MEASURING THE POWER OF ARITHMETICAL THEORIES

JÖRGEN SJÖGREN

Abstract

This thesis discusses the possibility to measure the power of extensions of Peano Arithmetic, PA. It consists of three parts, an introduction and two separately written papers. In the introduction we present the problem and briefly give an account of van Lambalgen's and Raatikainen's criticism of Chaitin's efforts to measure the power of theories. The first paper contains generalizations of two versions of Chaitin's incompleteness theorem, and reinforces the above mentioned criticism. The second paper is the main paper of the thesis, and here, using the modal logic GL, we design a measure of the power, in terms of the capacity to prove theorems, of an important set of extensions of PA.

Acknowledgements

First of all I want to thank Christian Bennet at the department of Philosophy at Göteborgs Universitet for introducing me to the problems discussed in this thesis. He has always been very encouraging, and he has also made valuable comments on earlier versions of the thesis.

I also want to thank Torkel Franzén at Luleå University of Technology. He has helped me to make some obscure arguments more transparent, and he has also helped me to eliminate some errors in a preliminary version of this thesis.

My dear, old friend Leif Segelström, and Cristian Bennet have helped me make my english more readable.

Finally, my colleague at University of Skövde, Stefan Karlsson, and my son Martin Sjögren have helped me with mysteries of LaTeX.

List of Corrections

Page 6	line 11 from the bottom	reads should read	$K^n(p) > 0$ $K^e(p) > 0$
Page 15	line 7/8	reads should read	The number of strings that ... The number of strings of length $n > c$ that ...
Page 32	line 2	reads should read	These constant sentences ... With $T = PA$ these constant ...
Page 36	line 7	reads should read	$x_i \cap x_j = 0$ $x_i \cap x_j = \emptyset$
Page 37	line 12	reads should read	$m_n((A \wedge \neg B) + m_n(A \wedge B))$ $m_n(A \wedge \neg B) + m_n(A \wedge B)$
Page 38	footnote 22	reads shuold read	$\ldots \Box p \to p$ has a fixed point p. $\ldots \Box p \vee p$ has a fixed point $\neg\bot$.
Page 40	line 6 from the bottom	reads should read	$\ldots Prf(\neg\rho, x))$ $\ldots Prf(\neg\rho, y))$

Section 3.9: To get the order between theories correct, the measure m on theories should be designed such that, if $m_n^*(\varphi) = x$, then the theory $PA+\varphi$ is assigned the measure $1-x$. This means that $m(PA) = 0$ and $m(PA+\bot) = 1$, etc. It is then valid that, if $S \dashv T$, then $m(S) \leq m(T)$. Note that if $\varphi \leq \psi$ in the Lindenbaum algebra of PA, then $PA + \psi \dashv PA + \varphi$.

Contents

1 Introduction

1.1 The Problem

This thesis for the licentiate degree deals with the possibility to design a measure of the power of extensions of elementary arithmetic. There are several ways to specify what is to be meant by 'the power of a theory', where a theory is understood as the set of its axioms. We shall see below that Chaitin equates 'power' with 'information content', but it is perhaps more basic to view the power of a theory as its capacity to prove theorems. A theory S is a subtheory of a theory T, if T proves everything that S proves, in symbols $S \dashv T$. One possibility, then, is to say that T is at least as powerful as S if and only if $S \dashv T$. Here we will be interested in elementary arithmetic, Peano Arithmetic, or PA for short, and it is then natural to study the set $\mathcal{T} = \{T : PA \dashv T\}$. This set is partially ordered by the subtheory relation, and we shall study it, considered as the Lindenbaum algebra of PA, in section 3.

We would like to but cannot expect that a measure m simultanously satisfies the following conditions:

1. m takes its values in a linearly ordered set of numbers.

2. m is computable for every $T \in \mathcal{T}$, i.e. m is recursive.

3. $S \dashv T$ if and only if $m(S) \leq m(T)$

If such a measure existed, the theories in $\mathcal{T}$ would be linearly ordered by the subtheory relation, and we know that this is not the case. We would, furthermore, have a decision procedure for theoremhood in e.g. PA. To see this let φ be any sentence, and let $PA + \varphi$ be PA with φ added as an extra axiom. If $m(PA) < m(PA + \varphi)$, then $PA \not\vdash \varphi$. If $m(PA) = m(PA + \varphi)$, then $PA \vdash \varphi$.

A natural question, then, is which of the conditions we must exclude. All three are problematic. The function m can perhaps not have as range a set of linearly ordered numbers since the set $\mathcal{T}$ is partially ordered. Thus, it may be more natural to let m be vector-valued. There are several reasons why m might not be recursive, if m is to have $\mathcal{T}$ as its domain. There are e.g. theories in $\mathcal{T}$ whose axioms are not recursively enumerable, which means that we cannot know which the axioms of these theories are. And, finally, we may perhaps have to stay content with the 'only if' clause of the third condition, because two theories with the same measure need not be comparable with the subtheory relation. What then can be done?

Most well-known of earlier attempts to create a measure of the power of theories use Chaitin's incompleteness theorem, and identifies the power with the information content, defined as the Kolmogorov complexity, of a theory. [1] Chaitin announced his celebrated incompleteness theorem in the early 1970's. The theorem states that for every sound, formal theory S containing elementary arithmetic, there is a constant c, depending only on S, such that S does not prove any true propositions of the form $K(n) > c$. Here $K(n)$ is the Kolmogorov complexity of the string n, and is a measure of the difficulty of specifying n. A string has low Kolmogorov complexity if it has a short description, and a high one if it has no short descriptions. The notions above can be made precise in different ways, and depending on how this is done, the theorem and its proof take different forms. The number c is a natural number and Chaitin and his adherents interpret this constant as a measure of the information content of the theory S, and claim that this is an adequate measure of the power of S.[2] In [Cha82] Chaitin says that

> I would like to measure the power of a set of axioms and rules of inference. I would like to be able to say that if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from these axioms.

In the same paper Chaitin says that

> traditional proofs of Gödel's incompleteness theorem show that formal axiomatic systems are incomplete, but they do not suggest ways to measure the power of formal axiomatic systems.

Chaitin's argument runs in two steps. In the first he proves the incompleteness theorem. In the second step he draws the extra-logical, or philosophical, conclusion that there is an intimate relationship between the information content of a formal system, and the constant in his theorem. This interpretation, or use, of Chaitin's theorem has been critized by van Lambalgen [vL89] and Raatikainen [Raa98].

The main goal in this thesis is to use other means to construct a measure of the power of an interesting class of extensions of elementary arithmetic.

[1] In [Cha82] Chaitin presents three other possibilities to measure the information content of a theory, but since we focus on the above mentioned idea, we will not discuss these other possibilities.

[2] See e.g. the introduction of [Raa98] for some comments on the reception of Chaitin's claims.

The structure of the thesis is as follows. The rest of the introduction presents van Lambalgen's and Raatikainen's arguments against Chaitin's interpretation. A brief discussion of the possibility to draw extra-logical conclusions from theorems of logic follows, and finally there is a summary of two previously unpublished papers, that constitute the main part of the thesis. The first of these papers reinforces the criticism of the standard interpretation of the constant in Chaitin's theorem. The second paper, the main paper of the thesis, is of a more positive nature, and there we show how to construct a measure of some extensions of Peano Arithmetic. These two papers are separately written, and they contain some overlappings. There are also some overlappings between the introduction and the two papers. This has the advantage that the introduction and the two papers can be separately studied, but the disadvantage of some tiresome repetitions. We apologize for that.

1.2 Some Arguments Against the Received Interpretation of Chaitin's Theorem

That the philosophical conclusion of Chaitin's theorem is problematic, is one of the themes in an excellent paper by van Lambalgen [vL89]. In this paper he assumes that to every formal system S, which contains elementary arithmetic, there is a minimal constant c_S, the characteristic constant of S, such that S does not prove any theorem of the form $K(n) > c_S$, in symbols $S \nvdash K(n) > c_S$, or more exactly a formula representing the relation $K(n) > c_S$. As above $K(n)$ is the Kolmogorov complexity of the string n.[3]

In the paper referred to above, van Lambalgen shows that there is no connection between the information content of a theory S and the characteristic constant c_S associated with the theory. Where Chaitin usually defines the Kolmogorov complexity with concepts like 'abstract computer' or 'Turing machine', van Lambalgen in his discussions uses partial recursive functions, but this change of concepts does not affect the argument.

Here we will briefly sketch a proof by van Lambalgen showing that the characteristic constant depends on the Gödel numbering of the partial recursive functions. Let φ_e be the partial recursive function with index e, and define $K_{\varphi_e}(n) = min\{l(p) : \varphi_e(p) = n\}$, where $l(p)$ is the length of the binary string p. It is easy to define a bijection between binary strings and natural numbers. This makes it possible to speak of the complexity of a

[3] Concepts, as e.g. different ways of defining Kolmogorov complexity, are explained in more detail in the sections following.

number. If the p in $K_{\varphi_e}(p)$ is a number, then $l(p) = p$.[4]

This function, then, determines a minimal p that inserted in φ_e outputs n. To define the Kolmogorov complexity of n, $K(n)$, we use a universal Turing machine U that takes inputs of the form $q = 0^e 1p$, i.e. a string of e zeros followed by a one followed by the string p. The machine U simulates the action of φ_e on p, and we define $K(n)$ as $K_U(n)$. As a lemma van Lambalgen notes that for any partial recursive function φ_e, and for all n, $K(n) \leq K_{\varphi_e}(n) + e + 1$.[5]

To prove the desired result, van Lambalgen presupposes a listing of the proofs in PA, and defines a partial recursive function φ_e by the following condition:

> $\varphi_e(m) = n$ if and only if n is the k in the first proof in PA of a sentence of the form $\varphi_m(m) \neq k$.

He then proves that $\varphi_e(e)$ is undefined. Suppose $\varphi_e(e) = n$ for some n. Then $PA \vdash \varphi_e(e) \neq n$, and since PA is sound $\varphi_e(e) \neq n$, and this is a contradiction. Thus $\varphi_e(e) \neq n$ for all n. From the definition of φ_e it follows that $PA \nvdash \varphi_e(e) \neq n$, and thus $PA + \{\varphi_e(e) = n\}$ is consistent for all n. But then, by the definition of K and K_{φ_e}, and the lemma mentioned above

$$K(n) \leq K_{\varphi_e}(n) + e + 1 \leq e + e + 1 = 2e + 1.$$

It follows that $PA + \{K(n) \leq 2e + 1\}$ is consistent, and consequently $PA \nvdash K(n) > 2e + 1$. The constant $c = 2e + 1$ then depends on the Gödel numbering of the partial recursive functions and

> (t)he above argument shows convincingly that there is no a priori reason to expect that c_S and the information content of S ... are related in some interesting way [vL89].

Now, elementary arithmetic, PA, is, according to Chaitin, associated with a constant c_{PA}. The arithmetical fragment of ZF, the standard axiomatization of set theory, is associated with a constant c_{ZF}. According to van Lambalgen, we do not even know whether $c_{ZF} > c_{PA}$ or $c_{PA} > c_{ZF}$. If $S \dashv T$, then $S + X$ is an axiomatization of T over S, if X is recursively enumerable, and T and $S + X$ have the same theorems, $T \dashv\vdash S + X$. Citing a theorem of Kreisel and Levy [KL68], saying that the arithmetical fragment of ZF is not finitely axiomatizable over PA, van Lambalgen concludes

[4] In section 2.1 we define a bijection between strings and numbers, but there $l(p)$, where p is a number, equals the number of binary digits in the associated string.

[5] A result akin to this is briefly discussed in section 2.3.

that there are infinitely many theories S_n, where $n = 0, 1, \ldots$, such that $PA \dashv S_n \dashv S_{n+1} \dashv ZF$, proper, and since both c_{PA} and c_{ZF} are finite natural numbers, infinitely many of the different theories S_n must be associated with the same constant.

Finally, van Lambalgen points out that Chaitin's misconceptions may be seen as a confusion between object language and metalanguage. What Chaitin's incompleteness theorem states, is that there is a constant c_S such that $S \nvdash K(n) > c_S$, and that $K(n) > c_S$ is true. For infinitely many n the string n is a representation of the Gödel number of a sentence φ, where φ typically is a finite conjunction of axioms of S, so $S \vdash \varphi$. The Kolmogorov complexity of φ might be large, while the complexity of $K(n) > c_S$ is small. Even if φ is a sentence that is not a conjunction of axioms, it may very well be the case that $S \vdash \varphi$ and $S \nvdash K(n) > c_S$ with φ having high and $K(n) > c_S$ low Kolmogorov complexity. Furthermore, nothing at all is said of the information content of the formula $K(n) > c_S$.[6]

This paper by van Lambalgen has been ignored by Chaitin, who has continued to argue for the claim, now known to be unfounded, that c_S is a measure of the information content of S.[7] In frustration over this Raatikainen has published a paper [Raa98], where he elaborates and extends the criticism of van Lambalgen, and presents van Lambalgens arguments even more clearly.

Where van Lambalgen is a little bit cautious and says e.g. that "Chaitin's matematics do not support his philosophical conclusions", Raatikainen is straightforward and claims that he shows "conclusively that the received view is false".

After a presentation of some preliminaries, the incompleteness theorem and Chaitin's philosophical claims, Raatikainen first constructs codings such that the characteristic constant c_S gets the value zero, respectively can be chosen arbitrarily large. The codings Raatikainen constructs are somewhat artificial, but they are acceptable in the sense that given a standard numbering φ_e, as e.g. in [Men87], a numbering ψ_e is acceptable, if there are recursive functions f and g such that $\varphi_{f(x)} = \psi_x$ and $\psi_{g(x)} = \varphi_x$. This means that it is possible to go effectively from the standard numbering to an acceptable one and vice versa. There are of course no preferred way to construct a coding, and Raatikainen shows that the constant c_S can be given any value whatsoever.

Raatikainen uses a slightly different definition of Kolmogorov complexity

[6] See below for further discussion of this observation.

[7] See e.g. [Raa01].

of a string p than both van Lambalgen and Chaitin, but this does not affect the main ideas of his arguments. Using $\varphi_e(p) \simeq n$ meaning that the partial recursive function with index e takes the value n with the argument p, Raatikainen defines $K(p) = \mu e(\varphi_e(0) \simeq p)$. Here $\mu x\theta(x)$ is the least number a such that $\theta(a)$.

His arguments are as follows. He uses the same numbering of Turing machines as for the partial recursive functions, i.e. the Turing machine T_e computes the partial recursive function φ_e. For this enumeration $T_0, T_1, T_2, \ldots$ he defines a permutation π^n on the indices with

$$\pi^n(x) = \begin{cases} 0, & \text{if } x = n \\ x+1, & \text{if } x < n \\ x, & \text{if } x > n \end{cases}$$

The algorithmic complexity in relation to this new coding, K^n, is

$$K^n(x) = \mu z(\exists y(\pi^n(y) = z \wedge T_y \downarrow = x)).$$

With $T_y \downarrow = x$ we understand that the Turing machine with index y outputs x and stops with empty input. This corresponds to the formula $\varphi_y(0) \simeq x$ used in the definition of $K(x)$ above. The constructions are arithmetical, i.e. they can be coded into arithmetic. It is then possible to effectively find a Turing machine T_m, in the initial coding, that searches for the least number x such that there is a number p and x is (the Gödel number of) a proof in S of (the Gödel number of) a formula representing $K^n(p) > 0$, and when it finds such an x, if there is one, T_m outputs p and stops.

The problem now is to find a machine that operates like this and have its initial code number as the parameter n in K^n. This can be accomplished with the fixed-point theorem of recursion theory, and we let e be such a number. It then follows that this Turing machine never stops, and so there is no proof in S of $K^n(p) > 0$ for any p. Consequently $c_S = 0$.

To show that c_S can be chosen arbitrarily large, and that the theory S can prove theorems that have complexity larger than the theory itself, Raatikainen, very sketchily, argues as follows. Let T_0 be a Turing machine that prints the axioms of S and then stops. The complexity of the axioms of this theory is thus zero. Then let the Turing machines $T_1, T_2, \ldots, T_n$ be 'uninteresting' machines for arbitrarily large n, and let T_{n+1} output 1. Then $K(1) > n$, and so the characteristic constant of S is even larger than n, while the complexity of the axioms is zero, i.e. $c_S > n$.

Raatikainen then proceeds to identify exactly how the constant c gets its value. We saw above how van Lambalgen showed that $K(n)$, with his definition, depends strongly on the enumeration of the partial recursive functions. Raatikainen shows that $K(n)$, with his slightly different definition, is just

> the smallest (by its code) Turing machine which does not halt, but for which this cannot be proved in S.

Let e be the smallest index such that $\neg\exists x T_e \downarrow= x$ is true and unprovable in S. Next let $n > e$ and m be arbitrary. It is then impossible to prove in S that

$$T_n \downarrow= m \wedge \forall z < n \neg T_z \downarrow= m$$

It is, thus, not provable in S that a number has a complexity larger than e.

Of course, there is no reason to believe that this code number has anything to say about the power of the theory S.

Like van Lambalgen, Raatikainen identifies the origin of the misuse of the incompleteness theorem by the confusing of object language and metalanguage, but he discusses this in terms of 'use' and 'mention'. In the following sentence about the dog Fido, the first occurrence of 'Fido' is used and the second is mentioned.

> Fido has four legs, and 'Fido' contains four letters.

Raatikainen says, concerning sentences $K(n) > m$, that

> the sentence expressing (used) that a particular object has a very large complexity, e.g. '$K(n) > m$' (for a very large m), may itself have a quite simple (when mentioned) syntactical form.

When Chaitin says that ten pounds of axioms cannot prove a twenty-pound theorem, he confuses the complexity of axioms as mentioned and the complexity asserted by a theorem when used.

In his criticism Raatikainen presupposes that the information content of a theory is closely related to its power to prove theorems. He points out that his criticism is valid even if the information content, or power, of a theory is the algorithmic complexity of its axioms. A theory S is called an extension of T, if $T \dashv S$. Now, start with e.g. PA, and add sentences $\{\varphi_1, \dots \varphi_n\}$ to PA such that $PA + \{\varphi_1, \dots \varphi_n\}$ proves the same theorems as PA, and choose the sentences φ_i such that they have high Kolmogorov complexity. Then the axioms of $PA + \{\varphi_1, \dots \varphi_n\}$ are more complex than the axioms of PA, but the two theories prove the same theorems.

Raatikainen's last argument is that if Chaitin's constructions give a minimal constant c_S, then the halting problem is solvable. As mentioned above, the value of c_S is the smallest Turing machine which does not halt, and for which this is unprovable in S. Start with e.g. PA. We can then find a constant c_0, and this is the smallest Turing machine, by its code, that does not halt, and for which this is unprovable in PA. Note that it is provable in PA, that Turing machines with smaller index than c_0 that do not halt, do not halt. Then, add to PA the true sentence $\neg\exists x \varphi_{c_0}(0) \simeq x$. This new theory is associated with a minimal constant c_1, and we proceed in the same way. In this way we get an enumeration of all Turing machines that do not halt. Since the set of Turing machines that do halt is recursively enumerable, this set would be recursive, so we would have an effective method for deciding the halting problem.

These critical discussions of van Lambalgen and Raatikainen should effectively silence the proponents of the received interpretation. But unfortunately this does not seem to be the case. In a recent review paper[8] Raatikainen, among other things, discusses Chaitin's unwillingness to respond to criticism, and says that it

> is regrettable that Chaitin does not respond to criticism of his work but simply evades difficult questions and keeps on writing as if they did not exist.

1.3 What Extra-Logical Conclusions Can Be Drawn from a Theorem of Logic?

Knowing that Chaitin's philosophical conclusions from his incompleteness theorem are not warranted, it is interesting to ask if it is possible to draw extra-logical conclusions from a theorem of logic. Here we just intend to formulate some stray remarks, because this is a problem that is too large for this thesis. It is not an exaggeration to say that Gödel's incompleteness theorems are the theorems mostly used, or misused, in this context.

To give a well-known example we can mention Hofstadter. In the last chapter of [Hof79] he discusses 'consequences' of Gödel's two incompleteness theorems, and he is very explicit in saying that these theorems do not prove anything in e.g. psychology. At the same time he thinks that the theorems can reveal new truths if they are used metaphorically. Hofstadter thinks that the brain and the thinking, the acitivity of the mind, can be considered from a high, 'soft ware' level that contains concepts that cannot be seen at lower, 'neuron' levels, and that this high level might have an explaining capacity that cannot exist at lower levels. In a way, he equates this with the

[8] In [Raa01] Raatikainen reviews two recent books by Chaitin.

translation of number theory into metamathematics, and postulates that it is something like this that gives rise to our unanalysable feelings of the I.

This example of use, or misuse for that matter, is of course just the tip of an iceberg,[9] but this is not the place to discuss all the purported conclusions from Gödel's incompleteness theorems. We will just present a totally different use of Gödel's first incompleteness theorem by Dummet.[10] He initially says that a Gödel sentence g for a formal system S, although unprovable is recognizable by us as being true. Since there are structures, or models, in which g is true or false respectively, and we can recognize g as true, in the standard model, we must have

> a quite definitive idea of the kind of mathematical structure to which we intend to refer when we speak of natural numbers.

A Gödel sentence is a sentence of the form $\forall x A(x)$ where $A(x)$ is a recursive predicate. To know that $\forall x A(x)$ is true we need to know that $A(0), A(1)$ etc. are true, and this we must know for every natural number, so we must have some possibility to know what 'natural number' means. But the concept 'natural number' cannot be fully characterized in first order Peano Arithmetic. In a way, not discussed by Dummet, we know exactly which structure this is anyhow, because the natural numbers can be characterized up to isomorphism in second order Peano Arithmetic, as was proved by Dedekind. A large part of Dummet's paper is devoted to discussions of the meaning of 'natural number', and he wants to know

> what light is thrown by Gödel's theorem on the meaning of 'natural number' in so far as understanding its meaning involves grasping the application of the predicate 'true' to arithmetical sentences.

After a digression into different ways of attaching meaning to 'natural number', and a short discussion of the bearing of the ideas presented on intuitionism, Dummet comes to the conclusion that

> (t)he intuitive conception of a valid mathematical proof ... cannot in general be identified with the concept of proof within some

[9] Views akin to Hofstadter's can be found in e.g. [Pen89] and [Ruc95]. Rucker's book was originally published in 1982. In [Web80] there are extensive discussions of arguments for and against using Gödel's theorems, Church's Thesis, and Church's and Turing's results on decidability as support for mentalism, mechanism, etc.

[10] The paper *The Philosophical Significance of Gödel's First Incompleteness Theorem* was originally published in Ratio 1963. It can be found in [Dum78].

> formal system, for it may be the case that no formal system can ever succeed in embodying all the principles of proof that we should intuitively accept; and this is precisely what is shown to be the case in regard to number theory by Gödel's theorem.

Finally, as a support for his intuitionistic claims, he says that

> the intuitionists are right in claiming that, if the sense of mathematical statements is to be given in the notion of mathematical proof, it should be in terms of the inherently vague notion of an intuitively acceptable proof, and not in terms of a proof within any formal system.

These two examples of 'conclusions' from Gödel's theorem, and the above 'consequence' of Chaitin's theorem, illustrate that one should be very careful when one uses theorems of logic as support for extra-logical claims. We will also emphasize that theorems of logic are mathematical theorems formulated in a precise context, and that conclusions from this domain into e.g. a domain like that treated by psychology or whatever only can be analogical. To use a theorem like Gödel's incompleteness theorem about the human mind or brain, the mind or brain must be, at least approximately, a formal system including elementary arithmetic. Whatever man, mind or brain might be, it is hard to understand what it means to say that man etc. is a formal system. It is, of course, something completely different that a man can perform computing tasks. The problem is akin to the problem of the applicability of mathematics to 'reality'. But the problem of the applicability of logic to 'reality' consists of at least one more dimension. While it is understandable to say e.g. that the orbit of Jupiter is approximately elliptic, or that the energy levels of the hydrogen atom is such and such in the Bohr model of the atom, it is hardly understandable what it means to say that the human mind, or brain, is approximately a formal system including arithmetic.

Dummet's claim is perhaps acceptable together with his intuitionism, which in a way identifies 'true' with 'provable', but this is of course not the only possible way to understand 'true'. Hofstadter's idea might be acceptable if one 'sees' the analogy, but Chaitin's claim is simply wrong.

When discussing what a 'proof' of a proposition in mathematics is, we think that it is fruitful to separate the concepts 'proof in a formalized system' and 'proof in mathematics'. Especially as we understand a proof of a sentence as an argument for the truth of this sentence. Some proofs in mathematics are possible to formalize in first order logic, but it is too restrictive to use 'proof' only of these. Assuming consistency, traditional proofs of Gödel's first incompleteness theorem establish the truth of the Gödel sentence. But it is not provable in a first order formalization of Peano Arithmetic that this sentence is true. What mathematicians have meant by 'proof' has varied through the years, although Euclid's Elements has been the long time standard. It is possible that the development of the infinitesimal calculus in the seventeenth and eighteenth centuries could not have taken place, if only rigorous proofs would have been accepted by the mathematical community.

To be clear, the concept of true used in this thesis is the ordinary Aristotelian, or Tarskian, 'true by correspondence'.

1.4 A Summary of the Two Papers of the Thesis

In the first paper we present two proofs of Chaitin's incompleteness theorem. These proofs are generalized, and the aim of our generalization of the first version of the theorem is to show that the constant c in the theorem depends strongly on how the Kolmogorov complexity is defined. This version of Chaitin's theorem reads as follows.

Theorem 1.1 *There is a constant c such that for every formal system $\langle T, p\rangle$, if $\langle T, p\rangle$ is sound and $\langle T, p\rangle$ proves $K(s) > n$, then $n < l(p) + c$.*

In the theorem p is the set of the axioms, and T are the rules of deduction of the formal system $\langle T, p\rangle$. The function $l(p)$ is the length of p, and $K(s)$ is the Kolmogorov complexity of the string s. The definition of K utilizes a fixed, universal Turing machine U, and K is defined as $K(x) = min\{l(p) : U(p) \downarrow = x\}$. The proof of the theorem utilizes a pair coding of a number and a string. It is shown that generalizing the length function used in the definition of K, and the pairing function gives a valid proof for a large class of choices. The dependence of c on the choices made is immediate.

We now turn to the second version of Chaitin's theorem.

Theorem 1.2 *There is a recursively enumerable set B with an infinite complement, such that for every axiomatized, sound theory T there are only finitely many n for which the formulae $n \notin B$ are both true and provable in T, although infinitely many such formulae are true.*

The set B is a simple set, that is the set B is non-recursive, recursively enumerable, and the complement of B is infinite and contains no recursively enumerable infinite subset. In fact $B = \{x : K(x) \leq f(x)\}$, where K is as above, and f is a function satisfying certain conditions to be specified later. In the generalization of this theorem we show that the length function in the definition of K can be replaced by a large class of functions giving rise to different simple sets B.

Our argument thus shows that also this form of Chaitin's incompleteness theorem may be shown under more general conditions. The construction, furthermore, gives rise to a wide class of simple sets.

In the second paper, which is the main paper of the thesis, we show that something positive can be said about the possibility of defining a measure of the power of extensions of Peano Arithmetic. In order to do this we use a fragment of the modal logic GL, the letterless modal sentences, and its Lindenbaum algebra. We define a probability-like measure of certain finite parts of this fragment, and thus assign numbers to the equivalence classes of this part of the Lindenbaum algebra of GL, or rather to representatives of the equivalence classes. Via a translation we uniquely embed this structure into the Lindenbaum algebra of Peano Arithmetic. The translation of a letterless sentence is called a constant sentence. Using the same measure we get a measure on the equivalence classes of the corresponding finite fragment of the Lindenbaum algebra of Peano Arithmetic. Further, we show how this measure can be extended to certain important non-constant sentences. Finally, we use the measure to define a measure also on some extensions of Peano Arithmetic of the type $PA + \Phi$, where Φ is a set of constant sentences.

We also discuss problems that must be dealt with if one tries to extend our measure to a larger class of extensions.

One conclusion of this paper, then, is that it is possible to design a measure m on certain extensions of PA. The constructed measure m takes rational numbers as values, and the measure is such that $m(S) \leq m(T)$, if $S \dashv T$.

2 A Note on Chaitin's Incompleteness Theorem

2.1 Introduction

Chaitin's theorem says roughly that for every sound formal system T there is a constant c such that T does not prove any true proposition of the form $K(x) > c$ although an infinitude of propositions of this form are true. The function K is the Kolmogorov complexity of the string x, and is to be defined in section 2.2. The result is not especially sensitive to how the complexity measure K is defined.[11] This suggests that the theorem can be generalized to a larger class of complexity measures. In this paper we show that this is indeed the case, and that the value of c depends on the measure chosen and not only on the theory T.

In section 2.2 we present a standard definition of Kolmogorov complexity. Section 2.3 presents one version of Chaitin's incompleteness theorem, and a generalized version of it using a more general complexity measure. In section 2.4 we present another version of the theorem where we make use of simple sets, and finally we show a more general version of this proof. The rest of this section presents some notation and terminology.

As usual we use r.e. for *recursively enumerable*. The complement of a set S is denoted S^c. A non-recursive set S is *simple* if it is r.e., S^c is infinite, and every r.e. set $D \subseteq S^c$ is finite. All sets considered are subsets of the natural numbers. A function is *recursive* if it is partial recursive and total. Throughout the paper the notation φ_i presupposes a fixed enumeration of the partial recursive functions. We write $\varphi_i(n)\downarrow$ ($\varphi_i(n)\uparrow$) when the partial recursive function φ_i converges (diverges) with input n, that is, if the function has respectively has not a value for input n, and $\varphi_i(n)\downarrow= m$, if $\varphi_i(n)$ converges and takes the value m. If φ and ψ are two partial recursive functions, then $\varphi(n) = \psi(n)$ ($\varphi(n) \leq \psi(n)$) is true if and only if $\varphi(n)\downarrow$, $\psi(n)\downarrow$, and $\varphi(n) = \psi(n)$ ($\varphi(n) \leq \psi(n)$), and false otherwise. The restriction of a function f to a set S is denoted $f \mid S$. We also use Turing machines in some of our arguments, but the arguments do not depend on any special feature of Turing machines, so any version will do. An arithmetical formula is Δ_0 if all of its quantifiers are bounded, and it is Σ_1 (Π_1) if it is logically equivalent to a formula of the form $\exists v\alpha$ ($\forall v\alpha$), where α is a Δ_0-formula. A formula is Σ_n (Π_n) if it is logically equivalent to a formula of the form $\exists v\alpha$ ($\forall v\alpha$) where α is Π_{n-1} (Σ_{n-1}). The *length l of a string x*, $l(x)$, is the number of symbols in it. We also assume that we have a bijection between the

[11] See for example Boolos and Jeffrey [BJ89], Odifreddi [Odi93] and Chaitin [Cha71, Cha74] for alternative ways of defining K.

natural numbers and the set of all binary strings $B^* = \{0,1\}^*$. The strings are ordered lexicographically, and the number 0 is paired with the empty string, 1 with the string 0, 2 with the string 1, 3 with the string 00, etc. This has the advantage that we can freely switch between reasoning about numbers and binary strings. In this way we can talk about the Kolmogorov complexity of a number as well as of a string. The length of a number is the length of the corresponding binary string. The base 2 logarithm is denoted log. The floor function, that is the largest number that is smaller than or equal to the argument, is denoted $\lfloor \cdot \rfloor$. This notation has as a consequence, that $l(n) = \lfloor \log(n+1) \rfloor$ where the first occurrence of n can be a string or a number, and the second is a number. In the proofs below we freely make use of elementary results from recursion theory as well as Church's thesis. Everything we use can be found in elementary texts on recursion theory as for example [Odi93]. The end of a proof is marked with a $\Box$.

2.2 Kolmogorov Complexity

Chaitin's theorem has achieved attention because it uses a new kind of construction for producing incompleteness results, and also because of the widespread misconceptions of the meaning of the constant c as some kind of measure of the information content of a formal system T. See for example van Lambalgen [vL89] and Raatikainen [Raa98] for devastating criticisms of these misunderstandings.

We start off by defining the Kolmogorov complexity of a binary string, and in our exposition we primarily follow Li and Vitànyi [LV93].

The idea behind the Kolmogorov complexity is that some binary strings have very short descriptions, while most strings have not. Consider for example the string $0101\ldots01 = (01)^{10^6}$, which contains two million binary digits, and can be described using just a few symbols. Then consider a random string consisting of two million bits. There seems to be no way of describing this random string in a shorter fashion than presenting the whole string. Strings which have short descriptions have low Kolmogorov complexity, while strings which only have long descriptions have high complexity.

Let x and p be binary strings. Any partial recursive function φ such that $\varphi(p) \downarrow= x$ is a *description of* x. The *Kolmogorov Complexity of a binary string* x is

$$K_\varphi(x) = \min\{l(p) : \varphi(p) \downarrow= x\}$$

where φ is a fixed universal partial recursive function. Since φ is fixed, we usually skip the subscript φ. The string p above can be thought of as a program that generates the string x.

We now proceed to state some properties of K. By the example discussed above, we see that there are strings with low Kolmogorov complexity, strings that are very compressible, in fact infinitely many. For every number n, there are 2^n binary strings of length n, and at most $2^n - 1$ shorter programs. Since there are only finitely many strings that are shorter than a constant c, there are infinitely many true sentences of the form $K(x) > c$. For each constant c, we call a string *c-incompressible*, if $K(x) \geq l(x) - c$. The number of strings that are c-incompressible is then $2^n - 2^{n-c} + 1$. The majority of the strings of length n, with $n > c$, are c-incompressible. For a set A of cardinality m, there are at least $m(1 - 2^{-c}) + 1$ elements x with $K(x) \geq \log(m) - c$, since there are $2^{\log(m)-c} - 1$ programs of length less than $\log(m) - c$. Furthermore, $K(x) > l(x)$ for infinitely many x, since φ not in general is the most effective description of x. We formulate some of these observations as a lemma.

Lemma 2.1 *(i) There are at least one binary string (number) x of length n such that $K(x) \geq n$.*

(ii) For every constant c there are infinitely many strings x such that $K(x) > c$.

(iii) Let c be a positive integer. Then every finite set A with cardinality m has at least $m(1 - 2^{-c}) + 1$ elements x with $K(x) \geq \log(m) - c$.

(iv) $K(x) > l(x)$ for infinitely many x.

As we noted in the introduction there are many alternative ways of defining the Kolmogorov complexity of a binary string or a number. To list a few we have according to Odifreddi [Odi93] the Kolmogorov complexity defined as $K(x) = \mu i(\varphi_i(0) \downarrow= x)$, where μ is the μ-operator. In [Cha71] Chaitin defines the Kolmogorov complexity of x as the least number of states a three tape Turing machine with empty input must have to produce x. In [Cha74] Chaitin uses the minimal length of a program that produces x on an abstract computer. Finally, in [BJ89] Boolos and Jeffrey define it as the least number of quadruples a one tape Turing machine with an empty string as input must have to produce x. There is not any intrinsic difference between the types of measures listed above. There are for example techniques for transforming multitape Turing machines into single-tape ones. When the index of a partial recursive function is known, it is possible to construct a Turing machine that computes the same function. It is thus no great wonder that the same type of results can be proved using the seemingly different definitions above.

2.3 A Generalized Version of Chaitin's Incompleteness Theorem

In this section we intend to prove a somewhat more general version of one form of Chaitin's incompleteness theorem. The generalization consists in letting the function used in defining the Kolmogorov complexity of a number be a general function with certain restrictions. We will first present the proof in [Cha74], and then discuss how this can be generalized. In his proof Chaitin uses the concept of an *abstract computer*, a concept which is essentially the same as that of a Turing machine. The definition of Kolmogorov complexity in [Cha74] is almost the same as in section 2.2, but instead of using a universal, partial recursive function Chaitin uses an abstract computer, and we will use, almost following Chaitin, a fixed universal Turing machine U.

$$K_U(x) = \min\{l(p) : U(p) \downarrow = x\}$$

We will suppress the subscript U, and use the same symbol for Kolmogorov complexity as above. The notation $U(p) \downarrow = x$ means that the Turing machine U with input p stops with output x. Also, as above, x is a binary string, $l(p)$ is the length of a program p, in binary, fed to the Turing machine U. But note that we could equally well think of x and p as numbers.

A universal Turing machine can simulate every Turing machine. What is needed is that a number of bits is reserved, e.g. in the beginning of the input, to identify the Turing machine that is to be simulated. How many bits that are needed depends on how the input is specified, and of how the coding of the Turing machines is made. This *cost* of simulating a Turing machine M on U is denoted $simM$. The exact value of $simM$ depends on the constructions, and we just note that a number such as $simM$ exists. As a pair coding of a number k and a string s, Chaitin uses the string 0^k1s. One standard way of specifying the input of a universal Turing machine U that simulates M, is to use a self-delimiting description of the code of M. The input then takes the form $0^{l(n)}1np$, where $M = T_n$ in an enumeration of the Turing machines. Then $simM = 2l(n) + 1$

In the proof below we also use the evident estimation $K(M(p')) = min\{l(x) : U(x) \downarrow = M(p')\} \leq l(p') + simM$, if $M(p') \downarrow$. Chaitin's theorem [Cha74] can now be formulated as follows.

Theorem 2.1 (Chaitin) *There is a constant c such that for every program p, if $U(p)$ generates a sentence of the form $K(s) > n$ only if it is true, and $U(p)$ outputs $K(s) > n$, then $n < l(p) + c$.*

In the formulation of the theorem, $U(p)$ is a Turing machine that successively generates strings with the program p as input.

Proof: Let C be a Turing machine that does the following with input $p' = 0^k 1p$. If $p' = 0^k$, C stops with no output. If $p' \neq 0^k$, then C simulates $U(p)$, and C lets $U(p)$ search for sentences of the form $K(s) > n$, where $n \geq l(p') + k$. If and when such a sentence is found, C outputs s and stops.

Now suppose that p and $U(p)$ satisfies the hypothesis of the theorem, and consider the computation $C(0^{simC}1p)$. If $C(0^{simC}1p) \downarrow= s$, we have

$$K(s) \leq l(0^{simC}1p) + simC = l(p) + 2simC + 1.$$

We also have

$$n \geq l(p') + k = l(0^{simC}1p) + simC = l(p) + 2simC + 1.$$

But

$$K(s) > n \geq l(p) + 2simC + 1$$

and we have a contradiction. Letting $c = 2simC + 1$ proves the theorem. □

As is obvious from the proof, the value of the constant c depends on the chosen coding of Turing machines, since the cost of simulating C on U is a term in c. Furthermore, the value of c also depends on the choice of complexity measure, based on the length function, and the coding of pairs.

The corresponding result can be proved, if we consider $\langle U, p \rangle$ as a formal system, where U are the rules of inference of a formal system, and p is a set of axioms. The theorem then reads: There is a constant c such that for all formal systems $\langle U, p \rangle$, if $\langle U, p \rangle$ is sound and $\langle U, p \rangle$ proves $K(s) > n$, then $n < l(p) + c$.

If we scrutinize the above proof, we see that a corresponding result can be proved if we define the complexity measure as

$$K_U^f(s) = \min\{f(p) : U(p) \downarrow= s\}$$

where f is a suitable recursive function. The restrictions we have to make on f is that K^f is unbounded, and that f does not grow too fast. We want the results in Lemma 2.1 still to be true. In the discussion below we use $\langle x, y \rangle$ as an abbreviation for a recursive function pairing the number x and the string y. The chosen pairing function must be invertible in both its arguments. The proof may now proceed as above.

As before we start a computation $C(\langle simC, p \rangle)$, and supposing that $C(\langle simC, p \rangle) \downarrow= s$ we get the following corresponding estimations. As before the exact value of $simC$ depends on the constructions chosen.

$$K^f(s) \quad = \quad K^f(C(\langle simC, p \rangle)) = min\{f(x) : U(x) \downarrow= C(\langle simC, p \rangle)\} \leq$$

$$\leq \quad f(\langle simC, p\rangle) + simC$$

and

$$n \geq f(\langle simC, p\rangle) + simC$$

Finally, because $K^f(s) > n$ is true, we have

$$K^f(s) > n \geq f(\langle simC, p\rangle) + simC$$

which is a contradiction. If the function f and the pairing function satisfy the condition

$$f(\langle simC, p\rangle) = f(p) + c'$$

for some constant c', we can choose $c = c' + simC$ and the generalization of Chaitin's theorem is proved. This also shows the dependance of c on f and the pairing function chosen. We are now in a position to formulate a generalization of Chaitin's theorem.

Theorem 2.2 *Let $K^f(s) = min\{f(p) : U(p) \downarrow= s\}$, where f and $\langle x, y\rangle$ satisfy the restrictions in the above discussion. There is then a constant c such that for every program p, if $U(p)$ generates a sentence of the form $K^f(s) > n$ only if it is true, and $U(p)$ outputs $K^f(s) > n$, then $n < f(p)+c$. Furthermore, the constant c depends on both the choice of f and the pairing function.*

Apparently Chaitin's theorem, in which $f(p) = l(p)$ and $\langle k, s\rangle = 0^k1s$, is a special case of this theorem. And the conclusion that c, as the constant comes out in the proof of this form of Chaitin's theorem, could measure the information content of a theory is not at all warranted, or is simply wrong.

It is not especially remarkable that it is possible to use other functions than the length of a binary string. As is well known, the proof rests on a syntactical version of Berry's paradox. The paradox is originally semantical and it can be derived considering the expression

> *the least natural number not nameable in fewer than 22 syllables.*

The syntactical counterpart, used in the proof of Chaitin's theorem, can be formulated

> *the least natural number not computable by a computer of complexity less than n.*

As is obvious, the paradox can be derived using a plentitude of formulations.

In [Raa98] Raatikainen discusses the constant c, and conclusively shows that c cannot measure the power of a formal system. Sometimes, reading Chaitin, one gets the impression that he means that the constant produced by his proof is a measure of the power of a formal system. In [Cha82] he for example contrasts Gödel's incompleteness theorem with his own and says that

> *(t)hese traditional proofs of Gödels's incompleteness theorem show that formal axiomatic systems are incomplete, but they do not suggest ways to measure the power of the formal axiomatic systems, to rank their degree of completeness or incompleteness.*

And it is understood that in Chaitin's incompleteness theorem this ranking is achieved.

Our result reinforces Raatikainen's criticism, and shows that the constant c, produced by Chaitin's proof, depends not only on the formal system, but on both the coding and the complexity measure chosen.

2.4 A Second Proof of Chaitin's Incompleteness Theorem and its Generalization

There is a variant of Chaitin's incompleteness theorem where the proof uses simple sets. In this section we present, following [LV93], a standard proof of this version of the theorem, and we then study how this proof can be generalized. We first list some of the properties of the Kolmogorov complexity K as it was defined in section 2.2.

Theorem 2.3 *(i) K is total, and $K(x) \leq l(x) + c$.*

(ii) The function K is not partial recursive.

(iii) The set $A = \{(x, a) : K(x) \leq a\}$ is r.e., but not recursive.

(iv) Every partial recursive function φ which is a lower bound of K is bounded.

(v) Let f be a recursive function with $g(x) \leq f(x) \leq \log(x)$ for all x, and some unbounded monotonic function g. Then the set $B = \{x : K(x) \leq f(x)\}$ is simple.

Proof: (i) Note that a string is always a description of itself. The constant c is the cost of using a fixed, universal partial recursive function in the definition of K instead of a possibly more effective one.

(ii) We show the stronger proposition that no partial recursive function φ, defined on an infinite set of points, can coincide with K over the whole of its domain.

Suppose that there is a partial recursive function φ such that $\varphi(x) = K(x)$ on an infinite set of points.

Since the domain of φ is r.e. and every infinite r.e. set contains an infinite recursive subset, we can select an infinite recursive subset S in $dom\varphi$, the domain of φ. Define

$$\psi(m) = \min\{x \in S : K(x) \geq m\}.$$

The function ψ is recursive, since $\varphi(x) = K(x)$ on S, and it is also unbounded. By the definition of ψ we have $K(\psi(m)) \geq m$, and by the definition of K we have $K(\psi(m)) \leq K_\psi(\psi(m)) + c_\psi$ for some constant c_ψ. Finally we have $K_\psi(\psi(m)) \leq l(m)$. Combining this gives us

$$m \leq K(\psi(m)) \leq l(m) + c_\psi = \lfloor \log(m+1) \rfloor + c_\psi$$

which is a contradiction for large enough m.

(iii) Let U be a universal Turing machine that computes the function φ in the definition of K. To decide if $(x,a) \in A$ for fixed (x,a), we run U for all p such that $l(p) \leq a$ successively in t steps, where $t = 1, 2, 3, \ldots$, and test whether $\varphi(p) \downarrow= x$. The procedure eventually stops in a finite number of steps if $(x,a) \in A$ since we need only test finitely many p. If $(x,a) \notin A$ the procedure never stops. The set A is thus r.e.

To show that A is not recursive, we suppose for contradiction that A is recursive. It is then possible to compute $K(x)$. By item (i) we know that $K(x) \leq l(x) + c$ for some fixed constant c. Using this bound we successively test whether $(x,0)$, $(x,1)$, etc belong to A until we find a value for $K(x)$. But this contradicts item (ii).

(iv) Let φ be a partial recursive function and define $D = \{x : \varphi(x) \leq K(x)\}$.

If D is finite, φ is surely bounded.

Suppose that D is infinite, and for contradiction that φ is unbounded. Recursively enumerate the domain of φ without repetition, and define a recursive function g by

$$g(n) = \text{the least } x \text{ in the enumeration such that } \varphi(x) \geq n.$$

By hypothesis there is such an x for every n. Since g is recursive there is an index k such that $g = \varphi_k$. We may then conclude that

$$n \leq \varphi(x) \leq K(x) \leq K(g(n)) = K(\varphi_k(n)) \leq l(n) + c,$$

where c is a constant depending on the definition of K and the cost of simulating φ_k on the universal, partial recursive function φ.[12] For large enough n we have a contradiction.

(v) That B is r.e. follows from item (iii).

That B^c, the complement of B, is infinite follows from the results presented in Lemma 2.1. With B^c finite only finitely many strings would have a complexity greater than $l(x) = \lfloor \log(x+1) \rfloor$.

Now suppose that D is a r.e. subset of B^c, and suppose for contradiction that D is infinite. The restriction of f to D, is a partial recursive lower bound for K. Therefore $f \mid D$ is bounded according to item (iv). Since f is unbounded and growing with x, we have a contradiction. The conclusion is that D is finite, and therefore B is simple. □

As a corollary of this theorem we have a version of Chaitin's theorem.

Corollary 2.1 *There is a r.e. set B with an infinite complement, such that for every axiomatized sound theory T there are only finitely many n for which the formula $n \notin B$ is both true and provable in T, although infinitely many such formulae are true.*

Proof: Let $B = \{x : K(x) \leq f(x)\}$ where f is the function in theorem 2.3, item (v), and let B^c be its complement. Define the set $D = \{n : T \vdash n \notin B\}$.

Clearly $D \subseteq B^c$, since T is sound. Since B is simple and D is r.e., D must be finite. □

Since the set B is r.e., the relation $n \notin B$, that is $f(n) < K(n)$, is Π_1. It is worth noting that this is just like the Gödel sentence constructed in the proof of Gödel's first incompleteness theorem.

To see how this result can be generalized we replace the length function with an increasing, recursive function f that does not grow too fast. As in section 2.2 we fix a universal partial recursive function φ and define the complexity of a number, or a binary string, as follows.

$$K^f(x) = \min\{f(p) : \varphi(p) \downarrow = x\}$$

It is essential that f is such that K^f satisfies the results discussed in connection with Lemma 2.1, so we assume that f are such that these conditions are fulfilled. This means e.g. that an infinity of strings have a low complexity, and that $K^f(x) > f(x)$ for infinitely many x.

[12] cf. the discussion in section 2.3 on simulating a Turing machine on a universal Turing machine.

The intention now is to formulate and prove some theorems that are parallel to the results in Theorem 2.3.

Theorem 2.4 *(i) The function K^f is total, and there is a constant c such that $K^f(x) \leq f(x) + c$ for every x.*

(ii) The function K^f is not partial recursive.

(iii) The set $A = \{(x,a) : K^f(x) \leq a\}$ is r.e. but not recursive.

Proof: (i) Let φ simulate identity function and let $p = x$. The constant c can then be seen as the cost of simulating the identity function on φ.

(ii) As above we prove that K^f cannot coincide with any partial recursive function φ_j on the whole of $dom\varphi_j$, if the domain is infinite. So, suppose that $K^f = \varphi_j$ on $dom\varphi_j$ for some index j, where the domain of φ_j is infinite. Further, let S be an infinite, recursive subset of $dom\varphi_j$. Define the recursive function

$$\psi(m) = \min\{x \in S : K^f(x) \geq m\}.$$

Since K^f is unbounded ψ is total, and since S is recursive ψ is too. According to the definition of ψ, $K^f(\psi(m)) \geq m$. But, letting φ simulate ψ, we have

$$m \leq K^f(\psi(m)) = \min\{f(p) : \varphi(p) \downarrow= \psi(m)\} \leq f(m) + c,$$

for some constant c. This gives a contradiction if f is chosen such that $f(m) + c < m$ for large enough m.

(iii) It is already clear that K^f is not recursive, and consequently A is not recursive. To show that A is r.e. we let, as above, U be a universal Turing machine that computes the function φ in the definition of K^f. To decide if $(x,a) \in A$ for fixed (x,a), we successively run U for all p such that $f(p) \leq a$ in t steps, where $t = 1,2,3,\ldots$, and test whether $\varphi(p) \downarrow= x$. The procedure eventually stops if $(x,a) \in A$ since we need only test finitely many p. If $(x,a) \notin A$ the procedure never stops. The set A is thus r.e.$\Box$

In the proof of Theorem 2.3, item (v), we used the fact that the function K is irregular, meaning that large arguments can have short descriptions, and that these occur, as it seems, at random. We now want a corresponding result in a more general version. What we wish to prove is that every partial recursive function θ that is a lower bound on K^f is bounded.

So we let θ be a partial recursive function such that $\theta(x) \leq K^f(x)$, and define $D = \{x : \theta(x) \leq K^f(x)\}$.

If D is finite we are finished, so suppose D is infinite, and assume that θ is unbounded.

Enumerate the domain of θ without repetition, and define a recursive function g.

$$g(n) = \text{the least } x \text{ in the enumeration such that } \theta(x) \geq n$$

Simulating g on φ gives

$$n \leq \theta(x) \leq K^f(x) \leq K^f(g(n)) = min\{f(p) : \varphi(p) \downarrow = g(n)\} \leq f(n) + c$$

for some constant c. To receive a contradiction we have to choose the function f such that $f(n) + c < n$ for large enough n. We then have a proof of the following theorem.

Theorem 2.5 *If the complexity measure f is chosen such that $f(n)+c < n$ for large enough n, then every partial recursive function θ that is a lower bound of K^f is bounded.*

The next result we want to prove is that the set $B = \{x : K^f(x) \leq h(x)\}$ is simple, if $g(x) \leq h(x) \leq f(x)$, where g is an unbounded, monotone function, and h is a recursive function.

That B is r.e. follows from Theorem 2.4 item (iii). Since we have chosen f such that $K^f(x) > f(x)$ for infinitely many x, then $B^c = \{x : K^f(x) > f(x)\}$ is infinite. Now let $D \subseteq B^c$, where D is r.e., and suppose for a contradiction that D is infinite. Then $h \mid D$ is a partial recursive lower bound to K^f. According to the above theorem $h \mid D$ is bounded. But since g is unbounded and monotone, and $g(x) \leq h(x)$ we have a contradiction. The following result is then clear.

Theorem 2.6 *The set $B = \{x : K^f(x) \leq h(x)\}$ is simple if $g(x) \leq h(x) \leq f(x)$ for some unbounded, monotone function g, and $f(x)+c < x$ for x large enough.*

We may finally claim that

Corollary 2.2 *Let $K^f(x) = \min\{f(p) : \varphi(p) \downarrow = x\}$, where K^f satisfies the conditions discussed in connection with Lemma 2.1, and that $f(x)+c < x$ for x large enough. Let $B = \{x : K^f(x) \leq h(x)\}$ where $g(x) \leq h(x) \leq f(x)$ for some unbounded, monotone function g. Then for every axiomatized sound theory T, extending PA, there are only finitely many n for which the formula $n \notin B$ is both true and provable in T, while infinitely many such formulae are true.*

Proof: As was shown in theorem 2.6 the set B is simple. Defining $D = \{n : T \vdash n \notin B\}$. We then have $D \subseteq B^c$, and since D is r.e., D must be finite. $\square$
Once again Chaitin's theorem is a special case. Choosing $f = l$ gives us $K = K^f$.

2.5 Conclusion

Our arguments show that incompleteness results of Chaitin's type can be proved under rather general conditions. Our result reinforces Raatikainens criticism of interpretations of Chaitin's theorem. Associated with a theory T is not one single constant c, but, depending on the complexity measure chosen there are lots of constants associated with a formal system T. And this of course means that c cannot measure the 'information content' of T.

3 Levels of Undecidability

3.1 Introduction

When we study extensions of elementary arithmetic, Peano arithmetic, or PA for short, we meet a very intricate structure. Efforts have been made to construct measures on for example the 'information content' of these extensions. To be interesting, a measure on a theory, or on the sentences of a theory, ought to be recursive, that is it should be possible to compute the measure of a theory or of a sentence. But if we succeeded in accomplishing this, we would have a decision procedure for theoremhood, and this we know is impossible.[13] The qustion then naturally arises if it is possible to say anything positive at all about measurements in complex structures such as these. In this paper we show that it is in fact possible to do so. We define a measure, which we call a provability measure, of an interesting fragment of PA. This measure is also used to create a measure on some extensions of elementary arithmetic. In doing this, we on one hand make something positive, on the other hand we more clearly than before state some problems that must be taken care of if we are to construct a more comprehensive measure than we do in this paper. We now turn to a well-known attempt to measure information content.

In the early 1970's Chaitin[14] announced his now famous incompleteness theorem saying roughly that for every sound formal system T there is a constant c such that T does not prove any true proposition of the form $K(n) > c$, although infinitely many propositions of this form are true. $K(n)$ is the Kolmogorov complexity of the string n. The Kolmogorov complexity of a string is usually described as the length of the shortest description of the string. The concepts 'length' and 'description' can be specified in many ways, and depending on how this is done we get different versions of the function K. Several variants of the proof appeared in the seventies. There also emerged an interpretation of the meaning of the constant c appearing in the theorem. The constant was supposed to measure the *information content* of the theory T. As the story goes, the constant c was thought to depend only on the theory T. This interpretation has been severely critized by van Lambalgen [vL89] and Raatikainen [Raa98], who showed that the value of the constant c depends on the Gödel numbering used. It can also be shown that the value of c depends on how the function K, measuring the Kolmogorov complexity, is defined. Furthermore, the constant c would

[13] This problem is also discussed in the introduction.
[14] See e.g. [Cha71].

generate a linear order of the extensions with respect to information content, and this is hardly reasonable. The unequivocal conclusion is that c cannot measure the information content of T.

As we said above, one aim of this paper is to show that the situation is not totally negative. A recent paper by Knight gave rise to an idea of how to use the modal logic GL to shed light on a fragment of PA.[15] Although our method is inspired by Knight's paper, it differs in its use of a modal logic, and in the results aimed at.

A modal, propositional logic is ordinary propositional logic extended with the symbol $\Box$, usually read as 'it is necessary that'. Additional axioms and rules of deduction take care of the use of $\Box$ in the calculus. Many logicians have been suspicious of modal logics, and it was not until the early 1970's, with the development of a reasonable semantics, that modal logic got a wider acceptance. The most prominent use of modal logic that has been developed since then is provability logic, usually named GL after Gödel and Löb. In this logic the box is read 'it is provable in PA that'.

With the use of the modal logic GL we identify levels in the Lindenbaum algebra of PA. The levels consists of equivalence classes of the so called constant sentences, the translations of letterless modal sentences. With the help of these levels it is possible to identify an order type. It is also possible to define a probability-like measure, which we will call a *provability measure*, on finite fragments of the set of equivalence classes. Finally, using the same ideas, we define a kind of measure on certain extensions of PA.

In this paper we connect some well-known facts in a new and interesting way. Sections 3.2 and 3.3 contain elementary results concerning PA and GL, and readers familiar with these results may skip these sections. The results are presented in order to make the paper somewhat more self-contained. In section 3.4 we discuss the letterless sentences of GL and their connection to the constant sentences of PA. We show how the set of traces of letterless sentences build up a Boolean algebra, which is isomorphically embeddable in the Lindenbaum algebra of PA. Section 3.5 presents a definition of a provability measure on constant sentences, and section 3.6 exhibits some ideas, including the fixed point theorem of GL, of how to extend this prov-

[15] In [Kni02] Knight discusses ways of measuring inconsistency. Using a finite propositional language he introduces a probability-like measure in order to measure how consistent a set of sentences is. A set containing an explicit inconsistency gets the measure zero, and a set that is not inconsistent gets the measure one. The other sets get intermediate values. In doing this Knight uses the disjunctive normal form for propositional sentences, and elementary results in probability theory. See e.g. [Par94] which is one of Knight's main sources. This book can also be consulted when reading section 3.5 of this paper.

ability measure to some non-constant sentences. Section 3.7 discusses some possible shortcomings in our construction. This discussion centers on partial conservativeness and partial Lindenbaum algebras. In section 3.8 we present what Boolos called 'extremely undecidable sentences', and show that these sentences cannot be caught in the structure we present. Finally, in section 3.9, we define a measure on some extensions of PA.

Concepts and notations are introduced as they are needed. At this point we only mention that we use φ, ψ etc. as meta-variables ranging over sentences of arithmetic, and A, B etc. as meta-variables ranging over sentences of the modal logic GL.

3.2 Peano Arithmetic

In this section we present most of the material concerning Peano Arithmetic needed for the sequel. The results are presented without proofs, but proofs of all of the results we cite may be found in Mendelson [Men87], or in some cases in Lindström [Lin97]. We presuppose a first-order logic with identity, and define the language of PA, $L_{PA} = \{S, +, \times, 0\}$. We also presume that we have the standard axioms directing the use of the symbols in L_{PA}, and furthermore the axiom scheme of induction. All this given, we can in the usual way, via a Gödel numbering, formalize the syntax of PA. To the relation $PRF(x, y)$, that is the relation

> y is the Gödel number of a proof in PA of a formula with the Gödel number x,

there corresponds in L_{PA} a formula $Prf(x, y)$ such that

$$PA \vdash Prf(m, n) \text{ iff } PRF(m, n) \text{ is true, and}$$

$$PA \vdash \neg Prf(m, n) \text{ iff } PRF(m, n) \text{ is not true.}$$

In the formula $Prf(m, n)$, the symbols m and n are formal numerals, i.e. $SS \ldots S0$, containing m respectively n occurrences of the symbol S, but we identify the numeral for the Gödel number of m, with the number m. The formula $Pr(x)$ is short for $\exists y Prf(x, y)$, and it expresses that x is (the Gödel number of) a formula that is provable in PA. In the sequel we will normally omit the phrase within parentheses above. We will also, following Lindström [Lin97], write e.g. $Pr(\varphi)$ instead of $Pr(\ulcorner \varphi \urcorner)$, where $\ulcorner \varphi \urcorner$ is the numeral for the Gödel number of the sentence, or formula, φ. This should not render any problem of reading. It should also be clear that we make a choice when we define the formula $Prf(x, y)$ since this formula depends on how the axioms of PA are enumerated. Different enumerations give different proof predicates. However, this choice does not affect the results discussed in this paper.

To symbolize a sentence, disprovable in PA, we use $\bot$, which can be understood as the sentence $0 = S0$. The sentence $\neg Pr(\bot)$ thus expresses that a contradiction is not provable in PA, i.e. that PA is consistent, and we use Con as short for $\neg Pr(\bot)$. On some occasions we will refer to other theories than PA, where a theory is understood as a set of its axioms, and we will then write $Pr_T(\varphi)$ expressing that φ is provable in the theory T, and analogically we write Con_T to express that the theory T is consistent. As above this involves a choice, since the proof predicate depends on how the axioms of T are enumerated.

We now proceed to state some facts concerning the provability predicate in PA. The results presented can be formulated in much stronger versions, but we will almost only work in PA, so we use the weaker variants stated below.

Theorem 3.1 (The Fixed Point Theorem of Arithmetic) *Let $\xi(x)$ be a formula with only one free variable x. There is then a sentence φ such that $PA \vdash \varphi \leftrightarrow \xi(\varphi)$.*

Using the fixed point theorem, we get a Gödel sentence γ for PA letting $\xi(x)$ be the formula $\neg Pr(x)$, that is $PA \vdash \gamma \leftrightarrow \neg Pr(\gamma)$. Thus, the Gödel sentence in a way states of itself that it is not provable in PA. In a problem, raised by Henkin in the early 1950's, he asked whether a sentence expressing its own provability is provable. Using the Hilbert-Bernays-Löb provability conditions Löb proved in the mid fifties what has become known as Löb's theorem. The provability conditions are the following propositions.

(P1) $PA \vdash \varphi \Rightarrow PA \vdash Pr(\varphi)$

(P2) $PA \vdash Pr(\varphi \to \psi) \to (Pr(\varphi) \to Pr(\psi))$

(P3) $PA \vdash Pr(\varphi) \to Pr(Pr(\varphi))$

Theorem 3.2 (Löb's Theorem) *For any sentence φ, if $PA \vdash Pr(\varphi) \to \varphi$, then $PA \vdash \varphi$.*

In fact, the following somewhat stronger arithmetical version of Löb's theorem is provable.

(L) $PA \vdash Pr(Pr(\varphi) \to \varphi) \to Pr(\varphi)$

To prove (L) Löb used the fixed point $PA \vdash \theta \leftrightarrow (Pr(\theta) \to \varphi)$ and the provability conditions.

A sentence φ is *provable* in PA, if $PA \vdash \varphi$, *disprovable*, if $PA \vdash \neg\varphi$, and *undecidable*, if $PA \nvdash \varphi, \neg\varphi$. A possible way to depict the relationship between sentences in L_{PA} is to study the Lindenbaum algebra of PA. Defining the relation $\leq$ on sentences by $\varphi \leq \psi$ iff $PA \vdash \varphi \to \psi$ we get a relation that is reflexive and transitive. Letting $\varphi \equiv \psi$ iff $\varphi \leq \psi$ and $\psi \leq \varphi$, we get an equivalence relation. The equivalence class, the *degree*, with representative φ is denoted $d(\varphi)$. Finally, letting the symbol $\leq$ also order the equivalence classes we get a partial ordering of the degrees. The strict ordering between degrees is defined $d(\varphi) < d(\psi)$ iff $d(\varphi) \leq d(\psi)$ and $d(\psi) \not\leq d(\varphi)$. Letting $d(\bot) = 0$, $d(\neg\bot) = 1$, $d(\varphi) \cup d(\psi) = d(\varphi \vee \psi)$ (the supremum or join) and $d(\varphi) \cap d(\psi) = d(\varphi \wedge \psi)$ (the infimum or meet) we get the Lindenbaum algebra of PA. The structure so defined is a dense, countable Boolean algebra, and is as such not especially interesting, since all such algebras are isomorphic. One idea behind the study of Lindenbaum algebras of theories was that different theories could have different Lindenbaum algebras. The Lindenbaum algebra of a consistent and complete theory consists of only two elements, but the algebras of incomplete extensions of PA are all isomorphic. Between the provable and the disprovable sentences, or more correctly between the degrees of $\bot$ and $\neg\bot$, are the undecidable sentences. We can e. g. formulate the facts that $0 = d(\bot) < d(\gamma) < d(\neg\bot) = 1$, where γ is a Gödel sentence for PA, and $0 = d(\bot) < d(Con) < d(\neg\bot) = 1$, stating that γ and Con are undecidable in PA. We will often, in the sequel, speak of representatives of degrees rather than the degrees. This means that we sometimes will write $\varphi \leq \psi$ for $d(\varphi) \leq d(\psi)$.

One aim of this paper is to show how to define levels at which, at least some, undecidable sentences reside, and how some undecidable sentences are related to these levels.

Sometimes we need to classify the complexity of arithmetical formulae, and following Kaye [Kay91] we define an arithemetical formula to be Δ_0 if all of its quantifiers are bounded. For convenience we also denote this class with Σ_0 and Π_0. A formula is Σ_{n+1} (Π_{n+1}) if it is provably equivalent in PA to a formula $\exists v\varphi$ $(\forall v\varphi)$, where φ is Π_n (Σ_n).[16] Saying that a formula is Γ_n, we mean that it is Σ_n or Π_n.

[16] This definition is not adequate for all purposes, but it is good enough for the discussions in this paper. In [Lin97] the Δ_0 formulae are identified with the *primitive recursive* arithmetical formulae. For details see [Lin97].

3.3 Provability Logic

The most impressive application of modal logic, originating in the early 1970's, though hinted at already by Gödel, is provability logic, the study of provability.[17] In this section we present the modal logic GL and some results relating GL and PA. Again, we present no proofs, but this time everything we use can be found in Boolos [Boo93].

We first present the syntax of GL. GL is a propositional modal logic, and we use p, q, r, etc as sentence letters. The well formed, modal sentences (ms) are defined as follows.

(1) $\perp$ is a ms,

(2) All sentence letters are ms,

(3) If A and B are ms, then $(A \to B)$ is a ms,

(4) If A is a ms, then $\Box A$ is a ms.

The other propositional connectives are introduced as abbreviations in the obvious way. Inductively we define $\Box^n A$ by $\Box^0 A = A$, and $\Box^{n+1} A = \Box\Box^n A$. The rules of deduction are Modus Ponens (MP), and Necessitation (Nec), which allow us to deduce $\Box A$ whenever A is deduced. Substitution instances of theorems can be shown to be theorems in GL. As axiom schemes we have

(A1) All instances of propositional tautologies,

(A2) $\Box(A \to B) \to (\Box A \to \Box B)$,

(A3) $\Box(\Box A \to A) \to \Box A$

If we compare the provability conditions $(P1) - (P3)$ with the above modal schemes, we see a structural similarity between (Nec) and $(P1)$, while $(A2)$ looks like $(P2)$. The arithmetical version of Löb's theorem is structurally like $(A3)$. In GL it is possible to prove $\Box A \to \Box\Box A$, which looks like $(P3)$. A *normal modal logic*, K, uses the inference rules MP and Nec together with the axiom schemes $(A1)$ and $(A2)$. In K it is also possible to show that substitution instances of theorems are theorems. Systems of propositional modal logic then differ in the additional axiom schemes used, and e.g. $GL = K + (A3)$. It is not possible to substitute the Löb scheme

[17] For some historical details on the emergence of provability logic, see [BS91].

$(A3)$ for $\Box A \to \Box\Box A$, since a normal system together with $\Box A \to \Box\Box A$ does not prove $(A3)$.

Using ordinary Kripke semantics for modal logic, it is possible to prove that GL is sound and complete with respect to appropriate frames, and furthermore, that these frames can be chosen finite. We thus have that $GL \vdash A$ iff A is valid in all finite, transitive and irreflexive frames.

We connect GL with PA in the following way. A *realization* is a function $*$ which to every sentence letter p in the language of GL assigns a sentence φ in the language of PA, i.e. $*(p) = \varphi$. Inductively we then define a *translation* A^* according to the following.

(1) $p^* = *(p)$ for all sentence letters p

(2) $\bot^* = \bot$

(3) $(A \to B)^* = A^* \to B^*$

(4) $(\Box A)^* = Pr(A^*)$

A translation thus preserves the connectives, and two realizations can only differ in the way they assign sentences of arithmetic to sentence letters of GL. For any translation $*$ we have $(\neg\Box\bot)^* = \neg Pr(\bot)$, that is Con. The strong connections between GL and PA are expressed in the following theorems.

Theorem 3.3 (Arithmetical Soundness) *For any realization $*$, if $GL \vdash A$, then $PA \vdash A^*$.*

Theorem 3.4 (Arithmetical Completeness) *For every A, there is a realization $*$ such that, if $GL \nvdash A$, then $PA \nvdash A^*$.*

The realization in the arithmetical completeness theorem may depend on A.

A modal formula is *letterless* if it does not contain any sentence letter. For letterless modal sentences the following immediate corollary holds.

Theorem 3.5 *For any letterless sentence A and any realization $*$, $GL \vdash A$ iff $PA \vdash A^*$.*

The translation of a letterless sentence is called a *constant* sentence. Examples of constant sentences are the *iterated consistency sentences* defined inductively.

$Con(1, T) = Con_T$

$$Con(n+1,T) = Con(1, T + Con(n,T))$$

These constant sentences are realizations of respectively $\neg\Box\bot$ and $\neg\Box^{n+1}\bot$. Note that $PA \vdash Con(k, PA) \rightarrow Con(m, PA)$, if $0 < m < k$, and compare this with $GL \vdash \Box^m\bot \rightarrow \Box^k\bot$, i.e. $GL \vdash \neg\Box^k\bot \rightarrow \neg\Box^m\bot$.

3.4 Letterless Modal Sentences

We now focus on the letterless modal sentences, and denote the set of these LMS. A sentence in LMS is in *normal form* if it is a truth-functional combination of sentences $\Box^i\bot$, $i \geq 0$. For letterless modal sentences we have the following normal form theorem.

Theorem 3.6 (Normal Form Theorem) *For every $A \in LMS$, there is a $B \in LMS$, such that B is in normal form, and $GL \vdash A \leftrightarrow B$.*

In a way, then, sentences $\Box^i\bot$ are a kind of building blocks, but since they are not independent they are not especially useful as such for the aim of this paper. The *(modal) degree* of a modal sentence A is the maximum number of nested occurences of the symbol $\Box$ in A. If, in the normal form theorem, A is of degree n, an analysis of the proof of the normal form theorem shows that B is of a degree less than or equal to n.[18]

We then consider the Lindenbaum algebra of GL restricted to sentences in LMS, and this algebra is constructed just like the Lindenbaum algebra of PA. As before, we identify sentences with their equivalence classes, and often write $A \leq B$ instead of $d(A) \leq d(B)$. We denote the set of equivalence classes $\overline{LMS}$, and for clarity we state the following properties.

$$d(A) \leq d(B) \text{ iff } GL \vdash A \rightarrow B$$

$$d(A) < d(B) \text{ iff } d(A) \leq d(B) \text{ and } d(A) \not\leq d(B)$$

$$d(A) = d(B) \text{ iff } d(A) \leq d(B) \text{ and } d(B) \leq d(A) \text{ iff } GL \vdash A \leftrightarrow B$$

$$d(\bot) = 0,\ d(\neg\bot) = 1$$

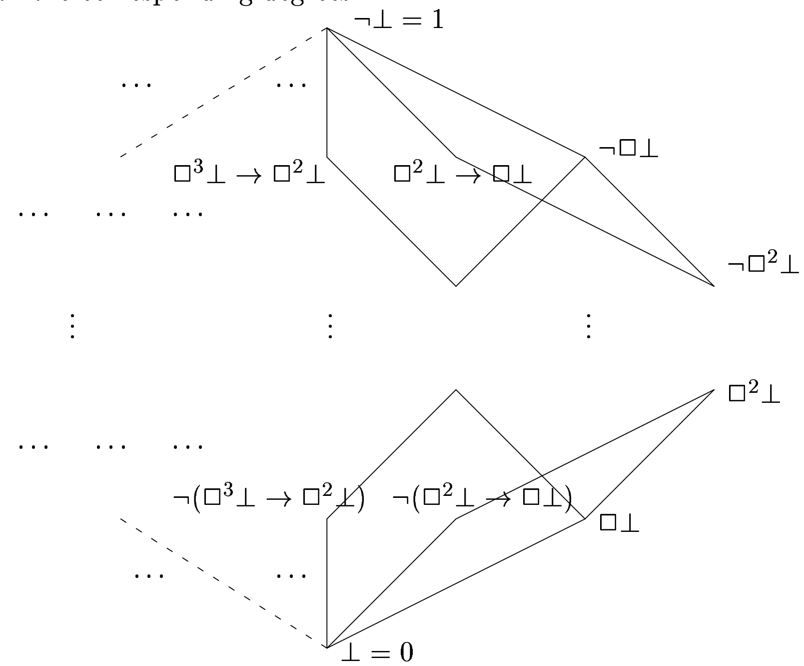
The supremum and infimum of two degrees are respectively $d(A) \cup d(B) = d(A \vee B)$ and $d(A) \cap d(B) = d(A \wedge B)$.

To compute e.g. $d(\Box\bot) \cap d(\neg(\Box^2\bot \rightarrow \Box\bot))$, we note that $GL \vdash (\Box\bot \wedge \neg(\Box^2\bot \rightarrow \Box\bot)) \leftrightarrow \bot$, and conclude that $d(\Box\bot) \cap d(\neg(\Box^2\bot \rightarrow \Box\bot)) = 0$. To compute $d(\Box\bot) \cup d(\neg(\Box^2\bot \rightarrow \Box\bot))$, we use the fact that $GL \vdash$

[18]See [Boo93] pp92f.

$(\Box\bot \vee \neg(\Box^2\bot \to \Box\bot)) \leftrightarrow \Box^2\bot$, and conclude $d(\Box\bot) \cup d(\neg(\Box^2\bot \to \Box\bot)) = d(\Box^2\bot)$.

In the figure below the Lindenbaum algebra of the letterless sentences of GL is hinted at. It is an infinite graph, and the dots and dashes indicate how to complete the graph. The vertices are, for brevity, labelled with sentences and not with the corresponding degrees.



As a kind of building blocks we can use the sentences $\Box^{n+1}\bot \wedge \neg\Box^n\bot$ where $n \geq 0$. The corresponding degrees are atoms of the Boolean algebra (the Lindenbaum algebra). Disjunctions and conjunctions of these sentences can be used to generate all sentences on the 'lower' half. Negations of these sentences give the sentences on the 'upper' half. In this way we have a kind of normal form for letterless modal sentences using as atoms the atoms of the Boolean algebra.

The Löb scheme is clearly provable in GL, that is $GL \vdash \Box(\Box A \to A) \to \Box A$. By normality, i.e. in K, it is possible to prove $GL \vdash \Box A \to \Box(\Box A \to A)$. Combining these results gives $GL \vdash \Box(\Box A \to A) \leftrightarrow \Box A$. Setting $\bot$ for A we thus can prove $GL \vdash \Box(\neg\Box\bot) \leftrightarrow \Box\bot$ that is $d(\Box(\neg\Box\bot)) = d(\Box\bot)$.

Using tricks like these it is possible to compute where a certain degree belongs in the Lindenbaum algebra, but this can be tedious, and also, there is no need for it, since there is a totally mechanical procedure of computation. To present this technique we first define the *trace of a letterless sentence.*

The trace of A, $t(A)$, is defined inductively via the following conditions, where we use S^c for the complement of S relative to ω, the natural numbers.

$t(\perp) = \emptyset$

$t(A \to B) = t(A)^c \cup t(B)$

$t(\Box A) = \{n : \forall i < n \; i \in t(A)\}$

We can, then, compute that e.g. $t(\neg A) = t(A \to \perp) = t(A)^c$, $t(\neg\perp) = \omega$, $t(A \vee B) = t(A) \cup t(B)$, $t(A \wedge B) = t(A) \cap t(B)$, $t(\Box^n \perp) = \{0, 1, \ldots, n-1\}$, $t(\Box^{n+1}\perp \to \Box^n \perp) = \{n\}^c$, $t(\neg(\Box^{n+1}\perp \to \Box^n\perp)) = \{n\}$, etc.

Furthermore, we have the following theorem on properties of the trace of a letterless sentence.

Theorem 3.7 *For letterless sentences A and B the following propositions are valid*

(1) $GL \vdash A$ *iff* $t(A) = \omega$

(2) $GL \vdash \neg A$ *iff* $t(A) = \emptyset$

(3) $GL \vdash A \to B$ *iff* $t(A) \subseteq t(B)$

(4) $GL \vdash A \leftrightarrow B$ *iff* $t(A) = t(B)$

(5) $t(A)$ *is either finite or cofinite*

If we consider the set

$$S^t = \{t(A) : A \in LMS\} = \{X \subseteq \omega : X \text{ is finite or cofinite}\},$$

the set of traces, and apply the usual set operations we get a Boolean algebra. From the theorem above and the definition of the Lindenbaum algebra of the fragment of GL that consists of the letterless sentences, we have that $A \leq B$ iff $t(A) \subseteq t(B)$. The supremum (infimum) of two sentences A and B corresponds to $t(A) \cup t(B)$ ($t(A) \cap t(B)$), $\perp$ ($\neg\perp$) corresponds to $\emptyset$ (ω). It is then clear that the Boolean algebra on S^t is isomorphic to the Boolean algebra on the set $\overline{LMS}$. Furthermore, this algebra is isomorphically embeddable in the Lindenbaum algebra of PA.[19]

With this we can have a clear picture of levels of undecidability in PA. Also, note that the constant sentences which are realizations of sentences in LMS are Boolean combinations of Σ_1 sentences, and, consequently, all are in B_1, the set of Boolean combinations of Σ_1 sentences. Hence, this Boolean algebra is isomorphically embeddable in any partial Lindenbaum algebra for PA that includes B_1.

[19]Boolos mentions the embeddability of $\overline{LMS}$ into PA in [Boo82].

3.5 On Measuring Levels of Undecidability

An *algebra over* X, $X \neq \emptyset$, is a non-empty set $\mathcal{A}$ such that $\mathcal{A} \subseteq \mathcal{P}(X)$, and $\mathcal{A}$ is closed under complementation and finite unions. An algebra is a σ-*algebra* if it is closed under countable unions as well. As an example we can mention that $S^t = \{X \subseteq \omega : \text{X is finite or cofinite}\}$ is an algebra but not a σ-algebra, since it is closed under finite, but not under countable unions. A *measure* on a set X, equipped with a σ-algebra $\mathcal{A}$, is a function $m : \mathcal{A} \to [0, \infty]$ such that

(m1) $m(\emptyset) = 0$

(m2) If $\{x_i\}_{i=1}^{\infty}$ is a sequence of pairwise disjoint sets, then $m(\bigcup_{i=1}^{\infty} x_i) = \sum_{i=1}^{\infty} m(x_i)$

The condition $(m2)$ implies finite additivity, that is

(m2') If $x_i \cap x_j = \emptyset$, then $m(x_i \cup x_j) = m(x_i) + m(x_j)$

The measure is a *probability measure* if $m : \mathcal{A} \to [0, 1]$ such that

(pm1) $m(X) = 1$

(pm2) $x_i \cap x_j = \emptyset \Rightarrow m(x_i \cup x_j) = m(x_i) + m(x_j)$

(pm3) $x_i \supseteq x_j, 0 < i < j, \bigcap_{i=1}^{\infty} x_i = \emptyset \Rightarrow \lim_{i \to \infty} m(x_i) = 0$

The condition $(pm3)$ is equivalent to $(m2)$. In measure theory one presupposes that measures are defined on σ-algebras, so there is no hope to apply measure theory to the set S^t defined above. To see where problems emerge, try to define a (probability) measure on S^t, i.e. a function $p : S^t \to [0, 1]$ that satisfies the conditions above, and note that the Boolean algebra on S^t is isomorphic to the Boolean algebra on $\overline{LMS}$. The function p then satisfies $m(\omega) = 1$, and the condition $(pm2)$ above. It seems natural to assign the same positive measure to all atoms. Letting $p(\{n\}) = \varepsilon > 0$ for all n, implies that $p(\{0, 1, \ldots, k\}) = \sum_{i=0}^{k} p(\{i\}) > 1$ for large enough k, i.e. $k > 1/\varepsilon$.

On the other hand, since every formula has a finite length, we could consider finite fragments of S^t, or, isomorphically, of $\overline{LMS}$. Thus, confining ourselves to finite fragments of S^t, we define $S^t_n = \{X : X \subseteq Z_n \text{ or } X^c \subseteq Z_n\}$, where $Z_n = \{0, 1, \ldots, n-1\}$. With standard set operations, S^t_n is a Boolean algebra, and it is furthermore, trivially, a σ-algebra, since it is finite and closed under complements and unions. For example we have for $n = 3$

$$S^t_n = \{\emptyset, \{0\}, \ldots, \{0, 1, 2\}, \omega, \{0\}^c, \ldots, \{0, 1, 2\}^c\}$$

Correspondingly we define $LMS_n = \{A : t(A) \in S_n^t\}$ and $\overline{LMS}_n = \{d(A) : t(A) \in S_n^t\}$

We can now define a probability-like measure on S_n^t, or equivalently on LMS_n or $\overline{LMS}_n$ in the following way

$$m_n : S_n^t \to [0, 1]$$

$$m_n(\omega) = 1$$

$$x_i \cap x_j = 0 \Rightarrow m_n(x_i \cup x_j) = m_n(x_i) + m_n(x_j)$$

Equivalently we can define

$$m_n : LMS_n \to [0, 1]$$

$$GL \vdash A \Leftrightarrow m_n(A) = 1$$

$$GL \vdash \neg(A \wedge B) \Rightarrow m_n(A \vee B) = m_n(A) + m_n(B)$$

We will call these measures *provability measures*,[20] because our intention is to measure how close to being provable a sentence is. It is also obvious that it cannot be a measure of closeness to truth, because e.g. sentences can be true without being provable in PA. Finally, we identify the provability measure of elements in $\overline{LMS}_n$, with the provability measures of its representatives. Since these three structures are isomorphic, we will freely use the idiom that most easily, in any particular situation, expresses the ideas we try to communicate. It is straightforward to prove the following theorem, and since it is not common to present the result in this context, we provide a proof.

Theorem 3.8 *For sentences $A, B \in LMS_n$ the following assertions are valid.*

(1) $m_n(\neg A) = 1 - m_n(A)$

(2) $GL \vdash \neg A \Leftrightarrow m_n(A) = 0$

(3) $GL \vdash A \to B \Rightarrow m_n(A) \leq m_n(B)$

(4) $GL \vdash A \leftrightarrow B \Rightarrow m_n(A) = m_n(B)$

(5) $m_n(A \vee B) = m_n(A) + m_n(B) - m_n(A \wedge B)$

[20] We make some further comments on the terminology used in the concluding remarks.

Proof: (1) Since $GL \vdash A \vee \neg A$ and $GL \vdash \neg(A \wedge \neg A)$, we have $1 = m_n(A) + m_n(\neg A)$.

(2) $GL \vdash \neg A$ iff $m_n(\neg A) = 1$ iff $m_n(A) = 0$

(3) Supposing $GL \vdash A \rightarrow B$ we get $GL \vdash \neg(A \wedge \neg B)$, and so $1 \geq m_n(A \vee \neg B) = m_n(A) + m_n(\neg B) = m_n(A) + 1 - m_n(B)$. Hence $m_n(A) \leq m_n(B)$.

(4) is now immediate.

(5) Tautologically we have $GL \vdash A \vee B \leftrightarrow (A \wedge \neg B) \vee B$, and so $m_n(A \vee B) = m_n((A \wedge \neg B) \vee B)$. Also, tautologically $GL \vdash \neg((A \wedge \neg B) \wedge B)$, and consequently, $m_n((A \wedge \neg B) \vee B) = m_n(A \wedge \neg B) + m_n(B)$. Furthermore, since $GL \vdash A \leftrightarrow (A \wedge \neg B) \vee (A \wedge B)$, we have $m_n(A) = m_n((A \wedge \neg B) \vee (A \wedge B))$. Finally, $GL \vdash \neg((A \wedge \neg B) \wedge (A \wedge B))$, and so $m_n((A \wedge \neg B) \vee (A \wedge B)) = m_n((A \wedge \neg B) + m_n(A \wedge B))$. When combining these observations, we get the desired result. $\Box$

It now seems reasonable to assign 'almost' disprovable sentences (the atoms of $\overline{LMS}_n$ or S^t_n) small values. But, note that with n singleton sets we can not define $m_n(\{i\}) = 1/n$, since then $m_n(Z_n) = 1$. Instead, we choose to make assignments such that $0 < m_n(Z_n) = a < 1/2$, and define $m_n(\{i\}) = a/n$ for $0 \leq i < n$, or equivalently $m_n(d(\neg(\Box^{i+1}\bot \rightarrow \Box^i\bot))) = m_n(\neg(\Box^{i+1}\bot \rightarrow \Box^i\bot)) = a/n$ for $0 \leq i < n$. This implies e.g. that $m_n(\neg\Box\bot) = 1 - a/n$.

Assigning measures to singleton sets in this way satisfies the axioms on the previous side, and the assignment thus constitutes a provability measure in this sense.

To get a measure on sentences of PA we let $LMS^*_n = \{A^* : A \in LMS_n\}$, and $\overline{LMS}^*_n = \{d(A^*) : d(A) \in \overline{LMS}_n\}$. We use m^*_n as a measure on the (degrees of the) constant sentences. Using the fact that $\overline{LMS}_n$ is embeddable in the Lindenbaum algebra of PA via a translation, and letting $m^*_n(d(A^*)) = m_n(d(A))$, we get a measure on the constant sentences. This also gives us a measure of Con close to 1. This seems reasonable enough since PA proves the consistency of every finite fragment of PA. Formally, $PA \vdash Con_{PA|k}$ for every k. The symbol $PA|k$ denotes the set consisting of axioms of PA with Gödel numbers less than k.

A *reflection* sentence for PA is a sentence $Pr(\varphi) \rightarrow \varphi$ where φ is a sentence in L_{PA}. The *local reflection principle* for PA is the set $Rfn = \{Pr(\varphi) \rightarrow \varphi : \varphi \text{ is a sentence in } L_{PA}\}$. Adding reflection to PA, adds soundness to PA and is a local, or piecemeal, way of saying that φ is true, if φ is provable.[21] It is an interesting result that the letterless reflection sentences $\Box^{i+1}\bot \rightarrow \Box^i\bot$, $0 \leq i < n$, all have the measure $1 - a/n$.

[21] See e.g. [Lin97] and [Bek97] on reflection.

Although it is not possible to define a measure, in the measure theoretic sense, on $\overline{LMS}$, or on the set S^t, it is possible to identify levels in the structure. Let the finite sets in S^t with the same cardinality be at the same level, and let the cofinite sets be at the same level if their complements have the same cardinality. For the letterless sentences we can define the levels in the corresponding way. The order type, with the obvious ordering, of these levels in $\overline{LMS}$ is $\omega + \omega^*$. This means that $\bot$ is at level 0, degrees in $\overline{LMS}$ that correspond to singleton sets in S^t are at level 1, degrees corresponding to n element sets are situated at level n etc. On the 'upper' half of the structure the situation is reversed. We thus have established a gradation on $\overline{LMS}$.

3.6 Computing Provability Measures

Our main technique to compute provability measures, or levels in the Lindenbaum algebra of PA, uses the fixed point theorem of modal logic. We use $\boxdot A$, often called the *strong box*, as short for $\Box A \wedge A$. A sentence A is *modalized* in p iff every occurance of p is within the scope of some $\Box$ symbol.

Theorem 3.9 (The Fixed Point Theorem of Modal Logic) *If A is modalized in p, then*

$$GL \vdash \boxdot(p \leftrightarrow A) \leftrightarrow \boxdot(p \leftrightarrow H)$$

where H only contains sentence variables occuring in A with the exception of p.

To illustrate the fixed point theorem, we use it to prove the wellknown fact that $PA \vdash \gamma \leftrightarrow Con$, where γ is a Gödel sentence for PA. We first note that if p is the only sentence letter in A, then H is letterless and H^* is a constant sentence. In this case it is easy to determine H with a truth-table-like method.[22] If e.g. A is the sentence $\neg\Box p$, then H can be chosen as $\neg\Box\bot$. Now, letting $*$ be such that $p^* = \gamma$, we have $PA \vdash (p \leftrightarrow \neg\Box p)^*$, since $PA \vdash \gamma \leftrightarrow \neg Pr(\gamma)$. Successively we get, $PA \vdash (\Box(p \leftrightarrow \neg\Box p))^*$, $PA \vdash (\boxdot(p \leftrightarrow \neg\Box p))^*$, $PA \vdash (\boxdot(p \leftrightarrow \neg\Box\bot))^*$, by the fixed point theorem. And finally $PA \vdash (p \leftrightarrow \neg\Box\bot)^*$, i.e. $PA \vdash \gamma \leftrightarrow Con$. As a consequence we have that $d(\gamma) = d(Con)$ in the Lindenbaum algebra of PA.

[22]See e.g. Boolos [Boo93] pp110f. In [SV82] a general method to compute fixed points to sentences A modalized in p is described. There are also sentences that are not modalized in p that have fixed points, e.g. $\Box p \to p$ has a fixed point p.

The sentence γ is not a constant sentence, but it is possible to extend m_n^* to some non-constant sentences (see below), and it is then natural to set $m_n^*(\gamma) = 1 - a/n$.

To illustrate the method of determining the letterless sentence H in a special case of the fixed point theorem, the case where the only sentence letter in A is p, we need some concepts. If $\langle W, R\rangle$ is a finite, transitive and irreflexive frame, there is for every $w \in W$ a greatest n such that for some sequence of worlds $w_n, \ldots, w_1, w_0 \in W$, where $w = w_n R \ldots R w_1 R w_0$. For each $w \in W$ we define the *rank* of w, $r(w)$, as the greatest such n. A sentence is called a *p sentence* if its only sentence letter is p. We now generalize the concept of trace and define the *A-trace* of B, $t_A(B)$, for each p sentence B. Choose one enumeration of all p sentences $B_0, B_1, \ldots$ in which p comes after A, and in which truth-functional compounds always come after their components. Inductively we define

$$t_A(\bot) = \emptyset$$

$$t_A(B \to C) = t_A(B)^c \cup t_A(C)$$

$$t_A(\Box D) = \{m : \forall i < m \; i \in t_A(D)\}$$

$$t_A(p) = t_A(A)$$

Facts concerning the concept A-trace parallels those of the concept trace. A fixed point of A is true iff $t_A(A)$ is cofinite, it is provable iff $t_A(A) = \omega$. To illustrate the technique we determine the fixed point of $A = \neg\Box p$.

	$\Box p$	$\neg\Box p$	p
0	$\top$	$\bot$	$\bot$
1	$\bot$	$\top$	$\top$
2	$\bot$	$\top$	$\top$

The lines of the table corresponds to ranks of worlds. In worlds of rank 0 all sentences $\Box D$ get the value $\top$. Truth-functional compounds inherit their truth-value on a line from their components. $\Box D$ gets the value $\bot$ on a line, if D has the value $\bot$ on an earlier line. p gets the same value as A. In the example above lines 1 and 2 are equal, so there is no need to proceed, nothing new can happen. We conclude that $t_A(\neg\Box p) = \{0\}^c = t(\neg\Box\bot)$, and that $\neg\Box\bot$ is provably equivalent to the fixed point of $\neg\Box p$. We give one more example of the method, and determine the fixed point of $A = \neg\Box p \wedge \Box\neg p$.

	$\Box p$	$\Box\neg p$	$\neg\Box p$	$\neg\Box p \wedge \Box\neg p$	p	$\neg p$
0	$\top$	$\top$	$\bot$	$\bot$	$\bot$	$\top$
1	$\bot$	$\top$	$\top$	$\top$	$\top$	$\bot$
2	$\bot$	$\bot$	$\top$	$\bot$	$\bot$	$\top$
3	$\bot$	$\bot$	$\top$	$\bot$	$\bot$	$\top$

Lines 2 and 3 are identical, and we conclude that $t_A(\neg\Box p \wedge \Box\neg p) = \{1\} = t(\neg(\Box^2\bot \rightarrow \Box\bot))$, and that H can be chosen as $\neg(\Box^2\bot \rightarrow \Box\bot)$.

With the technique described, it is possible to compute where translations of certain sentences are in the hierarchy, and at the same time we get the provability measure of the sentence. This is possible for p sentences A that are modalized in p and such that $GL \vdash p \leftrightarrow A$. This is an important type of sentences, since its translations are such that $PA \vdash \varphi \leftrightarrow \xi(\varphi)$, where $\xi(x)$ is built up using truth-functional compounds of the provability predicate $Pr(x)$.

We now make some remarks on the possibility of extending m_n^* to non-constant sentences. There is no problem in adding a sentence like γ to the domain of m_n^*, and the only reasonable measure to assign to γ is $m_n^*(Con)$. This situation is valid for any fixed point φ such that $PA \vdash \varphi \leftrightarrow \xi(\varphi)$, since in this case it is possible, with the help of the fixed point theorem for GL, to find a constant sentence H^* that is provably equivalent to φ. The measure assigned to φ is $m_n^*(H^*)$. The situation gets much more complicated if we try to extend m_n^* to arbitrary sentences φ such that $PA \vdash \varphi \leftrightarrow L^*$, where L is a letterless sentence. Formally we could do this, but we would loose the recursiveness of m_n^*, since there is no decision procedure for theoremhood in PA. We cannot simply in the general case compute the measure of such a sentence φ, since it is impossible to identify it. We will use the same function symbol m_n^* for recursive extensions of m_n^*, which originally was defined on LMS_n^*.

In some cases it is possible to relate more complex fixed points to the Lindenbaum algebra than those equivalent to a constant sentence. As an example, we discuss Rosser sentences. Let ρ be a Π_1-Rosser sentence for PA, i. e. a sentence ρ such that

$$PA \vdash \rho \leftrightarrow \forall x(Prf(\rho, x) \rightarrow \exists y \leq x Prf(\neg\rho, x))$$

It is a well-known fact that $PA \vdash Con \rightarrow \neg Pr(\neg\rho)$. Furthermore, since $PA \vdash \bot \rightarrow \neg\rho$, by the provability conditions we have $PA \vdash Pr(\bot) \rightarrow Pr(\neg\rho)$, that is $PA \vdash \neg Pr(\neg\rho) \rightarrow Con$. Hence $PA \vdash Con \leftrightarrow \neg Pr(\neg\rho)$. Also, since $\neg\rho \in \Sigma_1$, by provable Σ_1-completeness, $PA \vdash \neg\rho \rightarrow Pr(\neg\rho)$, and it is clear that $PA \vdash Con \rightarrow \rho$, that is $d(Con) \leq d(\rho)$. Supposing that $PA \vdash \rho \to Con$, we get $PA + \rho \vdash Con$. From the initially stated fact, it follows that $PA \vdash Con \to \neg Pr(\rho \to \bot)$, which implies that $PA + \rho \vdash Con_{PA+\rho}$, which contradicts Gödels second incompleteness theorem. Hence $PA \nvdash \rho \to Con$. We have thus proved the following theorem.

Theorem 3.10 $d(\gamma) = d(Con) < d(\rho) < 1$.

It is now possible to extend m_n^* to give ρ a measure. The Lindenbaum algebra of PA is, as we already mentioned, dense. As a consequence of this there are infinitely many degrees between $d(Con)$ and 1. Since the only thing we know about $d(\rho)$ is that it is strictly between those two degrees we can consistently assign any rational number between $1 - a/n$ and 1, and thus letting $1 - a/n < m_n^*(\rho) < 1$. Consequently this extension is not unique. This pointwise type of extension is always possible, but if we try do treat general sets of formulae in this way we will have problems with the recursiveness of our measure.

Compare this result with the 'unwitnessed' Rosser sentence, that is a sentence φ such that $PA \vdash \varphi \leftrightarrow (Pr(\varphi) \to Pr(\neg\varphi))$. Using the fixed point theorem, and the technique described above, we can prove that $d(\varphi) = d((\Box^2\bot \to \Box\bot)^*)$.

To conclude this section it is, thus, possible to compute the provability measure of constant sentences. It is also possible to extend this measure to sentences φ such that $PA \vdash \varphi \leftrightarrow \xi(\varphi)$, where $\xi(\varphi)$ is as above. In some cases, it is also possible to extend the provability measure to sentences, such as the Rosser sentence.

3.7 Partial Conservativity and Partial Lindenbaum Algebras

A sentence φ is Γ_n conservative over a theory T if for every Γ_n sentence θ, if $T + \varphi \vdash \theta$, then $T \vdash \theta$. It is a well-known fact that $\neg Con$ is Π_1 conservative over PA. This implies that the sentences $\neg Con(n, PA)$, for $n > 1$, also are Π_1 conservative over PA. Suppose that $PA + \neg Con(n, PA) \vdash \pi$ for $\pi \in \Pi_1$. Then, since $GL \vdash \Box\bot \to \Box^n\bot$, $PA + \neg Con \vdash \pi$. Thus $PA \vdash \pi$. With the same argument we can prove that for any sentence $A \in LMS$ such that $\{0\} \subseteq t(A)$, A^* is Π_1 conservative over PA. Thus Π_1 conservativity is inherited upwards in the structure.

This could be seen as a possible obstacle to the intuitive ideas presented above, since $\neg Con$ is highly unprovable, $m_n(\neg Con) = a/n$. On the other hand, the only sentence in LMS provable in GL whose translation is a Π_1 sentence is $\neg\bot$, so the ideas are consistent so far. It is also interesting that, in PA, the highly unprovable sentence $\neg Con$ does not add any force to PA concerning provability of Π_1 sentences.

Since PA is true, it is also clear that the Π_1 sentence Con is Σ_1 conservative over PA. Assume that $PA + Con \vdash \sigma$, where $\sigma \in \Sigma_1$. Since PA is true, $Con \to \sigma$ is true, and then σ is true because Con is true. Thus $PA \vdash \sigma$, since PA proves all true Σ_1 sentences. The same type of argument shows that the Π_1 sentences $Con(n, PA)$ also are Σ_1 conservative over PA. Furthermore, all of the sentences $\neg Con(n+1, PA) \to \neg Con(n, PA) = (\Box^{n+1}\bot \to \Box^n\bot)^* \in B_1$, for $n \geq 1$, are both Σ_1 and Π_1 conservative over PA. Assuming that $PA + \neg Con(n+1, PA) \to \neg Con(n, PA) \vdash \theta$, we get for $\theta \in \Pi_1$, that $PA + \neg Con(n, PA) \vdash \theta$, and finally $PA \vdash \theta$. For $\theta \in \Sigma_1$, $PA + Con(n+1, PA) \vdash \theta$, and thus $PA \vdash \theta$. The sentences not yet treated that are translations of sentences with cofinite trace are also Σ_1 conservative over PA. They are all true, since they are conjuncions of true sentences, and then the argument proceeds as before.

The sentences $\neg(Con(n+1, PA) \to Con(n, PA))$, $n \geq 1$, are neither Σ_1 nor Π_1 conservative over PA. This is so, because $PA + \neg(Con(n+1, PA) \to Con(n, PA)) \vdash Con(n+1, PA)$, but $PA \nvdash Con(n+1, PA)$. And also $PA + \neg(Con(n+1, PA) \to Con(n, PA)) \vdash \neg Con(n, PA)$, and $PA \nvdash \neg Con(n, PA)$. We summarize these observations in the following theorem.

Theorem 3.11 *We have the following results concerning conservativeness over PA for constant sentences. All sentences $Con(n, PA)$, $n \geq 1$, are Σ_1 conservative. The sentences $\neg Con(n+1, PA) \to \neg Con(n, PA)$, $n \geq 1$, are both Σ_1 and Π_1 conservative. Generally, sentences that are translations of sentences with cofinite traces are Σ_1 conservative. The sentences $\neg Con(n, PA)$, $n \geq 1$, are Π_1 conservative over PA, and, more generally, for any sentence $A \in LMS$ such that $\{0\} \subseteq t(A)$, A^* is Π_1 conservative. And, finally, the sentences $\neg(Con(n+1, PA) \to Con(n, PA))$, $n \geq 1$, are neither Σ_1 nor Π_1 conservative over PA.*

A standard technique to produce non-trivial, partially conservative sentences over PA is to use the fixed point $PA \vdash \varphi \leftrightarrow \exists x(\Gamma_n(x) \wedge Pr(\varphi \to x) \wedge \neg Tr_{\Gamma_n}(x))$. In this formula $\Gamma_n(x)$ is a formula expressing that x is a Γ_n sentence. The formula $Tr_{\Gamma_n}(x)$ is a partial truth predicate, and for $\varphi \in \Gamma_n$ we have $PA \vdash \varphi \leftrightarrow Tr_{\Gamma_n}(\varphi)$. In the fixed point φ is Σ_n if Γ_n is Π_n. To construct non-trivial, partially conservative sentences over theories T that are not true, requires more complex fixed points.[23] It is, however, not easy

[23] See [Lin97], pp65f.

to see how Γ_1 fixed points as constructed above could be incorporated in the structure that are under discussion.

In his discussion of partial Lindenbaum algebras, Bennet presents a concept *much smaller than (with respect to cup (cap))*, originally introduced by Lindström.[24] Following the notation of Bennet, we let $\underline{\Sigma}_n^{PA}$ ($\underline{\Pi}_n^{PA}$) be the partial Lindenbaum algebras where the sentences in the degrees are confined to sentences provably equivalent in PA to Σ_n (Π_n) sentences. We let a, b, c range over degrees and define

$a <<_\cup b$ iff $a < b$ and $\forall c(c \cup a \geq b \Rightarrow c \geq b)$

$a <<_\cap b$ iff $a < b$ and $\forall c(c \cap b \leq a \Rightarrow c \leq a)$

$a << b$ iff $a <<_\cup b$ and $a <<_\cap b$

From this definition it follows that the following facts are valid in $\underline{\Sigma}_n^{PA}$ ($\underline{\Pi}_n^{PA}$)

$0 <<_\cup d(\varphi)$ iff $0 < d(\varphi)$

$d(\varphi) <<_\cap 1$ iff $d(\varphi) < 1$

$0 <<_\cap d(\varphi)$ iff φ is Π_n (Σ_n) conservative over PA

$d(\varphi) <<_\cup 1$ iff $\neg\varphi$ is Σ_n (Π_n) conservative over PA

Since $Con \in \Pi_1$, and Con ($\neg Con$) is Σ_1 (Π_1) conservative over PA, both $d(Con) << 1$ and $0 << d(Con)$ are valid in $\underline{\Pi}_1^{PA}$. In $\underline{\Sigma}_1^{PA}$ we have that $0 << d(\neg Con)$ and $d(\neg Con) << 1$.

The reading of $a << b$ as 'much smaller than' is not to be taken at face value. Adding more structure, more degrees, as when considering e.g. $\underline{\Sigma}_2^{PA}$ or $\underline{\Pi}_2^{PA}$, has as a consequence that degrees that are far apart in $\underline{\Sigma}_1^{PA}$ or $\underline{\Pi}_1^{PA}$, need not be far apart any more.[25] One possible conclusion is that the concept *much smaller than* is of no relevance to the problems discussed in this paper.

3.8 Extremely Undecidable Sentences

In the arithmetical completeness theorem for GL the realization depends on the sentence A. In [Boo82] Boolos proves that there is a realization $*$ that works uniformly for every sentence A. In this context he introduces

[24] Bennets discussion can be found in [Ben86]. The concept was introduced by Lindström 1979 in his discussion of degrees of interpretability. See e.g. [Lin97].

[25] See [Ben86] pp65f.

a concept ‘extremely undecidable sentence’. The definition goes like this. As before a sentence A is a *p sentence* if its only sentence letter is p. A sentence φ over L_{PA} is *extremely undecidable* in PA, if for every p sentence A, if there is a realization $\#$ such that $PA \nvdash A^{\#}$, then $PA \nvdash A^*$, where $p^* = \varphi$. An alternative formulation is that φ is extremely undecidable in PA, if there is a realization $*$ such that $*(p) = \varphi$ and $PA \vdash A^*$ implies that $PA \vdash A^{\#}$ for every realization $\#$ and every p sentence A. This means, citing Boolos [Boo82], that

> roughly speaking, a sentence is extremely undecidable if it can be proved to have only those modal-logically characterizable properties that every sentence can be proved to have.

Boolos also proves that there are infinitely many extremely undecidable sentences, but note that neither the Gödel sentence γ nor the Rosser sentence ρ are extremely undecidable. It is now an interesting fact that the only (degrees of) constant sentences that are related to (the degree of) an extremely undecidable sentence are $d(\bot)$ and $d(\neg\bot)$. We formulate this observation as a theorem.

Theorem 3.12 *Let φ be an extremely undecidable sentence, and $L \in LMS$ such that $GL \nvdash L, \neg L$. Then $\varphi \nleq L^*$, $L^* \nleq \varphi$, $\neg\varphi \nleq L^*$, and $L^* \nleq \neg\varphi$*

Proof: Letting ρ be a Rosser sentence for PA, we will first prove that $PA \nvdash \rho \to L^*$. Suppose for a contradiction that $PA \vdash \rho \to L^*$. Since $t(L) \subseteq t(\Box^{n+1}\bot \to \Box^n\bot)$ for some $n \geq 0$, $GL \vdash L \to (\Box^{n+1}\bot \to \Box^n\bot)$. Therefore $PA \vdash \rho \to (\Box^{n+1}\bot \to \Box^n\bot)^*$. Since $PA \vdash Con \to \rho$, we conclude that $PA \vdash (\neg\Box\bot \to (\Box^{n+1}\bot \to \Box^n\bot))^*$, $n \geq 0$, but this contradicts that $t(\neg\Box\bot) \nsubseteq t(\Box^{n+1}\bot \to \Box^n\bot)$ for $n \geq 1$. The case when $n = 0$ is already clear, since $PA \nvdash \rho \to Con$. The conclusion is that $PA \nvdash \rho \to L^*$. By this argument it is also clear that $PA \nvdash L^* \to \neg\rho$.

Now, letting $A = p \to L$, and $\#$ a realization such that $p^{\#} = \rho$ it follows that $PA \nvdash (p \to L)^{\#}$. Since φ is extremely undecidable, $PA \nvdash \varphi \to L^*$, where $p^* = \varphi$, and so $\varphi \nleq L^*$. Letting $A = L \to p$, and $\#$ such that $p^{\#} = \neg\rho$, we conclude $PA \nvdash (L \to p)^{\#}$, and consequently $PA \nvdash L^* \to \varphi$, i.e. $L^* \nleq \varphi$. Proceeding in the same way we set $A = \neg p \to L$ and $\#(p) = \neg\rho$, which implies that $PA \nvdash A^{\#}$. Once again, since φ is extremely undecidable, $PA \nvdash \neg\varphi \to L^*$, where $*(p) = \varphi$, and hence $\neg\varphi \nleq L^*$. Finally, letting $A = L \to \neg p$ and $\#(p) = \rho$, and arguing in the same way it is clear that $L^* \nleq \neg\varphi$. $\Box$

This theorem gives us the interesting property of extremely undecidable sentences that they are not related to any constant sentences, with the exception of $\perp$ and $\neg\perp$. It is interesting in its own right that it is not possible to decide where these sentences are situated in the structure we study.

3.9 Comparing Theories

With the means developed above, there is a method to compare, and even to construct, a kind of measure on some extensions of PA in an illuminating way. Consider an extension T of PA, notationally $PA \dashv T$, obtained by adding constant sentences, or sentences provably equivalent to constant sentences, to PA. Letting $T = PA + \Phi$, where Φ is a set of constant sentences, two cases emerge. Each element $\varphi \in \Phi$ is the translation of some letterless sentence A. As before $t(A)$ denotes the trace of A.

At first we will deal with the case when Φ is finite. Determine the infimum of Φ as $t(\Phi) = \bigcap_{A^* \in \Phi} t(A)$, the *trace of* Φ, and associate this set with T. The set $t(\Phi)$ is in this case an element of S^t, the set of traces. Just as above we can identify levels of order type $\omega + \omega^*$ in the set of theories $\{PA + \Phi : \Phi \text{ is finite}\}$. To get a kind of two-dimensional measure, we can also identify a 'direction' where $PA + \Phi$ reside. If we restrict the sentences in Φ to be translations of sentences in LMS_n, or sentences provably equivalent to such translations, it is also possible to assign a probability-like measure to $PA + \Phi$ using the value assigned to $t(\Phi)$ as we did above. If e.g. $T = PA + Con$ we associate $\{0\}^c$ with T, and T can, just as in the case of the Lindenbaum algebras, be assigned the value $1 - a/n$. PA itself is associated with ω, or the value 1, and $PA + \perp$ with $\emptyset$ or the measure 0.

The case when Φ is an arbitrary set of constant sentences is much more intricate. As before we define the infimum of Φ as $T(\Phi) = \bigcap_{A^* \in \Phi} t(A)$, and associate this set with T. But $T(\Phi)$ need not be an element of S^t, and the complexity of sets $T(\Phi)$ ranges over sets that are e.g. not recursively enumerable, r.e. for short. Still, $T(\Phi)$ is an element of $\mathcal{P}(\omega)$, and this Boolean algebra is isomorphically embeddable in the set of all extensions of PA. Furthermore, the set

$$\{PA + \Phi : \Phi \text{ is a set of constant sentences}\}$$

is isomorphic to the Boolean algebra $\mathcal{P}(\omega)$. It is not possible to define levels in $\mathcal{P}(\omega)$ as we did in the finite case, and it is not possible to assign a probability-like measure, if we want all the atoms of $\mathcal{P}(\omega)$ to have the same positive value.

If Φ is r.e., we can use a theorem by Lindström[26] to get a characterization of $PA+\Phi$. Since Φ is a set of constant sentences it is a subset of B_1, and so it is also in Σ_2, and Π_2. We write $S \dashv_{\Pi_2} T$ to denote that S is a Π_2-subtheory of T, i.e. every Π_2 sentence provable in S is provable in T.

Theorem 3.13 *Let Φ be a r.e. set of constant sentences. There is then a Σ_2 sentence θ such that $PA + \Phi \dashv PA + \theta \dashv_{\Pi_2} PA + \Phi$.*

In the theorem we can choose θ to be a Π_2 sentence, and then $PA + \Phi \dashv PA + \theta \dashv_{\Sigma_2} PA + \Phi$. The sentence θ can be effectively constructed and is 'below' every sentence in Φ. Next we associate θ with the set $T(\Phi)$ as defined above. The condition that Φ is r.e. can, according to Craig's theorem, be weakened to Φ being primitive recursive. Craig's theorem says that for every r.e. theory S, that is a theory whose set of axioms are r.e., there is a primitive recursive theory T such that S and T have the same theorems.

In both cases there are, of course, infinitely many theories between two theories $T_1 \dashv T_2$, T_1 being a proper subtheory of T_2. We know that $PA \dashv PA+\rho \dashv PA+\gamma$, proper, so a measure of $PA+\rho$ should get a value between $1 - a/n$ and 1, which of course is reasonable.

In a way, then, we thus have accomplished at least something of what Chaitin aimed at in his interpretation of the constant c in Chaitin's incompleteness theorem.

3.10 Concluding Remarks

It is important to realize that the measures above not in any reasonable sense can measure 'closeness' to truth. Arithmetical sentences are true or false, nothing in between. We use the term 'probability-like' just to indicate that we use the axioms in measure theory that are normally used to design a theory of probability. But our term 'provability' measure may also be misleading. The sentence Con is not provable in PA. In one way, discussed above, it is almost provable, but the measure we have defined is thought to indicate where in the structure e.g. Con is situated. It is close to sentences that are provable. The measure, and the identified levels, give an ordering of closeness to the provable or disprovable. The measures we have constructed on extensions of PA do not measure closeness to truth either. Some of the theories are true, that is true in the standard model, some of them are not true. In this case it is hardly reasonable to call the measure a provability

[26]Theorem 4, p 66 in [Lin97].

measure. In a way we measure the 'proving capability', or power, of theories since theories high up in the structure can prove less than theories low down.

To conclude we have in this paper shown that it is possible to define a provability measure of sentences in an important fragment of the Lindenbaum algebra of Peano Arithmetic, the sentences in the set LMS_n^*. We have also shown how to define levels in a larger fragment, the set of constant sentences. Finally, we have said something positive on the possibility to compare some extensions of Peano Arithmetic, and even how these extensions can be given kinds of measures. We have also shown that we, using our construction, cannot catch e.g. extremely undecidable sentences. In order to construct a full measure of e.g. information content, as Chaitin has tried, the structures, problems and possibilities discussed here have to be taken into consideration.

References

[Bek97] Lev Beklemishev. Notes on Local Reflection Principles. *Theoria*, 63:139–146, 1997.

[Ben86] Christian Bennet. *On Some Orderings of Extensions of Arithmetic*. Dept. of Philosphy, Univ of Göteborg, 1986.

[BJ89] George S. Boolos and Richard C. Jeffrey. *Computability and Logic*. Cambridge University Press, 3 edition, 1989.

[Boo82] George Boolos. Extremely Undecidable Sentences. *The Journal of Symbolic Logic*, 47:191–196, 1982.

[Boo93] George Boolos. *The Logic of Provability*. Cambridge University Press, 1993.

[BS91] George Boolos and Giovanni Sambin. Provability: the Emergence of a Mathematical Modality. *Studia Logica*, 50:1–23, 1991.

[Cha71] Gregory J. Chaitin. Computational Complexity and Gödel's Incompleteness Theorem. *ACM SIGACT News*, 9:11–12, 1971.

[Cha74] Gregory J. Chaitin. Information-Theoretic Limitations of Formal Systems. *Journal of the ACM*, 21:403–424, 1974.

[Cha82] Gregory J. Chaitin. Gödel's Theorem and Information. *International Journal of Theoretical Physics*, 22:941–954, 1982.

[Dum78] Michael Dummet. *Truth and other Enigmas*. Duckworth, 1978.

[Hof79] Douglas R. Hofstadter. *Gödel, Escher, Bach: An Eternal Golden Braid*. Basic Books, 1979.

[Kay91] Richard Kaye. *Models of Peano Arithmetic*. Oxford University Press, 1991.

[KL68] G. Kreisel and A. Levy. Reflection Principles and their use for Establishing the Complexity of Axiomatic Systems. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 14:97–142, 1968.

[Kni02] Kevin Knight. Measuring Inconsistency. *Journal of Philosophical Logic*, 31:77–98, 2002.

[Lin97] Per Lindström. *Aspects of Incompleteness*. Springer, 1997.

[LV93] Ming Li and Paul Vitányi. *An Introduction to Kolmogorov Complexity and its Applications.* Springer, 1993.

[Men87] Elliot Mendelson. *Introduction to Mathematical Logic.* Wadsworth, 3 edition, 1987.

[Odi93] Piergiorgio Odifreddi. *Classical Recursion Theory.* Springer, 1993.

[Par94] J.B. Paris. *The Uncertain Reasoner's Companian: A Mathematical Perspective.* Cambridge University Press, 1994.

[Pen89] Roger Penrose. *The Emperor's New Mind.* Oxford University Press, 1989.

[Raa98] Panu Raatikainen. On Interpreting Chaitin's Incompleteness Theorem. *Journal of Philosophical Logic*, 27:569–586, 1998.

[Raa01] Panu Raatikainen. Exploring Randomness and The Unknowable, Review article. *Notices of the AMS*, 48:992–996, 2001.

[Ruc95] Rudy Rucker. *Infinity and the Mind.* Princeton University Press, 1995.

[SV82] Giovanni Sambin and Silvio Valentini. The Modal Logic of Provability. The Sequential Approach. *Journal of Philosophical Logic*, 11:311–342, 1982.

[vL89] Michiel van Lambalgen. Algorithmic Information Theory. *Journal of Symbolic Logic*, 54:1389–1400, 1989.

[Web80] Judson Chambers Webb. *Mechanism, Mentalism, and Metamathematics.* D. Reidel Publishing Company, 1980.