univocally determined. In conclusion, the absolute maximum degree (as the sum of all values of $g(i)$ for $1 \leq i \leq M$, to which we have to add $\bar{p}$ as above, Section 5.2) is:

$$\bar{p}(2^M),$$

where the constant M is the total number of tuples. This is exactly the initial absolute maximum degree, $\bar{p}^*$, that was taken in the inductive construction.

5.4. The complexity of the procedure. Let us now look at the metatheory employed in the consistency proof. The methods used by von Neumann can be certainly formalized in

Primitive Recursive Arithmetic (PRA), and indeed the much weaker Exponential Function Arithmetic (EFA) suffices (see below for details). The proof uses a double induction on natural numbers (an induction on the number of the groups of types of τ-terms and, inside it, an induction on the degrees of the closed expressions involved in the substitutions: see above).

We now give some details concerning the bounds involved in the proof, first about the function φ defined above (Step 6). Denoting by $A_P(i)$ the number of closed expressions with degree $\leq i$ (under any reduction rule P), it is easy to see (by definition of degree) that $A_P(i)$ satisfies the following equations.

$$A_P(0) = \bar{c}, \quad A_P(i+1) \leq \bar{c} + (\bar{o} + 3\bar{s})(A_P(i))^{\bar{n}}.$$

Setting

$$\varphi(0) = \bar{c}, \quad \varphi(i+1) = \bar{c} + (\bar{o} + 3\bar{s})(\varphi(i))^{\bar{n}},$$

we have $A_P(i) \leq \varphi(i)$.

The historical interest of the bounds computed by von Neumann can be assessed as follows. The reduction rule needs (for its use and the proof of its correctness) the totality of the exponential function, since the bound (as defined in the preceding subsection) is triply exponential: $2^{(\varphi(\bar{p}-1))^{\bar{n}}}$ is exponential, and the function φ occurring in its exponent is doubly exponential in its argument, as we have just shown. Thus, von Neumann's proof can be formalized in Exponential Function Arithmetic (EFA), but not below. The bound is analogous to the lower bound established by Fischer and Rabin (see, e.g., Ferrante–Rackoff 1979) for any decision method for Presburger Arithmetic, a theory whose consistency is provable in EFA; compare also the analogous bounds for theories with one successor (ibid.).[18] On the other hand, the best known syntactic proofs of consistency for first-order predicate calculus all require Super-Exponential Function Arithmetic (SEFA), which proves the totality of the super-exponential function. These are the proofs by cut-elimination (Gentzen 1934), the uses of the criterion for the consistency of open theories given by Shoenfield in his famous textbook (1967, Section 4.3),[19] and the applications of (analogs of) Herbrand's method (see Hilbert–Bernays 1939, Hajek–Pudlák 1998 and Moser–Zach 2006). But all the mentioned results are much stronger than the mere consistency result proved by von Neumann, although he treats a system with a successor function, going beyond the pure predicate calculus (but still weaker than Robinson Arithmetic Q, which again needs SEFA for its consistency proof).

What about heuristics? Is there a heuristic justification for the choice of the bounds? So far, we have proved that they are exactly the right ones only *post factum*. This univocal determination suffices, of course, but the reader can be a little dissatisfied. We do not presume here to enter the (notoriously inaccessible) mechanisms of von Neumann's mind, of course, but we only want to see whether there is some reasonable, independent, and a little more intuitive explanation for the choice of such bounds. On this point, we tentatively suggest what follows.

---

[18] In general, for the evaluation of upper (and lower) bounds (in the sense of computational complexity) on decision procedures for theories related to the one treated by von Neumann, we refer the reader to Ferrante–Rackoff (1979). It would be of at least historical interest to further investigate whether von Neumann's proof has tighter connections with decision procedures developed in the decades following his work.

[19] Although Shoenfield's method of instantiation, as we said above, is somewhat similar to Von Neumann's procedure.

Consider the fact that

$$L = 2^M = 2^{(\varphi(\bar{p}-1))^{\bar{n}}}$$

is the number of the equivalence classes of possible types on the tuples of reduced expressions, when types are viewed simply as functions of codomain $\{0, 1\}$ and two types are viewed as equivalent when they are extensionally equivalent as such functions. Recall that (by definition of degree) for all n the numeral of the number n has degree at most n, since the only numerical constant is 0, the only numerical function symbol is the successor symbol, and each application increases the degree by one.

This could be relevant for the determination of the maximum degree of the closed terms b to be substituted: there are L equivalence classes, and at most $\bar{p}$ occurring operators; thus, $\bar{p}$ is multiplied by L.

One could also recall the classic result according to which if a first-order formula in which only monadic predicates occur is satisfiable, then it is satisfiable in a domain of at most 2^k elements, where k is the number of distinct predicates occurring in the formula. If identity occurs, we have this result: if a first-order formula without function symbols, in which only monadic predicates and identity occur, with quantifier rank (maximum number of nested quantifiers) q, and m occurring monadic predicates is satisfiable, then it has a model of cardinality at most $q(2^m)$ (for a simple proof see Börger–Grädel–Gurevich 2001, 250). If identity does not occur, one can always take $q = 1$.

This result could also be relevant, since in the inductive construction above each substitution of a tuple of reduced expressions in a Ground-type yields, after removing the initial operator, a monadic predicate on the variable of the operator.

We leave it as an open problem to give a more precise and (if possible) deeper heuristic explanation of von Neumann's choice.

§6. Conclusion. We have analyzed von Neumann's consistency proof of 1925, published in 1927. We have argued that von Neumann's work can be considered the first rigorous syntactic treatment of the consistency of first-order predicate calculus, thus filling a gap between Hilbert's lectures of the early Twenties and Ackermann's 1924 dissertation. In fact, von Neumann's method works for a stronger system, namely the first-order theory of one successor function, and essentially gives the first published rigorous syntactic treatment of the consistency of open theories. Finally we have stressed how von Neumann's reduction is metamathematically optimal, in that it is formalizable in Exponential Function Arithmetic.

§7. Acknowledgements. I wish to thank Lorenzo Carlucci, with whom this work was begun and whose contribution was fundamental, Marco Forti and Enrico Moriconi for many helpful discussions and comments, Paolo Mancosu and Richard Zach for their interest in our work, and two anonymous referees for a lot of useful corrections and remarks.

BIBLIOGRAPHY

Ackermann, W. (1925). Begründung des 'Tertium non datur' mittels der Hilbertschen Theorie des Widerspruchsfreiheits. *Mathematische Annalen*, **93**, 1–36.

Ackermann, W. (1940). Zur Widerspruchsfreiheit der Zahlentheorie. *Mathematische Annalen*, **117**, 162–194.

Börger, E., E. Grädel, & Y. Gurevich (2001). *The classical decision problem*. Springer-Verlag.

Buchholz, W., G. Mints, & S. Tupailo (1996). Epsilon substitution method for elementary analysis. *Archive for Mathematical Logic*, **35**, 103–130.

Ferrante, J. & C. Rackoff (1979). *The computational complexity of logical theories*. Berlin: Springer-Verlag.

Gentzen. G. (1934). Untersuchungen über das logische Schließen. *Mathematische Zeitschrift*, **39**, 176–210, 405–431.

Hajek, P. & P. Pudlák (1998). *The metamathematics of first-order arithmetic*. Berlin: Springer-Verlag.

Herbrand, J. (1931). Sur la non-contradiction de l'arithmétique. *Journal für die reine und angewandte Mathematik*, **166**, 1–8.

Heyting, A. (1955). *Fondements des mathématiques*. Paris: Gauthier-Villars.

Hilbert, D. (1923). Die logischen Grundlagen der Mathematik. *Mathematische Annalen*, **88**, 151–165.

Hilbert, D. (1928). Probleme der Grundlegung der Mathematik. In: *Atti del Congresso internazionale dei matematici*, 135–141. Bologna: Zanichelli.

Hilbert, D. (1929). Probleme der Grundlegung der Mathematik. *Mathematische Annalen*, **102**, 1–9.

Hilbert, D. (2013). *David Hilbert's lectures on the foundations of arithmetic and logic*, 1917–1933, W. Ewald and W. Sieg (eds.). Berlin: Springer-Verlag.

Hilbert, D. & P. Bernays (1923). Logische Grundlagen der Mathematik. Winter-Semester 1922–23. Lecture notes by H. Kneser. In: Hilbert (2013), 599–635.

Hilbert, D. & P. Bernays (1923a). Logische Grundlagen der Mathematik. Winter-Semester 1922–23. Lecture notes by P. Bernays, with handwritten notes by D. Hilbert. In: Hilbert (2013), 528–549.

Hilbert, D. & P. Bernays (1939). *Grundlagen der Mathematik*, Vol. 2. Berlin: Springer-Verlag.

Leśniewski. S. (1929). Grundzüge eines neuen Systems der Grundlagen der Mathematik. *Fundamenta Mathematicae*, **14**, 1–81.

Mancosu, P., R. Zach, & C. Badesa (2009). The development of mathematical logic from Russell to Tarski: 1900–1935. In: L. Haaparanta (ed.), *The development of modern logic*, 318–470. Oxford: Oxford University Press.

Moser, G. (2006). Ackermann's substitution method (remixed). *Annals of Pure and Applied Logic*, **142**, 1–18.

Moser, G. & R. Zach (2006). The Epsilon calculus and Herbrand complexity. *Studia Logica*, **82**, 133–155.

Shoenfield, J. R. (1967). *Mathematical Logic*. Reading, Massachusetts: Addison-Wesley.

Tait, W. W. (1965). The substitution method. *Journal of Symbolic Logic*, **30**, 175–192.

Ulam, S. (1958). John von Neumann (1903–1957). *Bulletin of the American Mathematical Society*, **64**, 1–49.

Van Heijenoort, J. (ed.) (1967). *From Frege to Gödel*. Cambridge, Massachusetts: Harvard University Press.

Von Neumann, J. (1927). Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift*, **26**, 1–46.

Von Neumann, J. (1931). Bemerkungen zu den Ausführungen von Herrn S. Leśniewski über meine Arbeit 'Zur Hilbertschen Beweistheorie'. *Fundamenta Mathematicae*, **17**, 331–334.

Von Neumann, J. (1961). *Collected works*, Vol. 1. Oxford: Pergamon Press.

Wang, H. (1963). *A survey of mathematical logic*. Amsterdam: North Holland.
Zach, R. (2003). The practice of finitism: Epsilon calculus and consistency proofs in Hilbert's Program. *Synthese*, **137**, 211–259.

DEPARTMENT CFS (PHILOSOPHY)
UNIVERSITY OF PISA
VIA PAOLI 15, 56126 PISA, ITALY
E-mail: luca.bellotti@unipi.it