Kybernetika

Statistical Approach to Proof Theory

IVAN KRAMOSIL

ACADEMIA
PRAHA

The aim of this work is to survey various approaches, methods and results connected with application of probability theory and mathematical statistics in proof theory. A special attention is devoted to various methods of statistical deducibility testing and their applications. The historical background of this branch of applied mathematics as well as a short description of the particular chapters of this work and an intuitive motivation can be found in the introductory chapter.

CONTENTS

1. INTRODUCTION

Probably every mathematician and, perhaps, also a great part of laymen would agree if we proclaimed the domain of theorem-proving to play the basic and central role in all mathematical activity. Moreover, it is again theorem-proving when the creative, sophisticated and intelligent character of mathematician's work can be seen at the first sight and in the most persuasive way. A mathematical proof is something like a bridge enabling to transit from the domain of hypothesis, no matter how interesting and supported it may be, into the world of precisely formalized and mathematically verified truths, which can serve as a ground for a further reasoning, deciding, or acting.

Hence, it is nothing strange or surprising in the fact that theorem-proving was likely one of the first fields, in which a human being – mathematician tried to compare his abilities with those of his spiritual child – a machine, computer. Even the first and, from contemporary point of view rather old-fashioned computers proved themselves to be much more effective than a man in the domain of mechanical, routine work dominanting in the area of numerical computations. The following questions (and other ones) naturally arose: Is a computer able, at least potentially, to replace his creator – mathematician also in looking for a proof of a formula or is this domain exclusively reserved for human intellectual activity proving, in this way, his eternal and decisive supremacy to machines? ... "I propose the question:

Can machines think?" expresses a similar idea A. M. Turing in his classical paper [7]. And even if, after all, a computer is able to prove theorems, are its abilities in this direction comparable with those of a mathematician? Can a computer at least help a mathematician in his theorem-proving effort?

The idea of replacing theorem-proving by calculations was expressed for the first time by Descartes (Cartesius), the designer of the first mechanical computer, in [2]. It appeared again at the beginning of this century, when the necessity to solve the "crisis of mathematics" and namely, the "crisis of set theory" lead to the notion of formalized proof, based entirely on syntactic notions and operations and so much more adequate for a mechanical treatment. Important role in the direction toward a mechanized theorem proving was played also by various results and papers proving the algorithmical decidability or undecidability of various formalized theories starting from the most simple ones (e.g., propositional calculus), but including also some very complex and rich theories as, e.g., Presburger Arithmetics (cf. [4]). Even the well-known Hilbert program contained the idea of transforming all mathematics into calculation.

Gödel's famous results concerning the formalized arithmetic, in spite of their immense theoretical sense and importance discouraged, in a degree, further attempt to mechanize the mathematician's work, especially theorem proving. An activity in this direction has been renewed since the coming on scene of the first electronic computers. Thanks to their great effectivity in numerical and algorithmical operating even the semi-decision procedures, known to be only theoretically applicable for formalized theories in general, seemed to be "effective enough" to be worth of being implemented and applied.

Most of the procedures for automated theorem proving proposed in those days were based on Herbrand's theorem which profits of the semantical completeness of the first-order predicate calculus and transforms theorem-proving in this predicate calculus into a sequence of theorem-provings in the propositional calculus which is known to be decidable (cf. Chapter 3 or [1] for more details). An important step forward in this direction represents the well-known and already classical paper by J. A. Robinson [6] introducing the notion of resolution and resolution principle. This paper has been followed by a great number of other papers and books, examining the Robinson's basic ideas, modifying and improving them, dealing with implementation or with various applications, etc. From the quantitative point of view the decade 1965—1975 could be considered as the most successful in all the history of automated theorem-proving, however, taking into account also the qualitative aspects we ought to be more careful before claiming something like this. We shall come back to this question later.

The already mentioned Turing's paper [7] with his suggestive question "can machines think?" is usually taken as the birthpoint of a new branch of applied science, called artificial intelligence. Theorem proving has been considered as an important part of this new science since its very origins. The aim of artificial intelligence has

been, and still is, to build mathematical (in the broadest sense) formulations and theories for many processes, procedures and activities which are usually considered as intelligent, if performed or executed by a human subject. As mathematical logic can be seen as a metatheory of formalized mathematical theories, it is nothing strange in noticing that many attempts have been done to apply means and tools of mathematical logic in various subdomains of artificial intelligence. Such an approach is naturally influenced by the fact, that various devices designed for realizing the artificial intelligence procedures, algorithms or heuristics, can work only at the syntactic level, i.e., various data, inputs or commands can be taken only as syntactic configurations of symbols, it is beyond the power of the device (computer, automaton, robot) in question to "understand" the inputs semantically. Mathematical logic deals with the relations between the syntax and the semantic of formalized theories, hence, it has a wide range of applications in artificial intelligence.

Also the proof theory has been used in this connection. The problem of verifying whether an operator is or is not applicable in the actual stage of a problem solving as well as the problem of verifying whether a goal has been already reached or not can be converted into that of proving or disproving appropriate formulas describing the applicability conditions of operators or the desired goal. Even a plan itself for solving a problem can be excerpted, under some conditions, from the proof of an appropriate formula (cf. Chapter 8 for more details). Also many questions concerning formal representations of the environment and subject's knowledge can be expressed in and solved by the means of proof theory; some of them will be also mentioned in Chapter 8.

However, we can see at the first sight, that the role of formalized proofs in these applications is different from their role in the classical, pure mathematics. In pure mathematics the length of proof, time, effort and other possible expenses necessary for obtaining the proof, etc. do not play any role, a hypothesis had simply remained to be a hypothesis until it was proved or disproved. In artificial intelligence systems the situation is quite different; a decision must be taken and an action executed in time, in a real time comparable with other changes taking place in the environment. Even the best decision taken too late is useless, as the situation may be already decisively changed.

Hence, the necessity occured to investigate the theorem-proving procedures and algorithms also from the point of view of their time and storage pretensions. Almost simultaneously with the artificial intelligence, and also as a consequence of the birth of computers a new field of mathematics emerged, the theory of computational complexity. Various computation and decision procedures including theorem-proving algorithms became very soon the objects of investigation of this new theory and some interesting, but not too hopeful results have been achieved. Roughly speaking, theorem-proving algorithms have been proved to belong to the group of algorithms with the highest computational complexity, i.e. to those which are the most time and storage consuming. Expressed in a more detailed form: the theory of computational

complexity does not investigate the complexity of particular computations but rather of whole classes of computations with the same program (algorithm), but various inputs. The complexity or extent of these inputs are measured (e.g., by the length of the input taken as a word in a formal alphabet) and the computation cmoplexity is expressed as a function of this input complexity. If this function is of polynomial type (can be majorized by a polynomial function) the procedure is usually considered as practically applicable (a hypothesis justified informally by the practice of specialists dealing with computers). If this is not the case, the computation (algorithm, procedure) is called to be of exponential type and such procedures are usually taken as practically useless, even if, from the theoretical point of view they may be worth studying. This classification of computational procedures into those which are of polynomial type and the exponential ones has been proved to be very deep, sharp and stable in the sense that no scale or implementation change, no transformation of the formalism used for expressing the algorithm can transform an exponential procedure into a polynomial one, this difference lies very deeply in the internal character of the algorithm in question. On the other hand, the coefficients of the corresponding exponential or polynomial functions expressing or majorizing the quantitative computational complexity can be always modified by an approximate implementation.

What is of the crucial importance for our further reasoning is the fact that all theorem-proving algorithms or semi-algorithms have been proved to be of exponential type (in fact, many of them are of superexponential types) and, hence, not useful for a practical use in technical devices design with the intention to act in a real world and in a real time according to the dynamical character of the environment. Some experiments with robots using theorem-proving as theoretical basis of their decision making have proved this theoretical conclusion (cf. [3]).

Hence, we can see, that any theorem-proving procedure possesses at least two aspects going against each other – the demand of mathematical correctness and absolute reliability and that of a practical applicability. It is not possible to maintain both the demands simultaneously, something must be abandoned. The classical, pure mathematics strictly preferred the logical rigorousness and correctness, leaving the question of feasibility opened; in other words, no storage or time savings can justify the replacing of a correct theorem-proving procedure by an unprecise one, no matter how small the probability of error may be. The basic idea of this work may be expressed as "choosing the other outcome" from the dilema mentioned above, in other words, we shall admit the possibility, that the result of a theorem-proving procedure may be wrong, from time to time, but the probability of such a failure is "small enough" and if it is payed by a significant decrease of computational complexity, such a procedure may be admissible, even more admissible than another, precise, but too complicate one. In Chapter 4 we give a more detailed argumentation in favor of this approach. A much more general expression of the same idea by M. O. Rabin can be found in [5] in the form of a metamathematical hypothesis.

Probably the first attempts to introduce probability theory and statistics into the domain of theorem-proving were made by A. Špaček in 1959–60, his ideas have been developed and modified later by the author of this work who has confronted them with the demands and problems of other branches of artificial intelligence, namely with automated problem-solving and robotics. Also some works of W. van Vestrhenen and his research group in the Netherlands as well as some papers by S. Ju. Maslov and E. D. Rusakov, Leningrad, U.S.S.R., are devoted to the same or similar subjects. A survey of all known works and approaches dealing with application of statistics and probability theory in the domain of automated theorem-proving will be the main goal of this work. In order to facilitate the reading of what follows to a reader not familiar enough with mathematical logic and proof theory we explain some basic ideas and principles of these branches in Chapters 2 and 3. In Chapter 2 we describe, very briefly, the way of modern mathematics leading to the notion of formalized theory, the elementary stones necessary in order to build a formalized theory, we define formalized proofs and theorems and compare these notions with the semantically based notions of satisfiability and truth. By investigating the relations between validity and provability we shall elucidate the powers, but also limits of formalization in modern mathematics.

Chapter 3 explains the basic principles of the most developed method of automated theorem proving based on Robinson's resolution principle. We shall get familiar with the notion of semantical completeness of the first-order predicate calculus, with Herbrand's theorem and Herbrand's universum. We describe the resolution principle and resolution-based proving procedures, also some of their modifications and improvements are mentioned.

Then we discuss, at the beginning of Chapter 4, the theoretical as well as the practical limitations of automated theorem-proving procedures. We proceed by explaining the notions necessary for constructing a general model of statistical decision making and by suggesting of such a model. Statistical methods of theorem-proving are described as a special case of this general model, the statistical as well as the deterministic approaches to automated theorem proving are compared and confronted with each other from the theoretical and practical points of view. Some arguments are suggested favorizing the statistical methods of theorem proving when the applicational approach is considered as the dominating one.

Chapter 5 describes the basic Špaček's model of theoremhood testing in at random sampled extensions of a given formalized theory and some other modifications and improvements of this basic idea. The testing problem is transformed into that of a parametric hypothesis testing. The theoretical conditions under which these tests work as well as their implementational possibilities are discussed.

As a practical realization of a statistical theorem-proving method will not be limited by testing of one single formula, we may try to profit of the formulas already tested and proclaimed to be theorems or disproved as non-theorems, when testing another

formula. Some possibilities of such experience use and learning are discussed in Chapter 6.

The next, seventh, chapter briefly introduces other methods of statistical theorem proving, e.g., random sampling in resolution-based procedures or stochastic generation of formalized theories by random sampling of premises and deduction rules during the process of theorem-proving.

Various applications of statistical theorem-proving methods are mentioned in Chapter 8, namely those which are close to artificial intelligence, automated problem solving and automated plan formation, especially for uncertain or incomplete plans.

Chapter 9 deals with other conceptions how to introduce uncertainty into the formalized proofs (fuzzy logic, incorrect deduction rules), some more general considerations and results concerning the relations between the admission of a possibility that a procedure fails and significant time and storage savings (R. M. Karp, M. O. Rabin) are also discussed. The last Chapter 10 tries to evaluate the actual state of the surveyed branch of mathematics and outlines some possibilities for its further development.

Let us emphasize our intention to conceive this work as a surveyal one, with the aim to sketch the outlines of this new field of science and to offer a first insight to anybody interested in this field of mathematics but not having any special preliminary knowledge about it. In no case we would like to dupplicate or replace special papers and other sources dealing with matters which will be mentioned below. It is also why proofs of various assertions stated in what follows will be often omitted or restricted, giving, at the same time, a reference when a detailed proof can be found. As a rule, we shall introduce here only those proofs or parts of proofs which are, because of their ideas, techniques or partial results, of a metodological or illustrative value for the subsequent explanation. Precise formalizations of the given concepts and assertions will be offered only in case when the necessary effort and time and space expenses needed to this goal are proportional to the importance of the notion or statement in question in the used context. The same care as to formal preciseness will be devoted also to clearness and lucidity, the explanation will be enriched by illustrative examples, if possible. Preliminaries with which the reader is supposed to be familiar as well as the notation and symbolics used are mentioned by the occasion of their first occurence.

As far as the references are concerned, we give at the end of each chapter the list of references mentioned for the first time in this chapter. Because of the character of this work (an appendix appearing through whole volume of Kybernetika) we prefer this way of references listing to the usually adopted one, introducing the list of all references at the very end. When referring to an item of the list of the present chapter we use the single enumeration, e.g., "... as shown in [6]...", when mentioning a reference of another, as a rule, one of the preceding chapters, we make profit of double enumeration, the first numeral referring to the chapter, e.g., "... as can

be found in [6.2]..." refers to the second item of the reference list at the end of Chapter 6. Theorems, definitions, examples and relations are numbered by the usual double enumeration, the first numeral referring to the chapter in question.

REFERENCES

[1] C. L. Chang, R. T. C. Lee: Symbolic logic and mechanical theorem proving. Academic Press, New York and London 1973.

[2] R. Descartes: Discours sur la méthode. Paris 1637.

[3] N. J. Nillson: A mobile automaton: An application of artificial intelligence techniques. Proc. of the First Internat. Joint Conf. on Artificial Intelligence, Washington. D. C., 509—520.

[4] M. Presburger: Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation herrvortritt. Comptes rendus du I-er Congrès des Mathématiciens des Pays Slaves, Warsaw, 1930, 92—101.

[5] M. O. Rabin: Theoretical impediments to artificial intelligence. In: Information Processing 1974, North Holland Publ. Comp., Amsterdam 1974, 615—619.

[6] J. A. Robinson: A machine oriented logic based on the resolution principle. J. of the Assoc. for Comp. Machinery *12* (1965), 1, 23—41.

[7] A. M. Turing: Computing machinery and intelligence. Mind *59* (1950), 433—460, also in: Computers and thought (E. A. Feigenbaum and J. Feldman, Eds.), McGraw-Hill, New York 1967. (Russian translation: Moscow 1973.)

2. FORMALIZED THEORIES

Scientific theories are as old as the science itself. The very origins of science, usually dated to the times of antic Greece, are inseparably linked with the first attempts of human intellect to describe and systemize somehow one's experience based on the repeated observations of the surrounding world. From the most general point of view a scientific theory can be seen as a collection of sentences of a language, originally the natural language used by the community in question; the sentences are to describe the environment or the universe, at least from some aspects and are, more or less, supported by observations as well as by conclusions driven from immediately observable facts to more abstract ones according to some rules of reasoning, generally adopted as sound and corresponding to the common sense. In this, broadest sense a scientific theory can be formalized by a pair $\langle \mathscr{L}, \mathscr{T} \rangle$, where $\mathscr{L}$ is the used language considered as the set of all its sentences and $\mathscr{T} \subset \mathscr{L}$ is the subset of those sentences which are taken or claimed as the valid ones. However, at this stage of scientific development the collections $\mathscr{L}$ and $\mathscr{T}$ of sentences are usually not sharply defined, so that such a set-theoretic formalization would be only of a limited value.

This early period of science can be called descriptional and cummulative, as it describes the facts concerning the environment and these pieces of information are simply cummulated to each other. The further step in scientific development is to systemize somehow the knowledge according to some principles. First attempts in this direction were done, again, already in antic Greece by Euclides. His idea was

simple and genial: to choose a small number of basic and immediately observable sentences from which all the other valid sentences could be derived by some rules of reasoning. Euclides was successful in applying this idea to geometry, he chose five elementary geometric assertions, usually called *axioms* or *postulates* and derived from them all valid geometric sentences known in his age. A mathematical, or, in general, scientific theory described in this way is called *axiomatic theory*, it can be formalized by a pair $\langle \mathscr{L}, Ax \rangle$, where $\mathscr{L}$ is a language and $Ax \subset \mathscr{L}$ is the set of axioms, this set is strictly defined and is decidable in the sense that for any sentence of $\mathscr{L}$ we can effectively decide whether it is or is not an axiom.

Aristoteles axiomatized, in a similar way, also a part of mathematical logic and, for more than two thousand years, Euclidean geometry and Aristotelian logic served as ideal patterns to which all other branches of mathematics and other sciences should tend. Some successes in this direction have been really achieved the most important among them being, probably, the Newtonian mechanics.

Since the end of 19th century the set theory has been considered as the basic branch of mathematics due to fundamental works of Cantor, Frege and Dedekind, who recognized the basic role of set theory in the process of logical and systematical building of modern mathematics. Hence, a remarkable effort has been put forward in order to axiomatize the set theory. The result was surprising and threated to destroy all grounds of mathematics – the paradoxes occured.

Paradoxes were known already to old Greeks, but these paradoxes usually resolved from an illegal mixing of language and meta-language and could be explained after a short reasoning. Other paradoxes, as, e.g.. the Zenon's ones, were not paradoxal at all from the logical point of view, they were only hardly accessible to an intuitive imagination. However, the set-theoretic paradoxes (Russell's paradoxon, Buralli-Forti paradoxon, etc.) could not be removed from mathematics in a simple way and they proved the necessity to revise critically all the formal grounds of contemporary mathematics.

This revision showed the inevitable necessity to formalize precisely the language of mathematics or, particularly, of a mathematical theory and to formalize, as well, the rules of "sound reasoning", i.e., the deduction rules, as the paradoxa showed that our belief in the "common sense" had failed. All this effort led to the notion of *formalized theory* which is the crucial one of modern mathematixal logic. A very general definition of this notion was given, under the title of *simple type theory*, by Whitehead and Russell in their monumental monography [12]. In our days this formalization is considered as rather old-fashioned and over-dimensioned, preference is given to restricted systems fitted for particular theories (propositional calculus, first-order or second-order predicate calculi, etc.). However, simple type theory copes with the generality level on which all our explanations in this work will be formulated and it is, moreover, appropriate from the methodological point of view for our introducing of the principles of formalized theories.

Let us start with a formal explanation.

Definition. 2.1. Let $*$ (), be formal symbols different from all other symbols occuring below.

(a) $*$ is a logical type;

(b) if n is a positive integer and $c_1, c_2, \ldots, c_n$ are logical types, then $(c_1, c_2, \ldots, c_n)$ is also a logical type;

(c) there are no other types. The set of all logical types is denoted by τ.

Intuitively, logical types are ascribed to various mathematical objects, e.g., the elementary logical type $*$ is that ascribed to individuals, type $(*)$ belogs to set of individuals, $(*, *)$ to binary relations between individuals, $((*), (*))$ to binary relations between sets of individuals, etc. The definition of logical types can be generalized by admitting more than one elementary types, denoting them, e.g., $*_1, *_2, \ldots$.

The next definition describes the language of the simple type theory.

Definition 2.2. Let τ be the set of all logical types, let there be given, for each $c \in \tau$,

(a_1) an infinite sequence $x_1^c, x_2^c, \ldots$ of logical indeterminates of the type c (we use the expression "indeterminate" instead of the more often used one "variable" in order to reserve the later term to be used in the context "random variable" below);

(b_1) a finite or infinite sequence $\varrho_1^c, \varrho_2^c, \ldots, \varrho_n^c, \ldots$ of relational symbols of the type c (relational symbols of type $*$ are called *individual constants*).

Let $c = (c_1, \ldots, c_n) \in \tau$. Let $t_i^{c_i}$, $i = 1, 2, \ldots, n$, be an indeterminate or relational symbol of the type c_i, then $\varrho_i^c(t_1^{c_1}, t_2^{c_2}, \ldots, t_n^{c_n})$ and $x_i^c(t_1^{c_1}, t_2^{c_2}, \ldots, t_n^{c_n})$, $i = 1, 2, \ldots$, are *elementary formulas.*

Let $\forall$(general or universal quantifier), $\exists$ (existential quantifier), $\wedge$ (conjunction). $\vee$ (disjunction), $\neg$ (negation), $\rightarrow$ (implication), $\equiv$ (equivalence) be auxiliary symbols.

(a_2) Each elementary formula is a *well-formed formula* (of the simple type theory).

(b_2) If A, B are well-formed formulas and x_i is an indeterminate, then $\neg(A)$, $(A) \wedge (B)$, $(A) \vee (B)$, $(A) \rightarrow (B)$, $(A) \equiv (B)$, $(\forall x_i)(A)$, $(\exists x_i)(A)$ are well-formed formulas (some pairs of brackets will be often omitted if no danger of confusion threats).

(c_2) There are no other well-formed formulas. The set of all well-formed formulas will be denoted by $\mathscr{L}$ and called the *language of the simple type theory.*

In order to simplify the definition above we have omitted the function symbols using the fact that functions, being a special kind of relations, are not inevitable in logical formalisms. Nevertheless, in more simple formalized languages – fragments of the simple type theory language defined above – we distinguish among relations and functions and the later are used to generate terms (indeterminate and constants are terms, functions map terms into terms and terms of appropriate types

are connected with relational indeterminates or symbols to form elementary formulas).

The most important and most often met fragments of the language of the simple type theory are *propositional language, first-order predicate language* and *second-order predicate language.* Propositional language results from the simple type theory language by replacing elementary formulas by new, propositional indeterminates (denoted, as a rule, by $p, q, r, p_1, q_1, \ldots$) and by erasing of all occurences of $(\forall x_i)$ and $(\exists x_i)$. I.e., well-formed propositional formulas are generated from propositional indeterminates by propositional connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\equiv$. First-order predicate language contains only one infinite sequence of indeterminates, namely those of type $*$, and a sequence of relational symbols $\varrho_1, \varrho_2, \ldots, \varrho_k, \ldots$, each ϱ_i being of the type $*$ or $(c_1, c_2, \ldots, c_{n(i)})$, $c_j = *$, $j \leqq n(i)$, $i \leqq k$. As we have already mentioned, sometimes a particular sequence of functional constants $f_1, f_2, \ldots, f_l$ is considered, together with their arities $k_1, k_2, \ldots, k_l$. Individual indeterminates and relation symbols of type $*$ (individual constants) are terms. If $t_1, t_2, \ldots, t_{k(i)}$ are terms, then $f_i(t_1, \ldots, t_{k(i)})$ is also a term. If ϱ_i is a relational symbol of type $(c_1, c_2, \ldots$ $\ldots, c_{n(i)})$, $c_j = *$ for all $j \leqq n(i)$, and if $t_1, t_2, \ldots, t_{n(i)}$ are terms, then $\varrho_i(t_1, t_2, \ldots$ $\ldots, t_{n(i)})$ is an elementary formula. The construction of the first-order predicate language with functions is finished by closing the set of elementary formulas with respect to proposional connectives (functors) and quantifiers.

Second-order predicate language contains indeterminates and relational symbols of type $*$ as well as of all types $(c_1, c_2, \ldots, c_n)$, $n = 1, 2, \ldots$, $c_j = *$, $j \leqq n$, and finite or infinite sequences of relational symbols of all types in which the admitted indeterminates may immediately occur (if functions are treated separately, terms of different types must be defined). The creation of elementary formulas and well-formed formulas runs as in Definition 2.2.

It has been shown that, admitting as meaningful formulas of mathematics only those which are well-formed with respect to the rules of simple type theory language, we avoid from our formalism all the paradoxa known until now. E.g., the paradoxon based on the notion of "the set of all sets" cannot be obtained in simple type language as it inevitably requires to consider a formula of the form $x^c(x^c)$, x^c being an indeterminate of the type c. However, such an expression cannot be well-formed in the simple type language, as in this language the logical type of the "head" of an elementary formula always differs from the logical types of all arguments.

Since now we always suppose that $\mathscr{L}$ is either the simple type language or some of its fragments and we shall briefly say that $\mathscr{L}$ is a *formalized language.*

An occurence of an indeterminate x in a w.f.f. $A \in \mathscr{L}$ is called *bound*, if it is a part of a subformula of A, beginning with $(\forall x)$ or $(\exists x)$. An occurence of an indeterminate which is not bound is called *free.* A w.f.f. is called *closed*, if it does not contain any free occurence of any indeterminate, if a w.f.f. A contains free occurences of just the indeterminates $x_{i(1)}, x_{i(2)}, \ldots, x_{i(n)}$, then the w.f.f. $(\forall x_{i(1)}) \ldots (\forall x_{i(n)}) A$ is called the *universal* or *general closure of A.*

Having formalized the notion of acceptable formulas we may begin the next step of the process of creation of a formalized theory – the choosing of appropriate *axioms*. There are two kinds of axioms, the *logical* and the *extralogical* ones. The logical axioms are usually common to all formalized theories based on the same kind of language and express the most common patterns of reasoning formalizable in the language in question. In fact, there are many various formalization of these logical axioms for a given language, the word "common" used above should be understood in the sense that the sets of all logical consequences of these logical axioms are always the same. Various intuitionistic and other non-classical systems are not considered here.

Let us adopt an infinite system of logical axioms described by the following finite set of axiom schemata. In any axiom schema symbols A, B, C, ... are meta-language indeterminates; if they are replaced by particular w.f.f.s, particular axioms result (i.e., the set of all axioms is recursive, an inevitable condition which each axiomatic system is to satisfy).

(A1) $A \to (B \to A)$

(A2) $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$

(A3) $(\neg A \to \neg B) \to (B \to A)$

(A4) $(A \vee B) \to (\neg A \to B)$

(A5) $(\neg A \to B) \to (A \vee B)$

(A6) $(A \wedge B) \to \neg(\neg A \vee \neg B)$

(A7) $\neg(\neg A \vee \neg B) \to (A \wedge B)$

(A8) $((A \to B) \wedge (B \to A)) \to (A \equiv B)$

(A9) $(A \equiv B) \to ((A \to B) \wedge (B \to A))$

(A10) $((\forall x)(A \to B)) \to (A \to (\forall x) B)$, if x is any indeterminate not occuring freely in A

(A11) $(\forall \mathbf{x}) A \to S_b^{\mathbf{x}} A$, where $\mathbf{x}$ is an indeterminate, b is an indeterminate or relational symbol of the same type as $\mathbf{x}$ and no free occurence of $\mathbf{x}$ in A occurs in a subformula of A which begins with $(\forall b)$ or $(\exists b)$.

(A12) $((\exists x) A) \equiv (\neg((\forall x) \neg A))$, where x is an indeterminate.

This system of logical axioms does not pretend to be minimal or logically independent. It is a well-known fact that even our system of propositional connectives and quantifiers is superfluously rich. e.g., only $\neg$, $\to$ and $\forall$ suffice. In such a case just the axioms (A1)–(A3), (A10) and (A11) remain, the other axioms are either added as definitory axioms or the other connectives and the existential quantifier are taken as meta-language abbreviations for certain formulas.

The extra-logical axioms differ from case to case and express the attributes of the

theory which we are to formalize. As the identity relation occurs in almost all mathematical theories, the three axiom schemata governing this relation cam be seen as very often used examples of extra-logical axioms:

(I1) for all $c \in \tau$, $j = 1, 2, \ldots, (\forall x_j^c)(x_j^c = x_j^c)$

(I2) for all $c \in \tau$, $j, k = 1, 2, \ldots, (\forall x_j^c)(\forall x_k^c)((x_j^c = x_k^c) \to (x_k^c = x_j^c))$

(I3) for all $c \in \tau$, $j, k, l = 1, 2, \ldots$
$(\forall x_j^c)(\forall x_k^c)(\forall x_l^c)((x_j^c - x_k^c) \to ((x_k^c = x_l^c) \to (x_j^c = x_l^c)))$.

To be precise, we should express the identity relation by a systen $\{= (c),\ c \in \tau\}$ of binary relational symbols for various types, in (I1)–(I3) we omit the index c. Identity axioms are sometimes included among the logical ones, but we do not follow this pattern here. As other examples of extra-logical axioms can serve axioms of groups or Boolean algebras, if group theory or Boolean algebra theory are to be formalized.

Finally, we must formalize the deduction or reasoning rules enabling to derive new assertions from axioms and from the already derived ones. There are, again, several systems of such deduction rules equivalent from the point of view of the set of all deducible consequences. Let us introduce here, again in the form of deduction schemata, a simple system consisting of two deduction rules:

(R1) If A, B are w.f.f.s., then $A \to B$ and A yield B (modus ponens).

(R2) If A is a w.f.f. and x is an indeterminate, then A yields $(\forall x)(A)$ (generalization).

Hence, the basic notion of this chapter can be defined as follows.

Definition 2.3. A *formalized theory* is a triple $\langle \mathscr{L}, \mathscr{A}x, \mathscr{R} \rangle$, where $\mathscr{L}$ is a formalized language, $\mathscr{A}x \subset \mathscr{L}$ is a recursive set of axioms and $\mathscr{R}$ is a recursive set of deduction rules (formally, $\mathscr{R} \subset \{f : \mathscr{P}_{fin}(\mathscr{L}) \to \mathscr{L}$ partially$\}$, axioms can be seen as deduction rules with no premises). Namely, if $\mathscr{L}$ is the formalized language defined in Definition 2.2., $\mathscr{A}x = \{(A1), (A2), \ldots, (A12)\}$, $\mathscr{R} = \{(R1), (R2)\}$, the resulting formalized theory is called *simple type theory* (of course, this name is used also in case when equivalent systems of axioms and deduction rules are used).

The following definition formalizes one of the principal notions of this work, that of a formalized proof.

Definition 2.4. Let $\langle \mathscr{L}, \mathscr{A}x, \mathscr{R} \rangle$ be a formalized theory. A finite sequence $\langle a_1, a_2, \ldots, a_n \rangle$ of formulas from $\mathscr{L}$ is called *formalized proof*, if for every $j \leqq n$, a_j either is an axiom or follows from some formulas $a_{i(1)}, a_{i(2)}, \ldots, a_{i(k)}$, $i_l < j$ for all $l = 1, 2, \ldots, k$, with respect to a deduction rule. A formalized proof $\langle a_1, a_2, \ldots, a_n \rangle$ is called a proof of its last formula, i.e., of a_n. A formula of $\mathscr{L}$ is called a *theorem* (of the formalized theory in question) if there exists a proof of this formula.

The set of all theorems will be denoted by $\mathscr{T} = \mathscr{T}(\mathscr{L}, \mathscr{A}x. \mathscr{R}) \subset \mathscr{L}$, the set of all formalized proofs will be denoted by $\mathscr{D} = \mathscr{D}(\mathscr{L}, \mathscr{A}x, \mathscr{R}) \subset \mathscr{L}^* = \bigcup_{i=1}^{\infty} \mathscr{L}^i$. We write sometimes $\langle \mathscr{L}, \mathscr{T} \rangle$ instead of $\langle \mathscr{L}, \mathscr{A}x, \mathscr{R} \rangle$, if the set of theorems and not the specific way of its definition by $\mathscr{A}x$ and $\mathscr{R}$ lies in the medium of our attention.

Notice that all the notions defined above are of purely syntactic and combinatoric nature without any reference to semantics. The main problem of every mathematical theory, namely that of looking for valid assertions or deciding whether an assertion is or is not valid in the considered theory, is now converted into that of deciding whether a formula of a formalized theory is or is not a theorem, i.e., whether it is or is not deducible (provable). In which degree, under which conditions and by which means is this question decidable?

In every case, this question is semi-decidable in the sense that the set $\mathscr{T}$ of all theorems is recursively enumerable. This means that there exists an algorithmical procedure giving as its outputs one theorem after another and such that every theorem eventually occurs in this sequence – but the index of its first occurrence is not effectively computable. The set $\mathscr{T}$ is decidable, i.e., recursive, only in some cases of the most simplest (in a sense) theories, the *propositional calculus* being probably the best known case (propositional calculus is based on the propositional language, axioms (A1)–(A9) and deduction rule (R1), it is decidable, for example, by the well-known zero-one procedure). If $\mathscr{T}$ is decidable, the theory $\langle \mathscr{L}, \mathscr{T} \rangle$ itself is called *decidable*.

On the other hand, the set $\mathscr{D}$ of formalized proofs is always recursive, i.e., decidable. In fact, having a finite sequence of formulas of $\mathscr{L}$ we can effectively decide, for every formula in the sequence, whether it is or is not an axiom (the set of all axioms is recursive), In the negative case we can effectively examine all finite sets of formulas preceding the formula in question and try, whether they yield it or not with respect to a deduction rule (these rules are also supposed to be recursive, usually it suffices to consider just the rules (R1) and (R2)). Using the notation borrowed from the recursion theory we may say that the set $\mathscr{D}$ of formalized proofs is always a Σ_0-set (hence, at the same time, $\prod_0$-set), but the set $\mathscr{T}$ of all theorems is, in general, a Σ_1-set.

Mathematical logic aims not only to separate semantical notions from the syntactic ones and to translate the former into the later, but it also tries to formalize the semantical notions of satisfiability, validity, truth, etc. and to study their relations to the syntactical notions of proof, deducibility and theoremhood. As some assertions of this kind are important for the most known methods of automatic theorem-proving, namely in the first-order predicate theories (cf. the next chapter), we introduce these notions and results here, limiting ourselves, for the sake of simplicity, by the first-order languages.

Definition 2.5. Let $\mathscr{L} = \langle X, c_1, c_2, \ldots, \varphi_1^{n(1)}, \varphi_2^{n(2)}, \ldots, \varrho_1^{m(1)}, \varrho_2^{m(2)}, \ldots \rangle$ be a first-order predicate language, $X = \langle x_1, x_2, \ldots \rangle$ is the sequence of individual

indeterminates (i.e., intdeterminates of type $*$), $\langle c_1, c_2, \ldots \rangle$ is a finite of infinite sequence of individual constants, $\langle \varphi_1^{n(1)}, \varphi_2^{n(2)}, \ldots \rangle$ is a finite or infinite sequence of functional symbols, every $\varphi_i^{n(i)}$ being of the arity n_i, $\langle \varrho_1^{m(1)}, \varrho_2^{m(2)}, \ldots \rangle$ is a finite or infinite sequence of relational symbols, every $\varrho_i^{m(i)}$ being of the type $\langle c_1, c_2, \ldots, c_{m(i)} \rangle$, $c_j = *$, $j \leqq m(i)$ (having separated individual constants we may suppose that $n(i) > 0$, $m(i) > 0$, $i = 1, 2, \ldots$).

Relational structure or *model* $\mathscr{M} = \langle M, a_1, a_2, \ldots, f_1^{h_1}, f_2^{h_2}, \ldots, r_1^{k_1}, r_2^{k_2}, \ldots \rangle$ is a mathematical object such that M is a nonempty set, $\langle a_1, a_2, \ldots \rangle$ is a finite or infinite sequence of elements of M, $\langle f_1^{h(1)}, f_2^{h(2)}, \ldots \rangle$ is a finite or infinite sequence of functions over M, in general, partial with arities h_i, i.e., every f_i is defined in $M^{h(i)}$ and takes its values in M, $\langle r_1^{k(1)}, r_2^{k(2)}, \ldots \rangle$ is a finite or infinite sequence of relations over M with arities k_i, i.e., every r_i is a subset of $M^{k(i)}$. The model $\mathscr{M}$ is called *to be of the same signature as the language* $\mathscr{L}$, if the number of individual constants is the same as the number of separated elements $a_1, a_2, \ldots$, the number of functional symbols in $\mathscr{L}$ is the same as the number of relations in $\mathscr{M}$ and if, moreover, for all $j = 1, 2, \ldots, n_j = h_j$, $m_j = k_j$.

Let $\mathscr{L}$ and $\mathscr{M}$ be of the same signature. An evaluation I of the language $\mathscr{L}$ in the model $\mathscr{M}$ is a mapping defined on $X \cup \{c_1, c_2, \ldots\} \cup \{\varphi_1, \varphi_2, \ldots\} \cup \{\varrho_1, \varrho_2, \ldots\}$, taking its values in $M \cup \{a_1, a_2, \ldots\} \cup \{f_1, f_2, \ldots\} \cup \{r_1, r_2, \ldots\}$ and such that $I(c_i) = a_i$, $I(\varphi_i) = f_i$, $I(\varrho_i) = r_i$ for all $i = 1, 2, \ldots$. Two evaluations, I_1 and I_2, of $\mathscr{L}$ in $\mathscr{M}$ are called *equivalent*, $I_1 \approx I_2$ in symbols, if $I_1(y) = I_2(y)$ for all $y \notin X$, i.e., if they differ, at most, in the way in which indeterminates are evaluated.

Let $\mathscr{L}$ and $\mathscr{M}$ be of the same signature, let I be an evaluation of $\mathscr{L}$ in $\mathscr{M}$. Define the set $\mathscr{V}_I \subset \mathscr{L}$ of formulas which are *valid* in the evaluation I in the following inductive way:

(a) the mapping I is extended to a mapping of terms into M by setting

$$I(\varphi_i(t_1, t_2, \ldots, t_{n(i)})) = (I\varphi_i)(It_1, It_2, \ldots, It_{n(i)}),$$

where $t_1, t_2, \ldots, t_{n(i)}$ are terms;

(b) let $t_1, t_2, \ldots, t_{m(i)}$ be terms, then the atomic formula $\varrho_i(t_1, t_2, \ldots, t_{m(i)})$ is valid in I iff $(I\varrho_i)(It_1, It_2, \ldots, It_{m(i)})$ holds, i.e., iff

$$\langle It_1, It_2, \ldots, It_{m(i)} \rangle \in r_i = I(\varrho_i);$$

(c) a formula $(\forall x) A(x)$ is valid in I, iff $A(x)$ is valid in all evaluations I_1 of $\mathscr{L}$ in $\mathscr{M}$ such that $I_1 \approx I$,

(d) $\neg A$ is valid in I, iff A is not valid in I;

(e) $A \vee B$ is valid in I, iff either A is valid in I, or B is valid in I;

(f) $(\exists x) A(x)$ is valid in I, iff $\neg((\forall x) \neg A)$ is valid in I;

(g) $A \to B$ is valid in I, iff $(\neg A) \vee B$ is valid in I;

(h) $A \wedge B$ is valid in I, iff $\neg((\neg A) \vee (\neg B))$ is valid in I;

(i) $A \equiv B$ is valid in I, iff $(A \to B) \wedge (B \to A)$ is valid in I.

Clearly, $\approx$ is an equivalence relation in the set of all evaluations of $\mathscr{L}$ in $\mathscr{M}$. Any equivalence class with respect to $\approx$ is called an *interpretation* of $\mathscr{L}$ in $\mathscr{M}$, hence, an interpretation $\mathscr{I}$ is uniquely defined by restriction of an evaluation to objects which are not indeterminates. A formula A of $\mathscr{L}$ is called *valid* or *true in an interpretation* (of $\mathscr{L}$ in $\mathscr{M}$), iff A is valid in all evaluations belonging to $\mathscr{I}$. A formula A of $\mathscr{L}$ is called *invalid* or *false* in an interpretation $\mathscr{I}$ (of $\mathscr{L}$ in $\mathscr{M}$), iff A is valid in no evaluation belonging to $\mathscr{I}$. The set of all formulas valid in $\mathscr{I}$ will be denoted by $\mathscr{V}(\mathscr{I})$.

Corollary 2.1. Let $\mathscr{L}$, $\mathscr{M}$, $\mathscr{I}$ be as in Definition 2.5. Then every closed formula of $\mathscr{L}$ is either true or false in $\mathscr{I}$.

Proof. An immediate consequence of Definition 2·5. When ascribing, in what follows, the adjective "true", "valid", "false", "invalid" also to formulas containing free occurences of indeterminates, we refer them always to the universal closures of the formulas in question.

Notice, that the set $\mathscr{V}(\mathscr{I})$ of valid formulas is neither effectively decidable, nor recursively enumerable, e.g., to verify a formula beginning with a general quantifier requests to examine an infinite number of evaluations. The aim of a formalized theory is to replace, in the best possible way, this undecidable set $\mathscr{V}(\mathscr{I})$ by a set $\mathscr{T}$ of theorems, which is, in general, recursively enumerable.

A model $\mathscr{M}$ is called *a model of a theory* $\langle \mathscr{L}, \mathscr{T} \rangle$ *in the interpretation* $\mathscr{I}$, iff $\mathscr{T} \subset \mathscr{V}(\mathscr{I})$, i.e., iff every theorem is valid in $\mathscr{I}$. If the deduction rules are *truth-preserving*, this means, if they lead from valid premises to valid conclusions, then the necessary and sufficient condition for the theory $\langle \mathscr{L}, \mathscr{T} \rangle = \langle \mathscr{L}, \mathscr{A}x, \mathscr{R} \rangle$ to be valid in (or: to have) a model $\mathscr{M}$ in the interpretation $\mathscr{I}$ is that $\mathscr{A}x \subset \mathscr{V}(\mathscr{I})$, i.e., all axioms are valid formulas. Namely, the deduction rules (R1) and (R2) can be easily proved to be truth-preserving and logical axioms can be proved to be valid in all models of a first-order language $\mathscr{L}$.

A formalized theory is called *consistent*, iff $\mathscr{T} \neq \mathscr{L}$, it is equivalent with the condition that, for each $A \in \mathscr{L}$, $[A \wedge (\neg A)] \in \mathscr{L} - \mathscr{T}$, i.e., no contradiction can be proved, a theory is called *inconsistent* in the opposite case.

Let $\langle \mathscr{L}, \mathscr{T} \rangle = \langle \mathscr{L}, \mathscr{A}x, \mathscr{R} \rangle$ be a consistent formalized theory. This theory is called *syntactically complete*, if for each $x \in \mathscr{L} - \mathscr{T}$ the theory $\langle \mathscr{L}, \mathscr{A}x \cup \{x\}, \mathscr{R} \rangle$ is consistent, in other words, if enriching the set of axioms by any non-theorem enables to prove everything. The theory $\langle \mathscr{L}, \mathscr{T} \rangle$ is called *semantically complete*, if each formula, which is valid in all models in which all axioms are valid, is also provable, i.e., belongs to $\mathscr{T}$ (hence, in such a case the formalization of the theory

in the form $\langle \mathscr{L}, \mathscr{T} \rangle = \langle \mathscr{L}, \mathscr{A}x, \mathscr{R} \rangle$ describes and expresses precisely the set of valid formulas). As an example of syntactically complete theories we can mention the propositional calculus, in fact, joining any non-deducible formula of the propositional language to (A1)–(A9) as a new axiom would make the theory inconsistent. As an example of a semantically complete theory we can introduce the *pure first-order predicate calculus* (this theory is based on a first-order language with relational of functional constants, its axioms are just the logical ones (A1)–(A12), deduction rules are (R1), (R2)). Because of the importance of this fact in what follows, namely in the next chapter, let us formalize it in the form of the following theorem.

Theorem 2.1. A formula of the pure first-order predicate calculus is valid in all interpretations of this theory iff it is a theorem of this theory (Gödel's Completeness Theorem for the first-order predicate calculus).

Proof can be found, e.g., in [1] as well as in many other textbooks on mathematical logic some of them being referred below.

This theorem enables to convert the syntactic problems (deducibility testing) into semantical ones (validity checking) and vice versa and this transformation plays an important role in resolution-based theorem-proving. As Theorem 2.1 holds, in this strict sense, only for the pure first-order predicate theory we need a tool to transform the deducibility problem for a general first-order theory (with extra-logical symbols) into the deducibility problem for the pure first-order theory. Such a possibility is offered by the following theorem, often called Deduction Theorem, we present here a formulation adequate for the first-order theories.

Theorem 2.2. Let $\langle \mathscr{L}, \mathscr{A}x, \mathscr{R} \rangle$ be a first-order formalized theory, let $\mathrm{A}x_0 \subset \mathscr{A}x$ be the logical axioms, let $\mathscr{R} = \{(\mathrm{R1}), (\mathrm{R2})\}$. Write $\mathscr{A}_x \vdash A$, if a formula $A \in \mathscr{L}$ is deducible from axioms by the deduction rules. Let $A_1, A_2, \ldots, A_n$ be formulas from $\mathscr{L}$ then

$$\mathscr{A}x_0 \cup \{A_1, A_2, \ldots, A_n\} \vdash A \quad \text{yields} \quad \mathscr{A}x_0 \cup \{A_1, \ldots, A_{n-1}\} \vdash A_n \to A\,.$$

Proof can be found, e.g., in [1] as well as in many other textbooks on mathematical logic, some of them being referred below.

Hence, having a first-order predicate theory with a finite number of extra-logical axioms $A_1, A_2, \ldots, A_n$, we may replace the theoremhood testing of a formula A by the similar problem for the formula

$$A_1 \to (A_2 \to \ldots \to (A_n \to A)\ldots)$$

in the pure first-order predicate calculus (extra-logical axioms are supposed to be closed formulas). This way of reasoning will be, again, of great importance in what follows.

All the notions which we have formulated here for the case of first-order theories

can be generalized also for higher-order theories as well as for the simple type theory in whole. On the other hand, the completeness theorem as formulated above (Theorem 2.1) expresses a specific feature of the first-order predicate calculus and has no counterpart in higher-order theories. The famous Gödel's Incompleteness Theorem sounds that any consistent formalized theory, rich enough to enable to formalize the arithmetic, is necessarily semantically incomplete and cannot be completed by enriching the set of axioms (as far as this set of axioms remains recursive, as the metatheory of formalized theories requests). However, the higher-order analogies of the notions introduced above for the first-order case will not be necessary in what follows.

The fact that we close this chapter, dealing with some preliminaries from the domain of mathematical logic, just now in no case means that we have already exhausted all the notions and statements of this branch which will be necessary or useful in the rest of this work. Many notions and assertions will be introduced or mentioned at appropriate places, as a rule, immediately before their application. Here we have concerned only the most basic notions of mathematical logic which can be and need to be explained in a systematic way in order to illuminate and emphasize their internal connections and dependences.

As far as the references are concerned, we introduce below several well-known textbook on mathematical logic in which all the notions and assertions mentioned above can be found together with proofs and with more detailed explanations. We have concentrated our attention to textbooks available, sometimes in translations, in our country and to textbooks which have already proved their pedagogical qualities.

REFERENCES

[1] A. Church: Introduction to Mathematical Logic I. Princeton University Press, Princeton, New Jersey 1956. (Russian translation: IIL, Moscow 1960).

[2] H. B. Curry: Foundations of Mathematical Logic. McGraw-Hill Book Comp., New York—San Francisco—Toronto—London 1963. (Russian translation: Mir, Moscow 1969.)

[3] A. Grzegorczyk: Zarys logiki matematycznej. Second edition, PWN, Warszawa 1969.

[4] S. C. Kleene: Mathematical Logic. John Wiley and Sons, New York—London—Sydney 1967. (Russian translation: Mir, Moscow 1973.)

[5] S. C. Kleene: Introduction to Metamathematics. D. van Nostrand Comp., New York—Toronto 1952. (Russian translation: IIL, Moscow 1957.)

[6] A. Mostowski: Logika matematyczna. Warszawa, Wroclaw 1948.

[7] H. Rasiowa, R. Sikorski: The Mathematics of Metamathematics. PWN, Warszawa 1970.

[8] L. Rieger: Algebraic Methods of Mathematical Logic. Academia, Prague 1967.

[9] J. R. Shoenfield: Mathematical Logic. Addison-Wesley Publ. Comp., 1967. (Russian translation: Nauka, Moscow 1975).

[10] A. Tarski: Introduction to Logic and to the Methodology of Deductive Sciences. Oxford Univ. Press, 1965. (Czech translation: Academia, Prague 1966).

[11] Hao Wang: A Survey of Symbolic Logic. North-Holland Publ. Comp., Amsterdam 1962.

[12] A. N. Whitehead, B. Russell: Principia Mathematica, vol. 1—3. Cambridge Univ. Press, Cambridge 1925—27.

3. RESOLUTION-BASED THEOREM-PROVING

The aim of this work is to survey various statistical approaches and methods of proof theory and theorem-proving. As will be shown later, most of these methods contain a "usual", i.e., deterministic theorem-proving algorithm as an integral part and the qualities of this part influence very strongly the quality of the resulting statistical procedure. Moreover, an appropriate deterministic theorem-proving algorithm should play also the role of a "pre-selector", submitting to the subsequent statistical theorem-proving procedure only those formulas which are not decidable by the algorithm in question. It is why we have decided to devote this chapter to a brief explanation of the basic principles of automated theorem-proving, at the same time we have concentrated our attention to the so called resolution-based methods, as most of the modern theorem-proving procedures belong to this group.

Finding a general decision procedure to verify the validity of a formula was considered long ago. It was first tried by Leibniz and further revived by Peano around the turn of the century and by Hilbert's school in the 1920s. It was not until 1936 that this was proved impossible. Church [1] and Turing [4] independently showed that there was no general decision procedure to check the validity of formulas of first-order predicate theories. However, there are proof procedures which can verify that a formula is valid if indeed it is valid. For invalid formulas these procedures, in general, will never terminate. In view of the result of Church and Turing, this is the best we can expect to get from a deterministic proof procedure.

In 1930, Herbrand developed an algorithm to find an interpretation that can falsify a given formula, if it is not valid. Herbrand's method is the basis for all resolution-based proof procedures. Gilmore in [3] was one of the first persons to implement Herbrand's procedure on a computer. Since a formula is valid iff its negation is invalid (inconsistent), his program was designed to detect the invalidity of the negation of the given formula. If this negation is invalid, his program will eventually detect this fact. Gilmore's method was improved by Davis and Putnam in [2], however, their improvement was still not enough to overcome the inefficiency of the original procedure. Many valid and rather simple formulas of first-order predicate theories still could not be proved by computers in a reasonable amount of time.

A major breakthrough was made by Robinson in 1965, cf. [1.6], who introduced the resolution principle. Since the introduction of this principle several refinements have been suggested in attempts further to increase its efficiency. Some of them will, be, in a very brief way, mentioned below.

Since now, the object of our interest in this chapter will be the pure first-order predicate theory, also called first-order logic. Only the formulas which are in the so called prenex normal form with matrices in the so called conjunctive normal form will be tested for theoremhood, as will be immediately shown, there is no loss of generality in this assumption.

Definition 3.1.

(a) A *literal* is an atomic formula of the negation of an atomic formula.

(b) A finite disjuction of literals is called a *clause*.

(c) A formula F which contains no occurrences of a quantifier is said to be in the *conjunctive normal form* iff F has the form $F_1 \wedge F_2 \wedge \ldots \wedge F_n$, $n \geqq 1$, where each F_i, $i \leqq n$, is a clause.

(d) A formula F is said to be in a *prenex normal form* iff F has the form $(Q_1x_1) . (Q_2x_2) \ldots (Q_nx_n)(M)$, where every (Q_ix_i) is either $(\forall x_i)$ or $(\exists x_i)$, and M is a formula containing no quantifiers. $(Q_1x_1) \ldots (Q_qx_n)$ is called the *prefix* and M is called the *matrix* of the formula F.

Theorem 3.1.

(a) Let F be a formula of the first-order logic, let there be no occurrences of a quantifier in F. Then there is a formula F_1 in the conjunctive normal form such that $F \equiv F_1$ is provable in the first-order logic.

(b) Let F be a formula of the first-order logic, then there is a formula F_2 in the prenex normal form such that $F \equiv F_2$ is provable in the first-order logic.

(c) Let F be a formula of the first-order logic, then there is a formula F_3 in the prenex normal form and with matrix in the conjunctive normal form such that $F \equiv F_3$ is provable in the first-order logic.

Proof. The proofs of the assertions (a) and (b) as well as algorithms for finding the appropriate formulas F_1, F_2 can be found in [1.1] and in some textbooks on mathematical logic among them referred in Chapter 2. The assertion (c) immediately follows from the previous ones and justifies our idea to limit ourselves, in what follows, by formulas of this type.

Having already transformed a formula F in a prenex normal form $(Q_1x_1)(Q_2x_2) \ldots \ldots (Q_nx_n)(M)$, where M is in a conjunctive normal form, the existential quantifiers in the prefix can be eliminated by using Skolem functions, without affecting the inconsistency property.

Suppose Q_r, $1 \leqq r \leqq n$, is an existential quantifier in the prefix $(Q_1x_1) \ldots (Q_nx_n)$. If no universal quantifier appears before Q_r, we choose a new constant c, different from other constants occuring in M, replace all x_r appearing in M by c, and delete (Q_rx_r) from the prefix. If $Q_{s(1)}, \ldots, Q_{s(m)}$ are all the universal quantifiers appearing before Q_r, $1 \leqq s(1) < s(2) < \ldots < s(m) < r$, we choose a new m-ary function symbol f different from other function symbols, replace all x_r in M by $f(x_{s(1)}, x_{s(2)}, \ldots \ldots, x_{s(m)})$, and delete (Q_rx_r) from the prefix. After the above process is applied to all the existential quantifiers in the prefix, the last formula we obtain is a (*Skolem*) *standard form* of the formula F. The constants and functions used to replace the existential indeterminates are called *Skolem functions*.

Example 3.1. Obtain a standard form of the formula

$$(\exists x)(\forall y)(\forall z)(\exists u)(\forall v)(\exists w)\, P(x, y, z, u, v, w).$$

Here $(\exists x)$ is preceded by no universal quantifiers, $(\exists u)$ is preceded by $(\forall y)$ and $(\forall z)$, and $(\exists w)$ by $(\forall y)$, $(\forall z)$ and $(\forall v)$. Therefore, we replace x by a new individual constant c, u by a binary function symbol $f(y, z)$, and w by a ternary function symbol $g(y, z, v)$. Thus, we obtain the following standard form of the formula

$$(\forall y)(\forall z)(\forall v)\, P(a, y, z, f(y, z), v, g(y, z, v)).$$

Example 3.2. Obtain a standard form of the formula

$$(\forall x)(\exists y)(\exists z)((\neg P(x, y) \wedge Q(x, z)) \vee R(x, y, z)).$$

First, the matrix is transformed into a conjunctive normal form

$$(\forall x)(\exists y)(\exists z)((\neg P(x, y) \vee R(x, y, z)) \wedge (Q(x, z) \vee R(x, y, z))).$$

Then, since $(\exists y)$ and $(\exists z)$ are both preceded by $(\forall x)$, the indeterminates y and z are replaced, respectively, by unary function symbols $f(x)$ and $g(x)$. Thus, we obtain the following standard form of the formula.

$$(\forall x)((\neg P(x, f(x)) \vee R(x, f(x), g(x))) \wedge (Q(x, g(x)) \vee R(x, f(x), g(x)))).$$

When it is convenient, we shall regard a set of literals as synonymous with a clause. A clause consisting of r literals is called an *r-literal clause*. A one-literal clause is called a *unit clause*. When a clause contains no literal, we call it the *empty clause* and denote by $\square$. Since the empty clause has no literals that can be satisfied by an interpretation, the empty clause is always false. A set S of clauses is regarded as a conjunction of all clauses in S, where every indeterminate in S is considered to be governed by a universal quantifier. By this convention, a standard form can be simply represented by a set of clauses. E.g., the standard form of Example 3.2. can be represented by the set of clauses

$$\{\{\neg P(x, f(x)), R(x, f(x), g(x))\}, \{Q(x, g(x)), R(x, f(x), g(x))\}\}.$$

Theorem 3.2. Let S be a set of clauses that represents a standard form of a formula F. Then F is invalid (inconsistent), iff S is invalid (inconsistent).

Proof. Cf. [1.1], p. 48, Theorem 4.1 and its proof.

Before examining how to solve the problem of invalidity (inconsistency, i.e., unsatisfiability) of a set S of clauses let us resume all the process of transformation of the original problem into the current one. First of all, we have a first-order formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$ with extralogical axioms $A_1, A_2, \ldots, A_n$, we have a formula $A \in \mathscr{L}$ and we ask, whether $A \in \mathscr{T}$, i.e., whether $A_1, A_2, \ldots, A_n \vdash A$ or not. Using

deduction theorem we replace this problem by that whether $A_1 \to (A_2 \to \ldots$ $\ldots (A_n \to A) \ldots)$, hence, $(A_1 \wedge A_2 \wedge \ldots \wedge A_n) \to A$, is provable in the pure first-order predicate calculus. Completeness theorem converts this question to the question, whether $(A_1 \wedge \ldots \wedge A_n) \to A$ is valid, in other formulation, whether $\neg((A_1 \wedge \ldots \wedge A_n) \to A)$ and so also $A_1 \wedge \ldots \wedge A_n \wedge \neg A$, is invalid. Denoting by $S = S(A)$ the set of clauses representing the formula $A_1 \wedge \ldots \wedge A_n \wedge \neg A$, we obtain, due to Theorem 3.2., the problem whether S is unsatisfiable (hence, $A \in \mathscr{T}$) or not (hence, $A \in \mathscr{L} - \mathscr{T}$). It is why we restrict ourselves, since now, to the problem of unsatisfiability of sets of clauses.

By definition, a set S of clauses is unsatisfiable iff it is false under all interpretations over all domains. Since it is inconvenient and impossible to consider all interpretations over all domains, it would be nice if we could fix on one special domain H such that S is unsatisfiable iff S is false under all the interpretations over this domain. Fortunately, there does exist such a domain, which is called the *Herbrand universe* of S and defined as follows.

Definition 3.2. Let S be a set of clauses, let H_0 be the set of individual constants appearing in S. If no constant, appears in S, then H_0 is to consist of a single constant, say $H_0 = \{a\}$. For $i = 0, 1, 2, \ldots$ let H_{i+1} be the union of H_i and the set of all terms of the form $f(t_1, \ldots, t_n)$ for all n-ary function symbols occuring in S, where $t_1, t_2, \ldots, t_n$ belong to H_i. Then each H_i is called the *i-level constant set of* S, and $H = \bigcup_{i=0}^{\infty} H_i$ is called the *Herbrand universe of* S.

Example 3.3. Let $S = \{\{P(f(x), a, g(y), b)\}\}$. Then $H_0 = \{a, b\}$, $H_1 = \{a, b, f(a), f(b), g(a), g(b)\}$, $H_2 = \{a, b, f(a), f(b), g(a), g(b), f(f(a)), f(f(b)), f(g(a)), f(g(b)), g(f(a)), g(f(b)), g(g(a)), g(g(b))\}$, $H_3 = \ldots$.

In the sequel, by *expression* we mean a term, a set of terms, an atom, a set of atoms, a literal, a clause or a set of clauses. When no indeterminate appears in an expression, we call this expression a *groud expression* to emphasize this fact. A *subexpression* of an expression E is an expression that occurs in E.

Definition 3.3. Let S be a set of clauses. The set of ground atoms of the form $P(t_1, \ldots, t_n)$ for all n-ary predicate symbols P occurring in S, where $t_1, \ldots, t_n$ are elements of the Herbrand universe of S, is called the *atom set*, or the *Herbrand base* of S. A ground instance of a clause C of a set S of clauses is a clause obtained by replacing indeterminates in C by members of the Herbrand universe of S.

Definition 3.4. Let S be a set of clauses, $H = H(S)$ the Herbrand universe of S, and $\mathscr{I}$ an interpretation of S over H. $\mathscr{I}$ is said to be an *H-interpretation of* S if it satisfies the following conditions

(a) $\mathscr{I}$ maps all constants in S to themselves.

(b) Let f be an n-ary function symbol and $h_1, h_2, \ldots, h_n$ be elements of H. $\mathcal{I}$ assigns to f a function that maps $\langle h_1, \ldots, h_n \rangle \in H^n$ to $f(h_1, \ldots, h_n) \in H$.

Example 3.4. Consider the set $S = \{\{P(x), Q(x)\}, \{R(f(y))\}\}$. $H = H(S) = \{a, f(a), f((a)), f(f(a))), \ldots\}$. There are three predicate symbols: P, Q, and R. Hence, the atom set of S is $A = \{P(a), Q(a), R(a), P(f(a)), Q(f(a)), R(f(a)), \ldots\}$. Some H-interpretations for S are as follows:

$$\mathcal{I}_1 = \{P(a), Q(a), R(a), P(f(a)), Q(f(a)), R(f(a)), \ldots\},$$

$$\mathcal{I}_2 = \{\neg P(a), \neg Q(a), \neg R(a), \neg P(f(a)), \neg Q(f(a)), \neg R(f(a)), \ldots\},$$

$$\mathcal{I}_3 = \{P(a), Q(a), \neg R(a), P(f(a)), Q(f(a)), \neg R(f(a)), \ldots\}.$$

Theorem 3.3. A set S of clauses is unsatisfiable iff S is false under all the H-interpretations of S.

Proof. Cf. [1.1], p. 55, Theorem 4.2 and its proof.

An important consequence of this theorem sounds as follows:

Theorem 3.4. (A version of the so called Herbrand's Theorem) A set S of clauses is unsatisfiable iff there is a finite unsatisfiable set S' of ground instances of clauses of S.

Proof. Cf. [1.1], p. 61, Theorem 4.4 and its proof.

Finding a proof for a set of clauses can be described in the form of a semantic tree. The nodes correspond to ground instances of literals of the set of clauses and the two vertices leaving each node correspond to the two possible H-interpretations of this literal. Each branch of the tree is developed until an inconsistency occurs, in such a case this branch is closed and another, not yet closed branch is followed. Having closed all the branches, we come to the conclusion that the set of clauses is unsatisfiable and we may read, from the nodes, the finite unsatisfiable set S' of ground instances the existence of which is asserted in the Herbrand's Theorem above. Cf. [1.1] for details.

The first computer implementation of this procedure was proposed and executed by Gilmore in 1960 [3], as mentioned above. Davis and Putnam introduced (cf. [1.1]) four additional rules enabling to restrict the quantity of finite sets of ground instances of clauses of S which should be tested for unsatisfiability (cf. [1.1], p. 63). However, even under these conditions the method remained to be very inefficient; already for rather simple theorems the extent of the set of all ground instances was too high to be storaged in a computer, not to mention test its unsatisfiability.

As already mentioned, an important improvement in this direction was realized by J. A. Robinson's resolution principle [1.6]. It is based on the following idea: if there were, in a set S of clauses, the empty clause $\square$, then $S' = \{\square\}$ is, clearly,

a finite unsatisfiable set of ground instances, requested by the Herbrand's Theorem in order to prove the unsatisfiability of S. Hence, in such a case the testing of S could be reduced to the looking for $\square$ in S. Suppose to have at your disposal a procedure enabling to enrich the set S by new clauses without affecting the original problem of satisfiability or unsatisfiability of S. So to obtain, by this procedure, the empty clause $\square$ proves the unsatisfiability of S. In fact, resolution principle is nothing else than such a rule of looking for the empty clause. First, let us to explain this principle on the propositional level.

Consider two clauses, $C_1 = \{P\}$, $C_2 = \{\neg P, Q\}$ (i.e., $C_2 = \{\neg P \vee Q\}$). Suppose that there is an interpretation $\mathscr{I}$ such that C_1 and C_2 are valid in $\mathscr{I}$. Hence, P is valid in $\mathscr{I}$, $\neg P$ is not valid in $\mathscr{I}$, so Q must be valid in $\mathscr{I}$ in order to assure the validity of $\neg P \vee Q$ in $\mathscr{I}$. This means that we may add a new clause, $C_3 = \{Q\}$, to C_1, C_2 without affecting the original problem of unsatisfiability of $\{C_1, C_2\}$. Extending the above rule and applying it to any pair of clauses (not necessarily unit ones) we have the following rule, which is called the *resolution principle.*

Definition 3.5. For any two clauses C_1 and C_2, if there is a literal L_1 in C_1 that is complementary to a literal L_2 in C_2 (i.e., either L_1 is the negation of L_2 or vice versa), then delete L_1 from C_1 and L_2 from C_2, and construct the disjunction of the remaining clauses. The constructed clause is called a *resolvent* of C_1 and C_2.

Example 3.5.

(a) Consider clauses $\{P, Q\}$ and $\{\neg P, R\}$, their resolvent is $\{R, Q\}$.

(b) Consider clauses $\{\neg P, Q, R\}$ and $\{\neg Q, S\}$, their resolvent is $\{\neg P, R, S\}$.

Theorem 3.5. Given two clauses, C_1 and C_2, a resolvent C of C_1 and C_2 is a logical consequence of C_1 and C_2.

Proof. Cf. [1.1], p. 72, Theorem 5.1 and its proof.

Let us extend the resolution principle to the first-order logic. The most important part of applying this principle is to find a literal in a clause that is complementary to a literal in another clause. For clauses containing no indeterminates, this is very simple. However, for clauses containing indeterminates, the problem is more complicated. For example, consider the clauses $C_1 : \{P(x), Q(x)\}$, $C_2 : \{\neg P(f(x)), R(x)\}$. There is no literal in C_1 that is complementary to a literal in C_2. However, if we substitute $f(a)$ for x in C_1 and a for x in C_2, we obtain $C_1' : \{P(f(a)), Q(f(a))\}$, $C_2' : \{\neg P(f(a)), R(a)\}$, Now, we can obtain a resolvent $C_3' : \{Q(f(a)), R(a)\}$. This new clause can be added to C_1 and C_2 without affecting the consistency problem for $\{C_1, C_2\}$. Or, remember that indeterminates in clauses are supposed to be bound by universal quantifiers, i.e., if $\{C_1, C_2\}$ is a consistent set, it must be consistent also for all possible instances of C_1 and C_2. This means that finding an inconsistent pair

of instances proves the inconsistency of the original set. Immediately follows, that we are allowed, looking for the empty clause, to add to the set of clauses also those obtained by resolution from possible instances of other clauses. As can be easily seen, resolution principle for first-order predicate theories is nothing else than a combination of the two deduction rules (R1) and (R2).

Definition 3.6. A *substitution* is a finite set of the form $\{t_1/v_1, t_2/v_2, \ldots, t_n/v_n\}$, where every v_i is an indeterminate, every t_i is a term different from v_i, and no two elements in the set have the same indeterminate after the stroke symbol. When $t_1, \ldots, t_n$ are ground terms, the substitution is called *ground substitution*. The *empty substitution* ε consists of no elements. Let $\theta = \{t_1/v_1, \ldots, t_n/v_n\}$ be a substitution and E be an expression. Then $E\theta$ is an expression obtained from E by replacing simultaneously each occurrence of the indeterminate v_i in E by the term t_i. $E\theta$ is called an *instance* of E. Let $\theta = \{t_1/x_1, \ldots, t_n/x_n\}$ and $\lambda = \{u_1/y_1, \ldots, u_m/y_m\}$ be two substitutions. Then their *composition* $\theta \circ \lambda$ is the substitution that is obtained from the set $\{t_1\lambda/x_1, \ldots, t_n\lambda/x_n, u_1/y_1, \ldots, u_m/y_m\}$ by deleting any element $t_j\lambda/x_j$ for which $t_j\lambda = x_j$, and any element u_i/y_i such that y_i is among $\{x_1, x_2, \ldots, x_n\}$. A substitution θ is called a *unifier* for a set $\{E_1, \ldots, E_n\}$ of expressions iff $E_1\theta = E_2\theta = \ldots = E_n\theta$. The set $\{E_1, \ldots, E_n\}$ is said to be *unifiable* iff there is a unifier for it. A unifier σ for a set of expressions is a *most general unifier* iff for each unifier θ for this set there is a substitution λ such that $\theta = \sigma \circ \lambda$.

There exists a unification algorithm for finding a most general unifier for a finite unifiable set of nonempty expressions. When the set is not unifiable, the algorithm will also detect this fact. Let us sketch briefly such an algorithm, again borrowed from [1.1].

Definition 3.7. The *disagreement set* of a nonempty set W of expressions is obtained by locating the first symbol (counting from the left) at which not all the expressions in W have exactly the same symbol, and then extracting from each expression in W the subexpression that begins with the symbol occupying that position. The set of these respective subexpressions is the disagreement set of W.

Unification Algorithm.

Step 1. Set $k = 0$, $W_k = W$, $\sigma_k = \varepsilon$.

Step 2. If W_k is a singleton, stop; σ_k is a most general unifier for W. Otherwise, find the disagreement set D_k of W_k.

Step 3. If there exist elements v_k and t_k in D_k such that v_k is an indeterminate that does not occur in t_k, go to *Step* 4. Otherwise, stop; W is not unifiable.

Step 4. Let $\sigma_{k+1} = \sigma_k \circ \{t_k/v_k\}$ and $W_{k+1} = W_k\{t_k/v_k\}$, hence, $W_{k+1} = W\sigma_{k+1}$.

Step 5. Set $k = k + 1$ and go to *Step* 2.

Theorem 3.6. (Unification Theorem) If W is a finite nonempty unifiable set of expressions, then the unification algorithm will always terminate at *Step* 2, and the last σ_k is a most general unifier for W.

Proof. Cf. Theorem 5.2 in [1.1], and its proof.

Having introduced the unification algorithm, we can now consider the resolution principle for the first-order logic.

Definition 3.8. If two or more literals (with the same sign) of a clause C have a most general unifier σ, then $C\sigma$ is called a *factor* of C. If $C\sigma$ is a unit clause, it is called a *unit factor* of C. Let C_1 and C_2 be two clauses (called *parent clauses*) with no indeterminates in common. Let L_1 and L_2 be two literals in C_1 and C_2, respectively. If L_1 and $\neg L_2$ have a most general unifier σ, then the clause $\{C_1\sigma - L_1\sigma\} \cup \{C_2\sigma - L_2\sigma\}$ is called a *binary resolvent* of C_1 and C_2. The literals L_1 and L_2 are called the *literals resolved upon*. *A resolvent* of (parent) clauses C_1 and C_2 is one of the following binary resolvents.

a binary resolvent of C_1 and C_2,

a binary resolvent of C_1 and a factor of C_2,

a binary resolvent of a factor of C_1 and C_2,

a binary resolvent of a factor of C_1 and a factor of C_2.

Example 3.6. Let $C_1 = \{P(x), P(f(y)), R(g(y))\}$, $C_2 = \{\neg P(f(g(a))), Q(b)\}$. A factor of C_1 is $C_1' = \{P(f(y)), R(g(y))\}$. A binary resolvent of C_1' and C_2 is $\{R(g(g(a))), Q(b)\}$. Therefore, $\{R(g(g(a))), Q(b)\}$ is a resolvent of C_1 and C_2.

The *resolution principle*, or *resolution* for short, is an inference rule that generates resolvents from a set of clauses. It is more efficient than the earlier procedures as those by Gilmore and Davis and Putnam mentioned above. Furthermore, resolution is complete in the sense that it will always generate the empty clause $\square$ from an unsatisfiable set of clauses, as the next theorem shows.

Theorem 3.7. (Completeness of the Resolution Principle) A set S of clauses is unsatisfiable iff there is a deduction of the empty clause $\square$ from S.

Proof. C.f. Theorem 5.3, in [1.1], and its proof.

Example 3.7. Consider the following set of formulas:

$F_1 : (\forall x)(C(x) \rightarrow (W(x) \wedge R(x)))$,

$F_2 : (\exists x)(C(x) \wedge O(x))$,

$F_3 : (\exists x)(O(x) \wedge R(x))$.

Our problem is to show that F_3 is a logical consequence of F_1 and F_2. We transform F_1, F_2, and $\neg F_3$ into standard form and obtain the following five clauses.

(1) $\{\neg C(x), W(x)\}$,

(2) $\{\neg C(x), R(x)\}$, (1) and (2) from F_1,

(3) $\{C(a)\}$,

(4) $\{O(a)\}$, (3) and (4) from F_2,

(5) $\{\neg O(x), \neg R(x)\}$, from $\neg F_3$.

Let us add some new clauses derived by the resolution principle.

(6) $\{R(a)\}$, a resolvent of (3) and (2),

(7) $\{\neg R(a)\}$, a resolvent of (5) and (4),

(8) $\square$, a resolvent of (7) and (6).

Hence, the original set of clauses is unsatisfiable, therefore, F_3 is a logical consequence of F_1 and F_2.

In spite of its great effectivity, an uncontrolled application of resolution principle leads to a "population explosion" of possible resolvents which must be, at least potentially, taken into consideration. In examples, usually introduced in papers or books in order to illustrate the resolution principle, this danger is avoided by a sophisticated and goal-oriented choosing of appropriate parent clauses to be resolved. Some improvements have been suggested, how to minimize the number of resolvents to be generated under the condition that the completeness of the modified method (in the sense of Theorem 3.7) is preserved.

So, e.g., *deletion strategy* offers some rules enabling to delete, without any loss of generality, some clauses from the set of potential parent clauses for further resolution. The so called *set of support strategy* divides the clauses into two groups in such a way that only those resolvents, whose parent clauses belong to different groups need to be generated. *Linear resolution* implements an ordering into the set of clauses with the aim to minimize the number of necessary resolutions when the resolved clause are chosen according to this ordering. As other examples can serve *hyper-resolution*, *Lock resolution*, etc. Because of the limited extent of this chapter we refer to [1.1] for more details on various improvements of the resolution principle. As a rule, for each modification there exist some inputs, for which it runs very well and effectively, however, there are also input data for which the same modification runs very ineffectively. Some quantitative estimates of the quality of various improvements from the point of view of computational complexity exist only in very few special cases.

As a rule, let us close this chapter by mentioning some references. There is a great number of items dealing with resolution-based theorem-proving, however, most

of them are special papers devoted to particular problems and requiring rather large preliminary knowledge. Among the monographies of surveyal character we can sincerely recommend [1.1], which we used throughout all this chapter and from which we have borrowed all the assertions, most of the definitions and many other formulations. This book is written in a very clear and concise way, the necessary preliminaries are reduced as possible and all the methods and assertions are illustrated by many examples of various difficulty. In [1.1] also a large list of further references can be found for those wishing to study the resolution-based theorem-proving in more details.

REFERENCES

[1] A. Church: An Unsolvable Problem of Number Theory. Amer. J. Math. *58* (1936), 345—363.

[2] M. Davis, H. Putnam: A Computing Procedure for Quantification Theory. Journal of the Assoc. for Comp. Machinery *7* (1960), 3, 201—215.

[3] P. C. Gilmore: A Proof Method for Quantification Theory; its Justification and Realization. IBM Journal of Research and Development *1* (1960), 28—35.

[4] A. M. Turing: On Computable Numbers with an Application to the "Entscheidungsproblem". Proc. of the London Math. Soc. *42* (1963), 230—265.

4. A GENERAL MODEL OF STATISTICAL THEOREM PROVING

Logic and statistics ... From the first sight this connection seems to be, if not paradoxical, then at least rather strange and courageous. Since more than two thousand years, since Aristotelian times, logic is considered to be the most perfect and most genial image of precisity and correctness, which cannot be reached by other sciences. And statistics? The layman's opinion in which statistics are "the precise sums of items which are far from being precise", or "using statistics everything can be proved" are, as we believe, too spread to be negligible.

In this chapter we follow the two main goals. First, to introduce the basic ideas of statistical decision making, on which all particular methods explained below are based. Second, to show that our resignation from the logical precisity and our introducing to logic an "uncertainty" or "doubts" is not a step backward, a resignation in general, a return to a pre-logical state of mathematics. We would like to prove, that connections between statistics and logic represents a new, higher state of development of mathematical logic, enabling to overcome some limitations which are own to the classical logic and which do not allow to apply the methods of mathematical logic and formalized theories in all the possible and sometimes very perspective cases.

Let us consider a formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$; let us recall that $\mathscr{L}$ is the set of all well-formed formulas of a formalized language and $\mathscr{T} \subset \mathscr{L}$ is the set of all theorems. Usually, $\mathscr{T}$ is defined as the smallest set of formulas from $\mathscr{L}$ which contains some

special formulas, called axioms or postulates and which is closed with respect to some deduction rules (also the expression "inference rules" is used).

Now, imagine a mathematician investigating such a theory. His work, has, of course, many various aspects, but most of them can be reduced to a simple thing: the mathematician chooses or is given sentences of the theory $\langle \mathscr{L}, \mathscr{T} \rangle$ and he is to decide, whether a sentence is or is not a theorem, i.e., whether it belongs to $\mathscr{T}$ or to $\mathscr{L} - \mathscr{T}$. Which are the possible answers to this question? From the theoretic point of view there are just two, namely

(1) "the sentence is a theorem",

(2) "the sentence is not a theorem".

In case the investigated theory were decidable, we could limit ourselves to these two answers also from the practical point of view supposing we knew an algorithm how to solve the deducibility problem for the considered theory. However, only the most simple theories are algorithmically decidable in an a priori given number of steps. In other cases some heuristic effort of the mathematician will be necessary to come to the correct decision. And in this case it is possible that our poor mathematician will not be able to solve the problem and, after some effort, he gives up and says:

(3) "I do not know", or "I cannot decide",

or something like this, the verbal form is not important.

The necessity to introduce into our model also this third possible answer will become much more clear if we modify slightly the question posed to the mathematician and instead of "is the tested sentence a theorem of the considered theory or not" we ask him "decide, within a given time interval (five minutes, one month, ten years) whether the tested formula is a theorem or not". Now it is clear, that after finishing this time period one of the possible answers may be "I am sorry, but it is beyond my powers to decide". Moreover, we can easily see that if the problem is formulated in this modified way we are not allowed to avoid this third possible answer neither in the case of decidable theories.

In the following we try to describe, which is the approach of classical, it means non-statistical and pure mathematics toward these three possible answers.

The main goal of mathematical effort is to receive the decisive and correct answer to the question about the tested formula, as introduced above. Supposing such an answer is obtained it is considered as some positive contribution to the sum of the mathematician's knowledge (or to the knowledge of all mankind in general). Even the negative answer, i.e., "the tested formula is not a theorem" can be interpreted in such a way, because it is equivalent to the positive answer to the question whether the metasentence "the tested formula is not a theorem of the theory in question" is or is not a metatheorem of a metatheory over the theory $\langle \mathscr{L}, \mathscr{T} \rangle$. So it is understood that some positive profit has been reached. This profit is considered to be independent from the circumstances under which it has been reached, it is also

considered to be time-independent. This means that in the pure and classical mathematics we do not suppose that a sentence may be very useful in one instant but less useful in other instant. Moreover, in this case we do not consider the expenses, the costs, the time necessary to obtain the desired result and the question "is the value of the obtained result great enough to justify the expenses? is not admitted at all.

The possibility that the answer would be a decisive one, i.e., of the type (1) or (2) above, but wrong is taken for the worst evil which can happen to an unhappy mathematician. In this case the profit is considered to be zero, of course, and the loss is taken to be infinitely large. This agrees with the usual practice according to which no mathematician can defend his error by saying that he has proved or that he will prove many good results. The only way to eliminate the eventually loss is to revoke the wrong decision and to replace it by the correct one.

The third possibility, the answer "I cannot decide" is considered as a neutral one, not giving any profit, but also not connected with any loss as the expenses of the investigation, which has been useless and in vain, are not taken into account.

This simple survey immediately shows that in the pure and classical mathematics there is no reason to accept some decision about the tested formulas supposing that these decisions are connected with a possibility of error, no matter how small the probability of this error may be. It is always better to say "I cannot decide" and to expect that sometimes in future the correct decision will be found, than to risk by accepting some not quite sure decision immediately. This is substantially caused by the fact that in pure mathematics the decisions about formulas are supposed to be the final ones in the sense that no further decisions depend on them. However, in what follows, we shall describe some situations when the evaluation of answers given above is not more adequate and justifiable. Hence, also our way of reasoning when searching for a decision must be modified.

Let us consider the situation which is, in general, described and investigated in the so called automaton-environment systems theory. The expression "automaton" is not taken here in its purely theoretical sense, for our purposes it will be sufficient to take for an automaton every system which has some degree of autonomy according to the surrounding it world-environment, which has some possibility to receive pieces of information from the environment as well as some possibilities to influence or to change somehow the state of the environment. E.g., a robot may serve as an example of such an automaton. The more simple automata are given some goal and a series of instructions (a program) enabling to reach this goal. Here goal means a concrete state of the environment or a class of such states connected by some common property or properties and the aim is to change the present state of the environment into the (or into a) goal one. More sophisticated automata are given only the goal and they are able themselves to find a sequence of operations leading to this goal.

Let us concern our attention just to this type of automata. How does such an automaton work? Not penetrating into details we can say:

First, the automaton, being equipped by necessary equipment, receives some information from the enviroment. The automaton observes the environment, measures some physical values (temperature, light, radiation), detects the positions of various objects, etc. Second, the automaton works out the obtained data in such way as to transform in into a sequence of formulas (sentences) of an appropriate formalized language (usually a first order formalized language is used for this purpose). The aim is, of course, that these formulas should describe the situation of the environment "good enough" in the sense that every assertion on the environment (or at least every assertion "important" in a sense) which is valid in the environment, i.e., for which the environment can serve as a model, should be also derivable from the data. The number and complexity of the formulas expressing those input data is limited, of course, by parameters of the automaton equipment, hence, the situation described above is just the ideal one and usually we must be satisfied if the set of obtained formulas can serve only as an approximation of an actually exhaustive axiomatic system describing the state of the environment, but it is not the goal of this chapter to study this problem in more details (cf. [1], [2]).

The third stage consists in transforming the given goal into a formula of the used language in such a way that the goal is reached iff this formula is valid in the environment. The problem how to find an appropriate sequence of operations leading to the desired goal can be, roughly speaking, transformed into the question whether a formula (or formulas) of the used language is (or are) derivable from the axioms describing the present state of the environment. Because of the importance of this application we shall study it in more details in Chapter 8 where also some references will be given.

And now, what happens if the automaton is not able to prove the corresponding formula. The analogy with the case of pure mathematics would be: to stop the automaton, to interrupt his activity and, perhaps, to print out "I cannot decide". However, would this solution be the most reasonable one? To do nothing, it is also an action with which some profit or loss is connected, this profit or loss being optimal if the goal is reached as soon as possible and using operations which are as cheap as possible. From this point of view it might be better to risk and to sample the sequence of operations (or some members of this sequence) at random. If the probability of random sampling of a correct decision is greater than the probability that the passive expecting is a correct decision in the actual situation, then our risk approach is quite justifiable from the statistical point of view. However, according to what we have said the random decision about the operations can be formalized as taking at random a decision concerning the corresponding formula or formulas. Hence, this way of reasoning can serve as a justification of our effort to develop a statistical method of deducibility testing.

As example of another kind we can consider the controlling of random processes. Roughly speaking, random (stochastic) process is any process the developing of which is not deterministic, being influenced by some random aspects. The designer or the

user can obtain a profit which depends either on the value taken by the random process in a fixed future instant or on the whole behaviour of the process in question till this instant. The user has some possibilities to intervene into the process with the aim to maximize his profit. Hence, the user observes and measures various aspects of the random process and in some instants, when it is possible or appropriate, he decides how to intervene. He chooses the optimal decision according to the actual situation, i.e., according to the fact whether some conditions are or are not satisfied in the time instant when a decision is to be taken. As can be easily seen, from the theoretical point of view the decision which intervention is to be applied can be transformed to the decision whether some formula (or formulas) is (are) valid or not. Hence, the problem of optimal decision choosing can be reduced to that of deducibility testing for adequate formulas. And we can see, again, that in case we are not able to decide about some formula it may be better to apply the statistical point of view, to risk and to try the decision at random, than to do nothing and to let the process without any control. And this leads, again, to the problem of statistical deducibility testing.

We have mentioned only two possible applications justifying the investigation of statistical methods in automated theorem-proving, but it is possible to give a number of another ones, because their common feature is very clear and expresses the well-known idea of everyday life: Not knowing exactly what to do in a situation one would better to apply a decision being only with some (great enough) probability the optimal one than to sit down, to give up any activity and to do nothing.

Let us introduce an abstract model of statistical decision making abstract enough to cover all particular deducibility testing procedures which will be explained in the following chapters.

Definition 4.1. Let Ω be a nonempty set. A nonempty system $\mathscr{S}$ of subsets of Ω is called a *σ-field*, if

(1) for all $A \in \mathscr{S}$ also $\Omega - A \in \mathscr{S}$,

(2) if $A_i \in \mathscr{S}$, $i = 1, 2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathscr{S}$.

The pair $\langle \Omega, \mathscr{S} \rangle$ is called a *measurable space*. Let $\langle \Omega_1, \mathscr{S}_1 \rangle$, $\langle \Omega_2, \mathscr{S}_2 \rangle$ be two measurable spaces. A mapping f, defined in Ω_1 and taking values in Ω_2 is called *measurable*, if

$$\{\{\omega : \omega \in \Omega_1, f(\omega) \in A\} : A \in \mathscr{S}_2\} \subset \mathscr{S}_1 .$$

Definition 4.2. Let $\langle \Omega, \mathscr{S} \rangle$ be a measurable space. A real-valued function P defined on $\mathscr{S}$ and taking its values in the interval $\langle 0, 1 \rangle$ or reals is called *probability measure* or simply *probability*, if

(1) $P(\Omega) = 1$,

(2) for each sequence $\{A_1, A_2, \ldots\}$ of mutually disjoint elements from $\mathscr{S}$,

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) .$$

In such a case the triple $\langle \Omega, \mathscr{S}, P \rangle$ is called a *probability space*, elements of Ω are *elementary events* and elements of $\mathscr{S}$ are *random events*. For each $A \in \mathscr{S}$, the real number $P(A)$ is called the *probability of A*.

Definition 4.3. *Statistical decision problem* is the quadruple

$$\Delta = \langle \langle X, \mathscr{X}, \mu \rangle, \langle Y, \mathscr{J}, \{v_x\} \rangle, \langle D, \mathscr{D} \rangle, w \rangle .$$

where $\langle X, \mathscr{X}, \mu \rangle$ is a probability space over the parameter space X, every $\langle Y, v_x \rangle$, $x \in X$, is a probability space over the observation space Y, $\langle D, \mathscr{D} \rangle$ is a measurable space over the space of decision D, and w is the weight or loss function defined on the Cartesian product $X \times D$ and taking non-negative reals as its values, here w is supposed to be a measurable mapping of the measurable space $\langle X \times D, \mathscr{X} \times \mathscr{D} \rangle$ into the Borel line $\langle E_1, \mathscr{B} \rangle$.

The intuition behind this formalization is as follows. Values x of X represent possible states of the environment or Nature and they are not accessible to an immediate observation. In our case of statistical deducibility testing there are just two values of parameter, t and f, the first corresponding to the situation when the tested formula is a theorem, the other corresponds to the case of non-theorem. The *a priori probability* μ on this two-element parameter space expresses our knowledge with which probability theorems (non-theorems, resp.) come to the input of our statistical decision procedure.

The only values which we are able to observe are those of Y. They are connected with the values of parameters in a stochastic way, namely, for each $x \in X$ and each $B \in \mathscr{J}$ the value $v_x(B)$ is the probability that the observed value belongs to B under the condition that the actual parameter value is x. An appropriate choose of the observational space will be rather sophisticated and will be given in the next chapter (e.g. we observe the relative frequency of at random sampled extensions in which the tested formula can be proved, etc.).

Decision space D consists of the possible decisions which can be taken on the ground of observation of the value or values from Y. In our case, the decisions about the deducibility of the tested formula can be taken in the form of (1) and (2) above, hence, from the formal point of view, the decision space can be identified with the parameter space $X = \{\mathrm{t}, \mathrm{f}\}$. It is also possible to include (3) in D, this decision can be interpreted either as a resignation and giving up, or as a decision to make some additive observation enabling to decide with a more certainly. This approach leads to the sequential decision making, as explained in [4] or [5].

Finally, the loss function w expresses the consequences of various decision in various situations, namely $w(x, d)$, $x \in X$, $d \in D$, is the loss suffered when the decision d has been taken and the actual parameter value is x. Very often the most simple case of loss function is used, i.e., $w(x, d) = 0$, if the decision d is "appropriate" or "the best" which respect to x, and $w(x, d) = 1$ otherwise. Also in our case, when $X = D = \{\mathrm{t}, \mathrm{f}\}$ we shall often use this type of loss function, setting $w(\mathrm{t}, \mathrm{t}) = w(\mathrm{f}, \mathrm{f}) = 0$, $w(\mathrm{t}, \mathrm{f}) = w(\mathrm{f}, \mathrm{t}) = 1$.

A solution to a statistical decision problem can be formalized in the form of the so called *decision function*, it is a measurable mapping ascribing to each observation from Y a decision from D. Of course, not every decision function is of the same quality with respect to the statistical decision problem in question. We prefer such decision functions which minimize the loss in a sense, i.e. which minimize either the expected loss or the maximal loss.

A formal definition sounds:

Definition 4.4. Consider the statistical decision problem Δ from Definition 4.3. Decision function δ is a measurable mapping from $\langle Y, \mathscr{Y} \rangle$ into $\langle D, \mathscr{D} \rangle$. The value:

$$r(x, \delta) = \int w(x, \delta(y)) \, \mathrm{d}\nu_x$$

is called the *risk connected with Δ and δ under the condition that the value of parameter is x*. Set

$$r_B(\Delta, \delta) = \int r(x, \delta) \, \mathrm{d}\mu = \iint w(x, \delta(y)) \, \mathrm{d}\nu_x \, \mathrm{d}\mu \,,$$

$$r_M(\Delta, \delta) = \sup \{r(x, \delta) : x \in X\} \,,$$

then $r_B(\Delta, \delta)$ is called the *Bayes risk* and $r_M(\Delta, \delta)$ the *minimax risk* connected with the problem Δ and decision function δ. A decision function δ_0 is called a *Bayes solution* (a *minimax solution*, resp.) *to the statistical decision problem* Δ, if $r_B(\Delta, \delta_0) \leqq r_B(\Delta, \delta)$ ($r_M(\Delta, \delta_0) \leqq r_M(\Delta, \delta)$, resp., for all decision functions δ.

In case of the zero-one loss function the Bayes risk reduces to the probability of error weighted with respect to the apriori distribution and the minimax risk reduces to the maximum of probabilities of error of both types. For the sake of simplicity we do not take into consideration the randomized decision functions, when to each observation from Y a probability measure on the space $\langle D, \mathscr{D} \rangle$ is ascribed; the actual decision is then obtained by a random experiment organized with respect to the ascribed probability on $\langle D, \mathscr{D} \rangle$.

REFERENCES

[1] I. Kramosil: Random Axiomatic Systems. Research Rep., Institute of Information Theory and Automation, Prague 1973.

[2] I. Kramosil: Gentzen-Like Random Axiomatic Systems. Transactions of the 7th Prague Conference on Information Theory, ..., 1974", Academia, Prague 1977, 345—352.
[3] E. L. Lehmann: Testing Statistical Hypotheses. John Wiley and Sons, New York 1959. (Russian translation: Moscow, Mir 1964.)
[4] A. Wald: Sequential Analysis. John Wiley and Sons, New York 1947. (Russian Translation: Moscow, Mir 1960.)
[5] A. Wald: Statistical Decision Functions. John Wiley and Sons, New York 1960.

5. STATISTICAL DEDUCIBILITY TESTING IN RANDOM EXTENSIONS

The first idea coming into mind if one is to propose a statistical test of deducibility is to connect somehow the decision about the tested formula with the result of an appropriate random experiment. Let us start this chapter by an extremely primitive example which, nevertheless, shows some problems of automated deducibility testing from quite another and strange point of view.

Consider the random sample consisting in the tossing of a regular coin together with the following very simple decision rule: if "head" occurs, the tested formula is proclaimed to be a theorem, if "tail" occurs, it is proclaimed to be a non-theorem. Such a decision rule is very trivial, of course, and in no case we pledge for replacing theorem proving by coin tossing, however, it offers at least one great advantage if compared with deterministic deducibility testing procedures. Namely, every formula is given a positive probability (which equals to 0·5) to be decided correctly. No algorithmic decision procedure possesses this property supposing the theory in question is undecidable. For, in such a case, there exists always a nonempty (and, as a matter of fact, infinite) set of formulas for which the decision will not be the right one, hence, these formulas are always decided wrongly not having been given any change to be decided, at least sometimes, correctly.

However, there is a simple reason for which coin tossing cannot be seriously considered as a statistical deducibility testing procedure, namely, the decision about the tested formula is statistically independent from the actual state of world, i.e. from the fact whether the tested formula actually is or is not a theorem. Considering a probability space $\langle \Omega, \mathscr{S}, P \rangle$ and defining on it two random variables c (representing the tossing of a regular coin with results H (head) and T(tail)) and x (representing the random sampling of the tested formula) the mentioned statistical independence can be formally described as follows:

$$P(\{\omega : \omega \in \Omega,\ c(\omega) = H\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}) =$$
$$= P(\{\omega : \omega \in \Omega, c(\omega) = T\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}) =$$
$$= P(\{\omega : \omega \in \Omega, c(\omega) = H\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \mathscr{T}\}) =$$
$$= P(\{\omega : \omega \in \Omega,\ c(\omega) = T\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \mathscr{T}\}) = 1/2\,,$$

where $\langle \mathscr{L}, \mathscr{T} \rangle$ is the formalized theory in question. It is why the coin tossing procedure cannot be considered to be "intelligent" or "sophisticated" enough and it seems quite natural to impose our requests for a statistical deducibility testing procedure by the following two conditions:

$$P(\{\omega : \omega \in \Omega, x(\omega) \quad \textit{is proclaimed to be a theorem}\}/\{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}) <$$

$$< P(\{\omega : \omega \in \Omega, x(\omega) \textit{ is proclaimed to be a theorem}\}/\{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \mathscr{T}\}),$$

$$P(\{\omega : \omega \in \Omega,\ x(\omega) \quad \textit{is proclaimed to be a non-theorem}\} :$$

$$: \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \mathscr{T}\}) >$$

$$> P(\{\omega : \omega \in \Omega,\ x(\omega) \textit{is proclaimed to be a non-theorem}\}/\{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}).$$

This means that the probability of decision about the tested formula is greater under the condition that this decision is right than under the condition that it is wrong.

As far as the author knows, the first who submitted a non-trivial (in the sense just mentioned) statistical test of deducibility was A. Špaček in 1959, cf. [16] and [17]. Let us briefly describe and consider his basic model.

We are not able to decide immediately whether a tested closed formula $x \in \mathscr{L}$ is a theorem or not, i.e. whether $x \in \mathscr{T}$ or $x \in \mathscr{L} - \mathscr{T}$. Suppose, however, that there exists a sequence $\langle A_1, A_2, \ldots \rangle$ of subsets of $\mathscr{L}$ such that $\bigcap_{i=1}^{\infty} A_i = \mathscr{T}$. If we were able to decide for each $i = 1, 2, \ldots$ whether $x \in A_i$ or not, we should also decide the original problem. Of course, this way is not effective (with the exception of the trivial case when $\mathscr{T} = \bigcap_{i=1}^{n_0} A_i$ for an appropriate n_0). If $x \in \mathscr{L} - \mathscr{T}$, we can find, eventually, an index n_0 such that $x \in \mathscr{L} - A_{n_0}$ and we can proclaim x to be a non-theorem without any danger of error, however, such an n_0 is, in general, not effectively computable or at least majorizable a priori. If $x \in \mathscr{T}$, then $x \in A_i$ for all $i = 1, 2, \ldots$, but no matter how large n is, the fact that $x \in \bigcap_{i=1}^{n} A_i$ does not logically imply that $x \in \bigcap_{i=1}^{\infty} A_i = \mathscr{T}$.

Happy enough, in the last case the situation is not so hopeless from the statistical point of view. We feel that the fact that $x \in \bigcap_{i=1}^{n} A_i$ supports somehow our belief that $x \in \mathscr{T}$ and this conviction is the greater, the greater n is. So we can propose such a testing procedure: "Sample at random n elements $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ of the sequence $\{A_i\}_{i=1}^{\infty}$. If $x \in \bigcap_{j=1}^{n} A_{i_j}$, proclaim x to be a theorem, in the other case proclaim x to be a non-theorem".

This is the Špaček's basic idea and, as for as we have seen till now, this idea can be expressed in a purely set-theoretic sense not using any notions or assertions of mathe-

matical logic. To apply this idea to our theory $\langle \mathscr{L}, \mathscr{T} \rangle$ we have to find an appropriate sequence $\{A_i\}_{i=1}^{\infty}$ of sets.

Definition 5.1. A formalized theory $\langle \mathscr{L}', \mathscr{T}' \rangle$ is called an *extension* of a formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$, if $\mathscr{L} = \mathscr{L}'$ and $\mathscr{T}' \supset \mathscr{T}$. The extension is called *proper*, if $\mathscr{T}' \neq \mathscr{T}$.

An extension of the theory $\langle \mathscr{L}, \mathscr{T} \rangle$ can be obtained by joining one or more new formulas to the set of axioms as new axions. If the joined axioms are derivable from the former ones, the obtained extension is not proper, as the set of all theorems has not been changed. If at least one among the new axioms is not derivable from the former ones, the obtained extension is proper.

Trivially,

$$\mathscr{T} = \bigcap \{\mathscr{T}' : \langle \mathscr{L}, \mathscr{T}' \rangle \text{ is } \textit{an extension of } \langle \mathscr{L}, \mathscr{T} \rangle\}.$$

Moreover, Špaček proved:

Theorem. 5.1. The relation

$$\mathscr{T} = \bigcap \{\mathscr{T}' : \langle \mathscr{L}, \mathscr{T}' \rangle \textit{ is a proper extension of } \langle \mathscr{L}, \mathscr{T} \rangle\}$$

holds iff the conjunction of all axioms of the theory $\langle \mathscr{L}, \mathscr{T} \rangle$ is not an atom of the Boolean algebra (so called Lindenbaum-Tarski algebra) over $\langle \mathscr{L}, \mathscr{T} \rangle$.

Proof. Cf. [16], p. 611, more arguments can be found in [2.6] or [15].

Now, let us suppose to have at our disposal a random mechanism enabling to sample at random proper, but consistent extensions of the formalized theory in question. It is equivalent to a random generator of closed formulas from $\mathscr{L}$, which are neither theorems nor negations of theorems of $\langle \mathscr{L}, \mathscr{T} \rangle$. Having sampled an extension we investigate, whether the tested formula is a theorem of this extension or not. If not, we proclaim the tested formula to be a non-theorem (of the original theory $\langle \mathscr{L}, \mathscr{T} \rangle$), this decision is always right with no possibility of error. If it is a theorem of the extension in question, we sample another extension and repeat our investigation. Špaček supposes the proper extension to be decidable. i.e. this step of his procedure to be always realizable. Finally, having sampled an a priori given number of proper consistent extensions and having found that the tested formula is valid in all of them we proclaim this formula to be a theorem of the theory $\langle \mathscr{L}, \mathscr{T} \rangle$.

The main properties of this testing procedure can be expressed as follows.

Theorem 5.2. Let $\langle \mathscr{L}, \mathscr{T} \rangle$ be a consistent formalized theory, let $\langle \Omega, \mathscr{S}, P \rangle$ be a probability space, let $x, a_1, a_2, \ldots$ be random variables defined on $\langle \Omega, \mathscr{S}, P \rangle$, taking their values in the set $\mathscr{L}_0$ of all closed formulas, for $a_1, a_2, \ldots$ mutually independent and equally distributed, and such, that for all closed non-theorems x which are not negations of theorems:

$$P(\{\omega : \omega \in \Omega,\ a_1(\omega) = x\}) > 0\,. \tag{5.1}$$

Denote, for each set A of formulas, by $Cn(A)$ the set of all logical consequences of formulas from A, then

(1) $P(\{\omega : \omega \in \Omega,\ x(\omega) \in \bigcap_{i=1}^{N} Cn(\mathscr{T} \cup \{a_i(\omega)\})\}/\{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \mathscr{T}\}) \to 0$.

if $N \to \infty$.

(Verbally: any non-theorem will be, eventually, with probability 1 proclaimed to be a non-theorem supposing the number of random extensions increases.)

(2) $P(\{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}/\{\omega : \omega \in \Omega,\ x(\omega) \in \bigcap_{i=1}^{N} Cn(\mathscr{T} \cup \{a_i(\omega)\})\}) \to 1$,

if $N \to \infty$.

(Verbally: the probability of error connected with proclaiming a formula to be a theorem on the ground of its validity in all the first N random extensions tends to zero with N increasing.)

(3) $P(\{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \mathscr{T}\}/\{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \bigcap_{i=1}^{N} Cn(\mathscr{T} \cup \{a_i(\omega)\})\}) = 1$

for all $N = 1, 2, \ldots$

(Verbally: if the tested formula is proclaimed to be a non-theorem, then with probability one it actually is a non-theorem.)

Proof. Let x be a non-theorem, then there exists at least one extension of the type described above such that x is not provable in it. Let $\mathscr{F}_x \subset \mathscr{L}$ denote the nonempty set of all formulas, which are neither theorems nor negations of theorems and which posses the property that $x \in \mathscr{L} - Cn(\mathscr{T} \cup \{y\})$ for each $y \in \mathscr{F}_x$. Sampling a formula from $\mathscr{F}_x$ by an a_i implies the refutation of x, i.e. proclaiming x to be a non-theorem. Hence, according to the supposed statistical independence of the corresponding random variables,

$$P(\{\omega : \omega \in \Omega,\ x \in \bigcap_{i=1}^{N} Cn(\mathscr{T} \cup \{a_i(\omega)\})\}) \leqq$$

$$\leqq P(\{\omega : \omega \in \Omega,\ a_i(\omega) \in \mathscr{L} - \mathscr{F}_x,\ i = 1, 2, \ldots, N\}) =$$

$$= \prod_{i=1}^{N} (1 - P(\{\omega : \omega \in \Omega,\ a_i(\omega) \in \mathscr{F}_x\})) \to 0 \quad \text{for} \quad N \to \infty,$$

as

$$P(\{\omega : \omega \in \Omega,\ a_1(\omega) \in \mathscr{F}_x\}) = \sum_{y \in \mathscr{F}_x} P(\{\omega : \omega \in \Omega,\ a_1(\omega) = y\}) > 0,$$

according to (5.1) and the fact that $\mathscr{F}_x$ is nonempty. This reasoning immediately implies (1) and (2) of Theorem 5.2, assertion (3) is obvious. Q.E.D.

The original Špaček's proof is more complicated than our and it is based on the so called Neymann-Pearson theorem, well-known to everybody familiar with the foundations of statistical hypothesis testing theory. Besides the presented results concerning the limit values of probabilities of errors Neymann-Pearson theorem proves also the suggested test to be the optimal (from the point of view of minimization of probabilities of error) among all other tests (decision functions) based on the results of deducibility testing of the tested formula in a finite number of random extensions. However, this fact can be seen almost obviously, as proclaiming the tested formula to be a theorem also in case it is not valid in some random extension clearly makes the probability of error greater. The only alternative decision rule which can eventually compete that explained above is the trivial rule proclaiming any formula to be a non-theorem without a further testing. If the probability of sampling a theorem to be tested is small enough and if the number N of investigated random extensions is fixed, the trivial test may appear to have a smaller probability of error than that based on those random extensions. On the other hand, if the a priori probability of sampling a theorem is positive (and it is the real situation, as in the opposite case no deducibility testing procedure is needed), then we can always choose the parameter N large enough for the original test in random extensions to be the best one from the point of view of minimization of probabilities of error.

Špaček's basic idea of deducibility testing in random extensions has proved (as the following parts of this work show) to be very fruithful, however, his original formulation, as explained above, suffers from a certain inconsistency of assumptions. First, the problem of realization of a random extension generator with the requested properties arises. We have already said that such a generator is equivalent to a random generator of formulas which are neither theorems nor negations of theorems. But, such a set of formulas, is, in general, neither recursive nor recursively enumerable – and only such types of sets can serve as sets of possible outcomes of an effective sampling procedure. Hence, every random extension generator coping with Špaček's demands must contain a subalgorithm deciding for any sampled formula, whether it is a theorem or the negation of a theorem and, in case of the positive answer, preventing these formulas from entering the output. But this subalgorithm would solve the decidability problem for the theory in question, and such an algorithm is supposed not to exist or at least not to be available – it is just why we try to propose a statistical deducibility testing procedure for this theory. Hence, a contradiction occurs.

Another problem is hidden in Špaček's assumption that each extension of the given type is decidable. Either the formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$ is complete in the sense that joining any non-theorem to the set of axioms makes the theory inconsistent, i.e. for each $x \in \mathscr{L} - \mathscr{T}$, $Cn(\mathscr{T} \cup \{x\}) = \mathscr{L}$. However, in this case no extension exists which would satisfy Špaček's conditions, hence, his testing procedure cannot, be realized. Or, there is such a formula x that neither $x \in \mathscr{T}$ nor $\neg x \in \mathscr{T}$, hence, both the extensions $\mathscr{T} \cup \{x\}$ as well as $\mathscr{T} \cup \{\neg x\}$ satisfy Špaček's conditions and they are, in such a case, decidable. However, having an algorithm deciding the set

$Cn(\mathscr{T} \cup \{x\})$ and another algorithm deciding the set $Cn(\mathscr{T} \cup \{\neg x\})$ we would be able to propose an algorithm deciding the set $\mathscr{T}$ itself, hence, again a contradiction arises.

In what follows, we propose another formulation of a statistical deducibility testing procedure based on weakened assumptions. Namely, the random generator will be supposed simply to offer extensions of the considered theory (not excluding those which are not proper or those which are inconsistent) and each formula will be supposed only with a positive probability to be decidable in the at random sampled extension. Our further explanation will be based mainly on [11], where also other details and proofs can be found.

Let $\langle \mathscr{L}, \mathscr{T} \rangle$ be a consistent formalized theory, i.e. $\mathscr{T} \neq \mathscr{L}$. Suppose to have at our disposal an effective and deterministic theorem-prover T, e.g. an appropriate resolution-based theorem-proving computer program with some space and time limitations. Formally, T is a mapping defined on $\mathscr{L}$ and taking its values in the three-element set $\{1, x_0, 0\}$; $T(x) = 1$ is interpreted, for $x \in \mathscr{L}$, as „x is proclaimed to be a theorem", or briefly, "$x \in \mathscr{T}$", $T(x) = 0$ is interpreted as "$x \in \mathscr{L} - \mathscr{T}$" and $T(x) = x_0$ as "we cannot decide about x". Very often this last decision is joined with "$x \in \mathscr{L} - \mathscr{T}$" so that T maps $\mathscr{L}$ into $\{0, 1\}$, but in this case we must keep in mind the different qualitative character of the two remaining decisions. The mapping T is supposed to be recursive in order to assure the effectiveness of the corresponding testing procedure. This means that T enables to deduce certain theorems, let us denote their set by $\mathscr{T}_0$, hence, $\mathscr{T}_0 = \{x : x \in \mathscr{L}, T(x) = 1\}$, giving the negative and, clearly, for $x \in \mathscr{T} - \mathscr{T}_0$ the wrong, answer for other formulas. In other words, if T proclaims a formula to be a theorem we can be sure about it, but not every theorem can be discovered using the theorem-prover T.

Starting from these assumptions we meet immediately the problem which causes most of the difficulties during a mathematical formalization and investigation of statistical deducibility testing procedures. There are, as we have seen, at least two structures by which the set of well-formed formulas is equipped, namely the probabilistic (statistical) and the meta-theoretic (logical) ones, and they are hardly compatible with each other. To be more concrete, probabilistic structures are those generated by random variables, $a, x_1, x_2, \ldots$ which sample the tested formula and the auxiliary axioms, logical structures are those generated by the deducibility rules, axioms or by the theorem-prover T. So, for an α, $0 < \alpha < 1$, the set

$$\{y : y \in \mathscr{L},\ P(\{\omega : \omega \in \Omega,\ a_1(\omega) = y\}) \leqq \alpha\} \tag{5.2}$$

of formulas is a subset of $\mathscr{L}$, which can be easily defined in this "probabilistic" way, i.e. using the terms of probability space, probability measure and random variables, but it would be hardly possible to describe it in "logical" way, e.g., to find a small and recursive set of axioms such that (5.2) would be just the set of all logical consequences of those axioms. On the other hand, the sets $\mathscr{T}$, $\mathscr{T}_0$ of theorems can be easily defined in terms of mathematical logic or T, but the probabilities $P(\{\omega : \omega \in$

$\in \Omega,\ a_1(\omega) \in \mathscr{T}\}),\ P(\{\omega : \omega \in \Omega,\ a_1(\omega) \in \mathscr{T}_0\}),\ P(\{\omega : \omega \in \Omega, a_1(\omega) \to x(\omega) \in \mathscr{T}_0\})$, etc., are very difficult and often practically impossible to compute. It is why we often have to be satisfied only with some rather rough lower or upper bounds for these values.

A model of statistical deducibility testing based on Špaček's ideas but not suffering from the disadvantages of Špaček's original formulation was suggested and studied in [8] and [9]. The explanations in [6] are given in more abstract algebraic terms and they do not limit themselves only to the problem of logical deducibility, in [9] the formulations are more specific. Technical difficulties connected with the incompatibility of the two structures mentioned above force to introduce too extent and complicate formal apparatus, so we do not consider this model for the most appropriate one to be introduced here. As we promised, we shall explain here the model from [11] which we believe to be more open for an intuitive imagination and to be more close to the general model of parametric statistical hypothesis testing as known by statisticians.

Let $x,\ a_1, a_2, \ldots$ be the same random variables as in Theorem 5.2 with the only exception that (5.1) is now supposed to be valid for all closed formulas $x \in \mathscr{L}$. Consider the following conditional probabilities.

$$p_1 = P(\{\omega : \omega \in \Omega,\ T(a_1(\omega) \to x(\omega)) = 1\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}), \tag{5.3}$$

$$p_2 = P(\{\omega : \omega \in \Omega,\ T(a_1(\omega) \to x(\omega)) = 1\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{L} - \mathscr{T}\}), \tag{5.4}$$

and suppose that $p_1 > p_2$. Hence, p_1 is the probability that $Ax \cup \{a_i(\omega)\} \vdash x(\omega)$ can be proved by T under the condition that $x(\omega)$ is a theorem, p_2 is the conditional probability of the same random event under the condition that $x(\omega)$ is not a theorem. However, the occurrences of the random event just described can be observed using the theorem-prover T, so we have transformed the deducibility problem (whether $x(\omega) \in \mathscr{T}$ or not) into a classical parametric test of a simple hypothesis (probability of a certain random event is p_1) again a simple alternative (that this probability equals to p_2, say $p_2 < p_1$). It is a well-known fact that this testing problem can be decided, within an a priori given probability of error, on the base of a sufficiently large number of statistically independent repeated observations of the random event in question.

As far as the assumption $p_1 > p_2$ is concerned, the following theorem can serve as its justification.

Theorem. 5.3. Let the notations and conditions of Theorem 5.2 hold with the exception that (5.1) is supposed to be valid for all $x \in \mathscr{L}_0$. Let T be a theorem-prover such that, for all $x, y \in \mathscr{L}_0$, $T(y) = 1$ ot $T(\neg x) = 1$ imply $T(x \to y) = 1$. Let p_1, p_2 be the conditional probabilities defined by (5.3), (5.4). Let

$$p = P(\{\omega : \omega \in \Omega,\ T(x(\omega)) = 1\} \mid \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}),$$

$$p' = P(\{\omega : \omega \in \Omega,\ T(\neg a_1(\omega)) = 1\} \mid \{\omega : \omega \in \Omega,\ a_1(\omega) \in \mathscr{L} - \mathscr{T}\}),$$

$$P_a(T) = P(\{\omega : \omega \in \Omega,\ a_1(\omega) \in \mathscr{T}\}).$$

Then

$$p > \frac{(1 - p_a(\mathscr{T}))(1 - p')}{1 - (1 - p_a(\mathscr{T}))\, p'} \tag{5.5}$$

implies $p_1 > p_2$.

Proof. Let $x(\omega) \in \mathscr{T}$, then a sufficient (but not necessary) condition for $T(a_1(\omega) \to \to x(\omega)) = 1$ is that either $T(x(\omega)) = 1$, or $T(x(\omega)) = 0$ and, at the same time, $a_1(\omega) \in \mathscr{L} - \mathscr{T}$, $T(\neg a_1(\omega)) = 1$. These two random events are disjoint, first of them occurs with probability p, the other with probability $(1 - p)(1 - p_a(\mathscr{T}))\, p'$ because of the supposed statistical independence of corresponding random variables. Hence,

$$p_1 = P(\{\omega : \omega \in \Omega,\ T(a_1(\omega) \to x(\omega)) = 1\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\}) \geqq \tag{5.6}$$

$$\geqq p + (1 - p)(1 - p_a(\mathscr{T}))\, p'.$$

Let $x(\omega) \in \mathscr{L} - \mathscr{T}$, then a necessary (but not sufficient) condition for $T(a_1(\omega) \to \to x(\omega)) = 1$ is that $a_1(\omega) \in \mathscr{L} - \mathscr{T}$, and this random event occurs with probability $1 - p_a(\mathscr{T})$. Hence, $p_2 \leqq 1 - p_a(\mathscr{T})$, and for p satisfying the inequality (5.5) the relation $p_1 > p_2$ holds. Q.E.D.

Expressed more intuitively, Theorem 5.3 claims that if the theorem-prover T is "clever" of "able" enough, i.e. if the ratio p of theorems which it is able to prove is "high enough", then the assumption $p_1 > p_2$ is valid. The intuitive idea that stays behind all the procedure of statistical deducibility testing using the random extensions, i.e., the idea that there is a greater probability that the auxiliary axioms help us to prove a theorem than to "prove" a non-theorem, is in this way, by Theorem 5.3, quantitatively expressed and, under some conditions, also justified. In fact, the difference between p_1 and p_2 is greater than that computed in the proof of Theorem 5.3, as there are always some theorems which can help us to prove a theorem not provable by T. On the other hand, not every non-theorem possesses the property that enables to "prove" some other non-theorem. A more strong, but simpler sufficient condition for the inequality $p_1 > p_2$ to hold is that $p > 1 - p_a(\mathscr{T})$, as can be easily derived from (5.6).

Supposing, for a while, that we know the values p_1, p_2, and that $p_1 > p_2$, let us introduce some most elementary decision functions for testing the problem whether the probability of $T(a_1(\omega) \to x(\omega)) = 1$ is p_1 or p_2, i.e. whether $x(\omega)$ is a theorem or not.

Theorem 5.4. Let the notations and conditions of Theorem 5.3 hold, let $p_1 > p_2$, let $N \geqq 1$ be an integer, let $\mathfrak{t}$, $\mathfrak{f}$ be two abstract symbols, let $D = D(N, x, \{a_i\}_{i=1}^N)$ be a random variable, defined on the probability space $\langle \Omega, \mathscr{S}, P \rangle$, taking its values

in $\{\mathrm{t}, \mathrm{f}\}$ and such that

$$D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{t}\,,$$

if

$$(5.7) \qquad \left|p_1 - \left(\sum_{i=1}^N T(a_i(\omega) \to x(\omega))\right) N^{-1}\right| \leqq \left|p_2 - \left(\sum_{i=1}^N T(a_i(\omega) \to x(\omega))\right) N^{-1}\right|,$$

$D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{f}$ otherwise. Then

$$(5.8) \qquad P(\{\omega : D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{f}\}/\{\omega : \omega \in \Omega,\ a(\omega) \in \mathscr{T}\}) \leqq \leqq (N|p_1 - p_2|^2)^{-1}\,,$$

$$(5.9) \qquad P(\{\omega : D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{t}\}/\{\omega : \omega \in \Omega, a(\omega) \in \mathscr{L} - \mathscr{T}\}) \leqq \leqq (N|p_1 - p_2|^2)^{-1}\,.$$

Remark. In spite of its rather complicated formalized form the intuitive idea behind this assertion is rather simple. We sample at random formulas $a_1(\omega), a_2(\omega), \ldots, a_N(\omega)$ and test, for each $i \leqq N$, whether $T(a_i(\omega) \to x(\omega)) = 1$ or 0. Moreover, we compute the relative frequency of the cases when this value equals to 1, i.e., when $a_i(\omega) \to x(\omega)$ is provable by T. If this relative frequency is not closer to p_2 than to p_1, then we accept the hypothesis p_1, i.e., we proclaim $x(\omega)$ to be a theorem (using the decision function D we write this decision formally as $D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{t}$). In other case we accept the alternative p_2 and proclaim $x(\omega)$ to be a non-theorem (formally, $D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{f}$). There are two possibilities of error, their probabilities are estimated by (5.8) and (5.9).

Proof of Theorem 5.4. Let $x(\omega) \in \mathscr{T}$, in this case $T(a_i(.) \to x(.))$ is a random variable with the expected value p_1 and with dispersion not exceeding $1/4$. For different i, j, these random variables are mutually independent and equally distributed, hence, the well-known Tchebyshev inequality gives

$$(5.10) \qquad P(\{\omega : \omega \in \Omega,\ \left|N^{-1}\left(\sum_{i=1}^N T(a_i(\omega) \to x(\omega))\right) - p_1\right| > \varepsilon\}) \leqq (4N\varepsilon^2)^{-1}\,.$$

In case $D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{f}$ necessarily the value $N^{-1}\left(\sum_{i=1}^N T(a_i(\omega) \to x(\omega))\right)$ must differ from p_1 by more than $\frac{1}{2}|p_1 - p_2|$; replacing in (5.10) ε by this value we obtain (5.8). (5.9) can be proved in a similar way. Q.E.D.

The decision function D defined above can be easily rewritten in such a way that $D(N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{t}$ iff $\sum_{i=1}^N T(a_i(\omega) \to x(\omega)) \geqq M$ with $M \leqq N$ being easily computable from (5.7). We shall not perform the computation for this case, but for a more general, one, when the both probabilities of error are not taken as comparable. This situation is common in general statistical hypothesis testing theory and it

is solved as follows. The "more dangerous" probability of error is strictly requested to be kept below an a priori given treshold value, say $\alpha > 0$, and the second probability of error is minimized under this condition. In our case, according to the viewpoint accepted in other works dealing with statistical deducibility testing, we consider the error consisting in proclaiming a non-theorem to be a theorem for the more dangerous (because this event may cause the set of formulas proclaimed to be theorems to become inconsistent and so, in a sense, useless for a further use, see the next chapter for more details and references). Hence, having got N, our aim is to find an appropriate $M \leqq N$ as the following theorem precises.

Theorem 5.5. Let the notations and conditions of Theorem 5.4 hold with the exception that the random variable D is defined as follows. For a given M, $M \leqq N$, $D(M, N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{t}$, if $\sum_{i=1}^{N} T(a_i(\omega) \to x(\omega)) \geqq M$, $D(M, N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{f}$ otherwise. Let $\alpha > 0$, let u_α be the α-quantile of the normal distribution $N(0, 1)$, let

$$M_1 = [N \cdot (u_{1-\alpha} \sqrt{(N^{-1} p_2(1 - p_2))} + p_2] + 1 .$$

Then

$$P(\{\omega : \omega \in \Omega,\ D(M_1, N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{t}\} / \{\omega : x(\omega) \in \mathscr{L} - \mathscr{T}\}) \leqq \alpha , \tag{5.11}$$

$$P(\{\omega : \omega \in \Omega,\ D(M_1, N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{f}\} / \{\omega : x(\omega) \in \mathscr{T}\}) = \tag{5.12}$$

$$= \min_{M = M_1}^{N} \{P(\{\omega : \omega \in \Omega,\ D(M, N, x, \{a_i\}_{i=1}^N, \omega) = \mathrm{f}\} / \{\omega : \omega \in \Omega,\ x(\omega) \in \mathscr{T}\})\} .$$

Remark. Remember that the notion of α-quantile for $N(0, 1)$ is defined in the following way: let X be a random variable defined on the probability space $\langle \Omega, \mathscr{S}, P \rangle$, taking its values in the Borel line $\langle E, \mathscr{B} \rangle$ and obeying the normal $N(0, 1)$ distribution, i.e., such that for each real $y \in E$,

$$\Phi(y) = P(\{\omega : \omega \in \Omega,\ X(\omega) \leqq y\}) = \int_{-\infty}^{y} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t .$$

Then u_α is uniquely defined by the equality

$$P(\{\omega : \omega \in \Omega,\ X(\omega) \leqq u_\alpha\}) = \alpha .$$

Proof of Theorem 5.5. Consider the classical statistical hypothesis testing problem with hypothesis $p = p_1$ against the alternative $p = p_2 < p_1$. We want to choose $M \leqq N$ such that the probability of at least M events of the type $T(a_i(\omega) \to x(\omega)) = 1$, $i \leqq N$ were majorized by α supposing that $p = p_2$. Moreover, we look for the minimal M possessing this property in order to minimize the other probability of error. Hence, we look for the minimal M such that

$$\sum_{i=1}^{M} \binom{N}{i} p_2^i (1 - p_2)^{N-i} \geqq 1 - \alpha . \tag{5.13}$$

Denote $\bar{p} = \bar{p}(N, \omega) = N^{-1} \sum_{i=1}^{N} T(a_i(\omega) \to x(\omega))$. The well-known Central Limit Theorem of probability theory then sounds that $\bar{p}$ has, approximately, the normal $N(\mu, \sigma^2)$ distribution with $\mu = p_2$, $\sigma^2 = N^{-1}p_2(1 - p_2)$ (under the condition that $p = p_2$), i.e., the distribution function of $\bar{p}$ is of the form

$$\Phi\left(\frac{x - p_2}{(N^{-1}p_2(1 - p_2))}\right),$$

where Φ is the distribution function of the normal distribution $N(0, 1)$ defined above. Now, (5.13) can be transformed into the form

$$\Phi\left(\frac{(M/N) - p_2}{\sqrt{(N^{-1}p_2(1 - p_2))}}\right) \geqq 1 - \alpha,$$

hence

$$\frac{(M/N) - p_2}{\sqrt{(N^{-1}p_2(1 - p_2))}} \geqq u_{1-\alpha},$$

and an easy calculation gives the value M_1 as stated above. (5.11) and (5.12) follow immediately from the way in which M_1 has been chosen. The values of α-quantiles of the normal distribution $N(0, 1)$ are tabeled and can be found in statistical tabels (e.g., [5]). Q.E.D.

The problem can be solved also in a non-asymptotic way using the notion of incomplete β-function.

The methods explained above have one common feature, namely, their length is fixed, i.e., the number N of random samples which are to be made before a decision is taken is given a priori. As an alternative to these procedures, the general hypotheses testing theory offers the so called sequential tests. In this case the number of random samples necessary to take a decision is a random variable and only its expected value, moments or other statistical characteristics can be computed or estimated. Let us limit ourselves by describing a simple variant of the sequential test procedure for our case, when hypothesis is $p = p_1$ and alternative $p = p_2$. This procedure is described in [1], the underlying theoretical results can be found in [4.4] or [4.5].

Let $r > 0$ be such a real number that we want the sum of both the probabilities of errors not to exceed r. Set

$$k = \frac{\log((1 - p_2)/(1 - p_1))}{\log(p_1/p_2) + \log((1 - p_2)/(1 - p_1))},$$

$$q = \frac{\log((1 - r)/r)}{\log(p_1/p_2) + \log((1 - p_2)/(1 - p_1))}.$$

For each $m = 1, 2, \ldots$ set

$$L_1(m) = km + q, \quad L_2(m) = km - q.$$

Now, sample $a_i(\omega)$ and compute $T(a_i(\omega) \to x(\omega))$. If

$$L_2(m) < \sum_{i=1}^{m} T(a_i(\omega) \to x(\omega)) < L_1(m),$$

sample $x_{m+1}(\omega)$ and continue as above. If

$$\sum_{i=1}^{m} T(a_i(\omega) \to x(\omega)) \leqq L_2(m),$$

stop the sampling and take the decision that $p = p_2$, i.e., proclaim $x(\omega)$ to be a non-theorem. If

$$\sum_{i=1}^{m} T(a_i(\omega) \to x(\omega)) \geqq L_1(m),$$

stop the sampling and take $p = p_1$, i.e., proclaim $x(\omega)$ to be a theorem. Under some very general conditions a decision will be eventually taken with the probability one.

In the considerations and constructions above we have proceeded as if the values p, p', $p_a(\mathscr{T})$, resp. p_1, p_2 of the corresponding probabilities were perfectly known. We have already mentioned before that this was not usually the case, however, we might again turn ourselves to mathematical statistics to offer us a partial (and relative, as will be shown later) remedy. These probabilities can be statistically estimated on the ground of random samples, i.e., by an appropriate corresponding relative frequencies. We know very well that the only assertions which we are justified to claim on the base of a finite sample are of the form

$$(5.14) \qquad P(\{\omega : \omega \in \Omega,\ \bar{p}_{a,n}(\mathscr{T}, \omega) - \delta < p_a(\mathscr{T}) < \bar{p}_{a,n}(\mathscr{T}, \omega) + \delta\}) < 1 - \varepsilon,$$

taking $p_a(\mathscr{T})$ as an example and denoting by $\bar{p}_{a,n}(\mathscr{T}, \omega)$ the relative frequency of theorems among n formulas sampled at random and independently by random variables $a_1, a_2, \ldots, a_n$ (and similarly for p, p', p_1, p_2). Positive reals δ and ε serve as parameters in (5.14) and can be diminished when n increases. In this way we can obtain (either by an immediate statistical estimation or by a computation starting from estimates of p, p', and $p_a(\mathscr{T})$ two values $\bar{p}_{n,1}(\omega)$, $\bar{p}_{n,2}(\omega)$ such that, for $i = 1, 2$,

$$(5.15) \qquad P(\{\omega : \omega \in \Omega,\ p_i \in \langle \bar{p}_{n,i}(\omega) - \delta,\ \bar{p}_{n,i}(\omega) + \delta \rangle\}) < 1 - \varepsilon$$

Increasing appropriately the value of n we may choose δ in such a way that $\bar{p}_{n,1}(\omega) - \delta > \bar{p}_{n,2}(\omega) + \delta$. Clearly, replacing in our hypothesis testing problem the hypothesis H:

$$p = p_1 \quad \text{by} \quad H' : p = \tilde{p}_1 = \bar{p}_{n,1}(\omega) - \delta \quad \text{and the alternative } A:$$

$$p = p_2 \quad \text{by} \quad A' : p = \tilde{p}_2 = \bar{p}_{n,2}(\omega) + \delta, \quad \text{we have made } A' \text{ and } H'$$

"more close" to each other than A and H had been. I.e., it will be more difficult to distinguish between them. In other words, taking the number of random samples made by random variables x_i's large enough to assure the distinguishing between A' and H' within a priori majorized probabilities of error we can rely (with probability at least $1 - \varepsilon$) that this test distinguishes also between H and A with the same or smaller probabilities of errors. Considering the probabilities of errors with respect to the original testing problem ($x(\omega) \in \mathscr{T}$ or $x(\omega) \in \mathscr{L} - \mathscr{T}$?) we must enlarge the probabilities of errors connected with H' and A' by the probability with which (5.15) does not hold for at least one $i \leqq 2$. However, this probability is a continuous function of ε and can be diminished when ε decreases, i.e., when n increases. Hence, improving the used statistical estimates, we may replace the original statistical hypothesis testing problem by a similar one with probabilistic parameters replaced by values obtained by statistical estimations or computed from extimates. A more detailed description can be found in [11].

As we have already mentioned, the solution consisting in replacing the probability values by corresponding relative frequencies is only a relative outcome. Consider again, as an example, the value $p_a(\mathscr{T})$. In order to obtain the relative frequency $\bar{p}_{a,n}(\mathscr{T}, \omega)$ we must take an n-tuple $a_1, a_2, \ldots, a_n$ of mutually independent and equally distributed random variables, then realize the random sample giving a sequence $a_1(\omega), a_2(\omega), \ldots, a_n(\omega)$ of formulas from $\mathscr{L}$ and then, finally, *decide, for each* $i \leqq n$, *whether* $a_i(\omega) \in \mathscr{T}$ *or not* and compute the relative frequency $\bar{p}_{a,n}(\mathscr{T}, \omega)$ of the positive answers. Hence, the original decision problem arises again!

It is why our estimate $\bar{p}_{a,n}(\mathscr{T}, \omega)$ as well as other estimates mentioned above can be and must be called *relative*. Or, after all, we need some other theorem-prover T', better than T in the sense that, for all $x \in \mathscr{T}$, $T(x) = 1$ implies $T'(x) = 1$ and that there is at least one $y \in \mathscr{T}$ such that $T(y) = 0$ and $T'(y) = 1$ (of course, $T'(x) = T(x) = 0$ for all $x \in \mathscr{L} - \mathscr{T}$). So, computing $\bar{p}_{a,n}(\mathscr{T}, \omega)$, we set in fact,

$$\bar{p}_{a,n}(\mathscr{T}, \omega) = \frac{1}{n} \sum_{i=1}^{n} T'(a_i(\omega)). \tag{5.16}$$

It is necessary to estimate the values $p_a(\mathscr{T})$, p, p', resp. p_1, p_2, only once for a formalized theory the formulas of which are to be statistically tested, not particularly for each tested formula. Moreover, this estimation need not be performed in a real time, i.e. simultaneously with a real physical process, as the statistical deducibility testing is expected to do (e.g., if used as a part of robot plan formation procedure, cf. Chapter 4 and Chapter 8). Hence, we may assume that the theorem-prover T' used in the process of an a priori calibration of T is of a better quality than T. E. g. if T' and T are both resolution-based theorem-provers, then T' executes a greater number of possibly more complex or difficult resolutions or substitutions than T. In this case the estimates of probability values obtained in the same way as in (5.16) can be used in the process of statistical deducibility testing and this test brings a positive information concerning the original decision problem whether $x(\omega) \in \mathscr{T}$ or not.

We do not intend to go into more details as far as this and similar statistical deducibility testing procedures are concerned and we refer the reader to the following chapters of this work and to references. Nevertheless, still one problem rests, which is worth mentioning, namely, how to realize the random sampling of well-formed formulas of a formalized theory. We have used the phrase "sample at random a formula $a_1(\omega)$ or $x(\omega)$" without any further explanation how to do it an in the same way Špaček proceeds in his papers. The demand of computer implementability of statistical deducibility testing procedures necessitates, however, to propose an effective random generator of well-formed formulas. This can be done in several ways, let us suggest one which is rather simple and, as practical experiments have proved, very quick (see [7] for more details).

First of all, we shall profit of the well-known fact of mathematical logic according to which functors and quantifiers used in a formalized theory can be reduced to, e.g., implication, negation and general quantifier, the others being definable by them. For example, if $A, B \in \mathscr{L}$, then

$$A \wedge B = (\mathrm{df})\,(\neg A) \to B\,,$$
$$A \wedge B = (\mathrm{df})\,\neg((\neg A) \vee (\neg B))\,,$$
$$(\exists x)\,A = (\mathrm{df})\,\neg(\forall x)\,(\neg A)\,.$$

Moreover, introducing a new propositional constant F (falsehood), i.e., an identically false formula, we may eliminate also negation, setting

$$\neg A = (\mathrm{df})\,A \to F.$$

Now, having at our disposal just one functor ($\to$) and just one quantifier ($\forall$) we may omit special symbols for them writing simply $[A]\,[B]$ for $A \to B$ and $[x[A]]$ for $(\forall x)\,A$. Let us denote by $\mathscr{L}^*$ the formalized language resulting from $\mathscr{L}$ by this modification and reduction. We may limit ourselves to the construction of a random generator which samples formulas from $\mathscr{L}^*$, as their reformulation in $\mathscr{L}$ is a matter of quick and easily programmable routine.

Let all elementary symbols of $\mathscr{L}^*$, i.e., [,], F and individual indeterminates, as well as elementary formulas of $\mathscr{L}^*$ (which are the same as in $\mathscr{L}$) be numbered (enumerated). Different objects are supposed to have different numbers ascribed, on the other hand, the possibility that to an object more than one index is ascribed is not excluded, in fact, this possibility can be operatively used in order to control the statistical parameters of the resulting random generator.

Let G be a random number generator such that G produces positive integers and only those corresponding to elementary symbols or formulas of $\mathscr{L}^*$ in the sense of the assumed enumeration. This can be easily achieved by a simple modulation. In this way any finite sequence produced by G can be understood as a sequence of elementary objects of $\mathscr{L}^*$, hence, G can be converted into a random generator of formulas from $\mathscr{L}^*$, if enriched by a procedure transforming each sequence of

elementary objects (i.e., symbols or elementary formulas) of $\mathscr{L}^*$ into a well-formed formula of $\mathscr{L}^*$. Such a procedure can run as follows.

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the sequence of random numbers sampled by G and understood as a sequence of elementary objects of $\mathscr{L}^*$.

(I) If there is at least one occurence of F or of an elementary formula, go to (II), otherwise replace α_1 by an elementary formula, say, by this one with the smallest index.

Remark. This step may be omitted, but in this case the possibility of sampling the empty formula is not excluded. Under some circumstances this possibility may be desirable, as it enables to sample formulas without any a priori restriction of their lengths (c.f. [4.2] for details).

(II) Put all occurences of elementary formulas and F into brackets, i.e., if α_j, $j \leqq n$, is an elementary formula or F, replace α_j by $[\alpha_j]$.

(III) Put the left bracket before all occurences of an indeterminate, i.e., if α_j is an indeterminate, replace α_j by $[\alpha_j$.

(IV) Take an auxiliary variable S, put $S = 0$. Pass through the sequence obtained from $\alpha_1, \alpha_2, \ldots, \alpha_n$ by (I)–(III), from the left to the right and replace S by $S + 1$, when an occurence of $[$ is met, replace S by $S - 1$, when an occurence of $]$ is met. If $S = -1$, inscribe $[$ at the very beginning of the sequence, put $S = 0$ and go on. If the final value of S is positive, put S occurences of $]$ at the very end of the sequence.

(V) Pass again through the vector obtained by (IV) from the left to the right. Meeting with the left bracket, start with the procedure described in (IV). The right bracket by which, for the first time, $S = 0$, corresponds to the initial left bracket and these two brackets form a pair. If there is a pair of brackets not containing any occurence of elementary formulas of F, all the symbols between these brackets including the brackets themselves are erased.

(VI) Any occurence of an indeterminate preceded by the left bracket and not occuring in the scope of the general quantifier formed by these two symbols is erased.

(VII) If a pair of brackets occurs inside another pair of brackets and if there is no symbol between the two left brackets and no symbol between the two right brackets, one pair of brackets is erased (to avoid superfluous double bracketing).

This procedure is based on a reformulation of the notion of well-formed formula suggested in [3] and [14], a more detailed description of the procedure as well as proofs of the following assertions expressing the most important properties of the random generator resulting when G combined with (I)–(VII).

Theorem 5.6. (a) For each finite sequence of random numbers sampled by G, the result of application of (I)–(VII) to this sequence is a well-formed formula of the language $\mathscr{L}^*$. (b) Let A be a formula of $\mathscr{L}^*$ such that, for each elementary object

occuring in A at least one of the indices ascribed to this elementary object is sampled by G with positive probability. Then the corresponding random generator of formulas samples A with a positive probability.

Because of the limited extent of this work and its surveyal character we do not describe here other variants of the basic model for statistical deducibility testing as explained above. Some experience-based modifications of these tests considering also a rather sophisticated way of their repetitive use are studied in the next chapter. In no way this is to mean that the idea of at random sampled auxiliary axioms proposed by Špaček is the only way how to introduce and use probability and statistics in the deducibility testing. After all, theorem-proving can be always understood as a searching procedure in a non-empty space, in general, an infinite one. In the case of a classical proof we are looking for appropriate premises which are derivable from the theorems already proved and which enable to derive the tested formula. In the case of resolution-based theorem-proving we are looking for an appropriate substitution enabling to match two clauses, and for an appropriate pair of clauses to be resolved with the aim to obtain, if possible and as soon as possible, the empty clause. When other forms of application of Herbrand theorem are considered (we recall that this theorem lies in the grounds of the resolution-based as well as many other theorem-provers), the searching process consists in looking for a useful element (or a finite set of elements) of the Herbrand universum of terms. In all these cases each deterministic searching procedure can be proved to be very useful and economic in some cases, being at the same time very impractical and stupid in other cases, also easily demonstrable (cf. various refinements of resolution-based theorem-proving as mentioned in Chapter 3 and studied in more details, e.g., in [1.1]). It is why nondeterministic and heuristic decision theories together with statistics may serve at least as one of several alternative tools how to describe and effectively handle this indeterminism (random sampling of premises in proofs, cf. [10], candidate clauses for resolution, elements of Herbrand universum, etc.).

The notions and assertions used in this chapter and belonging to the domain of probability theory and mathematical statistics are of a very elementary level and can be found in each undergraduate textbook of these branches of mathematics. In the references below we introduce some more advanced textbooks on probability theory which cover not only this chapter but also all the probabilistic and statistical notions and assertions which will be used in the rest of this work.

REFERENCES

[1] V. Fabian: Základní statistické metody. NČSAV (Publishing House of the Czechoslovak Academy of Sciences), Prague 1963.

[2] W. Feller: An Introduction to Probability Theory and its Applications, vol. I and II. John Wiley and Sons, Chapman and Hall, New York—London, vol. I (second edition) 1957, vol. 2 (first edition) 1966. (Russian translation: Mir, Moscow 1964, 1967.)

[3] G. Gentzen: Untersuchungen über das logische Schliessen. Mathem. Zeitschirft *39* (1934—5), 176—210, 405—431.
[4] Б. В. Гнеденко: Курс теории вероятностей. Физматгиз, Москва 1961.
[5] J. Janko: Statistické tabulky. NČSAV (Publishing House of the Czechoslovak Academy of Sciences), Prague 1958.
[6] I. Kramosil: Statistical Estimation of Deducibility in Polyadic Algebras. Kybernetika *7* (1971), 3, 181—200.
[7] I. Kramosil: A Method for Random Sampling of Well-Formed Formulas. Kybernetika *8* (1972), 2, 133—148.
[8] I. Kramosil: Statistical Estimation of Deducibility in Formalized Theories. In: Proceedings of the Fourth Conference on Probability Theory, Braşov, 1971, Editura Academici RSR, 1973, 281—298.
[9] I. Kramosil: A Method for Statistical Testing of an at Random Sampled Formula. Kybernetika *9* (1973), 3, 162—173.
[10] I. Kramosil: Konstruktivní test délky formalizovaných důkazů. Research Report no. 869, Institute of Information Theory and Automation, September 1978.
[11] I. Kramosil, J. Šindelář: Statistical Deducibility Testing with Stochastic Parameters. Kybernetika *14* (1978), 6, 385—396.
[12] M. Loève: Probability Theory. D. van Nostrand Comp., Princeton, N. J. Toronto, New York, London 1960. (Russian translation: IIL Moscow, 1962.)
[13] A. Rényi: Probability Theory. Akadémiai Kiadó, Budapest 1970. (Czech translation: Academia, Prague 1972.)
[14] Ohama Shigeo: On a Formalism which Makes any Sequence of Symbols Well-Formed Nagoya Math. J. *32* (1968), 1—4.
[15] R. Sikorski: Boolean Algebras, Second Edition. Springer-Verlag, Berlin—Göttingen—Heidelberg—New York 1964. (Russian translation: Mir, Moscow, 1969.)
[16] A. Špaček: Statistical Estimation of Provability in Boolean Logics. In: Transactions of the Second Prague Conference on Information Theory, Prague 1959. NČSAV (Publishing House of the Czechoslovak Academy of Sciences), Prague 1960, 609—626.
[17] A. Špaček. Statistical Estimation of Semantic Provability. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics, 1960, vol. I, 655—688.

6. THE ROLE OF EXPERIENCE IN STATISTICAL DEDUCIBILITY TESTING

There is a common feature of all the deterministic as well as stochastic theorem-proving or deducibility testing methods explained in this work or in references mentioned above. Namely, this feature consists in the fact that these procedures are proposed to test one particular formula. If this formula is tested and another one is to be investigated the mentioned procedures are not able to take a profit of the information obtained during the testing of the first formula and from the result of this test. This information is neglected even in the case of a very simple logical dependence between the two successively tested formulas, so the later formula will be tested in the same way as the former one. However, in all actual applications of a theorem-proving or deducibility testing procedure we must suppose that a sequence of formulas will enter the input of the procedure to be tested, hence, a sophisticated

way how to use the already tested formula to help us when testing the subsequent ones may be of a great importance.

This fact implies the necessity to change the criteria used in the procedures which are of stochastic character in order to classify the statistical qualities of these procedures. There were two principal criteria, namely, probabilities of errors: probability of proclaiming a formula to be a theorem under the condition that it is a non-theorem and the probability of proclaiming a formula to be a non-theorem under the condition that it is a theorem (the tested formula is supposed to be sampled at random and both the probabilities are supposed to be defined). Now, other criteria will be used: (1) the probability that an a priori given formula will be found among the formulas proclaimed to be theorems under the conditions that this formula is a non-theorem and that just n formulas were tested, (2) the probability that an a priori given formula will be found among the formulas proclaimed to be non-theorems under the conditions that this formula is a theorem and that just n formulas were tested, (3) expected values and limit values of the probabilities mentioned in (1) and (2). In this chapter a method will be proposed how to take into consideration at least the most simple connections and logical dependences among the tested formulas.

Consider a formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$, say, a first-order one. When describing a statistical deducibility testing procedure we tacitly assumed that the formula, submitted to this procedure, had been submited, before, to a deterministic theorem-proving or deducibility testing procedure, but without any success, in other words deducibility testing of each formula consists of two stages – the deterministic and the stochastic ones. We may use symbols $+2$ and -2 in order to denote the positive (negative, resp.) deterministic decision about the theoremhood of the tested formula, in the same way we may use symbols $+1$ and -1 in order to denote the positive (negative, resp.) statistical decision about the tested formula. Formally, we suppose to have at our disposition random variables $T(N_0, p, \cdot)$ defined, for each $N_0 = 1, 2, \ldots$ and each $p \in \mathscr{L}$, on a probability space $\langle \Omega, \mathscr{S}, P \rangle$, taking their values in the set $\{-2, -1, 1, 2\}$ of integers and satisfying the following properties:

(1) For each $N_0 \geqq 1$ and each $p \in \mathscr{L}$, if $T(N_0, p, \omega) = 2$, then $p \in \mathscr{T}$ (i.e., p is a theorem).

(2) For each $N_0 \geqq 1$ and each $p \in \mathscr{L}$, if $T(N_0, p, \omega) = -2$, then $p \in \mathscr{L} - \mathscr{T}$ (i.e. p is a non-theorem).

(3) For each $\varepsilon > 0$ there exists an $N_0(\varepsilon)$ such that for each $N > N_0(\varepsilon)$ and each non-theorem p
$P(\{\omega : \omega \in \Omega,\ T(N, p, \omega) = 1\}) < \varepsilon$.

(4) For each $N_0 \geqq 1$ and each non-theorem p
$P(\{\omega : \omega \in \Omega,\ T(N_0, p, \omega) = 2\}) = 0$.

(5) For each $N_0 \geqq 1$ and each theorem p
$P(\{\omega : \omega \in \Omega,\ T(N_0, p, \omega) = -2\}) = 0$.

(6) If for *some* $N_0 \geqq 1$ and for *some* $\omega \in \Omega$, $T(N_0, p, \omega) = 2$ (or $T(N_0, p, \omega) = -2$, resp.), then for *all* $N_0 \geqq 1$ and *all* $\omega \in \Omega$, $T(N_0, p, \omega) = 2$ (or $T(N_0, p, \omega) = -2$, resp.).

(7) Let p be such a formula that there is no occurence of the existential quantifier followed by an occurence of the universal quantifier in the prenex normal form of p. If $p \in \mathscr{L} - \mathscr{T}$, then $T(N_0, p, \omega) = -2$ for all $N_0 \geqq 1$ and all $\omega \in \Omega$, if $p \in \mathscr{T}$, then $T(N_0, p, \omega) = 2$ for all $N_0 \geqq 1$ and all $\omega \in \Omega$.

The tested formula p is proclaimed to be a theorem, if $T(N_0, p, \omega) > 0$, and it is proclaimed to be a non-theorem, if $T(N_0, p, \omega) < 0$. (6) states that all formulas from $\mathscr{L}$, which are of the so called $A-E$ *form* (i.e., in the prenex normal form of which universal quantifiers, if any, preceed to existential ones, if any) are algorithmically decidable. This assumption agrees with the fact that this class of formulas actually is decidable, c.f., e.g., [2.12] for more details.

Lemma 6.1. Let $\langle \mathscr{L}, \mathscr{T} \rangle$ be a first-order predicate theory, let T be the random variable defined above, let p be the result of a substitution into a propositional tautology.

Denote

$$R_1 = \{p : p \in \mathscr{L},\ T(N_0, p, \omega) = 2\}\,, \quad R_2 = \{p : p \in \mathscr{L},\ T(N_0, p, \omega) = -2\}\,,$$
$$R = R_1 \cup R_2\,.$$

Then $p \in R_1 = R \cap \mathscr{T}$.

Proof. Cf. Lemma 1, in [1], and its proof.

Lemma 6.2. Let $\langle \mathscr{L}, \mathscr{T} \rangle$ and T be the same as above, let x, y, z be formulas without free indeterminates, not belonging to $R_1 \cup R_2$. If $x \to y$ as well as $y \to z$ belong to R_1, then also $x \to z$ belongs to R_1.

Proof. Cf. Lemma 2, in [1], and its proof.

Let us consider a probability space $\langle \Omega, \mathscr{S}, P \rangle$ with a sequence $a_1, a_2, \ldots$ of random variables taking their values in the set of all well-formed formulas of the investigated theory, mutually independent and equally distributed, with the property

$$P(\{\omega : \omega \in \Omega,\ a_1(\omega) = p\}) > 0 \tag{6.1}$$

for each $p \in \mathscr{L}$. Such a sequence of random variables can be approximated by an independent repeating of the algorithm for random sampling of well-formed formulas, explained at the end of the last chapter (see more details in [4.2] or [5.7]).

Define, for every set $D \subset \mathscr{L}$ of formulas and every $p \in \mathscr{L}$, a function $S(D, p)$ as follows:

$S(D, p) = 1$, if p has the form $(\forall x_i) A \to A^*$, where A^* is the result of the substitution of an indeterminate or term for x_i in A, or if p has the form $(A \to B) \to$

$\rightarrow (A \rightarrow (\forall x_i) B)$, where the indeterminate x_i does not occur freely in A, or if p has the form $(\forall x_i) A$ for some $A \in D$, or if p has the form $A \wedge (A \rightarrow B)$ for some $A, B \in D$. $S(D, p) = 0$ otherwise. (Intuitively speaking, $S(D, p) = 1$, iff p is either a predicate calculus axiom or an immediate logical consequence of one or two formulas from D with respect to generalization rule or modus ponens rule). Denote, for each $p \in L$, by $\bar{p}$ the universal closure of p (namely, $\bar{p} = p$ if no free indeterminates occur in p).

Now, let us describe the decision procedure. Here we do so in a verbal form, a flowchart can be found in [1]. First of all, we have to initialize the input values.

Set $n = 0$, $D_1^1 = A_1^1 = \emptyset$ (the empty set), $D_1^2 = \{\mathsf{AX}_1, \mathsf{AX}_2, \ldots, \mathsf{AX}_S\}$, $A_1^2 = \{\neg \mathsf{AX}_1, \neg \mathsf{AX}_2, \ldots, \neg \mathsf{AX}_S\}$. Here $\mathsf{AX}_1, \ldots, \mathsf{AX}_S$ are the specific axioms of the investigated theory.

(a) Put $n = n + 1$, put $p = a_n(\omega)$.

Is $T(p) = 2$? If yes, put $D_{n+1}^2 = D_n^2 \cup \{\bar{p}\}$, $A_{n+1}^2 = A_n^2 \cup \{\neg \bar{p}\}$, and return to (a).

Is $T(p) = -2$? If yes, put $A_{n+1}^2 = A_n^2 \cup \{\bar{p}\}$, and return to (a).

Is $S(D_n^2, p) = 1$? If yes, put $D_{n+1}^2 = D_n^2 \cup \{\bar{p}\}$, $D_{n+1}^1 = D_n^1 - \{\bar{p}, \neg \bar{p}\}$, $A_{n+1}^2 = A_n^2 \cup \{\neg \bar{p}\}$, $A_{n+1}^1 = A_n^1 - \{\bar{p}, \neg \bar{p}\}$, and return to (a).

Set $p = \bar{p}$ (i.e., bound all free indeterminates in p by universal quantifiers.

Is there any $q \in D_n^2$ such that $T(q \rightarrow p) = 2$? If yes, put $D_{n+1}^2 = D_n^2 \cup \{p\}$, $D_{n+1}^1 = D_n^1 - \{\bar{p}, \neg \bar{p}\}$, $A_{n+1}^2 = A_n^2 \cup \{\neg p\}$, $A_{n+1}^1 = A_n^1 - \{p, \neg p\}$, and return to (a).

Is there any $q \in A_n^2$ such that $T(p \rightarrow q) = 2$? If yes, put $D_{n+1}^1 = D_n^1 - \{p\}$, $A_{n+1}^2 = A_n^2 \cup \{p\}$, $A_{n+1}^1 = A_n^1 - \{p, \neg p\}$, and return to (a).

(b) Is there any $q \in D_n^1$ such that $T(q \rightarrow p) = 2$? If yes, put $D_{n+1}^1 = D_n^1 \cup \{p\}$, $A_{n+1}^1 = A_n^1 \cup \{\neg p\}$, and return to (a).

Is there any $q \in A_n^1$ such that $T(p \rightarrow q) = 2$? If yes, put $A_{n+1}^1 = A_n^1 \cup \{p\}$, and return to (a).

Now, finish the computation of $T(N_0, p, \omega)$ in order to decide whether this value is $+1$, or -1 (i.e., apply the statistical deducibility testing procedure in order to decide about the deducibility of p).

Is $T(N_0, p, \omega) = 1$? If yes, put $D_{n+1}^1 = D_n^1 \cup \{p\}$, $A_{n+1}^1 = A_n^1 \cup \{\neg p\}$, and return to (a).

Put $A_{n+1}^1 = A_n^1 \cup \{p\}$ and return to (a).

Hence, the procedure samples at random formulas $a_1(\omega), \ldots, a_n(\omega)$ and classifies them into four classes. If a formula is sampled once more, it can be re-classified. The set $D_n^1 \cup D_n^2$ represents the set of formulas which are stated to be theorems under the condition that just n formulas were sampled and tested, $A_n^1 \cup A_n^2$ contains the formulas which are, at the same instant, stated to be non-theorems.

It follows immediately from the algorithm, that D_n^1, D_n^2, A_n^1 and A_n^2 are random variables, defined on the probability space $\langle \Omega, \mathscr{S}, \mathsf{P} \rangle$ and taking their values in the

set of all finite subsets of formulas of the theory in question. Countability of the set $\mathscr{L}$ of formulas implies the countability of the set of all its finite subsets and this fact proves the measurability of the random variables D_n^i, A_n^i, $i = 1, 2$, $n = 1, 2, \ldots$.

Theorem 6.1. For all $n \geqq 0$ and for all $\omega \in \Omega$ the following assertions hold:

(I) $D_n^2(\omega) \subset D_{n+1}^2(\omega)$, $A_n^2(\omega) \subset A_{n+1}^2(\omega)$.

(II) $D_n^2(\omega) \subset \mathscr{T}$, $A_n^2(\omega) \subset \mathscr{L} - \mathscr{T}$.

(III) There are no x, y, z such that $x \in D_n^2$, $y \in \mathscr{L}$, $z \in A_n^2$ and that the relations $x \to y \in \mathscr{T}$, $y \to z \in \mathscr{T}$ would hold simultaneously.

(IV) There are no x, y, z such that $x \in D_n^1$, $y \in \mathscr{L} - (R_1 \cup R_2)$, $z \in A_n^1$ and that the relations $x \to y \in R_1$, $y \to z \in R_1$ would hold simultaneously.

(V) The sets D_n^1, D_n^2, A_n^1, A_n^2 are mutually pairwise disjoint.

Proof. Cf. Theorem 1, in [1], and its proof.

This theorem shows the basic properties of the proposed algorithm and its content can be intuitively explained as follows. In a given time instant n each of the formulas sampled until this instant to be tested is classified and it is classified uniquely, i.e. it belongs to just one of the four classes. This classification, of course, need not be the original one and, on the other hand, it is subjected to the possibility of an eventual re-classification in the future. Only formulas from D_n^2 and A_n^2 are classified definitely and they cannot be replaced into another class. Assertions (III) and (IV) express the fact that the classification is, in a sense, self-consistent, namely, no formula exists, which would be derivable from formulas proclaimed to be theorems and which would, at the same time, imply a formula proclaimed to be a non-theorem. Finally, the classification is "partially correct": in the sense that the formulas classified into A_n^2 or D_n^2 are decided correctly, i.e., non-theorems as non-theorems, theorems as theorems.

The last sentence, expressing verbally assertion (II) of Theorem 6.1, evoques the idea of the ideal classification of the sequence $a_1(\omega), a_2(\omega), \ldots, a_n(\omega)$ of the tested formulas which would consist in classifying all theorems into D_n^2 and all non-theorems into A_n^2 leaving the sets D_n^1 and D_n^2 empty. As the following theorem states, this ideal state can be reached only partially and asymptotically.

Theorem 6.2. Consider the formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$ and the classification algorithm as above. For every $p \in \mathscr{T}$, and every q such that $\neg q \in \mathscr{T}$,

$$\lim_{n \to \infty} P(\{\omega : \omega \in \Omega,\ p \in D_n^2(\omega)\}) = 1\,, \tag{6.2}$$

$$\lim_{n \to \infty} P(\{\omega : \omega \in \Omega,\ q \in A_n^2(\omega)\}) = 1\,. \tag{6.3}$$

Remark. As $D_n^2(\omega) \subset \mathscr{T}$, $A_n^2(\omega) \subset \mathscr{L} - \mathscr{T}$, the assertion of this theorem can be expressed in the following way: every theorem will be, sooner or later, proclaimed to be a theorem, every negation of a theorem will be, sooner or later, proclaimed to be a non-theorem.

Proof. Cf. Theorem 2, in [1], and its proof.

It follows immediately from Theorem 6.2 that for each theorem p of the considered theory $\langle \mathscr{L}, \mathscr{T} \rangle$

$$P(\{\omega : \omega \in \Omega,\ p \in (A_n^1(\omega) \cup A_n^2(\omega) \cup D_n^1(\omega))\}) \to 0\,, \quad n \to \infty\,, \tag{6.4}$$

which answers the question contained in the first criterion introduced above in order to judge the quality of the proposed test. The following theorems will provide an information about the connections between the test and the other criteria mentioned above.

Let the sets $E_1(p)$, $E_2(p)$ of formulas and real numbers $e_1(p)$, $e_2(p)$, be defined for each formula p, $p \in \mathscr{L} - (R \cup \mathscr{T})$, in the following way:

$$E_1(p) = \{x : x \in \mathscr{L} - (R \cup \mathscr{T}),\ x \to p \in R_1\}\,,$$

$$E_2(p) = \{y : y \in \mathscr{L} - (R \cup \mathscr{T}),\ p \to y \in R_1\}\,,$$

$$e_1(p) = P(\{\omega : \omega \in \Omega,\ a_1(\omega) \in E_1(p)\})\,,$$

$$e_2(p) = P(\{\omega : \omega \in \Omega,\ a_1(\omega) \in E_2(p)\})\,.$$

Then we can assert

Theorem 6.3. Consider the formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$ and the classification algorithm as above. For every formula p, $p \in \mathscr{L}$, the inequality

$$P(\{\omega : \omega \in \Omega,\ p \in \bigcup_{n=1}^{\infty} A_n^2(\omega)\}) > 0 \tag{6.5}$$

holds iff there exist an index m and formulas $q_0, q_1, \ldots, q_m$ from $\mathscr{L}$ such that $q_i \to q_{i+1} \in R_1$, $i = 0, 1, 2, \ldots, m-1$, $q_0 = p$, $q_m \in R_2 = R \cap (\mathscr{L} - \mathscr{T})$. If it is the case, the following holds:

$$P(\{\omega : \omega \in \Omega,\ p \in \bigcup_{n=1}^{\infty} A_n^2(\omega)\}) = 1\,. \tag{6.6}$$

Proof. Cf. Lemma 3, in [1], and its proof.

Theorem 6.3 can serve as a lemma in order to prove the following.

Theorem 6.4. Consider the formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$ and the classification algorithm as above. Let p be a non-theorem not belonging to R, i.e., $p \in \mathscr{L} - (\mathscr{T} \cup R)$, then the following assertions hold:

(I) Supposing there exist an index m and formulas $q_0, q_1, \ldots, q_m$, $q_0 = p$, $q_m \in R_2$, $q_i \to q_{i+1} \in R_1$, $i = 0, 1, \ldots, m-1$, then

$$P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}/\{\omega : \omega \in \Omega,\ p \in \bigcup_{i=1}^{n-1} \{a_i(\omega)\}\}) \to 0\,, \quad n \to \infty\,, \tag{6.7}$$

$$P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}) \to 0\,, \quad n \to \infty\,. \tag{6.8}$$

(II) Supposing that m and $q_1, q_2, \ldots, q_m$ with the properties mentioned in (I) do not exist, then for all indices n

$$P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}/\{\omega : p \in \bigcup_{i=1}^{n-1} \{a_i(\omega)\}\}) \leqq \tag{6.9}$$

$$\leqq \frac{e_1(p)\,\varepsilon}{e_1(p)\,\varepsilon + e_2(p)\,(1-\varepsilon)} \left(1 - (1 - e_1(p)\,\varepsilon - e_2(p)\,(1-\varepsilon))^{n-1}\right)$$

$$\lim_{n\to\infty} P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}/\{\omega : \omega \in \Omega,\ p \in \bigcup_{i=1}^{n-1} \{a_i(\omega)\}\}) \leqq \tag{6.10}$$

$$\leqq \frac{e_1(p)\,\varepsilon}{e_1(p)\,\varepsilon + e_2(p)\,(1-\varepsilon)}\,,$$

$$P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}) \leqq \tag{6.11}$$

$$\leqq \frac{e_1(p)\,\varepsilon}{e_1(p)\,\varepsilon + e_2(p)\,(1-\varepsilon)} \left(1 - (1 - e_1(p)\,\varepsilon - e_2(p)\,(1-\varepsilon))^{n-1}\right),$$

$$\lim_{n\to\infty} P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}) \leqq \frac{e_1(p)\,\varepsilon}{e_1(p)\,\varepsilon + e_2(p)\,(1-\varepsilon)}\,, \tag{6.12}$$

where

$$\varepsilon = \sup_{\bar{p} \in \mathscr{L} - \mathscr{T}} [P(\{\omega : \omega \in \Omega,\ T(N_0, \bar{p}, \omega) = 1\})]\,.$$

Proof. Cf. Theorem 3, in [1], and its proof.

When realizing the investigated classification algorithm we must, of course, stop the running after a finite number of steps, say n, proclaiming formulas from $D_n^1 \cup D_n^2$ to be theorems. The possibility that a non-theorem enters D_n^2 is exluded, as we already know, however, for D_n^1 it is, in general, possible. The theorem above estimates the probability that a non-theorem can be found in D_n^1 either under the condition that this non-theorem has been already sampled and tested, or in the absolute (unconditioned) sense. As can be seen, in all cases this probability of error can be majorized by a linear function of ε, where ε majorizes the probability of proclaiming a non-

theorem to be a theorem by the original statistical deducibility testing procedure T. In some particular cases even the probability for a non-theorem to belong to D_n^1 tends to zero.

Theorem 6.4 immediately gives, that

$$P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}/\{\omega : \omega \in \Omega,\ p \in \bigcup_{i=1}^{n-1} \{a_i(\omega)\}\}) \to 0\,, \quad \varepsilon \to 0\,,$$

hence, because of the property (3) of the statistical deducibility testing procedure T it follows, that for each $\varepsilon_0 > 0$ such an $\varepsilon_1 > 0$ exists, that

$$\frac{e_1(p)\,\varepsilon_1}{e_1(p)\,\varepsilon_1 + e_2(p)\,(1 - \varepsilon_1)} \leqq \varepsilon_0\,,$$

hence, if $T(N_0(\varepsilon_1), \mathrm{p}, \omega)$ is used,

$$P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}/\{\omega : \omega \in \Omega,\ p \in \bigcup_{i=1}^{n-1} \{a_i(\omega)\}\}) < \varepsilon_0\,.$$

Such an ε_1 and also $N_0(\varepsilon_1)$ depend, of course, on p. The aim of the following theorem is to state the existence of an "average" N_0 depending only on ε and "good enough" from the point of view of the quality of our classification procedure.

Theorem 6.5. There exists, for each $\varepsilon > 0$, such an index $N_0(\varepsilon)$ that for each $N_1 \geqq \geqq N_0(\varepsilon)$ the random variable $T(N_1)$ satisfies the following:

$$(6.13) \qquad \sum_{\bar{p} \in \mathscr{L} - \mathscr{T}} [P(\{\omega : \omega \in \Omega,\ p \in D_n^1(\omega)\}/\{\omega : \omega \in \Omega,\ p \in \bigcup_{i=1}^{n-1} \{a_i(\omega)\}\}) \,.$$

$$. P(\{\omega : \omega \in \Omega,\ a_1(\omega) = \bar{p}\}) < \varepsilon\,.$$

Proof. Cf. Theorem 4, in [1], and its proof.

It can be easily seen that the results presented in this chapter until now will hold also in case the random variable T is substituted by another random variable T' satisfying the demands (1)–(6) and such that the set R_1', defined by T' analoguously to R_1, satisfies Lemma 6.2. There is also another possibility of generalization, namely in such a way that the implications $p_1 \to p$ ($p \to p_2$, resp.) are not investigated for all $p_2 \in D_n^1 \cup D_n^2 (p_2 \in A_n^1 \cup A_n^2$, resp.), but only for some of them, say, chosen at random. Such a model would better describe, in our opinion, the heuristic feature in one's behaviour, when trying to derive a tested formulas from some already known theorems. Let us briefly investigate such a model in the rest of this chapter.

Consider the pretentions of our original classification procedure seen, e.g., from the point of view of the time spare necessary for its performing or from the point of view of the number of some unit operations needed in order to decide about the tested formula (said in other word, the time and space complexity). We can see that these

pretentions increase when the number of the decided formulas increases. If this number is "small enough", the classification procedure may be more appropriate than the simple statistical decision procedure even from the point of view of its pretentions, as the statistical deducibility testing procedure may be rather difficult and the use of the former results may enable to avoid it, at least in some cases. But, when the number of decided formulas increases, the use of all of them begins to be rather impracticable. Let us suggest a solution which would avoid, at least in a degree, this difficulty and let us investigate which of the results stated above and proved in [1] remain to be valid.

Let us modify the classification procedure described above as follows. Before starting with the n-th formula $a_n = a_n(\omega) = p$ to be tested we choose a subset M'_n,

$$M'_n \subset M_n = D_n^1 \cup D_n^2 \cup A_n^1 \cup A_n^2 \supset \bigcup_{i=1}^{n-1} \{a_i(\omega)\}\,, \tag{6.14}$$

such that $card\ M'_n \leqq L_1$, where L_1 is a priori given integer. Let

$$D'_{1,n} = D_n^1 \cap M'_n\,, \quad D'_{2,n} = D_n^2 \cap M'_n\,, \tag{6.15}$$

$$A'_{1,n} = A_n^1 \cap M'_n\,, \quad A'_{2,n} = A_n^2 \cap M'_n$$

and apply the classification procedure from above with the only modification – when looking for an appropriate auxiliary formula q (q', resp.) such that $q \to p$ ($p \to q'$, resp.) is a decidable theorem we do not range over all the formulas from $D_n^1 \cup D_n^2$ ($A_n^1 \cup A_n^2$, resp.) but only over the sets $D'_{1,n} \cup D'_{2,n}$ ($A'_{1,n} \cup A'_{2,n}$, resp.) which play the role of "representants" of the original larger sets. This modification implies that the pretentions of the modified decision procedure are limited and their upper bound is, roughly speaking, a linear function of the parameter L_1. If such an appropriate auxiliary formula is not found inside the set M'_n, the investigated formula p is tested by the statistical deducibility testing procedure. The sets D_{n+1}^1, D_{n+1}^2, A_{n+1}^1, A_{n+1}^2, are constructed according to the result of this decision in the same way as above, but the instructions consisting in erasing some formula or formulas from D_n^1 or A_n^1 are omitted. Also the step (b) is omitted for the reasons which will be shown later.

It can be immediately seen that the properties of such a procedure depend, in a substantial manner, on the way in which the phrase "We choose a subset $M'_n \subset M_n$" will be interpreted. There exist two principial approaches, here, the deterministic one and the statistical one.

The deterministic approach can be formally described by the mean of a mapping G ascribing to every natural n and to every $n + 1$ formulas $a_1, a_2, \ldots, a_n, p$, some subset containing at most L_1 elements from the set $\bigcup_{i=1}^{n} \{a_i\}$. In general, M'_n may depend on p,

but it is necessary lest this dependence should be too complicated. As an example we can give two mappings G_1 and G_2 defined as follows.

$$G_1(\{a_1, a_2, \ldots, a_n, p\}) = \bigcup_{i=1}^{s} \{a_i\}\,, \quad s = \min(n, L_1)\,, \tag{6.16}$$

$$G_2(\{a_1, a_2, \ldots, a_n, p\}) = \bigcup_{i=R}^{n} \{a_i\}\,, \quad R = \max(1, n - L_1 + 1)\,.$$

This means that we use only the first or the last L_1 formulas among $a_1, a_2, \ldots, a_n$.

The statistical approach can be formally described in such a way that the elements of the set M_n' are sampled at random from M_n. The probability space $\langle \Omega, \mathscr{S}, P \rangle$ is considered together with a system $\{p_{ij}\}$, $i = 1, 2, \ldots, j = 1, 2, \ldots, L_1$ of random variables which are mutually independent and such that every p_{ij} takes its values in the set $\{1, 2, \ldots, i\}$ of integers. Now, we set

$$M_n' = M_n'(\omega) = \bigcup_{j=1}^{L_1} \{a^*(p_{nj}^-(\omega))\}\,, \tag{6.17}$$

$$M_n = \{a^*(1), a^*(2), \ldots, a^*(\bar{n})\}\,,$$

clearly, $card\, M_n' \leqq L_1$. In general, the random variables $\{p_{ij}\}$ need not to be independent of the random variables $\{a_j\}$ (sampling the sequence in which formulas are to be tested) or of the random variables $T(N_0, p, \cdot)$ (representing the used statistical deducibility testing procedure) but we shall suppose, in the rest of this chapter, that both of these types of statistical independence take places. Moreover, we suppose the random variables $a_1, a_2, \ldots$ to be independent, equally distributed, and such that

$$P(\{\omega : \omega \in \Omega,\ a_1(\omega) = p\}) > 0$$

iff p is a closed formula (sentence) of the formalized theory $\langle \mathscr{L}, \mathscr{T} \rangle$ in question. The random variables $\{p_{ij}\}$ will be also supposed to be mutually independent, and, for a fixed j, equally distributed, i.e.,

$$P(\{\omega : \omega \in \Omega,\ p_{ki}(\omega) = j\}) = 1/k\,, \tag{6.18}$$

$$k = 1, 2, \ldots\,, \quad j = 1, 2, \ldots, k\,, \quad i = 1, 2, \ldots, L_1\,.$$

In what follows we shall profit of the explanation in [2], where there are two cases of the modified classification procedure investigated separately. In the first case (called Algorithm I) we test, first of all, whether a sampled formula has been already tested, i.e., whether $a_n(\omega) \in M_n$ or not, and if the answer is positive, we find the corresponding one of the four sets D_n^1, D_n^2, A_n^1, A_n^2, in which $a_n(\omega)$ is situated and we put $a_n(\omega)$ again in this set without any further testing. The two or more occurences of the same formula are treated separately, they possess two indices and this makes

the probability of sampling of such a formula into M_n' larger (what is substantial in what follows). In the other case (Algorithm II) we omit this step and every formula is tested in the given way and classified with respect to the result of this test no matter whether it has been or has not been already tested and with which result. Of course, in this case the possibility of finding a formula in two different classes in the same time is not excluded. In [2], the sequence M_n is called the *universal memory* at the step n, the sequence $M_n' \subset M_n$ is called the *instantaneous memory* at the step n. Let us denote

$$\bar{\mathscr{T}} = (\bigcup_{n=1}^{\infty} D_n^2) \cup ((\bigcup_{n=1}^{\infty} (D_n^1 - A_n^1)) - (\bigcup_{n=1}^{\infty} A_n^2)),$$

$$\bar{\mathscr{N}} = (\bigcup_{n=1}^{\infty} A_n^2) \cup ((\bigcup_{n=1}^{\infty} (A_n^1 - D_n^1)) - (\bigcup_{n=1}^{\infty} D_n^2)),$$

$$\bar{\mathscr{A}} = \bigcup_{n=1}^{\infty} (D_n^2 \cup D_n^1 \cup A_n^1 \cup A_n^2).$$

Theorem 6.6. Consider Algorithm I, then

(a) (6.19) $$\bar{\mathscr{T}} \cap \bar{\mathscr{N}} = \emptyset, \quad \bar{\mathscr{T}} \cup \bar{\mathscr{N}} = \bar{\mathscr{A}}$$

(b) $$D_n^i \subset D_{n+1}^i, \quad A_n^i \subset A_{n+1}^i, \quad i = 1, 2, \quad n = 1, 2, \ldots,$$

(c) $$\bigcup_{n=1}^{\infty} D_n^2 \subset \mathscr{T}, \quad \bigcup_{n=1}^{\infty} A_n^2 \subset \mathscr{L} - \mathscr{T}.$$

Proof. Cf. Lemma 2, in [2], and its proof.

Corollary. For the universal memory M_{n+1} the inequality $n^* \leqq card\,(M_{n+1}) \leqq 2n^*$ holds (in the case of Algorithm I as well as Algorithm II), $n^* = n +$ the number of specific axioms set into D_0^2.

Proof. In every step at least one and at most two formulas are joined to one or two classes, i.e., to M_n. Here M_n is taken as a set of occurences of formulas rather than as a set of formulas. Cf. Lemma 3, in [2], and its proof for more details.

Let $p \in \mathscr{L}$ be a closed formula, let $n_0, n, n_0 \leqq n$ be integers. Let $\alpha(p, n_0, n)$ denote the relative frequency of the occurences of p among $a_{n_0}(\omega), a_{n_0+1}(\omega), \ldots, a_n(\omega)$, i.e.,

$$\alpha(p, n_0, n) = \alpha(p, n_0, n, \omega) = card\,(\{i : n_0 \leqq i \leqq n,\ a_i(\omega) = p\}).$$

Denote, moreover,

$$\pi(p) = P(\{\omega : \omega \in \Omega,\ a_1(\omega) = p\}).$$

Lemma 6.3. For every closed formula $p \in \mathscr{L}$ and every integer $n_0 \geqq 1$

(6.20) $$P(\{\omega : \omega \in \Omega,\ \lim_{n \to \infty} n^{-1} \alpha(p, n_0, n) = \pi(p)\}) = 1.$$

Proof. The independence and equal distribution of the random variables $a_1, a_2, \cdots$ imply that the well-known Borel theorem (cf. [5.4] or [5.13]) can be applied. Cf. Lemma 4, in [2] and its proof for more details.

Theorem 6.7. Consider Algorithm I, then

$$P(\{\omega : \omega \in \Omega,\ \mathscr{T} = \bigcup_{n=1}^{\infty} D_n^2(\omega)\}) = 1\,. \tag{6.21}$$

Proof. Cf. Theorem 1, in [2], and its proof.

Corollary. When Algorithm I applied, then every theorem will be, eventually, with the probability 1 proclaimed to be a theorem.

When discussing about the basic motives leading to the modification of the classification procedure investigated at the beginning of this chapter we mentioned namely the impracticably increasing time and space pretentions of this procedure. Considering our Algorithm I from this point of view we must admit that this difficulty has not been completely avoided. Or, the decision instructions of Algorithm I ask to find, whether the tested formula has or has not been tested and classified before, and in the case of the positive answer, to classify this new occurence of the same formula in the same way. Hence, Algorithm I requests, again, to handle with all the formulas which have been already tested and decided – and it is just what we wanted to avoid. It is why we have proposed also Algorithm II to be studied separately; Algorithm II results from the Algorithm I by omitting these decision steps. All other notions and notations keep their former meanings.

Lemma 6.4. Consider Algorithm II, let $q \in \mathscr{T}$ be a theorem. Then there exist an integer $N_0(q) \geqq 1$ and a positive real $c(q)$ such that for each $i \geqq N_0(q)$

$$P(\{\omega : \omega \in \Omega,\ a_i(\omega) \in D_{i+1}^2(\omega)\}/\{\omega : \omega \in \Omega,\ a_i(\omega) = q\}) \geqq c(q)\,. \tag{6.22}$$

Proof. Cf. Lemma 5, in [2], and its proof.

Verbally, this lemma claims that there is, in each case when a theorem is sampled to be tested, a positive and only on the theorem in question depending probability that this theorem will be put into D_{i+1}^2, i.e., that it will be classified as theorem without any danger of error. From Lemma 6.4 almost immediately the following assertion and its corollary can be deduced.

Theorem 6.8. Consider Algorithm II, then

$$P(\{\omega : \omega \in \Omega,\ \mathscr{T} = \bigcup_{n=1}^{\infty} D_n^2(\omega)\}) = 1\,. \tag{6.23}$$

Proof. Cf. Theorem 2, in [2], and its proof.

Corollary. When Algorithm II applied, then every theorem will be, eventually, with the probability 1 proclaimed to be a theorem.

Let q be a theorem, let $q_1, q_2, \ldots, q_n = q$ be a formalized proof of q. Analyzing the proof of Lemma 5 in [2] we can say that the assertion of this Lemma will hold if

$$c(q) = \left(\prod_{i=1}^{n-1} \pi(q_i)^{a_i}\right)\left(\tfrac{1}{2}\right)^{n-1}(1-\varepsilon)^{\sum_{i=1}^{n-1} d_i}, \tag{6.24}$$

where $\varepsilon > 0$ is defined in Theorem 6.4 and d_i is a positive integer showing how many times the formula q_i is used as an antecedent in order to deduce some q_j, $j > i$. Of course, the value $c(q)$ depends not only on q but also on the proof $q_1, q_2, \ldots, q_n$.

When analyzing the proof of Theorem 1 in [2] we obtain that in the case of Algorithm I an assertion analogous to that of Lemma 6.4 would hold if

$$c(q) = \tfrac{1}{2}\,\pi(q_i)\,\pi(q_j)\,(1-\varepsilon)^2,$$

in case q followed from q_i, q_j by the modus ponens rule, or

$$c(q) = \tfrac{1}{2}\,\pi(q_i)\,(1-\varepsilon),$$

in case $q_i \to q$ is a theorem decidable by the deterministic theorem prover being at our disposal and serving as a part of statistical theorem prover $T(N_0, p, \cdot)$, see condition (7) at the beginning of this chapter.

The value $(\pi(q)\,c(q))^{-1}$ can serve as an upper bound for the conditional expected value of the number of steps, which are necessary for joining the theorem q with $\bigcup_{n=1}^{\infty} D_n^2$. We can say that in the case of Algorithm I this value is given by the inverse value of the probability of sampling the premise necessary for the immediate deriving of q (or by the inverse value of the product of these probabilities if there are two premises). The length and complexity of the proof $q_1, q_2, \ldots, q_n$ in its whole do not play any role. However, when considering Algorithm II, this conditional expected value is given by the inverse value of the product of the probabilities of sampling for all the formulas occuring in the considered proof. It follows immediately that this later expected value is much more greater than the former one and depends on the complexity and length of the considered proof in its whole. This greater speed of Algorithm I is caused by the fact that this procedure tries whether a formula submitted for testing has been already tested or not and in the positive case uses this information. Algorithm II is not endowed with this ability because of the reasons explained above.

Hence, we can say that even Algorithm II provides that every theorem will be, eventually, with the probability 1 proclaimed to be a theorem, of course, with the average number of steps much more greater than in the case of Algorithm I. This fact is caused by a common feature of both these Algorithms, namely by the fact that each formula is given, in each step, a positive probability to be sampled for

testing and tested, no matter whether this formula has been already tested or has not been tested yet. This repeating of testing seems to be quite natural supposing the tested formula was joined with D_n^1 or A_n^1. Such a decision is not necessarily correct, there is a probability of error, so it is quite reasonable to have a possibility of revoking our former decision by joining this formula with some D_m^2 or A_m^2, $m > n$.

We shall see, however, that even to repeat a decision on a formula having been already correctly decided and joined with D_n^2 or A_n^2 is of some worth. In the following theorem the situation is investigated, when a formula, having been once tested and joined with D_n^2 or A_n^2 is no more sampled and tested again. This theorem shows that in such a case neither Algorithm I nor Algorithm II assures that every theorem will be, eventually, proclaimed to be a theorem. An upper bound for the probability of this event, introduced below, tends to 0 if the number of applications of the modus ponens rule, necessary for deduction of a considered theorem increases. Or, when a formula may be joined with D_n^1 or A_n^1 only once, i.e., there is at most one occurence of this formula in the universal memory, then the probability of sampling this formula into the instantaneous memory tends to zero in a linear proportion to the increasing number of sampled and tested formulas. However, in an application of the modus ponens rule is necessary to deduce a conclusion, both the necessary premises have to meet each other in the instantaneous memory and the probability of such an event tends to zero in a quadratic proportion to the increasing number of sampled and tested formulas, hence, there is a positive probability that this random event will never occur.

Theorem 6.9. Consider Algorithm I or Algorithm II. Let the random variables $a_1, a_2, \ldots$ satisfy the following condition

$$P(\{\omega : \omega \in \Omega,\ a_i(\omega) = p\}) = 0\,, \quad \text{if} \quad p \in D_i^2 \cup A_i^2\,,$$

$$P(\{\omega : \omega \in \Omega,\ a_i(\omega) = p\}) > 0\,, \quad \text{if} \quad p \in \mathscr{L} - (D_i^2 \cup A_i^2)\,,$$

p is a closed formula. Let q be a theorem, let k be such an integer that there are at least k applications of the modus ponens rule in every proof of q from axioms and those theorems which are deterministically decidable (i.e., for which $T(N_0, p, \omega) \equiv 2$). Then

$$P(\{\omega : \omega \in \Omega,\ q \in \bigcup_{n=1}^{\infty} D_n^2(\omega)\}) \leqq \tag{6.25}$$

$$\leqq \prod_{i=3}^{k+2} \sum_{j=i}^{\infty} \left(1 + \left(1 - \frac{2}{j}\right)^{L_1} - 2\left(1 - \frac{1}{j}\right)^{L_1}\right),$$

$$P(\{\omega : \omega \in \Omega,\ q \in \bigcup_{n=1}^{\infty} D_n^2(\omega)\}) \leqq \tag{6.26}$$

$$\leqq \min\left\{1, \tfrac{1}{2}L_1(L_1 - 1))^{k-(1/2)L_1(L_1-1)} \cdot \frac{1}{(k+2)!} \cdot (\tfrac{1}{2}(L_1 - 1)\, L_1)!\right\}.$$

These two upper bounds are not, in general, mutually comparable. Let us recall that L_1 denotes the upper bound for the cardinality of the instantaneous memory and let us emphasize the fact that in this case the random variables $a_1, a_2, \ldots$ are neither statistically independent nor equally distributed.

Proof. Cf. Theorem 3, in [2], its corollary and the corresponding proofs.

The theorems introduced above were dealing rather with the asymptotic properties of the two investigated algorithms, i.e., properties given by the sets $\bigcup_{n=1}^{\infty} D_n^2$, $\bigcup_{n=1}^{\infty} A_n^2$, etc. However, each actual realization of Algorithm I or Algorithm II should be stopped after a finite number of steps. The following theorem, the last in this chapter, offers an upper bound for the probability of proclaiming a non-theorem to be a theorem when classifying all formulas from $\mathscr{F}$ as theorems.

Theorem 6.10. Consider Algorithm I or Algorithm II. Let a real $\varepsilon > 0$ and an integer $N_0 \geqq 1$ satisfy the condition

$$P(\{\omega : \omega \in \Omega,\ T(N_0, q, \omega) = 1\}) \leqq \varepsilon$$

uniformly for all non-theorems q. Then

$$P(\{\omega : \omega \in \Omega,\ q \in \mathscr{T}(\omega)\}) \leqq \varepsilon \tag{6.27}$$

uniformly for all non-theorems. If, moreover, for a non-theorem q such an integer n and non-theorems $q_1, q_2, \ldots, q_n$ exist, that $q_n = q$, $T(N_0, q_i \to q_{i+1}, \omega) \equiv 2$, $i = 1, 2, \ldots, n-1$, $T(N_0, q_1, \omega) \equiv -2$, then

$$P(\{\omega : \omega \in \Omega,\ q \in \mathscr{T}(\omega)\}) = 0\,. \tag{6.28}$$

Proof. Cf. Theorem 4, in [2], and its proof.

This theorem gives an answer to the question which could perhaps arise when the two algorithms are investigated, namely, why the formulas from $D'_{1,n}$ are not used as premises when testing a formula $p = a_n(\omega)$, i.e., why we do not investigate, whether there exists or does not exist a formula $q \in D'_{1,n}$ such that $T(N_0, q \to p, \omega) \equiv 2$. But, if we looked for such a formula and if we joined p with $D_{1,n+1}$ if this was the case, then the probability of error connected with this decision would be, in general, greater than the probability of error connected with the simple statistical deducibility testing of q and this probability of error could cummulate in such a way that no acceptable function of ε could serve as an upper bound for this probability. In this feature our last algorithms differ from the classification procedure investigated at the beginning of this chapter where such an upper bound existed. It is why we have omitted this case and we add a formula to D_n^1 only in case the statistical deducibility testing procedure decides in this way, i.e., proclaims the tested formula to be a theorem.

REFERENCES

[1] I. Kramosil: Statistical Estimation of Deducibility in a Random Sequence of Formulas. Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes. Academia, Prague 1973, pp. 449–463.

[2] I. Kramosil: A Statistical Model for Theorem Proving with a Limited Instantaneous Memory. Proceedings of the 9. European Meeting of Statisticians, Colloquia Mathematica Societatis János Bolyai, Budapest 1972, pp. 425–453.

7. OTHER STATISTICAL APPROACHES TO DEDUCIBILITY TESTING

Having presented, in the two foregoing chapters, one possibility how to apply probability theory and mathematical statistics in the domain of theorem proving, we would not like to claim that this approach is the only possible or the best one. From time to time we can meet some ideas or propositions which concern the uncertainty and approximations in deduction processes or at least can be understood in such a way. This chapter is devoted to an overview of those among such ideas which have been already formalized on mathematical level comparable with that accepted here. Let us start with the concept of probabilistic cannonical systems (or calculi) introduced by S. Ju. Maslov and E. D. Rusakov in [7].

The problem solving for a large class of creative tasks can be converted into deducibility testing in an appropriate, general enough, system (cf. [6] and the next chapter of this work). As a system of a general type we can take, e.g., the well-known Post system (cf. [9] or [3]); this system has arisen as a straightforward generalization of the notion of formalized theory and its basic features should be clear from what follows; at least in the extent necessary for our purposes.

When searching for a proof in a relatively complicated system, the main difficulty is connected with the necessity how to organize a sufficiently exhaustive search in a large space of possible proofs. In order to restrict this searching complexity we can adopt some reglementations, i.e., searching strategies. Another possibility, which will be discussed below, consists in resignation to completeness, i.e., we give up the request that in all cases when a proof of desired type exists it should be, eventually, discovered and we would be satisfied, if the desired proof were discovered, under the condition that it exists, with at least such and such apriori given probability. The methods and results explained below are, in a sense, close to that used and proved in the theory of probabilistic algorithms (cf. [10] or [8] for more details). The restriction of our considerations to the case of one-premise deduction rules is not so strong as it may seem; remember the so called First Reduction in [9] or § 3 in [6].

Let $\mathscr{K}$ be a canonical system over an alphabet A (words in A are called A-words), with m one-premise (or unary) deduction schemas $\pi_1, \pi_2, \ldots, \pi_m$ and n axioms $a_1, a_2, \ldots, a_n$, let the result of application of each schema to each axiom be defined

and be defined unambiguously (this is just a matter of technical convenience, see the normal systems in [9]). $b_1, b_2, \ldots, b_n(\pi) \vdash b$ means that b is the result of the application of π to $b_1, \ldots, b_n$; $b_1, b_2, \ldots, b_n \vdash b$ means that there exists π_j such that $b_1, \ldots, b_n(\pi_j) \vdash b$.

Definition 7.1. A sequence

$$a_1, a_2, \ldots, a_n, \quad a_{n+1}\langle k_1, s_1\rangle, \ldots, a_{n+r}\langle k_r, s_r\rangle \tag{7.1}$$

is called an *analyzed proof in* $\mathscr{K}$, if $r \geqq 0$, $a_1, a_2, \ldots, a_{n+r}$ are A-words, and

(1) for all i, $1 \leqq i \leqq r$ implies $1 \leqq s_i \leqq m$, and $1 \leqq k_i < n + i$,

(2) $a_{k_i}(\pi_{s_i}) \vdash a_{n+i}$ (i.e., $a_{n+i} = a_{k_i}$, if π_{s_i} is not applicable, in the usual sense, to a_{k_i}).

The sequence $a_1, a_2, \ldots, a_n, a_{n+1}, \ldots, a_{n+r}$ is called *dual* to the analyzed proof in question and its length will be denoted by $d(\xi)$ (letters ξ, η, etc. serve to denote analyzed proofs).

Let us ascribe, to each i, $1 \leqq i \leqq k$, and to each j, $1 \leqq j \leqq m$, non-negative real numbers q_i, p_j, r_{ij}, such that $\sum_{i=1}^{k} \sum_{j=1}^{m} r_{ij} = 1$, $\sum_{j=1}^{m} r_{ij} = q_i$, $\sum_{i=1}^{k} r_{ij} = p_j$ for all $i \leqq k$, $j \leqq m$. The number r_{ij} can be interpreted as the probability of sampling of a pair $\langle a_i, \pi_j\rangle$, q_i as the probability of sampling of a_i, and p_j as the probability of sampling of π_j. These numbers can be ordered in an $(k + 1) \times (m + 1)$ – tabel T, it is why we shall write sometimes $\{T\}_{ij}$ instead of r_{ij}.

Definition 7.2. *Probabilistic system* is an ordered pair $\langle \mathscr{K}, \mathscr{A}\rangle$, where $\mathscr{K}$ is a canonical system of the type described above and $\mathscr{A}$ is an algorithm applicable to each analyzed proof ξ of $\mathscr{K}$ and ascribing to ξ a tabel T as defined above with a_1, $a_2, \ldots, a_k$ being the sequence dual to ξ. Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A}\rangle$ be a probabilistic system, let us define, for each analyzed proof ξ in $\mathscr{K}$ its probability $p[\xi]$ as follows.

(1) if $\xi = a_1, a_2, \ldots, a_n$, then $p[\xi] = 1$;

(2) if $\xi = \eta$, $a\langle k, s\rangle$, then $p[\xi] = p[\eta] \cdot \{\mathscr{A}(\eta)\}_{k,s}$. For each proof B in $\mathscr{K}$ set

(3) $M_B = \{\xi : \xi$ is an analyzed proof for which B is its dual sequence$\}$;

(4) $p[B] = \sum_{\xi \in M_B} p[\xi]$.

For each A-word a and each $k \geqq n$ denote by $M^k(a)$ the set of all proofs in $\mathscr{K}$ of the length k which contain a; denote by M^k the set of all proofs of the length k in $\mathscr{K}$. If $M \subset M^k$, set $p[M] = \sum_{B \in M} p[B]$. If B is a proof of the length l in $\mathscr{K}$, denote by H_B^k, $k \geqq 0$, the set of all proofs of the length $l + k$ which begin with B.

Theorem 7.1.

(1) Let M_1, M_2 be sets of proofs (in $\mathscr{K}$) of the same length, let $M_1 \cap M_2 = \emptyset$, then $p[M_1] + p[M_2] = p[M_1 \cup M_2]$.

(2) For each $k \geqq 0$ and each B, $p[H_B^k] = p[B]$; $p[M^{n+k}] = 1$.

(3) For each $k \geqq n$ and each A-word a, $p[M^{k+1}(a)] \geqq p[M^k(a)]$.

Proof. The assertions follow immediately from the definitions above.

Denote, for each δ real,

(7.2) $\mathscr{M}_\delta = \{a : a$ is A-word and there exists a $k \geqq n$ such that $p[M^k(a)] > \delta\}$.

The following assertion is an analogy of the corresponding theorem for probabilistic Turing Machines (cf. [5], where also a proof can be found).

Theorem 7.2. The set $\mathscr{M}_\delta$ is enumerable for each real δ.

Definition 7.3. Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ be a probabilistic system, let Ψ be a total recursive function such that, for all $k \geqq 0$, $\Psi(k+1) \geqq \Psi(k)$, let $\Psi(k) \to \infty$, if $k \to \infty$. Algorithm $\mathscr{A}$ is called *Ψ-regular*, if for each analyzed proof ξ of the form (7.1) and for each k, s, $1 \leqq k \leqq \min\{\Psi(d(\xi)), d(\xi)\}$ $1 \leqq s \leqq m$, the following holds: if $\{\mathscr{A}(\xi)\}_{k,s} = 0$, then there exists k', $1 \leqq k' \leqq d(\xi)$, such that either $a_{k'} = a_k$ and $\{\mathscr{A}(\xi)\}_{k',s} > 0$, or $a_k(\pi_s) \vdash a_{k'}$. If, moreover, for each l, $\Psi(l) \geqq l$, then the algorithm $\mathscr{A}$ is called *correct* (a correct algorithm is Ψ-correct for all Ψ).

Theorem 7.3. Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ be a probabilistic system, let $\mathscr{A}$ be a Ψ-regular algorithm. If a is derivable in $\mathscr{K}$, then there exists $k \geqq n$ such that $p[M^k(a)] > 0$ (hence, $\mathscr{M}_0$ is identical with the set of all words derivable in $\mathscr{K}$).

Proof. The assertion is an immediate consequence of Definition 7.3.

Definition 7.4. Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ be a probabilistic system, let φ be a non-negative real-valued function defined for each natural l. We say, that the *algorithm $\mathscr{A}$ has the minorant φ*, if for each analyzed proof ξ of the length l and for each k, s, $1 \leqq k \leqq d(\xi)$, $1 \leqq s \leqq m$, the following holds: if $\{\mathscr{A}(\xi)\}_{k,s} > 0$, then $\{A(\xi)\}_{k,s} \geqq \varphi(l)$. The minorant φ is called *substantial*, if $\sum_{i=1}^{k} \varphi(i) \to \infty$ for $k \to \infty$.

Theorem 7.4. Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ be a probabilistic system, let $\mathscr{A}$ be a Ψ-correct algorithm with a substantial minorant φ, then for each word a, derivable in $\mathscr{K}$, $p[M^k(a)] \to 1$, if $k \to \infty$. I.e., for each $\delta < 1$, $\mathscr{M}_\delta$ is identical with the set of words derivable in $\mathscr{K}$.

Proof. Cf. the proof of Theorem 3 and the corresponding lemma in [7].

The algorithm $\mathscr{A}$ is called *equiprobable*, if for each analyzed proof ξ and each k_1, k_2, s_1, s_2, $1 \leqq k_1, k_2 \leqq d(\xi)$, $1 \leqq s_1, s_2 \leqq m$, the following holds: $\{\mathscr{A}(\xi)\}_{k_1,s_1} = \{\mathscr{A}(\xi)\}_{k_2,s_2}$.

Theorem 7.5. (Corollary of Theorem 7.4). Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ be a probabilistic system, let $\mathscr{A}$ be a correct equiprobable algorithm, then for all a derivable in $\mathscr{K}$, $p[M^k(a)] \to 1$, if $k \to \infty$ (as an equiprobable algorithm has always the substantial minorant $\varphi(m) = (km)^{-1}$).

Let us introduce some special types of probabilistic systems. Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ be a probabilistic system. If for each ξ of the type (7.1.) and for each $i, j \leqq d(\xi)$, $\{\mathscr{A}(\xi)\}_{i,j} = 0$ holds iff the schema π_{s_i} is not applicable to a_{k_i} (in the usual sense), then $\mathscr{Q}$ is called *context-free*. If each element of $\mathscr{A}(\xi)$ is positive, then $\mathscr{Q}$ is called *strongly context-free*. If $\{\mathscr{A}(\xi)\}_{i,j} = 0$ just in the cases when there is $i' < i$ such that $a_i = a_{i'}$, then $\mathscr{Q}$ is called *conservative*. If (1) for all i, $1 \leqq i \leqq r$, $\{\mathscr{A}(\xi)\}_{k_i,s_i} = 0$ and, moreover, all elements from $\mathscr{A}(\xi)$, which are not subjected to (1), satisfy the condition for conservative systems, then $\mathscr{Q}$ is called (*conservative*) *system with memory*. All these four types of probabilistic systems have correct algorithms and can be easily generalized in the case of probabilistic systems with Ψ-correct algorithms.

Let $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ be such a probabilistic system that for each ξ of the type (7.1) and each pair $\langle a_i, \pi_j \rangle$ the probability ascribed to this pair by $\mathscr{A}(\xi)$ equals the product of the probability of sampling a_i in $\mathscr{A}(\xi)$ by the probability of sampling π_j in $\mathscr{A}(\xi)$. Then $\mathscr{Q}$ is called *system with independent probabilities*. Moreover, such a system is called *schema-constant*, if the probability distribution on schemas is the same in all tables. Algorithm $\mathscr{A}$ of a probabilistic system $\mathscr{Q} = \{\mathscr{K}, \mathscr{A}\rangle$ is called *word-constant*, if for each ξ, η such that the number of positive elements in $\mathscr{A}(\xi)$ and $\mathscr{A}(\eta)$ is the same, the following condition holds: for each i, j, $1 \leqq i \leqq \min(d(\xi), d(\eta))$, $1 \leqq j \leqq m$, if $\{\mathscr{A}(\xi)\}_{i,j} > 0$ and $\{\mathscr{A}(\eta)\}_{i,j} > 0$, then $\{\mathscr{A}(\xi)\}_{i,j} = \{\mathscr{A}(\eta)\}_{i,j}$.

Algorithm $\mathscr{A}$ of a probabilistic system $\langle \mathscr{K}, \mathscr{A} \rangle$ is called *failure-stabile*, if for each ξ of the type η, $a\langle k, s\rangle$, for which the probability of sampling a in $\mathscr{A}(\xi)$ equals 0 and for all i, j, $1 \leqq i \leqq d(\eta)$, $1 \leqq j \leqq m$, $\{\mathscr{A}(\xi)\}_{i,j} = \{\mathscr{A}(\eta)\}_{i,j}$.

Let $\mathscr{K}$ be a Post canonical system, let a be a word derivable in $\mathscr{K}$. Denote $\mathscr{L}_a = \{b$: there exists a proof of b in $\mathscr{K}$, not containing $a\}$, $\mathscr{K}_a = \{b$: there exists a proof of b in $\mathscr{K}$ such that for all b' from this proof $\neg(b' \vdash a)$ holds$\}$.

Let $\mathscr{K}$ be a canonical system, let a be a word derivable in $\mathscr{K}$, which is not an axiom, let ε be a positive real. Consider the problem, whether it is possible to construct probabilistic systems $\mathscr{Q} = \langle \mathscr{K}, \mathscr{A} \rangle$ such that, for all $k \geqq n$, $p[M^k(a)] \leqq \varepsilon$.

Theorem 7.6. Let there exist i_0 and b such that $1 \leqq i_0 \leqq n$, $a_{i_0} \vdash b$, $b \neq a$, then it is possible to construct

(1) a context-free $\mathscr{Q}$ with word-constant and failure-stabile algorithm,

(2) a strongly context-free $\mathcal{Q}$ with the same algorithm,

(3) conservative $\mathcal{Q}$.

If, moreover, $a_{i_0} \vdash a$ does not hold, then all these probabilistic systems can be constructed as systems with independent probabilities. Under the supplementary condition that $\mathscr{Z}_a$ is infinite, we can construct (1) context-free $\mathcal{Q}$, (2) strongly context-free $\mathcal{Q}$, (3) conservative $\mathcal{Q}$, and (4) $\mathcal{Q}$ with memory, in all these cases the algorithms can be chosen to be word-constant and failure-stabile. If $\mathscr{K}_a$ is infinite, then the first three algorithms can be construct in such a way that the corresponding probabilistic systems are, moreover, systems with independent probabilities.

Proof. Cf. Theorem 4 in [7]. Algorithms of the systems with independent probabilities mentioned in Theorem 7.6. are schema-constant and the probability distribution on schemas can be given a priori. All the conditions can be proved to be necessary.

The model explained above can be generalized to the case with many-premise deduction schemas. Consider a canonical system $\mathscr{K}$ with axioms $a_1, a_2, \ldots, a_n$ and deduction schemas $\pi_1, \pi_2, \ldots, \pi_m$ such that π_i requests δ_i premises. Let us begin with the situation when the result of application of each schema to each ordered sequence of premises (of appropriate length) is uniquely defined.

Each sequence ξ of the form

$$a_1, a_2, \ldots, a_n, \tag{7.3}$$

$$a_{n+1} < k_{1,1}, \ldots, k_{1,l_1}; s_1\rangle, \ldots, a_{n+r}\langle k_{r,1}, \ldots, k_{r,l_r}; s_r\rangle,$$

where $r \geqq 0$, $a_1, a_2, \ldots, a_r$ are A-words and

(1) for each i, j, $1 \leqq i \leqq r$, $1 \leqq j \leqq l_r$ holds: $1 \leqq s_i \leqq m$ and $1 \leqq k_{i,j} \leqq n + i$ and $l_i = \delta_{s_i}$,

(2) either $a_{k_{i,1}}, \ldots, a_{k_{i,l_i}}(\pi_{s_i}) \vdash a_{n+i}$, or $a_{n+i} = a_{k_{i,1}}$, if π_{s_i} is not applicable to $a_{k_{i,1}}, \ldots, a_{k_{i,l_i}}$ (in the usual sense)

is called *analyzed proof* in $\mathscr{K}$.

Let $\gamma_1, \ldots, \gamma_{m'}$, $m' \leqq m$, be the sequence of all numbers from $\delta_1, \delta_2, \ldots, \delta_m$ without repetitions. Choose an algorithm which ascribes, to each finite sequence of A-words, the list of all possible γ_1-tuples, γ_2-tuples, ..., $\gamma_{m'}$-tuples of words from this sequence. For each ξ of the type (7.3) the list ascribed by our algorithm to the sequence $a_1, \ldots, a_n, a_{n+1}, \ldots, a_{n+r}$ will be called *$\mathscr{K}$-dual to ξ*. We suppose that in the list $\mathscr{K}$-dual to ξ all sequences containing only the words $a_1, \ldots, a_k$, $k < n + r$, precede all sequences containing at least one of the words $a_{k+1}, \ldots, a_{k+r}$. The number of sequences in the list which is $\mathscr{K}$-dual to ξ is denoted by $d_{\mathscr{K}}(\xi)$.

Using these notions we can easily generalize the notion of probabilistic tabel and probabilistic system, also the probabilities of analyzed proofs and proofs are defined

as above. An element of such a tabel is called *fictive*, if it corresponds to a schema with γ premises and, simultaneously, to a sequence of γ' words with $\gamma' \neq \gamma$, the fictive elements of a tabel are supposed to be zero. All conditions and restrictions occurring in definitions of various types of algorithms are supposed to be related only to non-fictive elements.

Now, Theorems 7.2 to 7.4 remain to be valid in their verbal form. Let us introduce an example.

Consider the system with alphabet $\{(\ ,\ )\}$, axiom $(\ )$ and only deduction schema $p,\ q \vdash (p, q)$. Construct a correct equiprobable algorithm which ascribes, to each ξ of the length k, the tabel, in which the probability of all pairs, already used in ξ, equals 0 and for all other pairs it equals $(k^2 - k + 1)^{-1}$. The obtained system will be a conservative one with memory; in a natural sense this algorithm is the best correct equiprobable algorithm. The word $((\ )\ (\ ))$ will be generated by this probabilistic system with the probability 1, however, taking the word $((\ )\ ((\ )\ (\ )))$, there is a positive probability, namely

$$\prod_{n=2}^{\infty} \frac{n^2 - n}{n^2 - n + 1} > \frac{2}{3} \cdot \prod_{k=2}^{\infty} \frac{k^2 - 1}{k^2} = \frac{1}{3},$$

that this word will never be derived.

Let us briefly mention the case of ambiguous deduction rules, i.e., the case when a deduction schema, applied to a particular appropriate sequence of premises can give more than one result (the Post's restrictions imposed to the form of possible deduction schemata guarantee that the number of such possible results will be finite). Ambiguous deduction rules can be eliminated by defining the so called "working" or "operating zones" in words and joining some new deduction schemata for appropriate transpositions of these zones. This method is well-known and often used, e.g., in the theory of normal algorithms (cf. [2]).

Such an inclusion needs several small changes. First, we add to the definition of analyzed proof this condition: for all i, $1 \leqq i \leqq r$, if $b_1, \ldots, b_k$ is the list (without repetitions and in the lexicographical order) of all words derivable from $a_{k_i,1} \ldots \ldots, a_{k_i,l_i}$ by an application of π_{s_i} then either $k = 0$ and $a_{n+i} = a_{k_i,1}$, or $k > 0$ and there exist i', $0 \leqq i' \leqq r - k$, such that for all j, $1 \leqq j \leqq k$, $b_j = a_{i'+j}$. Now, in the point (2) of Definition 7.2 we must prolong η by a finite number of words. Let us remark, that considering the definition of Ψ-regularity (and some other notions), the second disjunctive member in this definition guarantees, that we can find in ξ not only a'_k, but also other words derivable from the same premises by the same schema. As can be easily seen, Theorems 7.2 to 7.4 remain to be valid, however, Theorem 7.5 (a consequence of Theorem 7.4 in the case of equiprobable correct algorithms) need not to be valid neither for multiple-premised nor for the ambiguous deduction schemata.

The model of Maslov and Rusakov as explained above can be seen as a kind of machine which generates new and new theorems according to some probabilistic laws, and from this point of view the model is close to that of experience-based statistical deducibility testing as explained in Chapter 6. Now, let us briefly describe the method of statistical theoremhood testing proposed by S. C. van Westrhenen in [12] which is, in a sense, more close to the model explained in Chapter 5. For the sake of simplicity we begin with the propositional calculus.

Let $r, k, n_1, n_2, \ldots, n_k$ be natural numbers, let $p_1, p_2, \ldots$ be the sequence of all propositional indeterminates. Define

$$K(n_1, n_2, \ldots, n_k, r) = \{F : F = \bigwedge_{i=1}^{k} \bigvee_{j=1}^{n_i} \alpha_{ij}\}, \tag{7.4}$$

where $\alpha_{ij} \in A_r = \{p_1, \neg p_1, p_2, \neg p_2, \ldots, p_r, \neg p_r\}$, i.e. $K(n_1, \ldots, n_k, r)$ is the set of all propositional formulas in conjunctive normal form (sets of clauses, in terms of Chapter 3), in which only first r indeterminates and their negations may occur.

If there are, for each $i \leqq k$, such indices $j_1(i), j_2(i) \leqq n_i$, that $\alpha_{ij_1(i)}$ is $\neg\alpha_{ij_2(i)}$, or vice versa, then clearly $\bigvee_{j=1}^{n_i} \alpha_{ij}$ as well as $\bigwedge_{i=1}^{k} \bigvee_{j=1}^{n_i} \alpha_j$ are theorems (of the propositional calculus). Van Westrhenen studies and numerically solves the following problems:

(i) The determination of the probability that a formula, sampled from $K(n_1, \ldots, n_k, r)$ by the uniform probability distribution is provable;

(ii) for a given real number $\varepsilon > 0$ we shall determine a natural number $N(\varepsilon) \leqq k$ such that an at random sampled formula F of the form (7.4) (sampled by the uniform probability distribution) will be estimated provable, with the probability of error smaller than ε, if at least $N(\varepsilon)$ members of the conjunction contain at least one propositional indeterminate together with its negation.

Consider a triple sequence $\{X_{ijr}\}$, $i, j, r = 1, 2, \ldots$ of mutually independent random variables, defined on a probability space $\langle \Omega, \mathscr{S}, P \rangle$, taking their values, for each r, in $\mathscr{A}_r$ and such that $P(\{\omega : \omega \in \Omega, X_{i,j,r}(\omega) = \alpha\}) = (2r)^{-1}$ for each $\alpha \in \mathscr{A}_r$ and each $i, j, r = 1, 2, \ldots$ As an abbreviation we also introduce the random variables $C_{ir} = X_{i1r} \vee X_{i2r} \vee \ldots \vee X_{in_ir} = \bigvee_{j=1}^{n_i} X_{ijr}$, clearly,

$$P(\{\omega : \omega \in \Omega, C_{ir}(\omega) = \alpha_1 \vee \ldots \vee \alpha_{n_i}\}) = \tag{7.5}$$

$$= \prod_{j=1}^{n_i} P(\{\omega : \omega \in \Omega, X_{ijr}(\omega) = \alpha_j\}),$$

$\alpha_j \in \mathscr{A}_r$, $j = 1, 2, \ldots, n_i$. The random sample od an element from $K(n_1, \ldots, n_k, r)$ is formalized by the random variable.

$$\varphi_{k,n_1,\ldots,n_k,r} = \bigwedge_{i=1}^{k} \bigvee_{j=1}^{n_i} X_{ijr}. \tag{7.6}$$

The probability distribution of $\varphi_{k,n_1,\ldots,n_k,r}$ easily follows from the definition, i.e.,

$$P(\{\omega : \omega \in \Omega, \varphi_{k,n_1,\ldots,n_k,r}(\omega) = \bigwedge_{i=1}^{k} \bigvee_{j=1}^{n_i} \alpha_{ij}\}) =$$

$$\prod_{i=1}^{k} P(\{\omega : \omega \in \Omega, C_{ir}(\omega) = \bigvee_{j=1}^{n_i} \alpha_{ij}\}) = (1/2r)^{n_1+n_2+\ldots+n_k}$$

for each $\bigwedge_{i=1}^{k} \bigvee_{j=1}^{n_i} \alpha_{ij} \in K(n_1, \ldots, n_k, r)$.

A conjunction member (clause) of a formula $F \in K(n_1, \ldots, n_k, r)$ is called closed, iff it contains a propositional indeterminate together with its negation. Hence, F is provable (F is a theorem) iff all its clauses are closed. Let w be the characteristic function of the set of all theorems from $K(n_1, \ldots, n_k, r)$. Set

$$K_0(n_1, n_2, \ldots, n_k, r) = \{F : F \in K(n_1, \ldots, n_k, r), w(F) = 0\}, \tag{7.7}$$

$$K_1(n_1, n_2, \ldots, n_k, r) = \{F : F \in K(n_1, \ldots, n_k, r), w(F) = 1\}.$$

Consider the so called Stirling numbers of the second kind, denoted by $S(n, j)$ and defined by

$$S(n, j) = \frac{1}{j!} \sum_{m_1+\ldots+m_j=n, m_i>0, 1 \leqq i \leqq j} \left(\frac{n!}{m_1! \ldots m_j!}\right),$$

$$n = 1, 2, \ldots, j = 1, 2, \ldots, n,$$

(cf., e.g., [4], for more details about these numbers).

Theorem 7.7.

$$P(\{\omega : \omega \in \Omega, \varphi_{1nr}(\omega) \in K_0(n, r)\}) = N(n, r)/(2r)^n, \tag{7.8}$$

where

$$N(n, r) = \sum_{1 \leqq j \leqq \min(n,r)} S(n, j) \cdot (r)_j \cdot 2^j, \quad (r)_j = j! \binom{r}{j}.$$

Proof. Cf. Theorem 2.1, in [12], and its proof.

Theorem 7.8. Denote $p(n, r) = N(n, r)/(2r)^n$, then

(1) $\lim_{n \to \infty} p(n, r) = 0$ monotonically for each r.

(2) $\lim_{r \to \infty} p(n, r) = 1$ for each n.

Proof. Cf. Theorem 2.2 for (1) and Theorem 2.3 for (2), in [12], and their proofs

Theorem 7.9.

$$(7.9)\qquad P(\{\omega : \omega \in \Omega,\ \varphi_{k,n_1,\ldots,n_k,r}(\omega) \in K_1(n_1, \ldots, n_k, r)\}) = \prod_{j=1}^{k} (1 - p(n_j, r)),$$

verbally, a formula sampled at random and with respect to the uniform probability distribution from $K(n_1, \ldots, n_k, r)$ is provable with probability $\prod_{j=1}^{k} (1 - p(n_j, r))$.

Proof. Cf. Theorem 2.4, in [12], and its proof.

Clearly, when increasing n_i (the length of the i-th sequence) simultaneously for each i, the probability of sampling a theorem tends to 1, when r increases, this probability tends to zero. Both these facts seem to be quite intuitive.

The trivial decision procedure for a formula F, $F \in K_1(n_1, \ldots, n_k, r)$ is the inspection of the 1st, 2nd, ..., k-th clause for a closure. In order to avoid sets without provable formulas or formulas with disjunctions consisting of one propositional indeterminate, we shall assume $n_i \geqq 2$, $i = 1, 2, \ldots, k$. This means that if F is sampled at random from $K(n_1, \ldots, n_k, r)$ by the uniform probability distribution, then the values of the stochastic variables X_{ijr}, $j = 1, 2, \ldots, n_i$ of the clauses C_{ir}, $i = 1, 2, \ldots, k$, are inspected for a closure.

The procedure may be altered in such a way that not all the values of the random variables X_{ijr} of the clause C_{ir} are inspected but only the first s_i, $2 \leqq s_i \leqq n_i$, $i = 1, 2, \ldots, m$, $m \leqq k$; therefore we define $C'_{ir} = X_{i1r} \vee X_{i2r} \vee \ldots \vee X_{is_ir}$, $i = 1, 2, \ldots, k$.

This means that the sampled formula is estimated provable, if the values of the first s_i propositional indeterminates of the i-th clause contain a pair A_t, $\neg A_t$ for $i = 1, 2, \ldots, m$. It is clear that in this case an error can be made. The formal description of the estimation procedure will be given by a random variable $h_{ms_1\ldots s_mr}$ with two possible values 0,1 and such that

$$(7.10)\qquad P(\{\omega : \omega \in \Omega,\ h_{ms_1\ldots s_mr}(\omega) = 1\}) = \prod_{i=1}^{m} P(\{\omega : \omega \in \Omega,\ w(C'_{ir})(\omega) = 1\}),$$

$$P(\{\omega : \omega \in \Omega,\ h_{ms_1\ldots s_mr}(\omega) = 0\}) = 1 - P(\{\omega : \omega \in \Omega,\ h_{ms_1\ldots s_mr}(\omega) = 1\}),$$

the last probability is equal to the probability that at least one of the m checked clauses is not closed. In what follows we shall write also h_m and φ_k instead of $h_{ms_1\ldots s_mr}$ and $\varphi_{kn_1\ldots n_kr}$.

The purpose of the statistical procedure is to estimate the probability of the value of φ_k on the basis of the provability of the value of h_m. The procedure called H_m is defined as: if the value of $h_{ms_1\ldots s_mr}$ equals 1 (0, resp.) then the value of $\varphi_{kn_1\ldots n_kr}$ is estimated as provable (unprovable).

The error probability $q(H_m, r)$ reads

$$q(H_m, r) = P(\{\omega : \omega \in \Omega,\ h_m(\omega) = 0,\ w(\varphi_k)(\omega) = 1\}) +$$
$$+ P(\{\omega : \omega \in \Omega,\ h_m(\omega) = 1,\ w(\varphi_k)(\omega) = 0\}),$$

$m = 1, 2, \ldots, k$. If $s_i = n_i$ for $i = 1, 2, \ldots, k$, then

$$P(\{\omega : \omega \in \Omega,\ h_m(\omega) = 0,\ w(\varphi_k)(\omega) = 1\}) = 0\,, \tag{7.11}$$

$m = 1, 2, \ldots, k$; $w(\varphi_k) = 0$ means $\varphi_k \in K_0$, $w(\varphi_k) = 1$ means $\varphi_k \in K_1$. For fixed m, H_m is called *Bayes*, if

$$\begin{aligned} &P(\{\omega : \omega \in \Omega,\ w(\varphi_k)(\omega) = 1\} \mid \{\omega : \omega \in \Omega,\ h_m(\omega) = 1\}) \geqq \\ &\geqq P(\{\omega : \omega \in \Omega,\ w(\varphi_k)(\omega) = 0\} \mid \{\omega : \omega \in \Omega,\ h_m(\omega) = 1\})\,, \end{aligned}$$

$$\begin{aligned} &P(\{\omega : \omega \in \Omega,\ w(\varphi_k)(\omega) = 0\} \mid \{\omega : \omega \in \Omega,\ h_m(\omega) = 0\}) \geqq \\ &\geqq P(\{\omega : \omega \in \Omega,\ w(\varphi_k)(\omega) = 1\} \mid \{\omega : \omega \in \Omega,\ h_m(\omega) = 0\})\,. \end{aligned} \tag{7.12}$$

These inequalities express that $w(\varphi_k) = 1$ (0, resp.) is the most probable under the condition $h_m = 1$ (0, resp.). The following theorem expresses that for a properly chosen m the estimation procedure H_m is the best Bayes one. Denote

$$P_k = \prod_{j=1}^{k} (1 - p(n_j, r))\,, \quad P'_k = \prod_{j=1}^{k} (1 - p(s_j, r))\,, \quad \lambda_m = P'_m/P_m\,.$$

Theorem 7.10. If the (un) provability of the value sampled by the random variable $\varphi_{kn_1 \ldots n_k r}$ is estimated on the basis of the value of the random variable $h_{ms_1 \ldots s_m r}$, $n_i \geqq s_i \geqq 2$, $i = 1, 2, \ldots, m$, $n_j \geqq 2$, $j = m + 1, \ldots, k$, according to the procedure H_m and if $P_k - P'_k < 1 - P_k$ then there exists a natural number $m_0 \leqq k$ such that the procedure H_m is Bayes for $m = m_0, \ldots, k$. The error probability for such m reads:

$$q(H_m, r) = P_k - \lambda_m(2P_k - P_m)\,,$$

clearly, $q(H_m, r) < 1 - P_k$, $q(H_m, r) \leqq P_k$.

Moreover, there exists, if $s_i = n_i$, $i = 1, 2, \ldots, k$ for each real $\varepsilon > 0$ a natural $N(\varepsilon)$, $N(\varepsilon) \leqq k$, such that the procedure H_m is Bayes and $q(H_m, r) < \varepsilon$ for $n = N(\varepsilon)$, $N(\varepsilon) + 1, \ldots, k$.

Proof. Cf. Theorem 2.5, Corollary 2.1 and their proofs in [12].

Van Westrhenen has made an experiment consisting in implementation of his testing procedure on a computer and compared the results with those theoretically forecast ones; he also used these practical results in order to propose some optimal or appropriate values of free parameters in his test. About 20 000 formulas have been sampled and tested within an hour, more detailed information about these experiments and their results can be found in Chapter 2 of [12] and cannot be referred here because of the limited extend and rather theoretical character of this work.

In [12] also some attempts to extend this testing procedure to the case of the first-order predicate calculus can be found. Let us recall the well-known Herbrand

Theorem (cf. Chapter 3) in a form appropriate for our purposes: There exists a construction which assigns to every well-formed formula F of the first-order predicate calculus a sequence of well-formed formulas $S_1, S_2, \ldots$ of the propositional calculus, with the following property: F is provable iff there is a natural number n such that $S_1 \vee S_2 \vee \ldots \vee S_n$ is provable. The S_i, $i = 1, 2, \ldots$, are the (substitution) instances of F.

The way how to obtain such instances of F via the process of skolemization can be found in Chapter 3 of this work or in, e.g., [1.1]. Let us demonstrate the use of probabilistic methods on the relatively simple decision procedure consisting in searching for an appropriate provable disjunction. Namely, instead of generating the disjunctions in a systematic way, they are stochastically selected by the uniform probability distribution from the following finite sets (F is a given well-formed formula of the first-order predicate calculus)

$$X_1(F, N) = \{S_1, S_2, \ldots, S_N\}, \quad N = 1, 2, \ldots,$$

$$X_i(F, M) = \{D : D = S_{\alpha_1} \vee S_{\alpha_2} \vee \ldots \vee S_{\alpha_i};$$

are natural numbers such that $1 \leqq \alpha_1 \leqq \alpha_2 \leqq \ldots \leqq \alpha_i \leqq M\}$, $M = i, i + 1, \ldots$ $\ldots, i = 2, 3, \ldots$

Now, we introduce the sequence $X_{r_1}(F, N_1), X_{r_2}(F, N_2), \ldots$ naturals r_i, N_i, are selected in such a way that they satisfy the conditions

$$r_1 = 1, \quad r_{i+1} > r_i, \quad N_i > r_i, \quad \sum_{i=1}^{\infty} \binom{N_i}{r_i}^{-1} = \infty. \tag{7.13}$$

A sequence $\{r_i, N_i\}_{i=1}^{\infty}$ is called *sampling plan*. The stochastic sample of the disjunction D for a given F takes place as follows: s samples, $s \geqq 1$, are made, by the uniform probability distribution, from each set $X_{r_i}(F, N_i)$, $i = 1, 2, \ldots$

Formally, introduce mutually independent random variables φ_{ijF}, $i = 1, 2, \ldots$ $\ldots, j = 1, 2, \ldots, s$, F is a first-order predicate calculus formula. The range of φ_{ijF} is $X_{r_i}(F, N_i)$ for each $j \leqq s$, at the same time,

$$P(\{\omega : \omega \in \Omega, \ \varphi_{ijF}(\omega) = D\}) = \binom{N_i}{r_i}^{-1}, \quad j = 1, 2, \ldots, s, \tag{7.14}$$

$$i = 1, 2, \ldots, D \in X_{r_i}(F, N_i).$$

For the sake of an easy description we introduce the stochastic vector variable $\Phi_{iF} = \langle \varphi_{i1F}, \varphi_{i2F}, \ldots, \varphi_{isF} \rangle$ with the following s-valuation w_s of Φ_{iF}:

$w_s(\Phi_{iF}) = 0$, iff $w(\varphi_{ijF}) = 0$ for each $j = 1, 2, \ldots, s$,

$w_s(\Phi_{iF}) = 1$, iff there is at least one $j_0 \leqq s$ such that $w(\varphi_{ij_0F}) = 1$.

A finite sequence $\Phi_{1F}, \Phi_{2F}, \ldots, \Phi_{nF}$ is called a *sample of length n* with respect to F. Finally, we introduce a random variable, called the *value function* of a sample of length n with respect to F, namely, $\Psi_{nF} = \sum_{i=1}^{n} w_s(\Phi_{iF})$.

Clearly, $\Psi_{nF} \geqq 1$ iff at least one of the $n \times s$ sampled disjunctions is provable. I.e., F is estimated to be provable iff $\Psi_{nF} \geqq 1$, otherwise F is estimated to be unprovable. Only in the last case we can make an error.

The described sampling technique can be proved to be a "good one" in the sense that the provability of a given provable formula F can be estimated on the basis of a sample (with respect to F) of finite length. The condition laid upon the sampling plan $\{r_i, N_i\}$ will appear to be of crucial importance.

This is intuitively clear. In order to select one of the provable disjunctions D it is necessary that the relevant sets $X_{r_i}(F, N_i)$ be "big enough". Otherwise we could never draw a provable disjunction D. The divergence part of the condition (7.13) has been built in for technical reasons of the proof.

Theorem 7.11. If F is a provable formula, then its provability will be eventually discovered with probability one on the basis of a finite sample with respect to F.

Proof. Cf. Theorem 3.1, in [12], and its proof.

With respect to a given well-formed formula F we introduce the hypothesis $H_{r_a}(F)$, where r_a is equal to one of the natural numbers $r_1, r_2, \ldots, r_n$ of the truncated sampling plan $\{r_i, N_i\}$, $i = 1, 2, \ldots, n$; $H_{r_a}(F)$ means that the set $X_{r_a}(F, N_a)$ contains at least one provable r_a-disjunction. If $H_{r_a}(F)$ holds, then

$$P(\{\omega : \omega \in \Omega,\ \Psi_{nF}(\omega) = 0\}) \leqq \prod_{j=a}^{n} (1 - \beta_{r_j})^s , \tag{7.15}$$

where

$$\beta_t = \binom{N_t - r_a}{r_t - r_a} \Big/ \binom{N_t}{r_t} .$$

It can be shown (cf. [12]), that β_t can be approximated in the same way as the hypergeometric distribution. Hence, if $\Psi_{nF} = 0$, then the hypothesis $H_{r_a}(F)$ may be rejected with a risk probability (the first kind probability of error) smaller than $\prod_{j=a}^{n} (1 - \beta_{r_j})^s$. The second kind probability of error (i.e., $\Psi_{nF} \geqq 1$ and F unprovable) is clearly equal to zero. This result may be used as a heuristic aid for the determination of the m sets $X_{r_n+1}(F, N_{n+1}), \ldots, X_{r_n+m}(F, N_{n+m})$, from which the next random selections are to be made and the hypothesis $H_{r_x}(F, N_x)$, $n + 1 \leq x \leqq m$, to be tested next.

Let us apply this procedure to the case of provability estimation of at random sampled well-formed formulas from a set $\mathfrak{N}$ of well-formed formulas of the first-order

predicate calculus. Denote by $\mathfrak{N}_1$, the set of all theorems among $\mathfrak{N}$, $\mathfrak{N}_0 = \mathfrak{N} - \mathfrak{N}_1$ suppose that $\mathfrak{N}, \mathfrak{N}_0, \mathfrak{N}_1 \neq \emptyset$.

Random sampling of an element from $\mathfrak{N}$ is represented by a random variable φ which takes its values in $\mathfrak{N}$ and such that

$$0 < P(\{\omega : \omega \in \Omega,\ \varphi(\omega) \in \mathfrak{N}_1\}) = p < 1 .$$

The estimation procedure Γ_n is defined as follows: sample an element from $\mathfrak{N}$, realize a sample of length n and compute the value of the value function $\Psi_{n\varphi}$. If $\Psi_{n\varphi} \geqq 1$, proclaim the tested formula $\varphi(\omega)$ to be provable, if $\Psi_{n\varphi} = 0$, proclaim it to be unprovable. The error probability for this estimation procedure reads

$$q(\Gamma_n) = P(\{\omega : \omega \in \Omega,\ \Psi_{n\varphi(\omega)}(\omega) = 0,\ \varphi(\omega) \in \mathfrak{N}_1\}) + \tag{7.16}$$
$$+ P(\{\omega : \omega \in \Omega,\ \Psi_{n\varphi(\omega)}(\omega) \geqq 1,\ \varphi(\omega) \in \mathfrak{N}_0\}) ,$$

the last term being equal zero. Analogously we say that the estimation procedure is Bayesian iff

$$P(\{\omega : \omega \in \Omega,\ \varphi(\omega) \in \mathfrak{N}_1\}/\{\omega : \omega \in \Omega,\ \Psi_{n\varphi(\omega)}(\omega) \geqq 1\}) \geqq \tag{7.17}$$
$$\geqq P(\{\omega : \omega \in \Omega,\ \varphi(\omega) \in \mathfrak{N}_0\}/\{\omega : \omega \in \Omega,\ \Psi_{n\varphi(\omega)}(\omega) \geqq 1\}) ,$$
$$P(\{\omega : \omega \in \Omega,\ \varphi(\omega) \in \mathfrak{N}_0\}/\{\omega : \omega \in \Omega,\ \Psi_{n\varphi(\omega)}(\omega) = 0\}) \geqq$$
$$\geqq P(\{\omega : \omega \in \Omega,\ \varphi(\omega) \in \mathfrak{N}_1\}/\{\omega : \omega \in \Omega,\ \Psi_{n\varphi(\omega)}(\omega) = 0\}) .$$

Theorem 7.12. If the value of φ is estimated as (un)provable on the basis of $\Psi_{n\varphi}$ according to decision procedure Γ_n, then there exists, for each real $\varepsilon > 0$ a natural $N(\varepsilon)$ such that for all $n > N(\varepsilon)$ the procedure Γ_n is Bayes and $q(\Gamma_n) < \varepsilon$.

Proof. Cf. Theorem 3.2, in [12], and its proof.

Let us close this short review of basic ideas and results from [12] with an application to a special class of formulas. It is a well-known fact of mathematical logic, that the class of first-order predicate formulas of the form $(\forall x)(\exists y)(\forall z)\, M(x, y, z)$ with the matrix M containing no free indeterminates other than x, y, z, is decidable. Moreover, a formula F of this type is provable iff the disjunction

$$d_{NF} = \bigvee_{j=1}^{N} M(1, j, j + 1) , \quad N = 2^{\nu} ,$$

is provable, when ν is the sum of the weights of the different predicates appearing in F (the weight of an n-ary predicate A is equal to the number of different formulas of the form $A(u_1, \ldots, u_n)$ occurirng as elementary parts in $M(x, y, z)$ with the exception of $A(v, \ldots, v)$, which will not be counted). Cf. [2.1] for more details.

Theorem 7.13. Let

$$F = (\forall x)(\exists y)(\forall z)\bigwedge_{i=1}^{k} D_i(x, y, z)$$

be a given formula. If all the clauses determined by a sample of length s with respect to F are closed, then F is estimated provable with a risk probability smaller than $(m - s)/m$, $m = k^N$.

Proof. Cf. Theorem 3.4, in [12], and its proof.

We remark that analogous statistical procedures may be applied in many other decision procedures, provided the number of cases from which the selection have to be made is not too large. Cf. [1] for some informal comments in this direction. Some ideas from [12] are presented also in an earlier van Westrhenen's paper [11].

REFERENCES

[1] E. W. Beth: Formal Methods. D. Reidel Publ. Comp., Dordrecht 1962.

[2] В. С. Чернышевский: Об одном классе нормальных алгорифмов Маркова. Сборник „Логические исследования", Москва 1959, 263—299.

[3] M. Davis: Computability and Unsolvability. Mc Graw-Hill Book Company, New York 1958.

[4] C. Jordan: Calculus of Finite Differences. Chelsea, New York 1950.

[5] А. Н. Маслов: Вероятностные машины Тюринга и рекурсивные функции. Доклады АН СССР *203* (1972), 5, 1018—1020.

[6] С.Ю. Маслов: О поиске вывода в исчислении общего типа. Исследования по конструктивной математике и математической логике V., Наука, Ленинград 1972, 59—65.

[7] С.Ю. Маслов, Е. Д. Русаков: Вероятностные канонические исчисления. Исследования по конструктивной математике и математической логике V., Наука, Ленинград 1972, 66—76.

[8] A. Paz: Introduction to Probabilistic Automata. Academic Press, 1971.

[9] E. L. Post: Formal Reductions of the General Combinatorial Decision Problem. Amer. Math. J. *65* (1943), 2, 197—215.

[10] Б. А. Трахтенброт, Я. М. Барздинь: Конечные автоматы (поведение и синтез). Москва 1970.

[11] S. C. van Westrhenen: A Probabilistic Machine for the Estimation of Provability in the First Order Predicate Calculus. Z. für math. Logik und Grundlagen der Math. *15* (1969), 291—297.

[12] S. C. van Westrhenen: Statistical Studies of Theoremhood in Classical Propositional and First Order Predicate Calculus. J. of the Assoc. for Comp. Machinery *19* (1972), 2, 347—365.

8. APPLICATIONS OF STATISTICAL DEDUCIBILITY TESTING

At the very beginning of this chapter we have to say that is is beyond our powers to describe all possible applications of statistical deducibility testing procedures within our limited scopes. It is caused by the simple fact, that formalizing a decision problem, whatever its special features may be, we arrive always at the problem

of deducibility testing, i.e., we convert the original problem into that whether a formula is or is not a theorem of an appropriate formalized theory. Remember the two very general examples which we mentioned in Chapter 4 in order to realize that the class of problems which can be subsumed under the two models is large enough.

It is why we limit ourselves by a more detailed explanation of one particular application of theorem proving, namely, in the domain of the so called automated problem solving (or artificial intelligence, in general). The methods derived from such considerations are, first, general enough to cover a large class of situations and, second, they can be applied, say, in robotics. Our explanation here will be based on a very informal level in order not to make the reading of this chapter too difficult by introducing special formalisms (they can be found in references).

Consider a situation well-known in experimental robotics; a room with several boxes of various sizes, colours or forms and a robot which is to transform the configuration of the boxes into another, a priori prescribed one (e.g., to put them one onto another in a given order). In the most simple cases the robot is given an instruction or a sequence of instructions, in other words, a *plan* how to solve the problem, however, some more sophisticated experimental robots are able to find themselves an appropriate plan. Of course, the robot is expected to have at its disposal a scale of *operators* enabling to change the configuration of the boxes, more generally, to change the *state* of the environment with the aim to reach a (or the) goal state.

This example can serve as an illustration of the very general notion of *state space*. Formally, state space is a pair $\langle S, \Phi \rangle$, where S is a nonempty set the elements of which are called states, and Φ is a nonempty set of partial mappings defined in S and taking their values again in S, the elements of Φ are called operators. The partial character of mappings from Φ corresponds to the fact that there are, in general, for each operator, some states in which it is not applicable. If $\varphi \in \Phi$, $s \in S$, and $\varphi(s)$ is defined, then, clearly, $\varphi(s)$ denotes the state resulting when φ is applied in the state s.

As we have already mentioned a problem consists in transforming a state (a given or initial one) into another state (a or the goal one). Formally, a *problem* in a state space $\langle S, \Phi \rangle$ is a pair $\langle s_0, G \rangle$, $s_0 \in S$, $G \subset S$, where s_0 is the initial state and G is the set of goal states. A sequence $\langle \varphi_1, \varphi_2, \ldots, \varphi_n \rangle \in \Phi^n$ is called a (linear) solution to the problem $\langle s_0, G \rangle$ in the state space $\langle S, \Phi \rangle$, iff, for each $i \leqq n$, $\varphi_i(\varphi_{i-1}, \ldots$ $\ldots, \varphi_1(s_0) \ldots)$ is defined, and $\varphi_n(\varphi_{n-1}, \ldots, \varphi_1(s_0) \ldots) \in G$. Instead of solution we can speak also about a plan for solving the problem in question (i.e., about a linear plan in this case).

Sometimes a generalization of the notion of linear solution to that of a *generalized* or *branching solution* or *plan* can be useful. Let Γ be a finite set of finite sequences of operators, i.e., $\Gamma \in \Phi^*$, $\Gamma \in \mathscr{P}_{fin}(\Phi)$ in symbols. Identifying the identical initial segments of sequences from Γ we can consider Γ as a tree (a branching structure); if $\Gamma = \{\langle \varphi_{i_1}, \varphi_{i_2}, \ldots, \varphi_{i_{n(i)}} \rangle, i = 1, 2, \ldots, k, \varphi_{i_j} \in \Phi$ for each $i \leqq k, j \leqq n(i)\}$, then we can ascribe to each φ_{i_j} the set $Sc(\varphi_{i_j})$ of its successors, clearly, $Sc(\varphi_{i_j})$ may be

empty. Now, Γ is called a branching plan (this term will be used in what follows). if at least one of the operators φ_{i_i}, $i \leqq k$, is applicable to s_0, and if, having applied an operator φ_{i_j}, either a goal state is already reached, or at least one of the successors from $Sc(\varphi_{i_j})$ is applicable. Intuitively said, the notion of branching plan corresponds to the case when we must take into consideration several possibilities how to solve the problem, paralelly (simultaneously), as we are not able to choose apriori the adequate linear solution; it is just during the process of execution when we decide which way will be actually followed.

Having a problem, it may be, and usually is, a very complicated and difficult matter to find a solution to this problem or even to decide whether a solution exists or not. It is why automated problem solving is taken for an important, if not the key, branch in the domain of artificial intelligence. Let us briefly describe a possibility how to apply proof theory when searching for a solution of a problem.

Before all, we need a formalized language appropriate for our goals. Let $\mathscr{L}$ be a language describing the environment and expressing its properties. E.g., in the case of a room with boxes $\mathscr{L}$ must contain names of particular boxes (A, B, ...), names of their properties (green, black, iron, wooden, ...) and names of relations between them (greater than, to lie on,...).

However, such a language, no matter how rich it may be, is static in the sense that it is not able to reflect the changes of the environment caused by applications of operators. It is why we replace $\mathscr{L}$ by another two-sorted (or more than two sorted if $\mathscr{L}$ itself is many-sorted) language $\mathscr{L}^*$, enriching each formula of $\mathscr{L}$ by a new, situation indeterminate or constant (or term, in general); this term will be listed as the last one in the list of terms, indeterminates or constants occurring in a formula of $\mathscr{L}^*$. There is just one situation term in each formula of $\mathscr{L}^*$. Namely, s_0 is the unique situation constant (corresponding to the initial situation), s, s_1, s_2, ... are situation indeterminates. There is, in the alphabet of $\mathscr{L}^*$, a special functional symbol $f(\varphi)$ for each operator $\varphi \in \Phi$; if t is a situation term (indeterminate, constant) and $\varphi \in \Phi$, then $f(\varphi)(t)$ is again a situation term. For each formula $A \in \mathscr{L}$ and each situation term t we denote by $A[t]$ the formula of $\mathscr{L}^*$ resulting from A when the list of terms occurring in A is enriched by t.

Now, we are to express the formalism of state space by the means of the language $\mathscr{L}^*$. Each operator $\varphi \in \Phi$ will be represented by a pair $\langle C(\varphi), R(\varphi)\rangle$ of sentences from $\mathscr{L}$ (the *condition* of φ and the *result* of φ) and to φ a special formula of $\mathscr{L}^*$, called *operator* (or *transition*) *axiom* for φ will be ascribed, namely the formula

$$(\forall s)\,(C(\varphi)[s] \to R(\varphi)[f(\varphi)(s)])\,. \tag{8.1}$$

The intuition is as follows: the operator φ is applicable just in the states of environment satisfying certain formula (condition) $C(\varphi)$. Moreover, if φ is applicable in a state s (i.e., if $C(\varphi)[s]$ holds) and if it is actually applied (executed), then, no matter which state will be reached, we can be sure that a formula $R(\varphi)$ (result of φ) will

hold. The term $f(\varphi)(s)$ occurring in (8.1) can be understood as a *name* of *the* state resulting from s when φ applied, hence, the function symbols $f(\varphi)$, $\varphi \in \Phi$, play the role of Skolem functions (cf. Chapter 3).

Besides the operator axioms consider as axioms also logical axioms of appropriate kinds and state-independent assertions of $\mathscr{L}^*$, i.e., general assertions which are valid in all states (e.g., transitivity of the relations "greater than", "to lie on", etc.). Denote by T_I the formalized theory generated in $\mathscr{L}^*$ by axioms and the usual deduction rules (the so called *core theory*). If A is a formula of $\mathscr{L}^*$, denote by $T_I[A]$ the theory resulting from T_I when the set of axioms is enriched by A. Such theories $T_I[A]$ are called *images* and the space of such theories is called *image space I*.

The notion of problem can be formalized by the means of $\mathscr{L}$ as a pair $\langle X, Y\rangle$ of sentences with the following intuition: if the initial state satisfies X, we search for an operator (sequence of operators, branching tree of operators) which would bring us to a state satisfying Y. A sequence $\langle \varphi_1, \varphi_2, \ldots, \varphi_n\rangle \in \Phi^n$ is a (*linear*) *solution* (*plan*) to a problem $\langle X, Y\rangle$ in the given image space, if $T_I \vdash C(\varphi_1)[s_0]$, $T_I[R(\varphi_i\varphi_{i-1} \ldots \varphi_1(s_0))] \vdash C(\varphi_{i+1})[\varphi_i\varphi_{i-1} \ldots \varphi_1(s_0)]$ for each $i = 1, 2, \ldots, n-1$, and if $T_I[R(\varphi_n \ldots \varphi_1(s_0))] \vdash Y[\varphi_n \ldots \varphi_1(s_0)]$ (we write $\varphi_i \ldots \varphi_1(s_0)$ instead of $\varphi_i(\varphi_{i-1} \ldots (\varphi_1(s_0)) \ldots)$). The notion of solution can be, again, generalized to the branching one replacing the demand of derivability (in the corresponding image) of the condition of the following operator by the demand of derivability of the disjunction, consisting of Y and of conditions of all successors of the operator which has been applied as the last. We present the necessary formalism at a very rough level, the necessary details can be found in [6], [7].

Consider a certain formula $F_I(X, Y)$ of $\mathscr{L}^*$ proclaiming, in a sense, the solvability of $\langle X, Y\rangle$, namely

$$F_I(X, Y) =_{\mathrm{df}} X[s_0] \to (\exists s)\, Y[s]\,. \tag{8.2}$$

Now, under some conditions concerning the consistencies of the occurring theories (as a matter of fact, these conditions represent an important and serious theoretical problem, but we will not discuss it here) the formula $F_I(X, Y)$ is provable in T_I, i.e., $T_I \vdash F_I(X, Y)$, iff there exists a solution (a linear or branching one) to the problem $\langle X, Y\rangle$ in I. Even more can be said: if $T_I \vdash F_I(X, Y)$ holds and if $F_I(X, Y)$ is proved from T_I using the resolution principle, then, due to skolemization, we can always find in this proof a provable formula of this type:

$$X[s_0] \to [Y(f_{11}f_{12} \ldots f_{1n(1)}(s_0)) \vee Y(f_{21}f_{22} \ldots f_{2n(2)}(s_0)) \vee \ldots \vee Y(f_{k1}f_{k2} \ldots f_{kn(k)}(s_0))]\,, \tag{8.3}$$

(we write f_{ij} as an abbreviation for $f(\varphi_{ij})$). Clearly, $F_I(X, Y)$ follows from (8.3) by the deduction rule consisting in introduction of the existential quantifier. At the same time, $\{\langle \varphi_{i,n(i)}, \varphi_{i,n(i)-1}, \ldots, \varphi_{i,1}\rangle, i \leqq k\}$ can be proved to be a solution to the problem $\langle X, Y\rangle$ (possibly a linear one, if $k = 1$). In other words said, the solution

can be immediately read, in the reverse order, from the longest (and last) occurred during the resolution based theorem proving of $F_I(X, Y)$ situation term (or terms).

This application of theorem proving has all the features typical for the examples mentioned in Chapter 4. The theoremhood testing is not the final stage of the decision process, but is followed by other actions (solution or plan derivation and application) which have to be executed in real time and according to the changes taking place in the environment. Hence, we are in a position when an approximate, not quite sure, but quick statistical deducibility testing procedure may be of greater value than a correct but rather slow deterministic method.

Let us briefly mention at least three possibilities how to introduce probability and statistics into this domain. First, trying to verify whether $T_I \vdash F_I(X, Y)$, we may apply the method of statistical deducibility testing in at random sampled extensions. Consider the case when this test proclaims $F_I(X, Y)$ to be derivable from T_I. In this case the test gives, as a by-product, a sequence $\langle a_1, a_2, \ldots, a_m \rangle$, $m \geqq M$ (the threshold value of the test in question, cf. Chapter 5), of formulas from $\mathscr{L}^*$ together with proofs of formulas $a_i \to F_I(X, Y)$ from T_I. These proofs can be taken as proofs of $F_I(X, Y)$ from $T_I[a_i]$ $(= T_I \cup \{a_i\})$ and can be easily combined into a proof of $(\bigvee_{i=1}^{m} a_i) \to F_I(X, Y)$ from T_I, i.e., into a proof of $F_I(X, Y)$ from $T_{I'}$, of a new image space I', corresponding to a state space in which $\bigvee_{i=1}^{m} a_i$ is supposed to be valid and, hence, included among the axioms of I'. Now, we can transform the obtained proof of $F_I(X, Y)$ from $T_{I'}$ into a resolution-based form and derive a solution (plan) Γ from this proof in the way mentioned above. According to the Corresponding Theorems, Γ is a solution to the problem $\langle X, Y \rangle$, but in the image space I'. If we execute this solution in the state space corresponding to I, we may arrive at a failure, of course, as Γ supposes something (namely $\bigvee_{i=1}^{m} a_i$) to be valid in the environment, but in fact this may be invalid. Nevertheless, the probability of a failure of Γ in I is majorized by the probability of error connected with the statistical deducibility testing procedure applied to $F_I(X, Y)$ and T_I. Hence, Γ may serve as an approximate or "statistically good" solution to $\langle X, Y \rangle$ in I. At present, this way of approximating plans or solutions is studied in more details.

When trying to verify $T_I \vdash F_I(X, Y)$, we may use probability and statistics also in another way (cf. [2]). Clearly, the richer the core theory T_I is, the simpler may be to find the desired proof. So we may consider as operators also some actions which are not quite safe from the point of view of their results, i.e., which may fail when applied. This may be caused by technical failures, unprecisely known or simplified conditions, etc. Joining the operator axioms connected with such unprecise operators makes the looking for the desired proof (and solution or plan, after all) more easy, however, the obtained solution is not safe and is subjected to a possible failure, as it uses unprecise or unprecisely known operators. So, again, the obtained solution

may serve as an approximation of the originally desired correct or ideal solution. Some types of such stochastic approximations or plans are described and discussed in [2].

Finally, let us mention another probabilistically based approximation of solutions, specially the branching ones. In practical applications the number of possibilities which must be considered together when looking for a plan is usually great enough, for the resulting branching plan to be unpractically large (from the point of view of a computer storage, say). In everyday life we solve such a problem by neglecting those possibilities (i.e., branches) which are little probable. In the case when, against our expectation, such a little probable case occurred, we should have to consider the actually occurred situation and to take adequate measures, i.e., to find a new plan. It is just this consideration according to which we do not think of possibility of an earthquake when settling our plans for tomorrow, even if we are not able to exclude this possibility on the ground of a logical deduction. In [5] we studied the possibility of such a reduction of branching plans based on the idea of erasing all the nodes for which the probability of their execution during an actual application of this plan is below a given value. As can be shown, the reduction of the extent of the branching plan is essential; roughly speaking, if $\varepsilon > 0$ is the threshold value decisive for adhering the node in question into the restricted version of the original plan, then the extend of the restricted plan can be majorized by $(1/\varepsilon) \log_2 (1/\varepsilon)$. Let us recall that this majorant does not depend on the length of branches in the original plan in spite of the fact that the extent of this original plan is an exponential function of the lengths of branches. In [4] also some applications to hierarchic planning are studied and some estimations of the total extent of the corresponding hierarchic plans are derived.

Let us close this chapter by mentioning the fact that there are also other applications of probability theory and statistics in mathematical logic and proof theory which cannot be (or at least usually are not) expressed in the model of statistical deducibility testing as developed in this work. Some of such approaches and results will be very briefly mentioned in the next chapter. Some general remarks concerning the three types of probabilistically modified solutions or plans as discussed above can be found also in reviewal papers [1] and [3].

REFERENCES

[1] I. M. Havel, I. Kramosil: Probabilistic Methods in Robot Decision Making. In: Sborník prací celostátní konference o kybernetice, Praha 1976, 66—80.

[2] I. M. Havel, I. Kramosil: A Stochastic Approach to Robot Plan Formation. Kybernetika *14* (1978), 3, 143—173.

[3] I. Kramosil: Stochastické plány v problematice robotů s vyšším stupněm inteligence. In: "Aplikovaná robotika 77", Karlovy Vary 1977, 91—109.

[4] I. Kramosil: Pravděpodobnostní redukce větvících se plánů pro činnost automatu s cílovým chováním. Research Report, Institute of Information Theory and Automation, 1977.

[5] I. Kramosil: A probabilistic Restriction of Branching Plans. In: Mathematical Foundations of Computer Science, 1977, Lecture Notes in Computer Science 53, Springer-Verlag, Berlin—Heidelberg—New York 1977, 342—349.
[6] O. Štěpánková, I. M. Havel: A Logical Theory of Robot Problem Solving. Artificial Intelligence *7* (1976), 129—161.
[7] O. Štěpánková, I. M. Havel: Incidental and State-Dependent Phenomena in Robot Problem Solving. Kybernetika *13* (1977), 6, 421—438. (Cf. also the preliminary version In: Proceedings of the AISB Summer Conference, Edinburgh 1976, 266—278).

9. OTHER CONCEPTIONS OF STATISTICAL APPROXIMATIONS IN PROOF THEORY

In Chapters 5 to 8 above we suggested several possibilities how to implement statistical decision theory into the domain of theorem proving and theoremhood testing. All these investigations as well as the necessary formalisms and preliminaries explained in Chapters 2 to 4 have one important and restrictive, in a sense, common feature. Namely, we have always supposed, that the formalized theories which served as objects of our statistical samples, experiments or studies are based one an appropriate *classical two-valued logical calculus*. In other words, no intuitionistic, many-valued, probabilistic, fuzzy, or other non-classical logics have been taken into consideration until now. Under this general assumption we were entitled to consider the meta-property of theoremhood as a classical two-valued one. I.e., at least from the platonistic or Omniscient point of view, each formula of the investigated theory either was a theorem or not with no uncertainty admitted at this level. It is only our *subjective knowledge* about the actual state of affairs, concerning the tested formula, which can be wrong, unprecise or charged by an uncertainty. The priority given in our work to this special way of understanding and introducing uncertainty and probability into the domain of theorem proving seems to be sufficiently justified by the very title, if the words "proof theory" used in it are understood in the sense "classical proof theory" or "proof theory in classical logic". For the other possibilities how to introduce probability theory and statistics into theorem proving, i.e., for various non-classical probabilistically oriented proof theories we limit ourselves to several short remarks and comments in this chapter.

A great part of non-classical probabilistically oriented logics can be reduced to the basic notions and assertions of the theory of fuzzy sets. This theory was conceived by Zadeh in 1965 (cf. [10]) as a straighforward generalization of the naive set theory. Having a nonempty space (universe) X, any subset Y of X can be, clearly, identified with its characteristic function χ_Y; $\chi_Y(x) = 1$, if $x \in Y$, $\chi_Y(x) = 0$, if $x \in X - Y$. Zadeh's idea was to consider each function defined on X and taking its values in the set $\langle 0, 1\rangle$ of real numbers as a (generalized) subset of X, called *fuzzy set*. Hence, formally, a fuzzy set A is a pair $\langle X, \chi_A : X \to \langle 0, 1\rangle\rangle$. If $x \in X$, then the real value $\chi_A(x)$ can be understood as the degree in which x belongs to A or as the probability

with which x belongs to A (until now, the semantics of fuzzy sets has not been developed enough to investigate in more details the similarities and differences with respect to probability theory). The mappings χ_A, with A ranging over the space of fuzzy sets in X, are subjected to some requests which generalize the usual set-theoretic notions and operations. Namely, if A and B are fuzzy sets in X, then their complement A^c, union $A \cup B$ and intersection $A \cap B$ are, again, fuzzy sets in X, defined, for each $x \in X$, by the relations

$$
\begin{aligned}
&\chi_{A^c}(x) = 1 - \chi_A(x)\,, \\
&\chi_{A \cup B}(x) = \max\{\chi_A(x), \chi_B(x)\}\,, \\
&\chi_{A \cap B}(x) = \min\{\chi_A(x), \chi_B(x)\}\,.
\end{aligned} \tag{9.1}
$$

Some basic properties of set-theoretic operations remain to be valid, e.g., the de Morgan rules (as can be easily checked), however, some other properties of fuzzy sets seem to be rather counterintuitive, e.g., if A is the fuzzy set for which $\chi_A(x) = 1/2$ for all $x \in X$, then $A^c = A = A^c \cap A = A^c \cup A$.

The basic idea of fuzzy sets is usually implemented into mathematical logic in such a way that the set $\mathscr{L}$ of all well-formed formulas is taken as the universe of discourse and the set $\mathscr{T}$ of theorems is generalized to a fuzzy set in $\mathscr{L}$. Instead of $\chi_{\mathscr{T}}(x)$ we write often $T(x)$, $x \in \mathscr{L}$, hence, $T(x) \in \langle 0, 1\rangle$ is the degree or probability with which x is taken as a theorem. As a rule, the mapping T is supposed to satisfy these conditions:

$$
\begin{aligned}
&(1)\quad T(\neg A) = 1 - T(A)\,, \\
&(2)\quad T(A \wedge B) = \min(T(A), T(B))\,, \\
&(3)\quad T(A \vee B) = \max(T(A), T(B))\,, \\
&(4)\quad T((\forall x)\, A) = \inf\{T(A(x)) : x \in D\}\,, \quad \text{if } D \text{ is the domain of } x\,, \\
&(5)\quad T((\exists x)\, A) = \sup\{T(A(x)) : x \in D\}\,, \quad \text{if } D \text{ is the domain of } x\,.
\end{aligned} \tag{9.2}
$$

Hence, it suffices to define $T(x)$ for atomic formulas $x \in \mathscr{L}$, the rules (9.2) enable to extent T unambiguously to all $\mathscr{L}$.

The following theorem shows some connections between the truth-values (i.e., values of T) of premises and consequences in resolution-based theorem proving. Recall that, for each set S of clauses, $R^n(S)$ is the set of all resolvents of n-th level obtainable from S (cf. Chapter 3).

Theorem 9.1. Let S be a set of clauses, let $C_1, C_2, \ldots, C_M$ be clauses in S. Denote $b = \max\{T(C_1), T(C_2), \ldots, T(C_M)\}$, $a = \min\{T(C_1), T(C_2), \ldots, T(C_M)\}$, let $a > 1/2$. Then, for each $n \geqq 0$ and each clause $C \in R^n(S)$, $a \leqq T(C) \leqq b$.

Proof. Cf. Theorem 9, in [6], and its proof.

Theorem 9.1 shows that if every clause in S is something "more than a half-truth" and the most unreliable clause has truth-value a, then we are guaranteed that all the

logical consequences obtained by repeatedly applying the resolution principle will have truth-value at least equal to a, but never exceeding the truth-value of the most reliable clause.

The assertion of Theorem 9.1 seems to be of a great practical worth, but it depends substantially on the min-max property of the mapping T((2) and (3) in (9.2)) which is often subjected to a serious criticism, namely from specialists working in probability theory. Let A and B be two formulas the validity of which in the environment depends on one or two random experiments (e.g., A and B describe a particular result or results of these experiments). Then A and B can be identified with the corresponding sets of their models (i.e., relational structures of appropriate signatures in which A, resp. B, are valid) and these sets of models can be taken as random events in the classical probability theory (more details on these transformations can be found in [4]). Hence, A and B themselves can be seen as random events to which some probabilities may be ascribed. Clearly, the random events A and B may be statistically independent or dependent, so $P(A \cap B)$ may vary from 0 to $\min \{P(A), P(B)\}$ and $P(A \cup B)$ may vary from $\max \{P(A), P(B)\}$ to $P(A) + P(B)$. This means that the demands (2) and (3) from (9.2) are equivalent to the assumption that there is a strictly defined type of dependence between A and B, namely, that A is a sufficient condition for B or vice versa. However, such a restriction seems to be too strong to be adequate for expressing all the types of uncertainty in surrounding us world.

In [4.1] and [4.2] we made an attempt to obtain results similar to Theorem 9.1 without the request of validity of (9.2). We begin with the truth-values of axioms and try to examine in which degree these truth-values are preserved by the consequences obtained from the axioms using the deduction rules. First of all, and this seems to be quite intuitive, no uniform positive lower bound (like a in Theorem 9.1) can be found (omitting the trivial case when the truth-values of all axioms are 1). In other words, if at least one axiom is not "quite sure", then there always exists a theorem (i.e. formula provable from axioms) the truth-values of which is below an a priori given positive real number. It is why a desirable positive lower bound for the truth-values of the consequences derived from uncertain axioms can be given only in the form of an expected value with respect to an a priori given probability distribution over the set $\mathscr{L}$ of well-formed formulas. A general, but rather abstract result of this type can be found in [4.2] (Theorem 1), here we shortly present an application of this general result to the case of the so called *Gentzen-like random axiomatic systems.*

Consider a formalization of the first order predicate calculus with the following properties:

(1) No individual constants and no functional indeterminates occur, only a finite number of functional constants may occur.

(2) If $q_1, q_2, \ldots, q_n$ is a formalized proof and if q_j results from q_i, q_k, $i, k < j$ with respect to a deduction rule then to every occurrence of a subformula of q_i

or q_k an occurrence of the same subformula (up to differences in indeterminates or terms) in q_j can be shown (different occurrences in q_j for different occurrences in q_i or q_k). This is the so called *subformula property* or *Gentzen property*; for more details cf., e.g., [2.2] or the original Gentzen's paper [5.3]. This condition excludes, e.g., modus ponens from the set of deduction rules which are at our disposal.

(3) The only operators are those of implication and general quantifier, $A \to B$ being written as $[A]\,[B]$ and $(\forall x)\,A$ as $[x[A]]$. There is one propositional constant F having the semantic interpretation "falsehood", hence, $\neg A$ is written as $[A]\,[F]$. For more details cf. the end of Chapter 5 or [5.7].

Let us define a random variable G on a probability space $\langle \Omega, \mathscr{S}, P \rangle$, taking its values in the set of all well-formed formulas of the theory just described, which is a slight modification of the random variable F defined by operations (I)–(VII) in the final part of Chapter 5 or in [5.7].

Let K_0, M_0, N_0 be positive integers, let to the left bracket [integers $1, 2, \ldots, K_0$ be ascribed, to the right bracket] the integers $K_0 + 1, \ldots, 2K_0$, to F the integers $2K_0 + 1, \ldots, 2K_0 + M_0$. To every individual indeterminate x_i, $i \leqq N_0$, the integer $2K_0 + M_0 + i$ is ascribed, to every elementary formula containing only indeterminates among $x_1, x_2, \ldots, x_{N_0}$ one integer beginning with $2K_0 + M_0 + N_0 + 1$ is ascribed. Let N_2 be the greatest integer used in this enumeration, let $N_3 > N_2$ be an integer, let θ be an auxiliary symbol not occurring in the considered formalized theory; the indices $N_2 + 1, \ldots, N_3$ are ascribed to θ.

Let $\beta_1, \beta_2, \ldots$ be a sequence of random variables defined on $\langle \Omega, \mathscr{S}, P \rangle$, taking their values in the set $\{1, 2, \ldots, N_3\}$ of integers, mutually independent and equally distributed in such a way that

$$(9.3) \qquad P(\{\omega : \omega \in \Omega,\ \beta_j(\omega) = i\}) = N_3^{-1}, \quad j = 1, 2, \ldots, i = 1, 2, \ldots, N_3 .$$

Now, we define $G(\omega) = \Lambda$ (the empty formula) if no occurence of an elementary formula of F preceeds the first occurrence of θ in $\{\beta_1(\omega), \beta_2(\omega), \ldots\}$, $G(\omega) = F(\beta_1(\omega), \ldots, \beta_k(\omega))$, if there is no occurrence of θ and at least one occurrence of F or an elementary formula among $\beta_1(\omega), \ldots, \beta_k(\omega)$ and if, at the same time, $\beta_{k+1}(\omega) = \theta$, i.e., $\beta_{k+1}(\omega) > N_2$, where F is the mapping generated by operations (I)–(VII) mentioned above. It means that G generalizes F in such a way that formulas of all lengths as well as the empty formula have a positive probability to be sampled. Immediately follows that $G(\omega)$ is always the empty or a well-formed formula.

Definition 9.1. Let $\langle \Omega, \mathscr{S}, P \rangle$ be a probability space, let $T = \langle 0, \infty)$ be a set of parameters, let N_n denote the set $\{1, 2, \ldots, n\}$ of integers. Then *random axiomatic system of degree n, over the language $\mathscr{L}$ and with respect to the probability space* $\langle \Omega, \mathscr{S}, P \rangle$ is a mapping X of the Cartesian product $N_n \times T \times \Omega$ into $\mathscr{S}$ such that for every $i \leqq n$ and $t \geqq 0$ the mapping $X(i, t, \cdot)$ is a random variable defined on $\langle \Omega,$

$\mathscr{S}, P\rangle$ and taking its values in $\mathscr{L}$, i.e., because of the countability of $\mathscr{L}$, for every $p \in \mathscr{L}$, $\{\omega : \omega \in \Omega,\ X(i, t, \omega) = p\} \in \mathscr{S}$.

The parameter t, which can be interpreted as *time*, expresses the dynamics of a random axiomatic system, the possibility and necessity to modify the representation of the environment as the time passes with respect to its development and changes.

In order to be able to describe and judge somehow the quality of a random axiomatic system X we suppose that the state of the environment at the time instant t is represented by a subset $\mathscr{T}(t) \subset \mathscr{L}$, namely by the set of all formulas valid in this time instant. If $A \subset \mathscr{L}$, we denote by $Cn(A)$ the set of all formulas derivable from the set A of formulas by considered deduction rules, i.e., $A \subset Cn(A) \subset \mathscr{L}$. Using this notation we can easily see that a random axiomatic system X is an ideal representation of the environment iff, for each $t \in T$,

$$Cn(\{X(1, t, \omega),\ X(2, t, \omega), \ldots, X(n, t, \omega)\}) = \mathscr{T}(t)\,. \tag{9.4}$$

However, usually this is not the case, so we have to measure somehow such a situation.

Definition 9.2. Characteristic function of the i-th axiom $X(i, \cdot, \cdot)$ of a random axiomatic system X is defined as

$$p(i, t) = P(\{\omega : \omega \in \Omega,\ X(i, t, \omega) \in \mathscr{L} - \mathscr{T}(t)\})\,. \tag{9.5}$$

For example, if $X(i, t, \omega)$ is a logical axiom or an axiom describing the fundamental time-and-space relations (and such axioms are necessary in any formalized representation of an environment, cf. Chapter 8), then $X(i, t, \omega)$ does not depend on ω and t and $p(i, t) \equiv 0$. A more detailed classification of random axioms can be found in [4.1]. To be able to measure somehow the quality of a random axiomatic system as a whole we must have at our disposal a random variable measuring the importance of particular formulas.

Definition 9.3. Let X be a random axiomatic system with respect to the probability space $\langle \Omega, \mathscr{S}, P\rangle$, let α be a random variable defined on $\langle \Omega, \mathscr{S}, P\rangle$ and taking its values in $\mathscr{L}$. Reliability of X with respect to α is defined as the conditional probability

$$R(t, \alpha)(X) = P(\{\omega : \omega \in \Omega, \alpha(\omega) \in \mathscr{T}(t)\} | \{\omega : \omega \in \Omega, \alpha(\omega) \in Cn(\{X(1, t, \omega),\ X(2, t, \omega), \ldots, X(n, t, \omega)\})\})\,. \tag{9.6}$$

This means that $R(t, \alpha)$ is the probability that a formula sampled at random with respect to α is valid in the environment under the condition that it is derivable from the random axioms. Instead of deriving a general expression for $R(t, \alpha)(X)$ (cf., as mentioned above, Theorem 1 in [4.2]) we apply Definition 9.3 to the case of the random variable G defined above.

Theorem 9.2. Consider a formalization of the first order predicate calculus satisfying the conditions (1)–(3) above. Let X be a random axiomatic system over this language such that the random variables $G, X(i, t, .)$, $i \leqq n$, are mutually independent for each fixed $z \in T$. Let γ be a function defined on T, taking its values in $\langle 0, 1\rangle$ and such that for all $i \leqq n$ and all $t \in T$ the inequality $p(i, t) \leqq 1 - \gamma(t)$ holds. Suppose that $p(i, t) \equiv 0$, if $X(i, t, \omega)$ does not contain any elementary formula and that the non-validity of the empty formula can be proved without any possibility of an error. Then

$$(9.7) \quad R(t, G)(X) \geqq \frac{S}{e(1-\gamma(t))+S} + \left(\frac{e\,\gamma(t)}{e+S}\right)^n \cdot \left(1 - \frac{S}{e(1-\gamma(t))+S}\right) \geqq$$

$$\geqq \left(1 + \frac{e}{S}(1-\gamma(t))\right)^{-1},$$

where

$$e = P(\{\omega : \omega \in \Omega,\ 2K_0 + M_0 + N_0 < \beta_1(\omega) \leqq N_2\}) =$$

$$= (N_2 - 2K_0 - M_0 - N_0) N_3^{-1},$$

$$S = P(\{\omega : \omega \in \Omega,\ N_2 < \beta_1(\omega) \leqq N_3\}) = (N_3 - N_2) . N_3^{-1}.$$

If $\gamma(t) \to 1$ uniformly for all $t \in T$, then $R(t, G)(X) \to 1$ uniformly for all $t \in T$.

Proof. Cf. Theorem 2, in [4.2], its proof and its corollary.

The second inequality in (9.7) is rather interesting as it does not depend on the number of axioms and tends to 1 if γ does. This fact offers a simple strategy when a random axiomatic system is formed: it is better to have a great number of reliable axioms than a small number of less reliable ones. This agrees with the effort to atomize the data into the most detailed form. The model based on the notion of random axiomatic system as explained above has been used in order to formalize and handle the stochastic and dynamic character of the environment needed for an automaton (e.g., robot) to make senseful and goal-oriented decision and actions in this environment (cf. a series of papers [1], [2], [3]).

Theorems 9.1 and 9.2 serve as examples of such an approach when uncertainty or fuzziness are introduced in proof procedures by a fuzzification of assumptions (premises) leaving the deduction rules unchanged. However, the idea of fuzzy sets can be applied immediately to generalize the notion of deduction rule. Usually, n-ary deduction rule or deduction rule with n premises is a partial mapping defined in $\mathscr{L}^n$, i.e., in the set of ordered n-tuples of formulas and taking its values in the set of finite subsets of $\mathscr{L}$ (because of the fact that, in general, a deduction rule may be applied in more than one way to a given sequence of premises). Using the technique of working zones and trivial extensions as mentioned in Chapter 7 we may assume that each n-ary deduction rules is a mapping from $\mathscr{L}^n$ into $\mathscr{L}$. An easy generalization gives that a fuzzy deduction rule may be defined as a mapping ascribing to each

n-tuple of premises $\langle a_1, \ldots, a_n \rangle \in \mathscr{L}^n$ a function $\chi_{<a_1,\ldots,a_n>}$ taking $\mathscr{L}$ into $\langle 0, 1 \rangle$; $\chi_{<a_1,\ldots,a_n>}(a)$, $a \in \mathscr{L}$, defines the degree or probability with which a formula a can be considered as the consequence of $a_1, a_2, \ldots, a_n$ by the fuzzy deduction rule in question. This corresponds to the intuitive situation when the deduction rule may "fail" and give an uncorrect consequence (with respect to the "usual" deduction rules). As this uncertainty or risk may cummulate when the length of a proof increases, we arrive at a situation similar to that in random axiomatic systems; the reliability of a derived formula depends on the length of its proof. This version has been elaborated in details in [8].

Let us mention, before closing this chapter, two things. First, we do not take into consideration here a great number of papers dealing with the problems how to ascribe probabilities to well-formed formulas of a formalized theory in such a way that some more or less intuitive conditions of syntactical or semantical character were satisfied. We have taken such a decision because of the fact that, as a rule, these papers do not work with the notion of proof, i.e., they do not consider the dynamics of a formalized theory. Moreover, when discussing such problems we should evoke many problems penetrating into the most fundamental parts of probability theory and mathematical logic and such reasonings would bring us far beyond the planned scope and extent of this work.

Before closing this chapter let us mention briefly an interesting and perspective modification of the basic problem of statistical deducibility testing as studied in this work. Our interest has been always oriented to the problem whether there exists or does not exist a proof of the tested formula and we have completely neglected the length, complexity or other qualitative or quantitative characteristics of the potential proof. However, from the applicational point of view the length of proof may be an important aspect, remember, e.g., automated plan formation mentioned in Chapter 8 or the automated experiment planning in [3]. So it may seem quite useful to replace the hypothesis "p is a theorem" by another hypothesis "p is a theorem for which there exists a formalized proof of a length not exceeding a given natural R" (p is the tested formula). Of course, also the alternative must be appropriately changed in order to remain the logical complement of the hypothesis. Some recent results show that the testing procedure developed in Chapter 5 can be used also in this case with probability of error slightly enlarged. A special test of constructivistic character for these goals is proposed in [5.10]. At present, a special paper dealing with these results in under preparation and it is why we limit ourselves to this short note (cf. [5]).

In [7] and [9] the authors investigate a particular resolution-based theorem-proving algorithm and define the length of the potential proof by the number of resolution principle applications used in this proof. Supposing the candidates for resolution are sampled at random, the length proof becomes a random variable. The two mentioned works give some upper and asymptotic estimates for the expected value and dispersion of this random variable; however, they do so only under very special

and rather artificial conditions. In every case, the modified testing problem as mentioned above seems to be very often more realistic than the original one and its more detailed study may bring interesting results of practical as well as theoretical nature.

REFERENCES

[1] I. Kramosil: A Probabilistic Approach to Automaton-Environment Systems. Kybernetika *11* (1975), 3, 173—206.

[2] I. Kramosil: A Selection-Based Formal Representation of an Environment. Kybernetika *12* (1976), 3, 127—150.

[3] I. Kramosil: Mechanized Experiment Planning in Automaton-Environment Systems. Kybernetika *13* (1974), 4, 225—244.

[4] I. Kramosil: Some Remarks on Probabilities over Formalized Languages. In: Transactions of the Eight Prague Conference on Information Theory, ..., Academia, Prague 1978, vol. A, 371—382.

[5] I. Kramosil: Statistical Testing Procedure for Lengths of Formalized Proofs. Submitted for publication.

[6] R. T. C. Lee: Fuzzy Logic and the Resolution Principle. Journal of the Association for Computing Machinery *19* (1972), 1, 109—119.

[7] J. V. Mayega: Statistical Decidability of Theorems. In: Creation in Mathematics 7 (1974), (J. Reichbach, ed.), Tel-Aviv 1974, 3—11.

[8] J. Pavelka: On Fuzzy-Logic I, II, III. To appear in Zeitschrift für Mathematische Logik und Grundlagen der Mathematik.

[9] J. Reichbach: Generalized Models, Probability of Formulas, Decisions and Statistical Decidability of Theorems. Jokohama Math. Journal *20* (1972), 2, 79—98.

[10] L. Zadeh: Fuzzy Sets. Information and Control *8* (1965), 338—353.

10. CONCLUSIVE REMARKS

As usual, the aim of such concluding chapters usually is to review and survey what has been done and to confront these results with the intentions and goals having been promised in the introduction. However, such a recapitulation is always only of subjective and relative worth. Having finished a work, probably each author feels that everything should and could have been done better, more precisely, more understandingly ... But this constant feeling of dissatisfaction and disquiet – it is an external curse and blessing, blessing and curse of every intellectual and creative activity.

Our intention in this work has been to survey the possibilities of various applications of probability theory and mathematical statistics in the domain of classical proof theory. When considering the results presented here as answers to some questions about such possibilities we must admit that the answers offered in this work are only of partial and relative character. Moreover, each of these answers is followed by a number of new questions, sometimes more peculiar to answer than the original one. First of all, it is the author which is to be blamed for this situation and we are far from trying to refuse the corresponding responsibility. On the other

hand, the process of continual revoking and relativizing of the obtained results and answers as well as the arising of new and more sophisticated questions are eternal attributes of each science – as far as it is to be considered for a real science, not for a dogma or doctrine.

It is why, instead of a detailed revision of what has been said, we shall concentrate our attention to several aspects of the investigated domain which seems to be perspective for a further development, at least from our subjective point of view. The scope of such problems is very wide and it ranges from purely theoretic and even philosophic matters to questions of extremely applicational and implementational nature.

A serious theoretical problem consists in comparing the subjective and objective aspects of uncertainty connected with statistical theorem proving. Namely, either we can consider a classical two-valued logic (or a theory based on it) seen by the medium of statistical experiments which charge our knowledge by some degree of uncertainty, or we may take the world surrounding us as internally indeterministic and stochastic and a fuzzy logic as an adequate and true formalization of this world. The open question is, whether the preference of one of these two approaches is only a matter of technical and mathematical convenience or whether such a choice involves some deeper consequences as far as the corresponding stochastic approximations of formalized proofs are concerned. Even in case the two approaches are equivalent in the sense that they may be "translated" into each other by appropriate mathematical transformations the question of their relative adequacy from mathematical and implementational points of view arises.

The same problems of adequacy and appropriateness can be related to the used formalization of probability theory and statistics. Here we used the classical Kolmogorov conception, but perhaps some other may be better. As an extremely interesting case we take the so called Boolean-valued probability theory, namely when the abstract values of the corresponding probability measures are elements of the Lindebaum-Tarski algebra over a formalized theory (cf. [2.6]), i.e., classes of formulas. Such a probability theory would eliminate the difficulties with two incoherent and hardly comparable structures over the set of formulas – the logical and the probabilistic ones.

Another group of problems are those connected with various possibilities which particular parts of a theorem-proving procedure should be randomized. We have concentrated our attention mainly to the case when the choosing of auxiliary premises is subjected to a randomization, i.e., is replaced by a random sampling. An advantage of this approach consists in the fact, that we are allowed to make the best profit of the already existing deterministic theorem-provers (e.g., the resolution-based ones, it is why we have devoted all the Chapter 3 to an explanation of such algorithms). In fact, we replace the original deducibility problem by a sequence of such problems with antecedents enriched by an at random sampled auxiliary one. However, the process of randomization may penetrate much more deeply into the very process of resolution-based theorem proving, e.g., we may sample at random the

candidates for resolution (as mentioned in [9.7] and [9.9]) or the introducing of a resolvent into the class of resolvents of the corresponding level may be preceeded by a statistical test (i.e., only the resolvents passing successfully this test are placed into the corresponding class to be considered as possible candidates for a further resolution). Also other parts of various theorem-proving procedures can be subjected to a randomization and various such possibilities can be compared from the viewpoint of, e.g., probability or probabilities of an error. Another open and important problem of this branch is that of a reverse interaction between the stochastic and the deterministic theorem-proving methods, i.e., the question whether, and in which sense and degree, the statistical results may influence, perhaps in the form of appropriate heuristics, the deterministic theorem-provers giving arise, say, to some new refinements of resolution-based theorem proving.

A great deal of further effort in the field of statistical theorem proving should be devoted also to the problem of computational complexity of various procedures. These questions can be seen as special instances of more general problems connected with problems of computational (or algorithmical) complexity of statistical procedures and approximations in general. Or, when accepting a statistical decision rule or approximate computation we admit some risk, some possibility of failure, but we know (or at least are justified to expect) that the decision or computation will be "much more" simple or shorter and, under the particular external circumstances, we prefer this complexity saving to the possible risk of an error or failure. However, statistical decision theory, at least in its present state, does not describe and formalize these both sides of one problem at the same or similar level. As we have seen (the end of Chapter 4 or elsewhere in this work), the notions of risk or possibility of a failure are precisely formalized by the notion of probabilities of errors of the two possible kinds; these notions are strictly described, defined and handled within the formal framework of the classical (Kolmogorov) probability theory. On the other hand the argumentation in favour of statistical tests or approximations is based either on the theoretical impossibility of a deterministic and precise decision or computation procedure (this is the case of theoremhood testing in indecidable theories) or on argumentation of informal and intuitive kind (it is supposed to be "intuitively clear" that the statistical procedure is „much more easy and simple" than a deterministic one for the some problem. Perhaps the complexity theory of computational processes, possibly enriched by appropriate oracles in order to formalize the random sampling, seems to be an adequate background for a description and handling of computational complexity for various statistical procedure. Some positive and concrete results in this direction, i.e., certain expressions or estimations for computational complexity of particular statistical deducibility testing or theorem-proving methods, are necessary in order to be able to say the final word about these methods when compared with deterministic theorem-proving algorithms. Appropriate results on computational complexity will be useful also for mutual comparing of various statistical theoremhood testing procedures.

Last but not least, we have to mention the problems connected with the possibilities of computer-oriented implementations and realizations of statistical theoremhood testing procedures introduced in this work. As far as the method based on random extensions is concerned (cf. Chapter 5), the original implementation theoretical difficulties have been overcome and the method has been modified in such a way that it makes the greatest profit of deterministic theorem-proving algorithms (as statistical theoremhood testing is reduced to a sequence of time-and-space limited deterministic theorem-proving problems). The only necessary supplementary subprogram is that one realizing a random sampling of well-formed formulas, i. e., a random generator of formulas. Such a generator has been developed; it is based on a simple pseudo-random number generator and on the procedure transforming each sequence into a well-formed formula (cf. the end of Chapter 5). This algorithm seems to be relatively very quick (in average, one formula is sampled within less than one second), its disadvantages consist in unknown output probability distribution which can be estimated only in a very difficult and unprecise way, and in a special formalism of output formulas which must be, hence, transformed into a form more adequate for common, e.g., resolution-based theorem-proving methods. At present, a program is under construction which tries to combine this random formula generator with an appropriate theorem prover. In any case, the implementation of statistical theorem-proving methods will request still great effort of theoretical as well as of experimental nature in order to choose (or develop) programming language and other apparatus and tools the most adequate for the sake of such an implementation. There are many open problems here, as our statistical orientation may bring new adequacy or appropriateness criteria for judging the qualities of various theorem-provers, and these criteria may be quite different from and even contradictory to the commonly used ones.

The list of open questions, problems and possible problems of further development, as presented above is, of course, far from being exhaustive and it is even impossible to give an exhaustive survey. Moreover, each of these open problems will surely produce and involve many new problems and questions as soon as it is studied in details. This is the continual flow of scientific development and research beginning somewhere at the very roots of our civilization and tending beyond the horizons of our perspectives. Our work is nothing else than a short stopping and small looking out in this flow – and each surveyal work, no matter with which branch of science it deals, can be only something like this. Of course, this changes nothing on the fact that the qualities of such a work may be very various; as far as this aspect is considered, the author takes all the responsibility for the weak (in various sense.) points of this work which he knows very well – as well as for those more weaknesses which will be highly probable, discovered by careful readers. On the other hand, the author believes that there are at least few positive aspects in what he has written, that at least some of the readers have found in it certain help and, maybe, partial answers to their questions. In short, the author believes that this work can be considered as a contri-

bution. Very likely, this belief is too pretentious and not justified – but such a belief is a virtual essence and necessary attribute of each creative effort.

CONTENTS