Philosophical Communications
Red series No. 41
ISSN 0347-5794

Metamathematical fixed points

Rasmus Blanck

Thesis for the degree of Licentiate of Philosophy
Göteborg 2011

UNIVERSITY OF GOTHENBURG
PHILOSOPHY, LINGUISTICS & THEORY OF SCIENCE

Thesis for the degree of Licentiate of Philosophy
University of Gothenburg 2011

Metamathematical fixed points
RASMUS BLANCK

©Olof Rasmus Blanck, 2011

Philosophical communications
Red series No. 41
ISSN 0347-5794

Distribution:
Department of Philosophy, Linguistics & Theory of Science
University of Gothenburg
Box 200
SE-405 30 Göteborg
Sweden

Printed in Sweden by
Reprocentralen
Humanistiska fakulteten
University of Gothenburg
Göteborg, Sweden 2011

till Nina

Sammanfattning

Denna text handlar om metamatematiska fixpunkter. Efter en kort introduktion ger vi en översikt över det fält som kallas metamatematik, från förra sekelskiftet till idag. Vi intresserar oss särskilt för begreppet fixpunkter, satser som visar att vissa typer av fixpunkter existerar och deras tillämpningar inom metamatematik. Andra halvan av texten är en teknisk undersökning av fixpunktsmängder. Givet en rekursivt enumerabel, konsistent extension T av Peanos aritmetik definierar vi, för varje formel $\theta(x)$, mängden

$$\mathrm{Fix}^T(\theta) := \{\delta : T \vdash \delta \leftrightarrow \theta(\delta)\}.$$

Vi bevisar att alla sådana mängder är Σ_1-fullständiga. Vidare definierar vi för varje formel $\theta(x)$, mängden

$$\mathrm{Fix}^T_\Gamma(\theta) := \{\delta : T \vdash \delta \leftrightarrow \theta(\delta)\},$$

där δ är en Γ-sats. Med hjälp av metoder som härrör från Bennet, Bernardi, Guaspari, Lindström och Smullyan karaktäriserar vi dessa mängder för formler i $\Gamma' \supset \Gamma$, och bevisar delresultat för formler i Γ. Vi ger ett tillräckligt villkor för att en rekursiv mängd skall vara en fixpunktsmängd och visar att sådana mängder existerar. Vidare ger vi ett tillräckligt villkor för att en rekursivt enumerabel mängd av Γ-satser skall vara en fixpunktsmängd till en Γ-formel.

I nästa avsnitt studerar vi strukturen av fixpunktsmängder ordnade under mängdinklusion, och beskriver vissa egenskaper hos dessa strukturer. Slutligen kopplar vi våra resultat till ett annat öppet problem inom metamatematik, och föreslår vidare arbete.

Nyckelord: Aritmetiserad metamatematik, fixpunkt, rekursionsteori.

Abstract

This thesis concerns the concept of metamathematical fixed points. After an introduction, we survey the field of metamathematics, from *la fin du siècle* to present. We are especially interested in the notion of fixed points, theorems on the existence of various kinds of fixed points, and their applications to metamathematics. The second part of the thesis is a technical investigation of sets of fixed points. Given some recursively enumerable, consistent extension T of Peano arithmetic, we define for each formula $\theta(x)$ the set

$$\mathrm{Fix}^T(\theta) := \{\delta : T \vdash \delta \leftrightarrow \theta(\delta)\}.$$

Our main result on these sets is that they are all Σ_1-complete. Furthermore, we define for each formula $\theta(x)$, the set

$$\mathrm{Fix}^T_\Gamma(\theta) := \{\delta : T \vdash \delta \leftrightarrow \theta(\delta)\},$$

where δ is a sentence in Γ. Using methods of Bennet, Bernardi, Guaspari, Lindström, and Smullyan, we characterise these sets for formulas in $\Gamma' \supset \Gamma$, and provide partial results for formulas in Γ. We give a sufficient condition on recursive sets to be a set of fixed points, and show that such sets exists. We also present a sufficient condition for a recursively enumerable set of Γ-sentences to be a set of fixed points of a Γ-formula.

In the following section, we study the structure of sets of fixed points ordered under set inclusion, and prove certain properties on these structures. Finally, we connect our research to another open problem of metamathematics, and state some possible further work.

Keywords: Arithmetised metamathematics, fixed point, recursion theory.

Acknowledgements

First of all, I would like to thank my supervisors Christian Bennet, Fredrik Engström and Dag Westerståhl. I am especially grateful to Christian Bennet—who taught me everything I know about metamathematics—for letting me include a number of his unpublished results in this thesis. I want to thank the participants of the logic seminar at the Department of Philosophy, Linguistics and Theory of Science at the University of Gothenburg, for the possibility to discuss my work in a friendly *milieu*. Furthermore, I owe thanks to *Stiftelsen Anna Ahrenbergs fond för vetenskapliga m. fl. ändamål*, which has provided for me during the preparation of this thesis. I also wish to thank Petter Remen and Niklas Ottosson, who introduced me to logic, and Niklas Rudbäck for invaluable and confusing discussions on this and other topics.

Contents

Chapter 1

Introduction

"The Cretans, always liars."

In the 6th century B.C., Epimenides of Knossos accused every Cretan of always lying. Unfortunately, Knossos is located on Crete, so Epimenides himself was a Cretan. It is then reasonable to ask how to interpret this claim. If we suppose that Epimenides told the truth, evidently, he was not lying. As he was a Cretan, we will have to conclude that the claim he made was false, as he did not lie. But as he made a false claim, he cannot have told the truth. So we have established that Epimenides' claim was false, and nothing is particularly strange about that. It simply means that there is one Cretan that occasionally speaks the truth.

However, Epimenides' claim can be seen as a precursor of the liar paradox, originating with Eubulides of Miletus in the 4th century B.C.:

> This sentence isn't true.

The truth value of the sentence above is harder to decide. Indeed, suppose the sentence is true. Then the sentence expresses a true proposition, namely the proposition that the sentence at hand is false. Thus the sentence is false. But if the sentence is false, then it expresses the false proposition that the sentence is false. Thus, it is true. We can then conclude that the sentence above is both true and false. Eubulides' sentence is the first known example of a self-referential sentence, one that makes claims about itself.

Self-referential constructions of this kind are somewhat disturbing. They allow us to construct seemingly well-formed sentences that we cannot determine

the truth value of. Assuming self reference to be unproblematic, it seems that anomalities of this kind points out a deficiency in our understanding of, in this case, the concept of truth. Even though it should be perfectly straightforward to interpret the liar sentence, there is no coherent way to give it a truth value. There are many different versions of paradoxes in the literature, of which we will here present but a few examples. The first two are traditionally classified as semantical paradoxes, whereas the last falls under the category of set-theoretical paradoxes.

- Grelling's paradox: A word is heterological if it does not describe itself. For example, "long" is heterological, as it is not long. "English" is not heterological, as it is an English word. It follows that "heterological" is heterological if and only if it is not heterological.

- Berry's paradox: Consider the expression "the least number that cannot be referred to in less than fifteen words". Since there is a finite number of expressions less than fifteen words long, there are only finitely many numbers that can be referred to by such expressions. However, there are infinitely many integers, so there are numbers that cannot be referred to by such an expression. By the well-ordering principle, there is a least such number, whence this fourteen-word expression refers to a number that cannot be referred to in less than fifteen words.

- Russell's paradox: R is the set of all sets that are not members of themselves. A contradiction arises when one asks whether R is a set that is a member of itself.

In the analysis of this kind of phenomena, a variety of different approaches have been taken. As it turns out, self reference is closely related to the mathematical (in a most general sense) concept of fixed points. In mathematical practice, many functions and operators have fixed points, a notion that can be explained in the following vague way:

> If O is some operation defined on a class of objects X, then a fixed point of O is an element x in X such that $Ox = x$.

We consider a few examples from different parts of mathematics and logic:

- Real analysis: The function $f(x) = x$, defined on the real numbers, has infinitely many fixed points, as every real is a fixed point of f. On the other hand, the function $g(x) = x^2$ has only two fixed points: $x = 0$ and $x = 1$.

- Algebraic topology: Brouwer's fixed point theorem in its simplest formulation—in two dimensions—states that every continuous function from a disc onto itself has a fixed point. Rumour has it, that Brouwer discovered the theorem when stirring a cup of coffee, noticing that there was always a point on the surface that did not move.

- Recursion theory: The recursion theorem states that given an enumeration of all recursive functions of n variables $f_0, f_1, \ldots$, and a recursive function of $n+1$ variables $g(z, x_0, \ldots x_n)$, there is a number e such that $f_e(x_0, \ldots, x_n)$ gives the same result as $g(e, x_0, \ldots, x_n)$.

- Modal logic: Let $\boxdot A$ be the sentence $\Box A \wedge A$. We say that a sentence A is modalised in p if every occurrence of the sentence letter p in A is within the scope of some occurrence of a $\Box$. The fixed point theorem for Gödel-Löb provability logic then says that for every sentence A, modalised in p, there is a sentence H containing only sentence letters from A, not containing p and such that $\mathrm{GL} \vdash \boxdot(p \leftrightarrow A) \leftrightarrow \boxdot(p \leftrightarrow H)$.

- Category theory: Recently, Noson Yanofsky [72] presented a category theoretical construction, analysing self reference in a very general way. He obtains descriptions of a plethora of self-referential paradoxes, fixed points and incompleteness theorems as instances of this construction.

- Arithmetised metamathematics: Every formula $\varphi(x)$ expressed in the language of an arithmetical theory T has a fixed point, in the sense that there is a sentence δ such that $T \vdash \delta \leftrightarrow \varphi(\delta)$.

Although all of these fields are rich in beautiful theorems and clever problem-solving, we will not explore them all. At times, it may turn out that some concepts that may look entirely different are in some way closely connected, maybe even being different sides of the same coin. We will take the route of the arithmetised metamathematician, and only occasionally glance in another direction.

1.1 This thesis

This thesis consists of two distinct parts, both related to metamathematical fixed points. Chapter 2 is an historical survey of the area, from *la fin du siècle* to present. Chapter 3 is purely technical and deals with sets of fixed points.

The text is intended as an historical introduction to metamathematics in general, and to fixed-point constructions and their applications in particular. Alas, as the title may suggest, the area is a technical and mathematical one, and we have to cumber the text with notation and technical constructions. We give a detailed survey of the development of ideas and proofs, especially in variations of the fixed point theorem. Similar presentations are often streamlined in terms of proofs, and results are generalised to allow a wider use. We have not aimed to do so here, but rather to give an historically correct, readable exposé of the actual development of the field. Minor embellishments aside, the proofs presented are original. We have, for example, sometimes changed the ordering of variables in substitution functions, for greater readability. When notation differs from the original, we explicitly point this out.

As for the prerequisites, we suppose the reader to be familiar with first-order logic, the theory of first-order arithmetic, naive set theory, and the theory of recursive functions. In what follows, T is a consistent, axiomatisable arithmetical theory extending Peano arithmetic, PA.[1] These theories are formalised in the language of arithmetic, $\mathscr{L}_A$. $\mathrm{Th}(T)$ is the set of sentences provable in T. We will use k, m, n (possibly with subscripts) as variables for natural numbers. A, B, X, Y will denote sets of (Gödel numbers of) sentences. Sentences and formulas are denoted by lower case Greek letters (again, possibly with subscripts), where sentences are written as $\delta, \gamma, \varphi, \psi$ etc. and formulas as $\theta(x), \chi(x), \eta(x), \xi(x, y)$ etc. In the historical part of the text, we sometimes divert from these conventions, in order to allow the reader to consult the original texts for details.

One point where we are not completely historical is on notation of Gödel numbers in formulas. Generally, $\ulcorner\varphi\urcorner$ denotes the numeral for the Gödel number of the formula φ. When x is free in φ, we use Feferman's dot notation $\ulcorner\varphi(\dot{x})\urcorner$ to denote that x is to be treated as a free variable, so the Gödel number of that expression depends on the actual value of x. Thus $\ulcorner\varphi(x)\urcorner$ and $\ulcorner\varphi(\dot{x})\urcorner$ are different numerals. On the other hand, when the context so admits, we conform to the modern way of identifying formal expressions with the numerals of their Gödel numbers. Occasionally, we, for readability, allow ourselves to introduce new function symbols to the formal languages in question, even if they are not present in the original texts.

[1] In fact, we could almost everywhere do with a much weaker theory, e.g. Robinson's arithmetic Q, plus induction for Σ_1-formulas. We have not aimed to state the strongest versions of theorems in this respect, but instead rest on the intuition that the results hold for reasonably strong arithmetical theories. The reader interested in technicalities of this kind may consult e.g. Hájek and Pudlák's *Metamathematics of First-Order Arithmetic* [23].

Chapter 2

Metamathematics: An historical survey

2.1 Foundationalism in the late 19th century

In the beginning, the well-renowned mathematician, philosopher and physicist David Hilbert devised a grand scheme to formalise all of mathematics and prove by undoubtable methods that this body of mathematics was free of contradiction: that it was *consistent*. The key idea was to first identify what such undoubtable methods could be, and to establish that these methods were themselves free from contradiction. These methods would then lay the ground for any other more complex mathematical framework, in that nothing should be added to mathematics without a proof of non-contradiction, carried out by the above-mentioned methods. The plan was great, and Hilbert devoted many years to this project. See e.g. [28, 30, 31].

The crucial problem that Hilbert tried to resolve was how to relate mathematical infinities to the real world. It is clear that we can justify the use of (relatively) small numbers by alluding to counting discrete objects in the world, but it is possibly more uncertain how to justify the use of actual mathematical infinities. Hilbert's remedy was the following [30].

1. Identify and formalise an undisputable system T_1 of finitistic mathematics and methods. By the work of William Tait [68], this system can roughly be taken to be primitive recursive arithmetic. This system is to be under-

stood as true and contentful, and will serve as the foundation of all other mathematics.

2. Formalise all mathematical knowledge in a system T_2, freely using infinite sets, transfinite induction and any other non-finitistic methods. On Hilbert's view, this system is to be understood as pure syntax.

3. Prove that T_2 is *consistent* (free of contradiction) and *complete* (sufficient to decide every sentence expressible in the system), using only the system T_1.

4. Prove that T_2 is *conservative* over T_1, in the sense that every contentful mathematical statement that is provable in T_2 is provable already in T_1.

Could we accomplish this, we would have a reduction of infinitistic mathematics to finitistic, and a justification of the use of infinitistic methods. We will not delve into the technicalities of Hilbert's program here, but refer the reader to e.g. [33, 58, 68].

A related project of the same era is the one undertaken in *Principia mathematica* [56] by Russell and Whitehead. Their goal was to deduce all of mathematics from a set of axioms and inference rules formalised in a symbolic logic, leaving no place for doubts about their truth or plausibility. The similarity between these two projects lies in that both were concerned with mathematical truths, and their search for a formal method to establish once and for all which these mathematical truths were. It was a common belief of the time that such a project would indeed be possible to carry out.

2.2 Incompleteness

That this foundational endeavour cannot be successful was proved by Kurt Gödel in his seminal 1931 paper *Über Formal Unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme, I.* [19]. Gödel showed that many systems formalizing arithmetic suffer from *incompleteness*, in the sense that there are arithmetical sentences that these systems can neither prove nor refute. For a system to be subject to this incompleteness phenomenon it has to satsify a certain list of conditions: it should be a formal system formulated in first-order logic with a (primitive) recursive set of axioms, strong enough to

represent all recursive functions, and ω-consistent.[1] From this point on, every formal theory is supposed to be formalised in first-order logic.

We now allow ourselves a technical inspection of Gödel's first incompleteness theorem, as stated in the following form:

Theorem 2.2.1 (Gödel's first incompleteness theorem)**.** *Any recursive, sufficiently strong, ω-consistent formal system is incomplete.*

Gödel confined himself to a formal system closely related to the system of *Principia mathematica*, but extended to contain Peano arithmetic. This system, P, is recursive, sufficently strong and ω-consistent, so it is incomplete by the way of there being an arithmetical statement γ such that neither $\mathrm{P} \vdash \gamma$ nor $\mathrm{P} \vdash \neg\gamma$. The "sufficient strength" called for is to be understood in a certain technical manner. P is strong enough to represent every recursive relation in such a way that if $R(x_0, \ldots, x_n)$ is a recursive relation, then there is a formula $\rho(x_0, \ldots, x_n)$ such that:

1. If $R(k_0, \ldots, k_n)$ holds, then $\mathrm{P} \vdash \rho(k_0, \ldots, k_n)$, and
2. if $R(k_0, \ldots, k_n)$ does not hold, then $\mathrm{P} \vdash \neg\rho(k_0, \ldots, k_n)$.

The ingenuity of Gödels proof is threefold. For the first part, he constructed a codification of logical syntax in arithmetic— a *Gödel numbering*. By identifying every symbol in the arithmetical language with a certain integer, it is possible to obtain a decidable correspondence between numbers and arithmetical formulas. This is subsequently extended to a correspondence between numbers and sequences of formulas. In this way, as proofs are sequences of formulas, it becomes possible to, loosely speaking, "talk about" arithmetical proofs inside arithmetic itself. An essential feature of this codification is that "simple" properties of syntactical objects, such as being a formula or a proof, corresponds to "simple" (recursive) properties of the corresponding Gödel numbers. The second part involves defining the recursive relations, and showing that these are actually representable in P. The proof relation "y is a proof of x in P" is recursive, so it can be represented in P by a formula $\mathrm{Prf}(x, y)$. As for the third ingenious contribution, Gödel constructed a sentence γ, such that γ is equivalent in P to the statement $\neg\,\mathrm{Pr}(\gamma)$ (where $\mathrm{Pr}(x)$ is short for the formula $\exists y\,\mathrm{Prf}(x, y)$), in symbols

$$P \vdash \gamma \leftrightarrow \neg\,\mathrm{Pr}(\gamma).$$

[1] Gödel's term is *rekursiv*, which is what today is called primitive recursive, as opposed to general recursive (used by e.g. Kleene and Rosser). We will see several weakenings of these conditions later.

Gödel's method for constructing this *Gödel sentence* is the first example of self reference in arithmetic. In other words, γ is a *fixed point* to the formula $\neg \Pr(x)$. We give here an outline of the original proof, which is quite involved, and not entirely transparent. It will, however be of interest to compare the different methods for obtaining self-referential formulas in the sequel.

Let κ be any class of formulas. κ is said to be *ω-consistent* if for any formula $\varphi(x)$, if κ proves $\neg\varphi(k)$ for every k, then κ does not prove $\exists x\varphi(x)$. It follows that every ω-consistent theory is consistent. The formula $\Pr_\kappa(x)$ stands for x being provable from κ.[2] Let $\sigma(\varphi, n, y)$ be the result of substituting the numeral y for the variable with Gödel number n in φ.[3] Now, let $Q(x, y)$ be the expression $\neg \operatorname{Prf}_\kappa(\sigma(y, 19, \ulcorner y \urcorner), x)$, where, as defined in Gödel's paper, 19 is the Gödel number of a particular variable. This relation is recursive, and as every such relation is represented in P, there is a predicate symbol (relation sign) q such that:

1. $\mathrm{P} \vdash \neg \operatorname{Prf}_\kappa(\sigma(y, 19, \ulcorner y \urcorner), x) \to \Pr_\kappa(\sigma(q, 17, \ulcorner x \urcorner, 19, \ulcorner y \urcorner))$, and

2. $\mathrm{P} \vdash \operatorname{Prf}_\kappa(\sigma(y, 19, \ulcorner y \urcorner), x) \to \Pr_\kappa(\neg\sigma(q, 17, \ulcorner x \urcorner, 19, \ulcorner y \urcorner))$.

Again, 17 is the Gödel number of a variable, distinct from the variable whose Gödel number is 19. Now, let $G(k, \varphi)$ be the Gödel number of the formula obtained by universally quantifying the formula φ over the variable with Gödel number k.[4] Though G is strictly not a function symbol of the formal theory at hand, we will allow ourselves to introduce it, thus simplifying the notation. Let $p := G(17, q)$, and let $r := \sigma(q, 19, \ulcorner p \urcorner))$. It follows that

$$\sigma(p, 19, \ulcorner p \urcorner) = \sigma(G(17, q), 19, \ulcorner p \urcorner) = G(17, \sigma(q, 19, \ulcorner p \urcorner)) = G(17, R),$$

and

$$\sigma(q, 17, \ulcorner x \urcorner, 19, \ulcorner p \urcorner) = \sigma(r, 17, \ulcorner x \urcorner).$$

By substituting p for y in 1) and 2) above, we get

[2] Gödel's term is Bew(x), from the German word *Beweisbar*, meaning provable. The proof relation is originally denoted xBy.

[3] Gödel's notation is $\mathrm{Sb}(\varphi^n_y)$. The function is then extended to accept any odd number of arguments, in our notation e.g. $\sigma(\varphi, n, y, k, w)$ being the result of substituting the numeral y for the variable with Gödel number n and the numeral w for the variable with Gödel number k of φ.

[4] E.g., if the variable with Gödel number k is x, then $G(k, \varphi)$ is $\ulcorner \forall x \varphi \urcorner$.

1′. $\mathrm{P} \vdash \neg \mathrm{Prf}_\kappa(G(17,r),x) \to \mathrm{Pr}_\kappa(\sigma(r,17,\ulcorner x \urcorner))$, and

2′. $\mathrm{P} \vdash \mathrm{Prf}_\kappa(G(17,r),x) \to \mathrm{Pr}_\kappa(\neg\sigma(r,17,\ulcorner x \urcorner))$.

Now, reason in P and suppose that $G(17,r)$ is provable from κ. Then there is an n such that $\mathrm{Prf}_\kappa(G(17,r),n)$, and so, by 2′, it follows that $\mathrm{Pr}_\kappa(\neg\sigma(r,17,\ulcorner x \urcorner))$, while from the provability of $G(17,r)$ the provability of $\sigma(r,17,\ulcorner n \urcorner)$ follows, leading to a contradiction. As we have just established that $G(17,r)$ is unprovable, it follows that $\forall n \neg \mathrm{Prf}_\kappa(G(17,r),n)$ holds. Together with 1′, we get $\forall n\, \mathrm{Pr}_\kappa(\sigma(r,17,\ulcorner n \urcorner))$. But this, in conjunction with $\mathrm{Pr}_\kappa(\neg G(17,r))$ is incompatible with the ω-consistency of κ, so neither does κ prove $\neg G(17,r)$. Thus, $G(17,r)$ is a sentence undecidable in P.

Note that by choosing the relation Q (and subsequently, the predicate symbol q representing Q) in a different way, the construction would yield a fixed point to another formula. Thus, it is not at all unreasonable to credit the Fixed point theorem to Gödel, although he never states it in any generality.

Gödel presented a preliminary version of the first incompleteness theorem at the Second Conference for Epistemology of the Exact Sciences, which took place in Königsberg between September 5 and 7, 1930. The only one who seems to fully have understood the significance of the result is J. von Neumann, who shortly after the conference wrote to Gödel:

> I have recently concerned myself again with logic, using the methods you have employed so successfully in order to exhibit undecidable properties. In doing so I achieved a result that seems to me to be remarkable. Namely, I was able to show that the consistency of mathematics is unprovable.[...]
>
> In a formal system that contains arithmetic it is possible to express, following your considerations, that the formula $1 = 2$ cannot be the endformula of a proof starting with the axioms of this system—in fact this formulation is a formula of the formal system under consideration. Let it be called W.
>
> In a contradictory system any formula is provable, thus also W. If the consistency [of the system] is established intuitionistically, then it is possible, through a "translation" of the contentual intuitionistic considerations into the formal [system], to prove W also.[...] Thus with unprovable W the system is consistent, but the consistency is unprovable.[5]

[5] von Neumann, Letter to K. Gödel (Nov. 20, 1930), in [47], p. 123.

Gödel replied shortly, and while that letter is lost, von Neumann's reply is clear:

> As you have established the theorem on the unprovability of consistency as a natural continuation and deepening of your earlier results, I clearly won't publish on this subject.[6]

von Neumann's observation amounts to the second incompleteness theorem on the unprovability of consistency. Let Con_T be the sentence $\neg\exists y\,\mathrm{Prf}(0=1,y)$. In a natural sense, Con_T thus expresses that T is consistent.

Theorem 2.2.2 (Gödel's second incompleteness theorem)**.** *For every recursive, sufficiently strong, consistent formal system T, it holds that $T \nvdash Con_T$.*

For the publication of [19], Gödel sketched a proof of this result. His argumentation is rather different from von Neumann's, and proceeds by observing that the reasoning involved in the proof of the incompleteness theorem is "arithmetical enough" to itself be translated into the language of P. He meant to prove this in full detail in a later paper, but the argument was convincing, and the second part of *Über Formal Unentscheidbare Sätze* was never written. It was not until 1939 that Hilbert and Paul Bernays presented a fully detailed proof of the Second incompleteness theorem, in their *Grundlagen der Mathematik* [32].

A strengthening of the incompleteness theorems was presented by J. Barkley Rosser in 1936. In the paper *Extensions of some theorems of Gödel and Church* [54], incompleteness theorems with weakened premises were proved. Rosser was able to weaken the constraint on the sets of axioms and inference rules to recursive enumerability, rather than primitive recursiveness. Moreover, by employing what is today known as *Rosser's trick* for comparing witnesses, he could weaken ω-consistency to mere consistency.

Theorem 2.2.3 (The Gödel-Rosser incompleteness theorem)**.** *Any recursively enumerable, sufficiently strong, consistent formal system is incomplete.*

To obtain his result, Rosser modified the proof predicate specified by Gödel. Let $\mathrm{Prf}(x,y)$ be Gödel's proof predicate, let $\dot{\neg}x$ be the Gödel number of the negation of the formula with Gödel number x, and let

$$\mathrm{Prf}^R(x,y) := \mathrm{Prf}(x,y) \wedge \neg\exists z(z{\leq}y \wedge \mathrm{Prf}(\dot{\neg}x,z)).$$

Thus, y is a proof of x in Rosser's sense, if y is a proof of x in Gödel's sense, and there is no shorter proof of of $\neg x$.[7] So in this sense of provability, we accept

[6] von Neumann, Letter to K. Gödel (Nov. 29, 1930), in [47], p. 124.

[7] The phrasing "x has a shorter proof than y" here means that there is a proof of x whose Gödel number is less than the Gödel number of any proof of y. Equivalently, we sometimes speak of x being proved earlier than y.

only proofs of sentences that we have not yet refuted, thus incorporating the consistency criterion in the proof predicate. In $T + \mathrm{Con}_T$ we can readily prove $\mathrm{Prf}(x, y) \leftrightarrow \mathrm{Prf}^R(x, y)$ for all x and y, while this is not provable in T on its own, as the proof requires the hypothesis that T is consistent. Rosser claims that the formalisation of this provability predicate has properties that Gödel's provability predicate does not possess. As an example of this, he states that for every sentence φ, $T \vdash \mathrm{Pr}^R(\varphi) \to \mathrm{Pr}^R(\mathrm{Pr}^R(\varphi))$ and $T \vdash \mathrm{Pr}^R(\neg\varphi) \to \mathrm{Pr}^R(\neg\,\mathrm{Pr}^R(\varphi))$. As we will see later, the first of these properties is indeed possessed by Gödel's provability predicate—although this was probably not known until the publication of *Grundlagen der Mathematik* in 1939.

For the proof, Rosser only points out that "one can proceed as [...] Gödel".[8] We will give an outline of the argument. By Gödel's method for obtaining self reference, Rosser constructs a formula ρ such that:

$$T \vdash \rho \leftrightarrow \forall y(\mathrm{Prf}(\rho, y) \to \exists z{\leq}y\,\mathrm{Prf}(\dot{\neg}\rho, z)).^9$$

This is indeed equivalent to $T \vdash \rho \leftrightarrow \neg\,\mathrm{Pr}^R(\rho)$, i.e. a Gödel sentence for Rosser's modified provability predicate. We begin by proving that $T \nvdash \rho$. Suppose, for a contradiction, that T proves ρ, and that p is the Gödel number for such a proof. Since T is consistent, it follows that $T \nvdash \neg\rho$ and that for every k, $T \vdash \neg\,\mathrm{Prf}(\dot{\neg}\rho, k)$. Thus, $T \vdash z \leq p \to \neg\,\mathrm{Prf}(\dot{\neg}\rho, z)$, and since p is a proof for ρ, we get $T \vdash \exists y(\mathrm{Prf}(\rho, y) \wedge \forall z{\leq}y \neg\,\mathrm{Prf}(\dot{\neg}\rho, z))$. By the construction of ρ, it follows that $T \vdash \neg\rho$, contradicting the assumption that T is consistent.

Next, suppose that $T \vdash \neg\rho$, and that p is the Gödel number for such a proof. Since T is consistent, there can be no proof of ρ, so for every k, $\vdash \neg\,\mathrm{Prf}(\rho, k)$. Consequently, $T \vdash y < p \to \neg\,\mathrm{Prf}(\rho, y)$. Since p is a proof of $\neg\rho$ it follows that $T \vdash \mathrm{Prf}(\dot{\neg}\rho, p)$ and trivially $T \vdash p \leq y \to \exists z{\leq}y\,\mathrm{Prf}(\dot{\neg}\rho, z)$. But $T \vdash (k \leq m) \vee (m < k)$ for all k, m, so $T \vdash \neg\,\mathrm{Prf}(\rho, y) \vee \exists z{\leq}y\,\mathrm{Prf}(\dot{\neg}\rho, z)$. By construction of ρ, it follows that $T \vdash \rho$, contradicting the assumption that T is consistent. Note that we here need the hypothesis of consistency, rather than ω-consistency, to make sure that T does not simultaneously prove both ρ and $\neg\rho$.

Both incompleteness theorems have bearings on Hilbert's program and Russell's logicist reduction. Firstly, in the case that T is consistent, then both the Gödel sentence and the Rosser sentence are true Π_1-statements. Thus there are statements that should rightly be contained in the weaker system T_1 of Hilbert, but which we could not hope to prove by the finitistic methods embodied in T_1.

[8] Rosser [54], p. 90.

[9] The original (equivalent) formulation is $T \vdash \rho \leftrightarrow \neg\exists y(\mathrm{Prf}(\rho, y) \wedge \neg\exists z(z{\leq}y \wedge \mathrm{Prf}(\dot{\neg}\rho, z)))$, but the present variation excels in clarity.

Secondly, as T_1 is a mathematical system, it is contained in the stronger system T_2. Also, by the second incompleteness theorem, T_1 cannot even prove its own consistency, nor can it prove the consistency of T_2. Gödel himself did not support such a view, at least not at the time of publication of his theorems:

> I wish to note expressly that [the second incompleteness theorem] does not contradict Hilbert's formalistic viewpoint. For this viewpoint presupposes only the existence of a consistency proof in which nothing but finite means of proof is used, and it is conceivable that there exists finitary proofs that *cannot* be expressed in the formalism of P.[10]

von Neumann did not share Gödel's belief that Hilbert's program stood untouched by the second incompleteness theorem. In a letter to Rudolf Carnap, he made the following comment on this point:

> Thus I am today of the opinion that
>
> 1. Gödel has shown the unrealizability of Hilbert's program.
> 2. There is no more reason to reject intuitionism (if one disregards the aesthethic issue, which in practice will also for me be the decisive factor).
>
> Therefore I consider the state of the foundational discussion in Königsberg to be outdated, for Gödel's fundamental discoveries has brought the question to a completely different level. (I know that Gödel is much more careful in the evaluation of his results, but in my opinion on *this* point he does not see the connections properly.[11]

Today it is widely thought that the incompleteness theorems show the impossibility to carry out the full Hilbert program. There have been suggestions, however, for partial realisations of Hilbert's program, for example through the use of methods stemming from the research program of Reverse Mathematics [58].[12]

A theorem closely related to those of Gödel is Alfred Tarski's theorem on the undefinability of truth, published in *The concept of truth in formalised languages* [69].[13] Let a *truth definition for* T be a formula $\tau(x)$ such that, for every sentence φ,

$$T \vdash \varphi \leftrightarrow \tau(\varphi).$$

[10] Gödel 1931, translated in [21], p. 195.

[11] von Neumann, Letter to R. Carnap, (June 7, 1931), in [47], pp. 85-86.

[12] Cf. Feferman [13].

[13] Though Tarski's paper was not published until 1933 (in Polish), most of the investigations resulting in [69] was carried out in 1929. Before the paper was published in a language more

Theorem 2.2.4 (Tarski's theorem)**.** *No consistent, sufficiently strong theory has a truth definition.*

The existence of a truth definition could be used to construct a formal version of the Liar paradox: "This sentence isn't true". As we saw in the introduction, any such sentence is both true and not true, which evidently is a contradiction. Tarski's construction of this self-referential statement is closely related to the construction of Gödel's. Tarski draws the distinction between the object language and the metalanguage, of which the latter is used to study the former. He makes the claim that an adequate definition of truth should have as consequences

> [A]ll sentences which are obtained from the expression '$x \in \mathrm{Tr}$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question, and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage.[14]

The proof proceeds by first showing that in whatever way the class of true expressions, Tr, is defined, it is possible to derive the contradiction of one of the statements of the form described in the above convention. Suppose that we have defined the class of true sentences Tr in the metalanguage. Define, also in the metalanguage, an infinite sequence $\Phi := \varphi_0, \varphi_1, \ldots$, such that every expression of the object language occurs exactly once. This corresponds to the numbering of formulas used by Gödel, but Tarski is explicit in that this numbering is carried out in the metalanguage. Let ι_k be a formula with a free variable n, that expresses the relation $k = n$. Let $E(n, y)$ be the Gödel number of the existential quantification of y with respect to the variable with Gödel number n.[15] Now, consider the expression

$$E(3, \iota_n \wedge \varphi_n) \notin \mathrm{Tr} \tag{2.1}$$

Again, 3 is the Gödel number for a variable. By the Gödel numbering, we can construct an arithmetical formula which is equivalent to the previous expression for every value of n, denoted $\psi(n)$. As this expression is arithmetical, it occurs

generally accepted for scientific exchange (German, in 1935), both Carnap (as we will see later) and Gödel had published their related results. Tarski points out that these results are "quite independent" of each other, but admits that he owes the method of arithmetisation to Gödel. See the historical notes to [69] in [70], and footnote 1 on p. 247 of [70] for further details.

[14] Tarski [69], in [70], p. 188.

[15] Again, the function symbol E is not strictly in the formal language.

in the sequence Φ, so for some k, we will have $\psi(n) = \varphi_k$. By substituting k for n in (2.1), and using the equivalence of $\psi(n)$ with (2.1), we obtain

$$E(3, \iota_k \wedge \varphi_k) \notin \text{Tr if and only if } \psi(k). \tag{2.2}$$

On the other hand, the expression $E(3, \iota_k \wedge \varphi_k)$ is a sentence of the language under consideration, and by applying the criterion of an adequate truth definition above, we obtain a sentence of the form '$x \in Tr$ if and only if p'. But by substituting $E(3, \iota_k \wedge \varphi_k)$ for x and $\psi(k)$ for p we get

$$E(3, \iota_k \wedge \varphi_k) \in Tr \text{ if and only if } \psi(k), \tag{2.3}$$

which clearly contradicts (2.2).

2.3 A step ahead: The Fixed point theorem

In proving the First incompleteness theorem, Gödel presented the first example of a *metamathematical fixed point*, the concept that is our main interest in this text. The Gödel sentence γ (as well as the Rosser sentence ρ) is a sentence constructed to share provability conditions with another sentence—in this case the sentence stating that γ has no proof. While Gödel, Tarski and Rosser all explicitely constructed their fixed points, they do not seem to have considered a more general approach, e.g. proving that every arithmetical formula possesses a fixed point. It took until 1934, when Carnap published the first edition of *The Logical Syntax of Language* [6], for a general formulation of a fixed point theorem to surface. Carnap constructed a mathematical language called *Language II*, for which he proved that, for any syntactical property, there is a sentence that semantically interpreted means that the sentence itself has that syntactical property.[16] In the spring of 1934, Gödel gave a series of lectures [20] at the Institute for Advanced Study at Princeton, where he credits the Fixed point theorem to Carnap. The theorem is also stated in Rosser's 1936 paper [54], where he credits it to Gödel—although he mentions that Gödel does not state it explicitly, but rather presents a specific instance. In today's terminology, Carnap's theorem could be put like this:

[16] *Language II* is quite strong; apart from the purely arithmetical *Language I* also introduced in the text, it contains variables for predicates and functions, an induction axiom and a variation of the axiom of choice. Its full strength is not, however, needed to prove the Fixed point theorem.

Theorem 2.3.1 (The Fixed point theorem)**.** *For any arithmetical formula $\theta(x)$ with one free variable, there is a sentence δ such that*

$$T \vdash \delta \leftrightarrow \theta(\delta).$$

Prima facie, this theorem states that we can construct a sentence δ that is provably equivalent to the formula $\theta(x)$ instantiated with the sentence δ itself. Loosely speaking, δ says about itself that it possesses the property expressed by $\theta(x)$. Of course, arithmetical sentences rarely speaks at all—after all, they are just sequences of symbols. These fixed point are really just constructed to fulfil the provable equivalence stated above. This can be accomplished after specifying a Gödel numbering, and by a carefully conceived substitution function which can be put in the following wording:

> The result of substituting the quotation of "The result of substituting the quotation of x for 'x' in x has property P." for 'x' in "The result of substituting the quotation of x for 'x' in x has property P." has property P.[17]

If we allow us to carry out the operation specified in the sentence above, it turns out the result is just the sentence itself. Thus, the sentence "says about itself" that it possesses the property P.

Carnap's proof of the Fixed point theorem proceeds as follows. Let φ be a formula with one free variable x, expressing the property P. Let $\sigma(x, s, y)$ express "the result of in x replacing all occurrences of s with y". Then let $\psi = \sigma(\varphi, x, \sigma(x, x, x))$. But ψ has a specific Gödel number, k, say, that we can substitute for x in ψ. Letting $\delta = \sigma(k, x, k)$, we obtain a formal version of the operation described by the quotation above, and δ, semantically interpreted, means that δ itself has the chosen syntactical property. This construction might require an intellectual effort to embrace. Spelled out, δ is

$$\sigma(\sigma(\varphi, x, \sigma(x, x, x)), x, \sigma(\varphi, x, \sigma(x, x, x))).$$

[17] This kind of self-referential contructions in natural languages is usually credited to W.V.O. Quine. The form

> "yields a sentence with property P when appended to its own quotation." yields a sentence with property P when appended to its own quotation.

can be extracted from examples in *The ways of paradox* [51] from 1962, but Smullyan [63] credits this construction to Quine as early as 1957. The present phrasing is from Franzén [15], p. 41.

Bearing in mind the definitions of $\sigma(x, s, y)$ and φ, it should be possible to realise that this arithmetical formula actually captures the meaning of the quotation above.

2.4 Further development: Henkin, Löb

While the results of Gödel, Tarski, Rosser and Carnap undoubtably attracted some attention, it took more than a decade before any further development of ideas took place. The method of constructing fixed points was sparsely used, and the rapidly developing field of recursion theory drew more attention and also provided similar incompleteness results in an abstract fashion. It was not until 1952 that Leon Henkin [27] asked the rather natural question of whether a fixed point to the provability predicate is provable or not. Applying the Fixed point theorem, he simply constructed a sentence η such that

$$T \vdash \eta \leftrightarrow \mathrm{Pr}(\eta).$$

It is not immediately clear if such a fixed point is true in the standard interpretation, nor if it is provable, refutable or undecidable.

In 1955, M. H. Löb [44] presented an answer to the question.

Theorem 2.4.1 (Löb's theorem)**.** [18]
Any sentence φ such that $T \vdash Pr(\varphi) \to \varphi$, is provable in T.

In order to prove this theorem, Löb identified three abstract derivability conditions, extracted from conditions stated by Hilbert and Bernays to hold for a provability predicate specified in [32].

Definition 2.4.2 (The Hilbert-Bernays-Löb derivability conditions)**.** *For any formulas φ, ψ:*

L1. $T \vdash (Pr(\varphi \to \psi) \wedge Pr(\varphi)) \to Pr(\psi)$,

L2. $T \vdash Pr(\varphi) \to Pr(Pr(\varphi))$,

L3. if $T \vdash \varphi$, *then* $T \vdash Pr(\varphi)$.

[18] This is the first of the presented results which the theory Q is *not* strong enough to prove. It is partly because of this result that we have chosen to let T be an extension of PA.

Now, take any arithmetical formula ψ that is such that $T \vdash \Pr(\psi) \to \psi$ (e.g. Henkin's formula above). Then apply the Fixed point theorem to the formula $\Pr(x) \to \psi$ obtain a formula λ such that:

$$T \vdash \lambda \leftrightarrow (\Pr(\lambda) \to \psi).$$

The reasoning proceeds as follows. By construction, we have

$$T \vdash \lambda \to (\Pr(\lambda) \to \psi) \tag{2.4}$$

and by L3 above,

$$T \vdash \Pr(\lambda \to (\Pr(\lambda) \to \psi)). \tag{2.5}$$

By L1

$$T \vdash \Pr(\lambda \to (\Pr(\lambda) \to \psi)) \wedge \Pr(\lambda) \to \Pr(\Pr(\lambda) \to \psi), \tag{2.6}$$

so, by 2.5 and 2.6,

$$T \vdash \Pr(\lambda) \to \Pr(\Pr(\lambda) \to \psi). \tag{2.7}$$

Again by L1,

$$T \vdash \Pr(\Pr(\lambda) \to \psi) \wedge \Pr(\Pr(\lambda)) \to \Pr(\psi), \tag{2.8}$$

so, by 2.7 and 2.8,

$$T \vdash \Pr(\lambda) \wedge \Pr(\Pr(\lambda)) \to \Pr(\psi), \tag{2.9}$$

but, by L2, it follows that

$$T \vdash \Pr(\lambda) \to \Pr(\psi). \tag{2.10}$$

Since we have supposed that ψ is such that

$$T \vdash \Pr(\psi) \to \psi, \tag{2.11}$$

it follows from 2.10 that

$$T \vdash \Pr(\lambda) \to \psi \tag{2.12}$$

but, by construction of λ, it follows that

$$T \vdash \lambda. \tag{2.13}$$

Finally, by L3,

$$T \vdash \Pr(\lambda), \tag{2.14}$$

so, by 2.12, we get the desired result

$$T \vdash \psi. \tag{2.15}$$

By this theorem any Henkin fixed point is indeed provable, and thus true. It also follows that the fixed points of $\mathrm{Pr}(x)$ are exactly the sentences provable in T. We return to this observation, and related questions, in Chapter 3.

Löb's theorem is also a direct consequence of the second incompleteness theorem. It is interesting, and somewhat lucky, that noone realised this before Löb published his proof, for both the proof in itself, and the derivability conditions have turned out to be of importance, e.g. spawning the field of provability logic. The simpler proof proceeds as follows:

Suppose $T \vdash \mathrm{Pr}_T(\psi) \to \psi$.[19] Then $T + \neg\psi \vdash \neg\,\mathrm{Pr}_T(\psi)$. Moreover, for every sentence φ, $T \vdash \neg\,\mathrm{Pr}_T(\varphi) \leftrightarrow \mathrm{Con}_{T+\neg\varphi}$, so $T + \neg\psi$ proves its own consistency. By the second incompleteness theorem, it follows that $T + \neg\psi$ is inconsistent, so $T \vdash \psi$.

2.5 Graduation day: Arithmetisation of metamathematics in a general setting

The term *metamathematics* can roughly be understood as referring to the study of mathematics by using mathematical methods. Or, in other words, we raise the level of abstraction by making mathematics itself the object of study, rather than being a method for investigating properties of numbers. Hilbert was the first to use the word metamathematics with some regularity—the first known use of the term is from 1922:

> [I]n addition to this proper mathematics, there appears a mathematics that is to some extent new, a *metamathematics* which serves to safeguard it by protecting it from the terror of unnecessary prohibitions as well as from the difficulty of paradoxes. In this metamathematics—in contrast to the purely formal modes of inference in mathematics proper—we apply contentual inference, in particular, to the proof of the consistency of the axioms.[20]

The term was used synonymous to proof theory, but proof theory was indeed the tool intended to carry out the finitistic reduction suggested in Hilbert's program. Though the concept has broadened today, it is clear that Hilbert's research even today is regarded as metamathematical.[21]

[19] In writing $\mathrm{Pr}_T(x)$, we emphasise that we are considering the provability predicate of T, even if we are using this predicate in the context of another theory.

[20] Hilbert [29], p. 212.

[21] Earlier work can certainly be said to be metamathematical to its nature, e.g. the investigations of Gottlob Frege, especially his *Begriffsschrift* [16].

A shift in attention occured when Gödel presented his arithmetisation of syntax. For the first time, it was possible to carry out metamathematical investigations with purely arithmetical methods. By the codification of decidable properties, and through the Gödel numbering of sentences and proofs, it became possible to, as seen earlier, talk about arithmetic within arithmetic itself.

With the publication of Feferman's *The arithmetization of metamathematics in a general setting* in 1960, the field of metamathematics was placed on new, stable grounds. The majority of the results were obtained as Feferman was working on his doctoral thesis, and mainly encompasses generalisations and developments of Gödel's arithmetisation methods. There are also quite a few new notions in the paper, e.g. the distinction between extensional and intensional formulas, the concepts of (bi)numerations of sets, and an example of why we need to be cautious when interpreting incompleteness theorems informally. We briefly look into each of these three topics.

In today's practise, a formula $\theta(x)$ is said to be *extensional* if $T \vdash \theta(\delta) \leftrightarrow \theta(\gamma)$ whenever $T \vdash \delta \leftrightarrow \gamma$. This means that an extensional formula cannot distinguish between different sentences, but only between different equivalence classes of sentences. An intensional formula, on the other hand, might take into account the orderings of proofs, or the Gödel numbers of sentences. Feferman informally introduces the distinction in the introduction of his paper:

> In broad terms, the applications of [arithmetization] can be classified as being *extensional* if essentially only numerically correct definitions are needed, or *intensional* if the definitions must more fully *express* the notions involved, so that various of the general properties of these notions can be formally derived.[22]

On this analysis, e.g. Rosser's proof predicate expresses an intensionally incorrect notion of proof, though from a metaperspective (and extensionally), Rosser's and Gödel's proof predicates coincide. Though we can show that if the theory in question is consistent, a sentence is Rosser-provable iff it is Gödel-provable, Rosser's notion intensionally expresses something else than Gödel's.

Two other key concepts made explicit by Feferman are those of *numerations* and *binumerations* of a relation. The idea was not entirely new—it can be seen as an explication of representing relations in a theory.

[22] Feferman [11], p. 35.

Definition 2.5.1. *$\xi(x_1, \ldots, x_n)$ numerates an n-ary relation R in a theory T if, for all $k_1, \ldots, k_n$, $R(k_1, \ldots, k_n)$ holds iff $T \vdash \xi(k_1, \ldots, k_n)$. If ξ also satisfies the condition that $R(k_1, \ldots, k_n)$ does not hold iff $T \vdash \neg\xi(k_1, \ldots, k_n)$, then ξ binumerates X in T.*

As stated by Feferman, the concept "binumerate" coincides with what other authors called "define", "strongly define", "strongly represent" and "numeralwise express". If R is a one-place relation it can be viewed as a characteristic function of a set, and we can consider binumerations of not only relations, but also of sets. It follows that it is precisely the recursive relations (and sets) that can be binumerated by both a Σ_1- and a Π_1-formula in T. Similarily, "numerate" corresponds to the earlier terms "represent", "weakly represent" and "weakly define", and the recursively enumerable relations (sets) are those having Σ_1-numerations in T.

Feferman further associates with a theory T the class of all formulas $\tau(x)$ that (through a Gödel numbering) numerates the set of axioms of T. Such a formula is said to numerate the theory T. This approach makes it possible to relativise the notion of e.g. "y being a proof of x in T" to different τ numerating the axioms of T. To each such numeration there is a corresponding consistency statement, Con_τ. It is made explicit that the Gödel-Rosser incompleteness theorem holds for a wide class of theories, but that some restrictions need to be imposed on what numerations are allowed. Feferman then proves the following strengthening of the Gödel-Rosser incompleteness theorem.

Theorem 2.5.2. *If T is a sufficiently strong, consistent theory, and $\tau(x)$ is any Σ_1-formula numerating (an extension of) T in T, then $T \nvdash Con_\tau$.*[23]

This theorem, Feferman's version of the Gödel-Rosser incompleteness theorem, is often informally stated along the lines of "if T satisfies certain not-very-interesting conditions, then T does not prove its own consistency". There is, however, reason to be careful when presenting this kind of technical results in such an informal way. That the restriction to Σ_1-numerations cannot be dropped is shown by the following theorem.[24]

[23] The present phrasing is from Lindström [43]. Feferman's original notation is "RE-formula", but in modern usage it is uncommon to talk about a formula being recursively enumerable. Strictly, an RE-formula is an existentially quantified primitive recursive formula, allowing no bounded universal quantification. By the MRDP theorem, due jointly to Y. V. Matijasevič, J. Robinson, M. Davis and H. Putnam, any Σ_1-formula is equivalent in PA to an existential formula—thus our present formulation of Theorem 2.5.2. See e.g. Davis [9] for more on this matter.

[24] See also Franzén [15], for a more thorough study of what incompleteness theorems do *not* say.

Theorem 2.5.3. *If T is recursive, and $\tau(x)$ is a primitive recursive formula that binumerates T in T, then the formula*

$$\tau^*(x) := \tau(x) \wedge Con_{\tau|x}$$

binumerates T in T, and $PA \vdash Con_{\tau^*}$.[25]

The numeration τ^* is a Π_1-formula, and the sentence Con_{τ^*} an intensionally incorrect way of expressing the consistency of T. While the formula τ^* extensionally corresponds to the set of axioms of T, as $T \vdash \tau^*(\varphi)$ if and only if φ is an axiom of T, it is not a correct description of φ being an axiom of T—rather of being an axiom of a subsystem of T which is always consistent, regardless whether T is consistent or not. Accordingly, the consistency of T is rather formulated as Con_τ, i.e. as the consistency of a binumeration $\tau(x)$ of T. In *The lattice of bi-numerations of arithmetic* [24, 25], Hájková shows that there is in fact no natural distinguished choice of such a binumeration.

A related concept is that of *relative consistency proofs*. While the incompleteness theorems rules out most systems as unable to prove their own consistency, it is still possible to prove a system to be consistent, *relative to some other system*. Consistency statements based on specific binumerations, and the possibility to binumerate one theory inside another, provides a powerful tool for establishing such relative consistency results. Examples of this may be found in Bennet's *On some orderings of extensions of arithmetic* [1]. Another way of relating theories is by interpreting one theory in another. S is said to be *interpretable* in T if there is a function $t(x)$ (translating the language of S into the language of T) such that if $S \vdash \varphi$, then $T \vdash t(\varphi)$. The connection between interpretability and consistency statements can be formulated as the following theorem, due to Feferman:

Theorem 2.5.4. *S is interpretable in T if and only if there is a formula $\sigma(x)$ numerating S in T such that $T \vdash Con_\sigma$.*

Actually, the main interest in Theorem 2.5.3 is to be found in this context. We will see more of interpretability in Section 2.7. All in all, it can be rightly said that Feferman's work gave the formal grounds for doing metamathematics in an organised way.

[25] If X is a set numerated by $\tau(x)$, then $X|k = \{n \in X : n \leq k\}$, and $\tau|y(x)$ is the formula $\tau(x) \wedge x \leq y$. For a proof of Theorem 2.5.3, see Feferman [11] or Lindström [43].

2.6 Variations on the fixed point theme

In the early 1960s, quite a few variations on the theme of fixed-point constructions was developed, each with different applications. Closely related to the investigations in *Arithmetization of metamathematics...* are the results of Ehrenfeucht and Feferman in *Representability of recursively enumerated sets in formal theories* [10]. Their main result is that certain theories T that represents all recursive functions, also represents all r.e. sets. To prove this theorem, they needed a parametrical version of the Fixed point theorem:

Theorem 2.6.1 (Ehrenfeucht and Feferman's Fixed point theorem)**.** *For any arithmetical formula $\varphi(x_0, x_1)$ with two free variables, there is a formula $\delta(x)$ with one free variable such that, for all numbers k,*

$$T \vdash \delta(k) \leftrightarrow \varphi(k, \delta(k)).$$

As expected, the proof is a variation of the proof of Carnap's Fixed point theorem. Instead of just substituting the Gödel number of a formula for x in the formula itself, we simultaneously make another substitution as well. Let $\sigma'(i, \varphi, k, \gamma)$ represent "γ is the result of substituting k for x_i in φ", and let θ be the following formula:

$$\exists z \exists w \big(\sigma'(1, x_1, x_1, z) \wedge \sigma'(0, x_0, z, w) \wedge \varphi(x_0, w)\big).$$

Thus z is the result of substituting x_1 for x_1 in x_1, and w is the result of substituting x_0 for x_0 in z. Let δ be the formula $\theta(x_0, \theta)$. It follows from this construction that

$$T \vdash \forall z \big(\sigma'(1, \theta, \theta, z) \leftrightarrow z = \delta\big),$$

and, for each number k,

$$T \vdash \forall w \big(\sigma'(0, k, \delta, w) \leftrightarrow w = \delta(k)\big).$$

Note the differences between this kind of substitution and the ones used before. Gödel, Tarski and Carnap constructs functions, whose results are new formulas resulting from carrying out the substitution. Ehrenfeucht and Feferman instead formalises the relation between the formula being substituted in, and the resulting formula. This is what yields e.g. the two equivalences above.

Another variation of a fixed-point construction is Richard Montague's "formulas in second person" from the paper *Theories incomparable with respect to relative interpretability* [46]. Consider the case where we do not only want a

sentence to express that it possesses a given property, but rather two or more different sentences prescribing different properties to each other. This can be formulated as

Theorem 2.6.2 (Montague's Second person fixed point theorem)**.** *For any arithmetical formulas $\varphi_0, \ldots, \varphi_n$ whose free variables include $v_0, \ldots, v_n$, there are formulas $\delta_0, \ldots, \delta_n$ whose free variables are among those of $\varphi_0, \ldots, \varphi_n$ but not among $v_0, \ldots, v_n$, and which are such that, for $i = 0, \ldots, n$,*

$$T \vdash \delta_i \leftrightarrow \varphi_i(\ulcorner \delta_0 \urcorner, \ldots, \ulcorner \delta_n \urcorner).$$

Here it is understood that the Gödel number for each formula δ_i is to be substituted for the variable v_i. As one might expect, the techniques needed for this construction is more complicated than the ones we have encountered before. In order to keep some readability and transparence of ideas, we use the notation $\ulcorner \varphi \urcorner$ for the numeral denoting the Gödel number of φ. For the proof let, for $i = 0, \ldots, n$, the function d_i be defined as

$$d_i(\chi_0, \ldots, \chi_n) = \chi_i(\ulcorner \chi_0 \urcorner, \ldots, \ulcorner \chi_n \urcorner).$$

Each d_i is clearly recursive, and is thus represented by an $n+1$-place formula σ_i in T. For $i = 0, \ldots, n$, let

$$\chi_i = \forall x_0 \ldots \forall x_n \big(\sigma_0(v_0, \ldots, v_n, x_0) \wedge \cdots \wedge \sigma_n(v_0, \ldots, v_n, x_n) \rightarrow \varphi_i(x_0, \ldots, x_n)\big),$$

where $x_0, \ldots, x_n$ are distinct new variables, and let

$$\delta_i = d_i(\chi_0, \ldots, \chi_n).$$

Note that no δ_i can contain any free occurrence of the variables $v_0, \ldots, v_n$, since in each χ_i any free variable is substituted by a χ_k for some $k \leq n$. By construction of χ_i, we also see that none of the x_i's are free, but that there might still be free occurrences of variables from φ_i.

It follows that for each $i \leq n$,

$$T \vdash \delta_i \leftrightarrow \varphi_i(\ulcorner \delta_0 \urcorner, \ldots, \ulcorner \delta_n \urcorner).$$

The last example of a general fixed point theorem is Montague's free variable variation, or the diagonalisation theorem.

Theorem 2.6.3 (The Diagonalisation theorem)**.** *For $n>0$ and any arithmetical formula φ whose free variables are $z, x_0, \ldots, x_n$, there is a formula δ with only $x_0, \ldots, x_n$ free (and no free occurrence of z), such that*

$$T \vdash \delta(x_0, \ldots, x_n) \leftrightarrow \varphi(\ulcorner \delta(z) \urcorner, x_0, \ldots, x_n).$$

Let $d(\chi)$ be the formula $\forall z((z = \ulcorner \chi \urcorner) \rightarrow \chi)$. As earlier, this function is recursive, so it is represented by some formula σ. It follows that for each χ,

$$T \vdash \sigma(\ulcorner \chi \urcorner, y) \leftrightarrow (y = \ulcorner d(\chi) \urcorner).$$

For each φ, let

$$\chi_\varphi = \forall y(\sigma(z, y) \rightarrow \varphi(z, x_0, \ldots, x_n)),$$

and let $\delta(\varphi) = d(\chi_\varphi)$. Then

$$T \vdash \delta(\varphi) \leftrightarrow \varphi(\ulcorner d(\chi_\varphi) \urcorner, x_0, \ldots, x_n).$$

Similar to the case of the Second person fixed point theorem, it follows that $x_0, \ldots, x_n$, but not z, are free in $\delta(\chi_\varphi)$ and $\varphi(\ulcorner d(\chi_\varphi) \urcorner, x_0, \ldots, x_n)$. Finally, since $\delta(\varphi) = d(\chi_\varphi)$, the theorem is proved.

In his beautiful exposé of fifty years of self reference, Craig Smoryński points out that

> Montague states the final result—the Diagonalisation Theorem [...] This is precisely the form analogous to Kleene's original formulation of the Recursion Theorem back in 1938.[26]

Finally, let us for a second turn our attention to the field of *recursion theory*. The theory of recursive functions originates with the works of e.g. Turing, Church, Kleene, Herbrand and Gödel in the 1930s. The idea was to explicate the informal notion of a computable function in a formally acceptable way. For the present purposes we define a partial recursive function as a recursive function that need not be defined for all arguments. Given a recursive enumeration of all partial recursive functions, let $\{e\}$ be the eth function of this listing. The recursion theorem can then be stated as follows:

Theorem 2.6.4 (The recursion theorem (Kleene [35], 1938))**.** *For each $n>0$, for any partial recursive function $\varphi(z, x_0, \ldots, x_n)$, a number e can be found which defines $\varphi(e, x_0, \ldots, x_n)$ recursively, i.e. such that*

$$\{e\}(x_0, \ldots, x_n) \equiv \varphi(e, x_0, \ldots, x_n),$$

[26] Smoryński, [60], p. 358.

where the equivalence means that if either of the functions is defined, then the other function is defined and they both take the same value; if either of the functions is undefined, then the other also is undefined.

We omit the proof. The interested reader is directed to e.g. Rogers [53] or Kleene [36]. The correspondence between this theorem and Montagues's diagonalisation theorem above should, however, be evident.

2.7 Reaching Maturity: Conservativity and interpretability

There are (at least) two different ways to compare the strength of two theories. One approach is to ask whether a theory T is at least as strong as a theory S (formulated in the same language as T) in the sense that T proves every theorem of S ($T \vdash S$), i.e. $\mathrm{Th}(S) \subseteq \mathrm{Th}(T)$. As we have seen in section 2.5, another way of regarding the question is whether S can be interpreted in T ($S \leq T$). This means, roughly speaking, that the concepts and the range of the variables of S can be expressed in T in such a way as to turn every theorem of S into a theorem of T.

When presented to the question of provability strength, the first of the notions mentioned above, it is reasonable to ask whether a theory can be extended in a *partially conservative* way, i.e. if there exists extensions S of T such S proves the same sentences within a certain class as T does. It is clear that every sentence that is provable in T is conservative over T, as a provable sentence gives no new information when added to the theory. In his 1979 paper *Partially conservative extensions of arithmetic*, David Guaspari formally introduced the concept of partial conservativity:

Definition 2.7.1. *Let Γ be either Σ_n or Π_n.*[27] *A sentence is Γ-conservative over T if and only if any Γ-sentence provable in $T + \varphi$ is provable in T already. The set of sentences Γ-conservative over T is denoted $Cons(\Gamma, T)$.*

A folklore example of a partially conservative sentence is $\neg\mathrm{Con}_T$ which is Π_1-conservative over T. The sentence $\neg\mathrm{Con}_T$ is Σ_1, and $T + \neg\mathrm{Con}_T$ proves exactly the same Π_1-sentences as T does. For suppose that $T + \neg\mathrm{Con}_T \vdash \varphi$, where φ is a Π_1-sentence. Then it follows that $\mathrm{PA} \vdash \mathrm{Pr}_T(\neg\varphi) \rightarrow \mathrm{Pr}_T(\mathrm{Con}_T)$, so $\mathrm{PA} \vdash$

[27] In the rest of this chapter, we occasionally allow ourselves to use unexplained technical terminology, most of which is basic metamathematics or explained in Chapter 3. E.g., the definition of the classes Σ_n and Π_n can be found in Definition 3.2.2 below.

$\mathrm{Pr}_T(\neg\varphi) \to \neg\mathrm{Con}_{T+\neg\mathrm{Con}_T}$. Since $\neg\varphi$ is a Σ_1-sentence, we have $\mathrm{PA} \vdash \neg\varphi \to \mathrm{Pr}_T(\neg\varphi)$. Finally, $\mathrm{PA} + \mathrm{Con}_T \vdash \mathrm{Con}_{T+\neg\mathrm{Con}_T}$. Combining these observations, we obtain $\mathrm{PA} \vdash \neg\varphi \to \neg\mathrm{Con}_T$, so $T \vdash \varphi$, and $\neg\mathrm{Con}_T$ is Π_1-conservative over T.

Guaspari shows in his paper that for every complexity class Γ, there is a sentence in Γ^d which is Γ-conservative over T, and yet unprovable in T. This application requires a more sophisticated fixed-point construction than seen before, as well as the use of a *partial truth definition for* T ($\mathrm{Tr}_\Gamma(x)$), which is a formula such that, for every sentence γ in Γ, $T \vdash \gamma \leftrightarrow \mathrm{Tr}_\Gamma(\gamma)$. We will use the following fixed-point contruction to yield the result, which is not Guaspari's original, but rather a slight variation due to Per Lindström. Let φ be such that

$$T \vdash \varphi \leftrightarrow \exists y \exists u{\leq}y \exists v{\leq}y \big(\Gamma(u) \wedge \mathrm{Prf}_{T+\varphi}(u,v) \wedge \neg\mathrm{Tr}_\Gamma(u) \wedge \forall z{\leq}y \neg\mathrm{Prf}_T(\varphi,z)\big),$$

where $\Gamma(x)$ is a binumeration of the set of Γ-sentences.

First we show that $T \nvdash \varphi$.

1. Suppose that $T \vdash \varphi$. Then $T \vdash \mathrm{Prf}_T(\varphi,p)$, for some p. But $T + \varphi \vdash \forall u,v{\leq}m(\Gamma(u) \wedge \mathrm{Prf}_{T+\varphi}(u,v) \to \mathrm{Tr}_\Gamma(u))$, where $m = \max(\varphi,p)$. This implies, by construction, that $T + \varphi \vdash \neg\varphi$. Thus $T \nvdash \varphi$.

2. It remains to show that φ is Γ-conservative over T. This follows if $T+\varphi \vdash \theta$ implies that $T+\neg\theta \vdash \varphi$, for every $\theta \in \Gamma$. So suppose that θ is a Γ-sentence such that $T + \varphi \vdash \theta$. By the reasoning above, it follows that $T + \neg\theta$ proves that θ is a false Γ-sentence that is provable in $T + \varphi$, and since $T \nvdash \varphi$, $\neg\,\mathrm{Prf}_T(\varphi,k)$ is provable for every k. Thus, again by construction, $T + \neg\theta \vdash \varphi$, whence $T + \neg\varphi \vdash \theta$. As we have supposed that $T + \varphi \vdash \theta$, we get $T \vdash \theta$, as desired.

Returning to the concept of relative interpretability, notice that if $T \vdash S$, then S is interpretable in T via the identity function. The beautiful relationship between the different ways of measuring the strength of theories is presented as the theorem below. The equivalence of statements 1 and 2 was shown by Stephen Orey [48]; the equivalence of 1 and 3 by Guaspari [17] and Lindström [38] (independently); the remaining equivalence is due to Feferman [11], using the construction in Theorem 2.5.3.

Theorem 2.7.2. *The following statements are equivalent:*

1. $S \leq T$,

2. $S|k \leq T$ *for every* k,

3. $S \dashv_{\Pi_1} T$,

4. There is a formula $\tau(x)$ *(bi)numerating* S *in* T *such that* $T \vdash Con_\tau$.

Let us for a moment return to the foundational program of Hilbert. It has been argued that what Hilbert took to be "contentful" mathematics is precisely the Π_1-sentences.[28] Remember that we called Hilbert's contentful finitistic system, supposedly being the foundation of all of mathematics, T_1, and the purely formal, stronger system T_2. Even though the incompleteness theorems show that Hilbert's program can not be fully realised, one could argue that a different perspective (based on e.g. interpretability) might be successful. By the theorem above, this cannot be true. A realisation of Hilbert's program would then consist of finding an interpretation of T_2 in T_1, since Hilbert's original approach was to prove the consistency of T_2 in T_1 (i.e. proving $T_2 \dashv_{\Pi_1} T_1$). However, since the stronger theory is supposed to encompass all of mathematics, including the finitary system, T_2 properly extends T_1. So if T_1 proves T_2 to be consistent, then T_2 proves its own consistency.

The 1980s were rich in results in conservativity and interpretability. Investigations were performed by e.g. Lindström [39, 40, 41], Švejdar [67], and Visser [71], building on earlier work by Feferman, Kreisel and Orey (e.g. [11, 48]). A deep presentation is available in Chapters 6 and 7 of Aspects of Incompleteness [43].

2.8 Topics and open problems

To sum up this historical survey, we take time to present a few recent topics in metamathematics. They are chosen for at least one of the following three reasons: to examplify the strength of the methods of metamathematics; to show how a recent application have called for greatly increased complexity of fixed-point constructions; to present an open problem related to the area developed in the following chapter.

Lindenbaum algebras and partial conservativity The Lindenbaum algebra for a first-order theory T is the structure $(X, \oplus, \otimes, 0, 1)$, where X is the set of sentences modulo provable equivalence in T, $\oplus$ and $\otimes$ are disjunction and conjunction, respectively, 0 is the equivalence class of sentences refutable in T, and 1 is the equivalence class of sentences provable in T. This structure is the

[28] See e.g. Tait [68], Simpson [58].

countable, atomless Boolean algebra. A *partial Lindenbaum algebra* Γ^T is the Lindenbaum algebra for T restricted to equivalence classes of Γ-sentences. This structure is a countable, dense, distributive lattice with 0 and 1, but it is not a Boolean algebra. Let the reduction principle for Γ^T be the following sentence:[29]

$$\forall a_0, a_1 \exists b_0, b_1 (a_0 \oplus a_1 = b_0 \oplus b_1 \wedge b_0 \otimes b_1 = 0 \wedge b_0 \leq a_0 \wedge b_1 \leq a_1).$$

It is not difficult to see that the reduction principle holds in all Σ_n^T. That Σ_n^T is not isomorphic to Π_m^T for any n, m, is shown by the following theorem. It is due to Bennet [1], as are the rest of the results of this paragraph.

Theorem 2.8.1. *The reduction principle is false in* Π_n^T.

We will not give the full proof, but rather present the complicated fixed-point construction used in proving the following lemma, which is in turn used to prove the theorem.

Lemma 2.8.2. *If X is an r.e. set, then there are Σ_n-formulas $\theta_0(x), \theta_1(x)$ such that, for $i = 0, 1$,*

1. *$T + \theta_i(k) \vdash \neg\theta_{1-i}(k)$, for all k,*
2. *if $k \in X$, then $T \vdash \neg\theta_i(k)$,*
3. *if $k \notin X$, then $\theta_i(k) \in Cons(\Pi_n, T + \theta_{1-i}(k))$.*

To prove the lemma, let $G(x, y)$ be a p.r. relation such that $X = \{k : \exists m G(k, m)\}$ and let $\gamma(x, y)$ be a p.r. binumeration of G. For $i = 0, 1$, let $\xi_i(x)$, $\rho_i(x)$, and $\theta_i(x)$ be such that the following is provable in T for all k

$$\xi_i(k) \leftrightarrow \exists z \big(\exists u, v \leq z \big(\Pi_n(u) \wedge \mathrm{Prf}_{T+\theta_i(k)}(u, v) \wedge \neg \mathrm{Tr}_{\Pi_n}(u)\big) \wedge \forall u \leq z \gamma(k, u)\big)$$
$$\xi_i(k) \leftrightarrow \exists z \rho_i(k, z)$$
$$\theta_i(k) \leftrightarrow \exists z \big(\rho_i(k, z) \wedge \forall y \leq z \neg\rho_{1-i}(k, y)\big)$$

We omit the details of the proof, but it should be clear that this application of double self reference is far more complicated than any we have seen so far. It is an open question whether Σ_n^T is isomorphic to Σ_m^T for any $n, m > 1$.

Guaspari has raised the following question on partial conservativity: given an r.e. sequence of theories, are there always Γ-sentences which are Γ^d-conservative over each theory in the sequence and unprovable in each of them? For hereditary partial conservativity, the answer is given by the following theorem.

[29] See Rogers [53].

Theorem 2.8.3. *Given two theories T_0, T_1, the following are equivalent:*

1. $\Gamma \cap \mathit{HCons}(\Gamma^d, T_0) \setminus \mathit{Th}(T_1) = \emptyset$,
2. $\Gamma \cap \mathit{HCons}(\Gamma^d, T_0) \setminus (\mathit{Th}(T_0) \cup \mathit{Th}(T_1)) = \emptyset$,
3. $\mathit{Th}_\Gamma(T_1)$ *is inconsistent with* T_0.

For partial conservativity, less is known. In the Π_n-case, Bennet [1] gives the following answer

Theorem 2.8.4. *Given two theories T_0, T_1, the following are equivalent:*

1. $\Pi_n \cap \mathit{Cons}(\Sigma_n, T_0) \setminus \mathit{Th}(T_1) = \emptyset$,
2. $\Pi_n \cap \mathit{Cons}(\Sigma_n, T_0) \setminus (\mathit{Th}(T_0) \cup \mathit{Th}(T_1)) = \emptyset$,
3. $\mathit{Th}_{\Pi_n}(T_0) \subseteq \mathit{Th}_{\Pi_n}(T_1)$ *and* $\mathit{Th}_{\Pi_n}(T_1)$ *is inconsistent with* T_0.

In the Σ_n-case, it is an open question, to which we will return in Chapter 3, whether we can have

$$\Sigma_n \cap \mathrm{Cons}(\Pi_n, T_0) \setminus \mathrm{Th}(T_1) = \emptyset.$$

Some results on interpretability The relation of mutual interpretability is an equivalence relation, and its equivalence classes are called degrees (of interpretability). The lattice $\mathbf{D}_T$ of degrees of interpretability (of extensions of T) was introduced by Lindström in 1979 [38]. In the same paper, Lindström shows that this lattice is isomorphic to the lattice $\mathbf{V}_T$ of finite extensions of T, introduced by Švejdar in 1978 [67]. There are a number of open problems in relation to this lattice.

1. If T is Σ_1-sound, but S is not, then $\mathbf{D}_T$ and $\mathbf{D}_S$ are not isomorphic. It is not known whether Σ_1-soundness is a sufficient condition for $\mathbf{D}_T$ and $\mathbf{D}_S$ to be isomorphic.

2. It is known that there are non-isomorphic intervals of $\mathbf{D}_T$, but the total picture is still lacking.

3. Let $\mathbf{G}_T$ be the set obtained from the Σ_1- and Π_1- degrees by closing under join and meet. It is an open problem whether $\mathbf{G}_T$ is isomorphic to $\mathbf{D}_T$.

4. We say that a cups to b if there is a $c < a$ such that $a \cup c = b$, where $\cup$ denotes join in the lattice. For every Π_1-degree $a > 0_T$, there is a Σ_1- (and a Π_1-) degree which cups to a. It is not known whether this holds for every degree.

5. If b is the greatest element of the set $\{c : a \cap c = 0\}$, then b is the pseudo-complement of a. Lindström has shown that every Σ_1-degree is the pseudo-complement of some degree. It is an open question whether the converse is true, that is, if the Σ_1-degrees can be characterised in a purely algebraic way as those degrees that are pseudo-complements.

Rosser sentences As we have seen, it is a consequence of Gödel's second incompleteness theorem that every Gödel sentence is provably equivalent to a consistency statement for T. In the case of Rosser sentence, no similar result is available. It is not even clear whether the set of Rosser sentences is contained in a single equivalence class. Guaspari and Solovay have shown that these properties of the set of Rosser sentences are in fact dependent on choice of proof predicate. In their 1979 paper [17], they construct two different proof predicates, each being extensionally equivalent on natural numbers to the "usual" proof predicate, of which the first has fixed points in only one equivalence class, while the second has fixed points in at least two different equivalence classes. Which of the two cases that holds for the "usual" proof predicate is an open question. It is not even clear how "usual" is to be defined in this context, and a thorough discussion of this problem is found in Chapter 6 in Smoryński's book *Self-reference and Modal Logic* [62].

On the relation *provable equivalence* and on partitions in effectively inseparable sets Smullyan proved in *The theory of formal systems* [64] that the set of provable sentences is effectively inseparable from the set of refutable sentences. In 1981, Bernardi [3] generalised this result to show that for any two sentences φ, ψ such that $T \nvdash \varphi \leftrightarrow \psi$ the sets $[\varphi] = \{\delta : T \vdash \delta \leftrightarrow \varphi\}$ and $[\psi] = \{\delta : T \vdash \delta \leftrightarrow \psi\}$ are effectively inseparable. To see this, let A and B be any disjoint, r.e. supersets of $[\varphi]$ and $[\psi]$, respectively, and let $\xi(x)$ be a formula such that if $k \in A$, then $T \vdash \xi(k)$ and if $k \in B$, then $T \vdash \neg\xi(k)$. By the Fixed point theorem, let δ be such that

$$T \vdash \delta \leftrightarrow \big((\xi(\delta) \wedge \psi) \vee (\neg\xi(\delta \wedge \varphi)\big).$$

Such a δ can be effectively found. Suppose that $\delta \in A$. Then $T \vdash \delta \leftrightarrow \psi$, contradicting that $[\psi] \subseteq B$. Conversely, if $\delta \in B$, then $T \vdash \delta \leftrightarrow \varphi$, contradicting

that $[\varphi] \subseteq A$. Thus we have effectively found a $\delta \in (A \cup B)^c$, and shown that $[\varphi]$ and $[\psi]$ are effectively inseparable.

On a new notion of partial conservativity In 1984, Petr Hájek presented a new notion of partial conservativity [22]. It is related to his earlier work on the length of (non-standard) proofs of consistency and inconsistency. Let $\mathcal{M}$ be a non-standard model of PA, let c be a non-standard element of $\mathcal{M}$, and let PA_c be $\mathrm{PA} + \{c \geq n : n \in \mathbb{N}\}$. A bounded formula $\varphi(x)$ of one free variable is $(2^{2^c}, c)$-conservative if for each bounded formula $\psi(x)$, $\mathrm{PA}_c + \varphi(2^{2^c}) \vdash \psi(c)$ implies $\mathrm{PA}_c \vdash \psi(c)$.

Theorem 2.8.5 (Hájek, 1984 [22])**.** *There is a formula φ such that $\varphi(x)$ implies "beneath x there is a proof of a contradiction in PA", and φ is $(2^{2^c}, c)$-conservative.*

Corollary 2.8.6. *For each $\mathcal{M} \models PA$, and each non-standard element a of $\mathcal{M}$, there is a $\mathcal{K} \models PA$ which is identical with $\mathcal{M}$ up to a, and such that in $\mathcal{K}$ there is a proof of contradiction beneath 2^{2^a}.*

This corollary can be seen as a strengthening of the second incompleteness theorem, in that it shows that there is a "rather *short*" proof of contradiction.[30] We will again omit the proof, and just state the fixed-point construction used to prove the result. The notation is quite involved, but the details are not of importance here; the point being that this is an entirely new kind of self reference.

In what follows $\varphi^{\leq x}$ is a conversion of the sentence φ into a bounded formula of one free variable, by replacing each $+$ and $\cdot$ by formulas of the form $w = u+v$ and $w = u \cdot v$ and restricting all quantifiers to x. Let $x \Vdash y$ express (in a certain technical way, beyond the scope of this text) that x satisfies y. $\Gamma(n)$ expresses certain properties of the relation $\Vdash$. Let c and $\hat{c}$ be distinct non-standard elements, and let as before, $\mathrm{PA}_{\hat{c}}$ be an expansion of PA. Finally, let λ be such that:

$$\mathrm{PA}_c \vdash [2^{2^c} \Vdash \lambda] \leftrightarrow [2^{2^c} \Vdash \exists s \exists z (\Gamma(\neg s) \wedge \mathrm{Prf}_{\mathrm{PA}_{\hat{c}}}(\lambda^{\leq \hat{c}} \rightarrow s^{\leq c}, z))].$$

That this construction is at all admissible is shown by modifying the Fixed point theorem. The interested reader is directed to Hájek [22].

[30] Hájek [22], p. 218.

Partially generic formulas in arithmetic In 1988, Lindström [42] exhibited a general type of fixed-point construction with numerous applications. Let T be an extension of PA with a new monadic predicate G. Formulas containing G will be written $\varphi(G;x)$. φ is $\Gamma[G]$ if $\varphi(\xi;x)$ is Γ whenever $\xi(x)$ is p.r. If X is any set of natural numbers, then $X|q = \{k \in X : k \leq q\}$, and if X is finite, then $[X](x) = \bigvee\{x = k : k \in X\}$. The notation $\tilde{x}$ is short for $x_0, \ldots, x_{n-1}$. If $\xi(x)$ is any formula, then $\varphi(\xi;\tilde{x})$ is obtained from φ by replacing G by $\xi(x)$, avoiding clashes of variables. When confusion may arise, the notation $\lambda x \xi(x)$ will be used.

Definition 2.8.7. *$\xi(x)$ is χ-generic in T if for all $\tilde{k}$, if $T \vdash \chi(\xi;\tilde{k})$, then there is a q such that $T \vdash \chi([X|q];\tilde{k})$. Here X is the set numerated by $\xi(x)$ in T.*

The proof that a χ-generic numeration exists uses an intriguing fixed-point construction. Let τ be a p.r. binumeration of T, and $\kappa(x,y)$ be a p.r. formula such that $X = \{k : \exists m \mathrm{PA} \vdash \kappa(k,m)\}$, and let $\xi_0(x)$ be such that, for all k,

$$\mathrm{PA} \vdash \xi_0(k) \leftrightarrow \exists y\big(\kappa(k,y) \wedge \forall z y{\leq}k{+}y\big(\mathrm{Prf}_\tau(\chi(\xi;z),u) \rightarrow \chi(\lambda w \exists v(v{+}w{\leq}u \wedge \kappa(w,v));z)\big)\big).$$

From the existence of certain kinds of χ-generic numerations, Lindström provides simple proofs for e.g. the following theorems:

Theorem 2.8.8. *If S is r.e., then there is a Σ_1-numeration $\sigma(x)$ of S in T such that $Pr_\sigma(x)$ numerates $Th_T(S)$ in T.*

Note that this is only of interest in the case where T is not Σ_1-sound.

Theorem 2.8.9 (Guaspari [18])**.** *For any r.e. set X, there is a Γ-formula $\xi(x)$ such that*

1. *If $k \in X$, then $T \vdash \xi(k)$,*
2. *if $k \notin X$, then $\neg\xi(k) \in Cons(\Gamma, T)$.*

Theorem 2.8.10 (Smorynski [61])**.** *Let X_0 and X_1 be disjoint r.e. sets. Then there is a Π_{n+1}-formula $\xi(x)$ such that*

1. *If $k \in X_i$, then $T \vdash \xi_i(k)$,*
2. *if $k \notin X_0 \cup X_1$, then $\xi(k)$ and $\neg\xi(k)$ are Σ_{n+1}- and Π_{n+1}-conservative over T, respectively.*

There are further examples of related applications of this general fixed-point construction, mainly concerned with numerations of r.e. sets and partial conservativity. Lindström also refines a result of H. Lessan [37] on non-standard models of arithmetic.

The Lindenbaum fixed point algebra is undecidable A fixed point algebra is a pair of Boolean algebras (A, B), where the elements of B are mappings $A \to A$, satisfying the following conditions (elements of A are denoted by latin letters and those of B by Greek ones):

1. $\alpha = \beta$ iff $\forall a(\alpha a = \beta a)$,
2. $(\alpha \# \beta)a = \alpha a \# \beta a$, where $\#$ is any Boolean operation,
3. $\forall a \exists \alpha \forall b(\alpha b = a)$,
4. $\forall \alpha \exists a(\alpha a = a)$.

A fixed point algebra associated with an r.e., consistent theory containing PA will be called a Lindenbaum fixed point algebra.

Theorem 2.8.11 (Shavrukov, 1991 [57])**.** *The first-order theory of a Lindenbaum fixed point algebra is hereditarily undecidable.*

The proof proceeds by defining a recursive procedure during which Gödel numbers of some sentences may be "painted" black. The arithmetical formula $B(x)$ shall be read as "x is eventually painted black", and the procedure is defined by stages using the Gödel number of $B(x)$, justified by the formalised recursion theorem. The idea is not entirely new—the procedure is an example of a *Solovay function*, a type of construction used by e.g. Guaspari & Solovay in [17]. In view of the connection between the Fixed point theorem and the recursion theorem, it is clear that this proof can be seen as a fixed-point construction.

2.9 Closing time

Central to the discussion of this chapter is the notion of fixed points. In the following chapter we will concentrate on the set of fixed points of an arbitrary arithmetical formula. Being able to describe the fixed points of a given formula in more detail might cast some light on e.g. the open problem on Rosser sentences, as stated above. As we will see later, there are also some connection to problems on partial conservativity stated by Guaspari and partially answered by Bennet [1].

We hope to have shown that the field of metamathematics is an interesting one, rich in deep results. In particular, the method of constructing fixed points lends itself to a diverse number of applications, though most of foundational character. Indeed, metamathematics is a foundational study of mathematics, seeking answers to what may be proved by formal means, and what may not. It investigates relations between different theories and axiomatic systems, indicating whether your formal system is adequate for your purposes. It has implications on how the notion of truth can be handled in a formal way. It might not change the way a non-foundationalist mathematician works, but it secures and tests the ground for any mathematical enterprise. We also hope to have paved the way for placing our own research in this context. While the area might not be as active as in, say, the 1970s, there are still interesting and difficult questions left unanswered.

Chapter 3

Sets of fixed points

3.1 Introduction

Up to this point we have discussed metamathematics in general, and a few recent topics in the field. In this chapter we turn to investigating sets of fixed points of an arithmetical formula. It is clear from Gödel [19] that each fixed point to $\neg \Pr_T(x)$ is provably equivalent in T to the statement Con_T. If we consider the set of Henkin fixed points, i.e. the set

$$\{\delta : T \vdash \delta \leftrightarrow \Pr(\delta)\},$$

it was shown by Löb [44] that this set is equal to the set of all sentences that are provable in T. From Smullyan [64] we have the result that for a large class of formal systems, the set of fixed points of $\Pr(x)$ is recursively inseparable from the set of fixed points of $\neg \Pr(x)$, and thus that both are Σ_1-complete. A later consideration is from di Paola [49] and the alternative proof by Bernardi [4], where a formula is constructed that has, among others, all elements of a given recursive set as fixed points.

Our first approach will be to study the set of *all* fixed points of a given formula. This project can be viewed in (at least) two different ways. On the one hand, if we are given a formula, we may describe its set of fixed points. On the other hand, if we are given a set of sentences, we may ask whether there is a formula with exactly the elements of this set as its fixed points. The first part of this chapter is devoted to finding partial solutions to these two questions. In the second part, we also study structural properties of sets of fixed points

ordered under set inclusion. In the final section we relate the present research to other areas, and state of some open questions.

3.2 Preliminaries

In what follows, every theory is assumed to be a consistent, p.r. extension of Peano arithmetic, formulated in the arithmetical language $\mathscr{L}_A$.[1] Such a theory is denoted $T, S, T_0, T_1, \ldots$. As noted in the introduction, the assumption that theories contain PA is everywhere too strong; for most of the proofs we could do with extensions of Robinson's arithmetic Q. Occasionally we will need enough induction to handle partial truth definitions, and in such cases Q plus induction for Σ_1-formulas would suffice. We have, however, not ventured to achieve the best possible results in this respect; indeed, our focus is on properties shared by all theories containing a sufficient amount of arithmetic. The terminology if this paper is closely related to that of Lindström [43], but we explicitly state some definitions that are used in the sequel. Initially, we presume any standard Gödel numbering of terms and formulas, and identify formulas with the numerals for their respective Gödel numbers.

Definition 3.2.1. *A formula* $\xi(x_0, \ldots, x_n)$ numerates *a relation* $R(k_0, \ldots, k_n)$ *in* T *if for all* $k_0, \ldots, k_n$,

$$R(k_0, \ldots, k_n) \text{ iff } T \vdash \xi(k_0, \ldots, k_n).$$

In particular, $\xi(x)$ *numerates a set* X *in* T *if for every* k,

$$k \in X \text{ iff } T \vdash \xi(k).$$

A formula $\xi(x_0, \ldots, x_n)$ binumerates *a relation* $R(k_0, \ldots, k_n)$ *in* T *if for all* $k_0, \ldots, k_n$,

$$R(k_0, \ldots, k_n) \text{ iff } T \vdash \xi(k_0, \ldots, k_n) \text{ and}$$
$$\text{not } R(k_0, \ldots, k_n) \text{ iff } T \vdash \neg\xi(k_0, \ldots, k_n).$$

Similar to the notion of numerating a set, we say that $\xi(x)$ *binumerates a set* X *if for every* k,

$$k \in X \text{ iff } T \vdash \xi(k) \text{ and}$$
$$k \notin X \text{ iff } T \vdash \neg\xi(k).$$

[1] Craig [7] has shown that for each r.e. set X, there is a p.r. set which is deductively equivalent to X. Thus we can freely use r.e. extensions of PA.

The *primitive recursive* (p.r.) functions are the functions constructed from constant functions, projection functions, and the successor function, by means of composition and primitive recursion. Gödel proved that every such function is definable by a formula in first-order arithmetic. The set of p.r. formulas can roughly be understood as the least set containing the formulas δ_f defining p.r. functions, closed under propositional connectives and bounded quantification. We will not linger on the technicalities of this point, and refer the reader to e.g. Lindström [43] and Rogers [53] for details, but note that these formulas contain no unbounded quantifiers. $\mathrm{Prf}(x, y)$ is a p.r. binumeration of the relation "y is a proof of x in T".

Definition 3.2.2 (The arithmetical hierarchy[2])**.** *Σ_n and Π_n are the least sets containing the p.r. formulas, closed under $\wedge$, $\vee$, and bounded quantification, and such that*

1. *$\Sigma_n \cup \Pi_n \subseteq \Sigma_{n+1} \cap \Pi_{n+1}$,*
2. *if ξ is Σ_n (Π_n), then $\neg\xi$ is Π_n (Σ_n),*
3. *if ξ_0 is Σ_n (Π_n) and ξ_1 is Π_n (Σ_n), then $\xi_0 \to \xi_1$ is Π_n (Σ_n),*
4. *if ξ is Σ_n (Π_n) and $\delta_f(x_0, \ldots, x_n, y)$ defines the function f, then $\exists z(\delta_f(x_0, \ldots, x_n, z) \wedge \xi)$ and $\forall z(\delta_f(x_0, \ldots, x_n, z) \to \xi)$ are Σ_n (Π_n).*
5. *Σ_n is closed under existential quantification,*
6. *Π_n is closed under universal quantification.*

It follows that $\Sigma_0 = \Pi_0 = \Delta_0$, which is the set of p.r. formulas. B_n is the set of boolean combinations of Σ_n-formulas. Δ_n^T is the set of Σ_n-formulas, that are provably in T equivalent to a Π_n-formula. Δ_n is the set Δ_n^{PA}. In what follows, Γ is either Σ_{n+1} or Π_{n+1}, and the dual of Γ (Γ^d) is Π_{n+1} if $\Gamma = \Sigma_{n+1}$, and conversely. Each set Γ is p.r., so there is a p.r. formula $\Gamma(x)$ binumerating this set in T. In writing Σ_n (Π_n, Δ_n, B_n) we almost always omit the (obvious) assumption that $n > 0$.

[2]It is common in other texts to use definitions of the type "$\exists\varphi$ is a Σ_{n+1}-formula if φ is a Π_n-formula", and letting the sets Σ_n and Π_n be closed under provable equivalence. Taking such an approach does not serve our purposes. As we will see later, we do *not* want Γ to be closed under provable equivalence. On the other hand, we want e.g. conjunctions of Σ_n-formulas to be Σ_n. While this these problems can be solved in other ways, e.g. by the use of normal form theorems, we have chosen the present, somewhat cumbersome definition, to provide for greater readability of the text.

Theorem 3.2.3 (Hilbert & Bernays [32]). *There is a* partial truth definition *for* Γ*-sentences, i.e. a* Γ*-formula* $Tr_\Gamma(x)$ *such that*

$$T \vdash \delta \leftrightarrow Tr_\Gamma(\delta) \text{ for all } \delta \in \Gamma.$$

Definition 3.2.4. *Let* $\theta(x)$ *be a formula with one free variable.* δ *is a* fixed point of $\theta(x)$ in T *if* $T \vdash \delta \leftrightarrow \theta(\delta)$.

We will construct fixed points by appealing to the following two variations of the Fixed point theorem:

Lemma 3.2.5 (Gödel [19]/Carnap [6]). *For any* Γ*-formula* $\theta(x)$*, we can effectively find a* Γ*-sentence* δ *such that*

$$T \vdash \delta \leftrightarrow \theta(\delta).$$

Lemma 3.2.6 (Ehrenfeucht & Feferman [10]). *For any* Γ*-formula* $\theta(x, y)$*, we can effectively find a* Γ*-formula* $\delta(x)$ *such that, for every* k*,*

$$T \vdash \delta(k) \leftrightarrow \theta(k, \delta(k)).$$

The concept of (hereditarily) partial conservativity, and the following lemma, are due to Guaspari.

Definition 3.2.7. *A sentence* φ *is* Γ-conservative over T *if, for every* Γ*-sentence* γ*,*

$$T + \varphi \vdash \gamma \Rightarrow T \vdash \gamma.$$

The set of such sentences is denoted $Cons(\Gamma, T)$*. Moreover,* φ *is* hereditarily Γ-conservative over T *(*$\varphi \in HCons(\Gamma, T)$*), if* φ *is* Γ*-conservative over every* S *such that* $T \vdash S \vdash PA$.

Lemma 3.2.8 (Guaspari [18]). *Let* X *be any r.e. set. Then there is a* Γ*-formula* $\xi(x)$ *such that*

$$k \in X \Rightarrow T \vdash \xi(k)$$
$$k \notin X \Rightarrow \neg\xi(k) \in Cons(\Gamma, T) \setminus Th(T).$$

Note that any such formula $\xi(x)$ *numerates* X *in* T.

Next we let $[\varphi]$ denote the (T-) equivalence class of a sentence φ, i.e. $[\varphi] := \{\psi : T \vdash \varphi \leftrightarrow \psi\}$. $[0]$ denotes the equivalence class of the refutable sentences, and $[1]$ the equivalence class of provable sentences. In contexts where we are

restricted to some set Γ we will (when needed for clarity) use the notation $[\varphi]_\Gamma$ for the set $[\varphi] \cap \Gamma$. Whenever X is a subset of Γ, we let X^c be $\Gamma \setminus X$.

Finally, we borrow some concepts and results from elementary recursion theory. For more information, see e.g. Rogers [53] or Soare [65].

Definition 3.2.9. *Two sets A, B are* effectively inseparable *if, for every disjoint pair (A', B') of r.e. supersets of A and B, we can effectively find an element of $(A' \cup B')^c$.*

A set X is *Turing reducible* to Y, $X \leq_T Y$, if there is a recursive function $f(x)$ such that $k \in X$ iff $f(k) \in Y$. If $X \leq_T Y$ and $Y \leq_T X$, then X and Y are *Turing equivalent* ($X \equiv_T Y$). Moreover, if A and B are infinite sets that differ on a finite set (i.e. $(A \setminus B) \cup (B \setminus A)$ is finite), then $A \equiv_T B$.

Definition 3.2.10. *A set Y is* Σ_1-complete *if for each r.e. (Σ_1-) set X, $X \leq_T Y$.*

It follows easily from these two definitions that if A and B are disjoint, effectively inseparable sets, then both A and B are *creative*. We will not give a definition of creative set, as the only aspect of creativeness we are interested in is that it implies Σ_1-completeness. On occasion, we also encounter creativity when using the following theorem to establish Σ_1-completeness:

Theorem 3.2.11 (Jockusch/Mohrherr)**.** *Let A be any r.e. set except $\mathbb{N}$. A is creative iff for every r.e. set B disjoint from A, it follows that $A \equiv_T A \cup B$.*

3.3 Recursion theoretic complexity

We now start investigating the recursion theoretic complexity of sets of fixed points. We define, for each formula $\theta(x)$ with exactly one free variable x, the set

$$\mathrm{Fix}^T(\theta) := \{\delta : T \vdash \delta \leftrightarrow \theta(\delta)\}.$$

This is the unbounded *set of fixed points of $\theta(x)$*, containing fixed points in every level of the arithmetical hierarchy. Note that we only allow sentences to be fixed points, i.e. δ contains no free variables. When it it is understood from the context which theory T we are discussing, we write $\mathrm{Fix}(\theta)$ for brevity.

Moreover, we define a bounded set of fixed points of a formula $\theta(x)$ by

$$\mathrm{Fix}^T_\Gamma(\theta) := \{\delta \in \Gamma : T \vdash \delta \leftrightarrow \theta(\delta)\},$$

The elements of Fix_Γ^T are denoted "Γ-fixed points (of $\theta(x)$)". Again, we often omit the reference to the theory T. We may also restrict the complexity of $\theta(x)$ and consider the set $\mathrm{Fix}_\Gamma(\theta)$ for a formula $\theta(x) \in \Gamma$. When doing so, this will be clear from the context.

Note that it is essential for the present results how the sets Γ are defined, as in this setting, $\mathrm{Fix}(\theta)$ contains sentences of arbitrarily high quantifier complexity. If we had chosen Γ to be Γ^T, i.e. closed under provable equivalence in T, the definitions of $\mathrm{Fix}(\theta)$ and $\mathrm{Fix}_\Gamma(\theta)$ would coincide in the following pathological way: Suppose $T \vdash \delta \leftrightarrow \theta(\delta)$. If $\theta \in \Gamma$ then $\delta \in \Gamma$. So, if $\theta \in \Gamma$ then $\mathrm{Fix}(\theta) = \mathrm{Fix}_\Gamma(\theta)$.

The two folklore examples of sets of fixed points, $\mathrm{Fix}(\mathrm{Pr})$ and $\mathrm{Fix}(\neg \mathrm{Pr})$, are Σ_1-complete sets. Thus one may wonder whether there are other formulas with Σ_1-complete sets of fixed points. A first proposition, providing another example of a Σ_1-complete set of fixed points, and initiating the research of the present chapter, is the following:

Proposition 3.3.1 (Bennet)**.** *The set of Rosser sentences of T is Σ_1-complete.*

Proof. Let X be any r.e. set, let $\rho(x)$ be the Rosser witness comparison construction $\forall z\big(\mathrm{Prf}(x,z) \to \exists u{\leq}z\,\mathrm{Prf}(\neg x,u))\big)$, and let, by Lemma 3.2.8, $\xi(x)$ be such that:

$$k \in X \Rightarrow T \vdash \neg\xi(k)$$

$$k \notin X \Rightarrow \xi(k) \in \mathrm{Cons}(\Pi_1, T).$$

By the Fixed point theorem, let $\delta(x)$ be such that, for all k,

$$T \vdash \delta(k) \leftrightarrow \rho(\delta(k)) \vee \xi(k).$$

Suppose that $k \in X$. Then $T \vdash \delta(k) \leftrightarrow \rho(\delta(k))$. Thus $\delta(k)$ is a Rosser sentence. Next, suppose that $k \notin X$ and, for a contradiction, that $T \vdash \delta(k) \leftrightarrow \rho(\delta(k))$. Then $T + \xi(k) \vdash \rho(\delta(k))$, and since $\xi(k)$ is Π_1-conservative over T, it follows that $T \vdash \rho(\delta(k))$. Thus, by construction, $T \vdash \delta(k)$, contradicting the assumption that $\delta(k)$ is a Rosser sentence for T.

We have shown that

1. if $k \in X$, then $\delta(k)$ is a Rosser sentence,
2. if $k \notin X$, then $\delta(k)$ is not a Rosser sentence,

and since X is an arbitrary r.e. set, the set of Rosser sentences is Σ_1-complete. □

Let us begin our general study of sets of fixed points by focusing on the sets $\mathrm{Fix}(\theta)$, where we lay no restriction on neither the formula $\theta(x)$, nor its set of fixed points. Such a set is r.e.: as the theory T is itself r.e., we can easily enumerate the proofs of $\delta \leftrightarrow \theta(\delta)$ and thus enumerate the sentences δ for which these equivalences are provable. A first result, stated here, mainly as an example of methodology, shows that every such set of fixed points is non-recursive and thus infinite:

Proposition 3.3.2. *For any formula $\theta(x)$, Fix(θ) is not recursive.*

Proof. Suppose, for a contradiction, that $\mathrm{Fix}(\theta)$ is recursive, and that $\xi(x)$ binumerates $\mathrm{Fix}(\theta)$ in T. By the Fixed point theorem, let δ be such that

$$T \vdash \delta \leftrightarrow \big(\big(\theta(\delta) \wedge \neg\xi(\delta)\big) \vee \big(\neg\theta(\delta) \wedge \xi(\delta)\big)\big).$$

Suppose now that $\delta \in \mathrm{Fix}(\theta)$. As we have supposed that $\xi(x)$ binumerates $\mathrm{Fix}(\theta)$, it follows that $T \vdash \xi(\delta)$. Thus $\neg\xi(\delta)$ is refutable in T, so $T \vdash \delta \leftrightarrow \neg\theta(\delta)$. This implies that $\delta \notin \mathrm{Fix}(\theta)$, contradicting our assumption. Thus $\delta \notin \mathrm{Fix}(\theta)$.

Again, as $\xi(x)$ binumerates $\mathrm{Fix}(\theta)$, $T \vdash \neg\xi(\delta)$. $\xi(\delta)$ is refutable, so $T \vdash \delta \leftrightarrow \theta(\delta)$. Thus $\delta \in \mathrm{Fix}(\theta)$, and we reach a contradiction.

Since $\mathrm{Fix}(\theta)$ is not recursive, it cannot be finite. □

We refine this method to show that every unrestricted set of fixed points is indeed Σ_1-complete. By using the variation of the Fixed point theorem due to Ehrenfeucht and Feferman, we can construct a formula reducing every r.e. set to a set of fixed points.

Theorem 3.3.3. *For any formula $\theta(x)$, Fix(θ) is a Σ_1-complete set.*

Proof. Let, by the Fixed point theorem, $\delta(x)$ be such that, for all k,

$$T \vdash \delta(k) \leftrightarrow \big(\big(\theta(\delta(k)) \wedge \xi(k)\big) \vee \big(\neg\theta(\delta(k)) \wedge \neg\xi(k)\big)\big),$$

where $\xi(x)$ is a numeration of an arbitrary r.e. set X. We show that

1. If $k \in X$, then $\delta(k) \in \mathrm{Fix}(\theta)$, and
2. if $k \notin X$, then $\delta(k) \notin \mathrm{Fix}(\theta)$.

Then $\delta(x)$ is a recursive function reducing any r.e. set X to $\mathrm{Fix}(\theta)$, so $\mathrm{Fix}(\theta)$ is Σ_1-complete.

Accordingly, suppose that $k \in X$. Since $\xi(k)$ numerates X, it follows that $\xi(k)$ is provable in T, and we get $T \vdash \delta(k) \leftrightarrow \theta(\delta(k))$. Thus $\delta(k) \in \mathrm{Fix}(\theta)$.

Now, suppose that $k \notin X$ and, for a contradiction, that $k \in \mathrm{Fix}(\theta)$. Since $\xi(k)$ numerates X, it follows that $T \nvdash \xi(k)$, so $T + \neg\xi(k)$ is consistent. By the construction of δ, $T + \neg\xi(k) \vdash \delta(k) \leftrightarrow \neg\theta(\delta(k))$, but by assumption, $T \vdash \delta(k) \leftrightarrow \theta(\delta(k))$. Thus $T + \neg\xi(k)$ is inconsistent, a contradiction. □

The converse is not true, i.e. there are many examples of Σ_1-complete sets of sentences that are not sets of fixed points in T. The following two results are due to Bennet. Here, a set X is *sufficiently closed* if $\psi \in X$ implies $\psi \vee \gamma \in X$, for all sentences γ. E.g., all deductively closed sets are sufficiently closed.

Proposition 3.3.4 (Bennet)**.** *If X is a sufficiently closed r.e. superset of $Th(T)$, then there is no $\theta(x)$ such that $X = Fix^T(\theta)$.*

Proof. Let X be a sufficently closed r.e. supserset of $\mathrm{Th}(T)$. Further suppose that $\psi \in X$ and that $X = \mathrm{Fix}(\theta)$, for some formula $\theta(x)$. By the Fixed point theorem, let δ be such that:

$$\mathrm{PA} \vdash \delta \leftrightarrow \neg\theta(\psi \vee \delta).$$

Since $\psi \in X$ and X is sufficiently closed, it follows that $\psi \vee \delta \in X$. But each sentence in X is an element of $\mathrm{Fix}^T(\theta)$, so $T \vdash (\psi \vee \delta) \leftrightarrow \theta(\psi \vee \delta)$. By construction of δ, it follows that $T \vdash \psi \vee \delta \leftrightarrow \neg\delta$. Then $T \vdash \psi$. □

Proposition 3.3.5 (Bennet)**.** *If X is a deductively closed r.e. subset of $Th(T)$, then there is no $\theta(x)$ such that $X = Fix^T(\theta)$.*

Proof. Let X be a deductively closed r.e. subset of $\mathrm{Th}(T)$. Further suppose that $T \vdash \psi$ and that $X = \mathrm{Fix}(\theta)$, for some formula $\theta(x)$. By the Fixed point theorem, let δ be such that:

$$\mathrm{PA} \vdash \delta \leftrightarrow \theta(\psi \wedge \delta).$$

Since $T \vdash \psi$, it follows that $T \vdash (\psi \wedge \delta) \leftrightarrow \theta(\psi \wedge \delta)$. But each sentence that is a fixed point of $\theta(x)$ in T is an element of X, so $\psi \wedge \delta \in X$. Since X is deductively closed, it follows that $\psi \in X$. □

It follows that if S is any theory other than T, then $\mathrm{Th}(S)$ is not a set of fixed points over T. This concludes our discussion of unbounded sets of fixed points.

Let us now regard bounded sets of fixed points. Any such set is r.e., but not necessarily infinite. We have the following characterisation, which is stated as an exercise (2.28c) in Lindström [43]. The present proof is due to Bennet.

Theorem 3.3.6. *If X is an r.e. subset of Σ_n, then there is a formula $\theta(x) \in \mathrm{B}_n$ such that $X = Fix_{\Sigma_n}(\theta)$. Dually, if X is an r.e. subset of Π_n, there is a formula $\theta(x) \in \mathrm{B}_n$ such that $X = Fix_{\Pi_n}(\theta)$.*

Proof. Let X be a r.e. subset of Σ_n, let $\gamma(x) \in \Sigma_n$ numerate X as stated in Lemma 3.2.8, let $\xi(x, z)$ be a p.r. formula such that $X = \{k : \exists m T \vdash \xi(k, m)\}$, and let $\theta(x)$ be such that, for all $\delta \in \Sigma_n$,

$$T \vdash \theta(\delta) \leftrightarrow \neg\gamma(\delta) \vee \big(\mathrm{Tr}_{\Sigma_n}(\delta) \wedge \exists z\big(\xi(\delta, z) \wedge \forall u{\leq}z \neg \mathrm{Prf}(\delta \leftrightarrow \theta(\delta), u)\big)\big).$$

Note that the complexity of $\theta(x)$ is B_n.

Suppose $\delta \in X$ and $\delta \notin \mathrm{Fix}_{\Sigma_n}(\theta)$. Then $T \vdash \theta(\delta) \leftrightarrow \delta$, a contradiction, so $X \subseteq \mathrm{Fix}_{\Sigma_n}(\theta)$.

Next, suppose $\delta \notin X$ and $\delta \in \mathrm{Fix}_{\Sigma_n}(\theta)$. The second disjunct is then refutable, so $T \vdash \neg\gamma(\delta) \leftrightarrow \theta(\delta)$. But $\delta \in \mathrm{Fix}_{\Sigma_n}(\theta)$, hence $T \vdash \neg\gamma(\delta) \leftrightarrow \delta$. As $\neg\gamma(\delta)$ is Σ_n-conservative over T, $T + \neg\gamma(\delta) \vdash \delta$ implies $T \vdash \delta$. Thus $T \vdash \neg\gamma(\delta)$, contradicting our choice of the numeration $\gamma(x)$.

The dual case follows by changing all Σ_n to Π_n. □

Noting that $\mathrm{Fix}_\Gamma(\theta)$ obviously consists only of Γ-sentences, we get the following characterisation.

Corollary 3.3.7. *X is a r.e. set of Γ-sentences iff there is a $\theta \in \mathrm{B}_n$ such that $X = Fix_\Gamma(\theta)$.*

Having characterised the sets $\mathrm{Fix}_\Gamma(\theta)$, where $\theta(x)$ is a formula in $\Gamma' \supset \Gamma$, we now restrict the set of fixed points to the same quantifier complexity as the formula in question. From this point on, when we write $\mathrm{Fix}_\Gamma(\theta)$ or speak of Γ-fixed points of $\theta(x)$, it is understood that $\theta(x) \in \Gamma$. Any set $\mathrm{Fix}_\Gamma(\theta)$, where $\theta(x) \in \Gamma$, is infinite and r.e.[3] It is evident that few of the proofs used in the previous sections apply directly, as the fixed points defined are always of a higher complexity than $\theta(x)$. E.g., in the proof of Theorem 3.3.2, we construct a sentence δ that can neither be in nor outside of $\mathrm{Fix}(\theta)$. The complexity of δ is B_n if $\theta(x)$ is Σ_n or Π_n. So, if we try to apply the proof to $\mathrm{Fix}_\Gamma(\theta)$ instead, this δ cannot serve as a counterexample, as $\mathrm{Fix}_\Gamma(\theta)$ contains no sentences in $\mathrm{B}_n \setminus \Gamma$. Also note that if we limit ourselves to Δ_n- or B_n-formulas, the ordinary proof of Σ_1-completeness (3.3.3) goes through, since the complexity bound no longer limits the use of negation in the diagonalisation.

[3] This is stated as Exercise 2.28b of [43].

Trying to prove that all sets of Γ-fixed points are Σ_1-complete, one would want to avoid the complexity problem by modifying the proof of Theorem 3.3.3 in the following way. Let

$$T \vdash \delta(k) \leftrightarrow \big(\big(\theta(\delta(k)) \wedge \xi(k)\big) \vee \big(\chi(\delta(k)) \wedge \neg\xi(k)\big)\big),$$

where $\xi(x)$ is a numeration of any r.e. set X. Then we would only need to construct a $\chi(x) \in \Gamma$ that is equivalent to $\neg\theta(x)$ on X^c. Such a $\chi(x)$ can, however, only be found in some cases, which we will discuss in the next section.

Now, there is a simple special case, namely when the formula in question is extensional. A formula $\theta(x)$ is *extensional* if it is such that $T \vdash \theta(\varphi) \leftrightarrow \theta(\psi)$, whenever $T \vdash \varphi \leftrightarrow \psi$. We prove that every set of Γ-fixed points of an extensional formula is effectively inseparable from its complement relative to Γ, except in the case when this complement is empty. Thus it follows that any such set is Σ_1-complete. We will use the following lemma to prove the proposition below:

Lemma 3.3.8 (Putnam & Smullyan [50])**.** *If X_0, X_1 are disjoint r.e. sets, then there is a Σ_1-formula $\xi(x)$ such that $\xi(x)$ numerates X_0 in T and $\neg\xi(x)$ numerates X_1 in T. By symmetry, there is also a Π_1-formula with these properties.*

Proposition 3.3.9. *If $\theta(x) \in \Gamma$, and X is an r.e. subset of Γ containing an equivalence class, and disjoint from $Fix_\Gamma(\theta)$, then $Fix(\theta)$ and X are effectively inseparable.*

Proof. Suppose that X is an r.e. subset of Γ, disjoint from $\mathrm{Fix}_\Gamma(\theta)$, and containing an equivalence class $[\psi]$. Suppose also that $\Gamma \neq \Pi_1$. By Lemma 3.3.8, let $\xi_0(x)$ be a Π_1-formula and $\xi_1(x)$ a Σ_1-formula such that, for $i = 0, 1$, $\xi_i(x)$ numerates $\mathrm{Fix}_\Gamma(\theta)$ and $\neg\xi_i(x)$ numerates X. By the Fixed point theorem, let δ be such that:

$$T \vdash \delta \leftrightarrow \big(\big(\theta(\delta) \wedge \neg\xi_0(\delta)\big) \vee \big(\psi \wedge \xi_1(\delta)\big)\big).$$

Note that such a δ can be effectively found, and that δ is a Γ-sentence.

Suppose, for a contradiction, that $\delta \in \mathrm{Fix}_\Gamma(\theta)$. Then $T \vdash \xi_i(\delta)$, so $T \vdash \delta \leftrightarrow \psi$. But $\mathrm{Fix}_\Gamma(\theta)$ is disjoint from $[\psi]$, and we have a contradiction. Analogously, suppose that $\delta \in X$. Then $T \vdash \neg\xi_i(\delta)$, so $T \vdash \delta \leftrightarrow \theta(\delta)$. But by supposition X is disjoint from $\mathrm{Fix}_\Gamma(\theta)$, a contradiction.

If $\Gamma = \Pi_1$, we have to change the complexity of the numerations $\xi_i(x)$ to ensure that δ is a Π_1-sentence. In this case, we choose $\xi_0(x)$ to be a Σ_1-formula, and $\xi_1(x)$ to be a Π_1-formula, and use the same construction as above. □

Note that the proof of this proposition does not depend on extensionality of $\theta(x)$. However, the proposition may be used to prove the following theorem for extensional formulas.

Theorem 3.3.10. *If $\theta(x)$ is an extensional Γ-formula such that $Fix_\Gamma(\theta) \neq \Gamma$ then $Fix_\Gamma(\theta)$ is Σ_1-complete.*

The proof is simple and follows directly from Proposition 3.3.9: Suppose that $\theta(x)$ is an extensional formula in Γ. If $\Gamma \setminus \mathrm{Fix}_\Gamma(\theta)$ is non-empty, then $\Gamma \setminus \mathrm{Fix}_\Gamma(\theta)$ contains a whole equivalence class, since $\theta(x)$ is extensional. By the proposition, it follows that $\mathrm{Fix}_\Gamma(\theta)$ is Σ_1-complete. As in the case of unbounded sets of fixed points, the converse is not true. If we restrict Proposition 3.3.5 to any deductively closed r.e. subset of $\mathrm{Th}(T) \cap \Gamma$, we get an example of a Σ_1-complete set of Γ-sentences that is not a set of Γ-fixed points.

Furthermore, if $\mathrm{Fix}_\Gamma(\theta) = \Gamma$, then $T \vdash \delta \leftrightarrow \theta(\delta)$ for all $\delta \in \Gamma$, so it follows that $T \vdash \theta(\delta) \leftrightarrow \mathrm{Tr}_\Gamma(\delta)$ for all $\delta \in \Gamma$. So, by this theorem, there is, modulo provable equivalence, only one extensional Γ-formula with a recursive set of fixed points.

3.3.1 Non-extensional formulas

In contrast to extensional formulas, a *non-extensional* formula may distinguish between distinct elements of an equivalence class, and the set of fixed points of such a formula may contain parts of equivalence classes. For example, the set of fixed points of the formula $\varphi \wedge x = \varphi$ (where φ is a non-refutable formula) is the set of sentences refutable in T in addition to φ. Thus the proof of Theorem 3.3.10 can not be modified to yield a similar result for non-extensional formulas. However, a partial result is available. It can be seen as a variation of either of Proposition 3.3.9, Bernardi's Theorem 1 [3] or the results in Chapter III of Smullyan [64].

Theorem 3.3.11. *If $\theta(x)$ and $\chi(x)$ are Γ-formulas, and X is an r.e. subset of Γ containing $Fix_\Gamma(\chi)$, and disjoint from $Fix_\Gamma(\theta)$, then $Fix_\Gamma(\theta)$ and X are effectively inseparable.*

Proof. The proof is similar to that of Proposition 3.3.9. Suppose that $\theta(x)$ and $\chi(x)$ are Γ-formulas, and that X is an r.e. subset of Γ, disjoint from $\mathrm{Fix}_\Gamma(\theta)$, and containing $\mathrm{Fix}_\Gamma(\chi)$. Suppose also that $\Gamma \neq \Pi_1$. By Lemma 3.3.8, let $\xi_0(x)$ be a Π_1-formula and $\xi_1(x)$ a Σ_1-formula such that, for $i = 0, 1$, $\xi_i(x)$ numerates $\mathrm{Fix}_\Gamma(\theta)$ and $\neg\xi_i(x)$ numerates X. By the Fixed point theorem, let δ be such that:

$$T \vdash \delta \leftrightarrow \big(\big(\theta(\delta) \wedge \neg\xi_0(\delta)\big) \vee \big(\chi(\delta) \wedge \xi_1(\delta)\big)\big).$$

Such a δ can be effectively found.

Suppose that $\delta \in \mathrm{Fix}_\Gamma(\theta)$. Then $T \vdash \xi_i(\delta)$, so $T \vdash \delta \leftrightarrow \chi(\delta)$. But $\mathrm{Fix}_\Gamma(\theta)$ is disjoint from $\mathrm{Fix}_\Gamma(\chi)$, and we have a contradiction. Analogously, suppose that $\delta \in X$. Then $T \vdash \neg\xi_i(\delta)$, so $T \vdash \delta \leftrightarrow \theta(\delta)$. But, by supposition, X is disjoint from $\mathrm{Fix}_\Gamma(\theta)$, again a contradiction.

If $\Gamma = \Pi_1$, we let $\xi_0(x)$ be a Σ_1-formula, $\xi_1(x)$ a Π_1-formula, and use the same construction. □

Corollary 3.3.12. *Given $\theta(x) \in \Gamma$, there is no $\chi(x) \in \Gamma$ such that $\Gamma \setminus Fix_\Gamma(\theta) = Fix_\Gamma(\chi)$*

Proof. Since every set of fixed points is r.e., $\Gamma \setminus \mathrm{Fix}_\Gamma(\theta) = \mathrm{Fix}_\Gamma(\chi)$ implies that both sets are recursive. But by the theorem it follows that both sets are Σ_1-complete. □

Moreover, every equivalence class $[\psi]$ is the set of fixed points of the non-extensional formula $\psi \wedge (x = x)$. Thus we can freely substitute any equivalence class $[\psi]$ for $\mathrm{Fix}_\Gamma(\chi)$ in the proof above. The above theorem is of course also equivalent to the statement that every set of Γ-fixed points that is not Σ_1-complete intersects all sets of Γ-fixed points (and all equivalence classes).

We also note that the previous results, in a way are the best possible one could hope for, applying this particular method to the problem at hand. In order to establish inseparability, we need an r.e. set that we can guarantee is disjoint from the set of fixed points. And it does not suffice to put a numeration $\xi(x)$ of an arbitrary r.e. set in place of ψ in the proof above, for nothing guarantees that $\delta \in \mathrm{Fix}_\Gamma(\theta)$ implies that $T \nvdash \delta \leftrightarrow \xi(\delta)$, even if the two sets are disjoint.[4] In fact, the set in question must be "generated" by some condition of provable equivalence, e.g. being a fixed point to another formula, or being in some particular equivalence class. With this in mind, it seems the solution must be sought elsewhere.

As we have established Σ_1-completeness of only *some* set of Γ-fixed points, there is reason to ask whether there are counterexamples or if our methods are not sophisticated enough to prove a full result. As it turns out, there are indeed recursive sets of fixed points other than Γ, though only to non-extensional formulas.

Proposition 3.3.13. *If $X = Fix_\Gamma(\theta)$ for some $\theta(x) \in \Gamma$, and Y is a recursive set such that $Fix_\Gamma(\tau) \cap Y = \emptyset$ for some $\tau(x) \in \Gamma$, then we can construct a formula $\chi(x) \in \Gamma$ such that $Fix_\Gamma(\chi) = X \setminus Y$.*

[4] Unless, of course, we suppose that the set numerated by $\xi(x)$ contains e.g. no provable sentences, in which case the conditions of Theorem 3.3.11 is fulfilled, and no strength is gained.

Proof. Let $X = \mathrm{Fix}_\Gamma(\theta)$, and let Y be a recursive set such that $\mathrm{Fix}_\Gamma(\tau) \cap Y = \emptyset$ for some $\tau(x) \in \Gamma$. Further, let $\eta(x)$ binumerate Y and let

$$\chi(x) := (\theta(x) \wedge \neg\eta(x)) \vee (\tau(x) \wedge \eta(x)).$$

Suppose that $\delta \in Y$. Then $T \vdash \chi(\delta) \leftrightarrow \tau(\delta)$. Suppose further that $\delta \in \mathrm{Fix}_\Gamma(\chi)$. Then $T \vdash \delta \leftrightarrow \tau(\delta)$, but Y is disjoint from $\mathrm{Fix}_\Gamma(\tau)$, a contradiction. Now suppose $\delta \notin Y$. Then $T \vdash \chi(\delta) \leftrightarrow \theta(\delta)$ so $\delta \in \mathrm{Fix}_\Gamma(\chi)$ iff $\delta \in \mathrm{Fix}_\Gamma(\theta)$. Thus $\mathrm{Fix}_\Gamma(\chi) = \mathrm{Fix}_\Gamma(\theta) \setminus Y$. □

Should X be a recursive set of fixed points, and Y a recursive subset of X satisfying the conditions of the proposition, then we can construct a new recursive set of fixed points, by removing Y from X. By successively applying this method, starting from the set Γ, we can construct infinitely many recursive sets of fixed points. We only need to make sure that Y is disjoint from some set of fixed points, but by choosing Y to be e.g. a recursive subset of an equivalence class, we guarantee that this set has the needed properties. Should we be given a set of fixed points, we can add any recursive set to it by a similar construction:

Proposition 3.3.14. *If $X = Fix_\Gamma(\theta)$ for some $\theta(x) \in \Gamma$, and Y is a recursive subset of Γ, then we can construct a formula χ such that $Fix_\Gamma(\chi) = X \cup Y$.*

Proof. Let $X = \mathrm{Fix}_\Gamma(\theta)$, and let Y be a recursive set, binumerated by $\eta(x)$. Let

$$\chi(x) := (\theta(x) \wedge \neg\eta(x)) \vee (\mathrm{Tr}_\Gamma(x) \wedge \eta(x)).$$

Suppose that $\delta \in Y$, then $T \vdash \eta(\delta)$, so $T \vdash \delta \leftrightarrow \mathrm{Tr}_\Gamma(\delta) \leftrightarrow \chi(\delta)$, so $\delta \in \mathrm{Fix}_\Gamma(\chi)$. If $\delta \notin Y$, then $T \vdash \neg\eta(\delta)$, so $T \vdash \chi(\delta) \leftrightarrow \theta(\delta)$ and $\delta \in \mathrm{Fix}_\Gamma(\chi)$ iff $\delta \in \mathrm{Fix}_\Gamma(\theta)$. □

In a sense, every set constructed from Γ by these means is a trivial example. We would like to find necessary and sufficient conditions for a recursive set to be a set of fixed points. The two following propositions are the weakest sufficient conditions we have. Note that they differ somewhat in flavour.

Proposition 3.3.15. *If X is a recursive subset of Γ and there is a formula $\theta(x) \in \Gamma$ such that $Fix_\Gamma(\theta) \subseteq X$, then we can construct a formula $\chi(x) \in \Gamma$ such that $Fix_\Gamma(\chi) = X$.*

Proof. Let X be a recursive subset of Γ such that $\mathrm{Fix}_\Gamma(\theta) \subseteq X$, for some $\theta(x) \in \Gamma$, and let X be binumerated by $\xi(x)$. Let

$$\chi(x) := (\mathrm{Tr}_\Gamma(x) \wedge \xi(x)) \vee (\theta(x) \wedge \neg\xi(x)).$$

If $\delta \in X$, then $T \vdash \xi(k)$, so $T \vdash \delta \leftrightarrow \mathrm{Tr}_\Gamma(\delta) \leftrightarrow \chi(\delta)$. If $\delta \notin X$, then $T \vdash \neg\xi(\delta)$, so if $\delta \in \mathrm{Fix}_\Gamma(\chi)$ it follows that $T \vdash \delta \leftrightarrow \theta(\delta)$, contradicting our assumption on X. □

Proposition 3.3.16. *If X is a recursive subset of Γ and there is a sentence $\psi \in \Gamma$ such that $[\psi] \cap X^c$ is recursive, then X is a set of fixed points.*

Proof. Let X be as in the statement of the proposition. Let X' be $X \cup ([\psi] \cap X^c)$. By Proposition 3.3.15, X' is a recursive set of fixed points, since it is a recursive set that contains an equivalence class. But $[\psi] \cap X^c$ is a recursive set disjoint from some equivalence class (i.e. every other equivalence class, since it is a subset of $[\psi]$), so by Proposition 3.3.13, $X' \setminus ([\psi] \cap X^c)$ is a recursive set of fixed points. □

This last result differs in an important respect from the other results of this section. In these, we may freely interchange "set of fixed points" and "equivalence class", and use virtually the same proof to prove a different result. In this case, however, it seems we need the fact that equivalence classes partition Γ. This is to make sure that the recursive part of X^c is disjoint from some equivalence class, so that this set may actually be removed. We could modify the premises of the proposition to obtain the following result:

Proposition 3.3.17. *If X is a recursive subset of Γ and there is a formula $\theta(x) \in \Gamma$ such that $Fix_\Gamma(\theta) \cap X^c$ is recursive, and this intersection is disjoint from some set of fixed points, then X is a set of fixed points.*

Here we may use the same proof as above, since the set $\mathrm{Fix}_\Gamma(\theta) \cap X^c$ is disjoint from some set of fixed points, and may thus be removed from the set X'.

Let us now tie these observations to our earlier results. When we try to settle the question of complexity of a set of fixed points, we readily run into the question of how such a set intersects other sets of fixed points (or equivalence classes). For suppose that a set of fixed points has a recursive intersection with some other set of fixed points (and that this intersection is indeed disjoint from some set of fixed points). Then we can use Proposition 3.3.13 to remove this intersection, and obtain a new set of fixed points which is now disjoint from

a set of fixed points and thus Σ_1-complete.[5] But by the Jockusch-Mohrherr theorem, the union of two disjoint r.e. sets of which one is creative is itself a creative set. Thus, our original set of fixed points is also Σ_1-complete. This argument improves Theorem 3.3.11.

We can use a similar argument to show that the intersection between a recursive set of fixed points and an equivalence class is non-recursive. Theorem 3.3.11 shows that a recursive set of fixed points intersects every equivalence class. But suppose that the intersection with $[\psi]$ is recursive. Then we can use Proposition 3.3.13 to remove this corresponding part of $[\psi]$, obtaining a set of fixed points that is disjoint from $[\psi]$. But since the intersection was recursive, the new set of fixed points is both recursive and Σ_1-complete, a contradiction.

These two arguments are not completely dual, as in the first one we establish the complexity of a given set, and in the second we establish some property of a set with a given complexity. It should, however, be clear how intersections of equivalence classes enter the discussion. Also note that we have no non-trivial examples, neither of Σ_1-complete, nor recursive sets of Γ-fixed points.

Consider the question for which r.e. sets $X \subseteq \Gamma$ we can construct a formula with exactly the elements of this set as fixed points. It is clear that for each such set, we can find a formula, i.e. $\mathrm{Tr}_\Gamma(x)$, whose set of fixed point contains X. What remains is to find means to make sure that nothing outside of X can be a fixed point of the formula we are trying to construct.

In [4], Bernardi briefly mentions the set of all fixed points of the formula $\theta(x) := \neg \Pr(x \neq c)$, where c is the Gödel number of $\neg \Pr(0 \neq 0)$. He notes that the set of fixed points of this formula equals the set of all refutable formulas together with c. In our present setting, where the formula $\theta(x)$ and its set of fixed points is restricted to Γ, we can acquire a stronger result from a simpler construction. Let $\theta(x)$ be the formula $\mathrm{Tr}_\Gamma(x) \wedge \xi(x)$, where $\xi(x)$ binumerates any recursive set X of Γ-sentences. Then $\mathrm{Fix}_\Gamma(x) = [0] \cup X$. Of course, this can also be seen as an application of the even more general Proposition 3.3.14.

We now present a similar criterion, applicable not only to recursive sets, but also to r.e. sets satsifying another condition.

Definition 3.3.18. *A set X of sentences has a* lower bound *if there is a non-refutable sentence φ such that $T \vdash \varphi \to \psi$, for all $\psi \in X$.*

Proposition 3.3.19. *Given an r.e. set $X \subseteq \Gamma$ such that X^c has a lower bound, there is a Γ-formula $\theta(x)$ such that $X = Fix_\Gamma(\theta)$.*

[5] Actually, the set is creative.

To prove the proposition, we use another definition and a lemma. Note also, that for an r.e. set to have a lower bound, it has to be disjoint from [0].

Definition 3.3.20. *A set X of sentences is* monoconsistent with T *if $T + \varphi$ is consistent for every $\varphi \in X$.*

Lemma 3.3.21 (Lindström [38])**.** *Suppose X and Y are r.e., and Y is monoconsistent with Q. Then there is a Σ_1- (and a Π_1-) formula $\xi(x)$ such that, for every k, if $k \in X$, then $Q \vdash \xi(k)$, and if $k \notin X$, then $\xi(k) \notin Y$.*

Proof of Proposition 3.3.19. Let X be any r.e. subset of Γ such that X^c has a lower bound φ. Since $T \nvdash \neg\varphi$, it follows that $T + \varphi$ is consistent. Let, by Lemma 3.3.21, $\xi(x)$ be a Σ_1-formula such that if $k \in X$, then $T \vdash \xi(k)$, and if $k \notin X$, then $\xi(k) \notin \mathrm{Th}(T + \varphi)$. Further, let $\theta(x) := \mathrm{Tr}_\Gamma(x) \wedge \xi(x)$.

Suppose $\psi \in X$. Then $T \vdash \xi(\psi)$, and $T + \psi \vdash \mathrm{Tr}_\Gamma(\psi) \wedge \xi(\psi)$, so $T \vdash \psi \to \theta(\psi)$. Moreover, $T + \theta(\psi) \vdash \mathrm{Tr}_\Gamma(\psi)$, so $T \vdash \theta(\psi) \to \psi$. Thus every $\psi \in X$ is a fixed point of $\theta(x)$.

Now, suppose that $\psi \notin X$ and, for a contradiction, that $\psi \in \mathrm{Fix}_\Gamma(\theta)$. Then $T \nvdash \xi(\psi)$, so $T + \neg\xi(\psi)$ is consistent and proves $\neg\theta(\psi)$. ψ is a fixed point of $\theta(x)$, whence $T \vdash \psi \to \xi(\psi)$. Since φ is a lower bound of X^c, we have $T \vdash \varphi \to \psi$, so it follows that $T + \varphi \vdash \xi(\psi)$. But by our choice of φ and $\xi(x)$, this is a contradiction. Thus we have shown that no $\psi \notin X$ is a fixed point of $\theta(x)$. □

Let X be any r.e. subset of Γ. For X^c to have a lower bound, it is clear that X must contain [0], for if X^c contains any refutable sentence, it can evidently have no lower bound. Bennet [1] observes that an r.e. set that is disjoint from [0] has a lower bound, so for any recursive $X \subseteq \Gamma$, X^c has a lower bound iff $[0] \subseteq X$. This means that for recursive sets, Proposition 3.3.19 yields nothing beyond Proposition 3.3.15. Finally, we note that the condition stated in the proposition is not necessary for an r.e. set to be a set of fixed points. If we pick an undecidable sentence ψ, it is clear that $[\psi]$ is a set of fixed points, but its complement has no lower bound.

For a brief summary of our results on the recursion theoretic complexity of sets of fixed points, we have the following facts.

1. For any formula $\theta(x)$, the set $\mathrm{Fix}(\theta)$ is Σ_1-complete. It is not the case that every Σ_1-complete set of sentences is a set of fixed points.

2. For every r.e. subset X of Σ_n (or Π_n), there is a formula $\theta(x) \in \mathrm{B}_n$ such that $\mathrm{Fix}_\Gamma(\theta) = X$. This is a complete characterisation of such sets in terms of sets of fixed points.

3. If two sets of Γ-fixed points of Γ-formulas are disjoint, then these sets are effectively inseparable, and are thus Σ_1-complete. This includes the sets of Γ-fixed points of extensional formulas, except for $\mathrm{Tr}_\Gamma(x)$.

4. There are recursive sets of Γ-fixed points, though only for non-extensional Γ-formulas. Any such set intersects every other set of Γ-fixed points non-recursively.

3.3.2 Δ_0-formulas

Here, we only briefly cover the special case where we restrict the complexity to Δ_0-sentences, as an example of a decidable fragment of a theory. For the rest of this discussion, A, B will be subsets of $[1]_{\Delta_0} := \{k : T \vdash k\} \cap \Delta_0$ and $[0]_{\Delta_0} := \{k : T \vdash \neg k\} \cap \Delta_0$, respectively. As every Δ_0-formula is decidable in Q, the set of provable formulas and the set of refutable formulas together make up the whole of Δ_0. By inspection, we see that every extensional Δ_0-formula is provably equivalent to either of $x = x$ or $x \neq x$. Note that the only sets having binumerations in Δ_0 are the p.r. sets.

Proposition 3.3.22. *If $A \subseteq [1]_{\Delta_0}$ and $B \subseteq [0]_{\Delta_0}$, then the following four statements are equivalent:*

1. $Fix_{\Delta_0}(\theta) = A \cup B$
2. $\theta(x)$ *binumerates* $A \cup ([0]_{\Delta_0} \setminus B)$
3. $\neg\theta(x)$ *binumerates* $([1]_{\Delta_0} \setminus A) \cup B)$
4. $Fix_{\Delta_0}(\neg\theta) = ([1]_{\Delta_0} \setminus A) \cup ([0]_{\Delta_0} \setminus B)$

The proof is routine, and is left to the interested reader. A simple argument also shows that no set of Δ_0-fixed points can be p.r. For suppose, for a contradiction, that $\mathrm{Fix}_{\Delta_0}(\theta)$ is p.r. Then this set is binumerated by a p.r. formula $\xi(x)$. By the Fixed point theorem, let δ be such that

$$T \vdash \delta \leftrightarrow \big(\big(\theta(\delta) \wedge \neg\xi(\delta)\big) \vee \big(\neg\theta(\delta) \wedge \xi(\delta)\big)\big).$$

It follows that $\delta \in \mathrm{Fix}_{\Delta_0}(\theta)$ iff $\delta \notin \mathrm{Fix}_{\Delta_0}(\theta)$, a contradiction.

Proposition 3.3.23. *If $Fix_{\Delta_0}(\theta) = A \cup B$, then A is p.r. iff $[0]_{\Delta_0} \setminus B$ is p.r.*

Proof. Let $\mathrm{Fix}_{\Delta_0}(\theta) = A \cup B$. Then $\theta(x)$ binumerates $A \cup ([0]_{\Delta_0} \setminus B)$. Suppose $\xi(x)$ binumerates A $([0]_{\Delta_0} \setminus B)$, and let $\eta(x) := \theta(x) \wedge \neg\xi(x)$. Then $\eta(x)$ binumerates $[0]_{\Delta_0} \setminus B$ (A). □

We have not outruled the possibility that $A \cup ([0]_{\Delta_0} \setminus B)$ could be p.r., but neither of A nor $[0]_{\Delta_0} \setminus B$. However, should none of these sets be p.r., they are in a sense inseparable by p.r. sets.

Proposition 3.3.24. *If $Fix_{\Delta_0}(\theta) = A \cup B$ and neither of A nor $[0]_{\Delta_0} \setminus B$ is p.r., then there is no p.r. set A' such that $A \subseteq A' \subseteq [1]_{\Delta_0}$.*

Proof. Suppose $\mathrm{Fix}_{\Delta_0}(\theta) = A \cup B$, that neither of A nor $[0]_{\Delta_0} \setminus B$ is p.r., and that there is a p.r. set A' (B') such that $A \subseteq A' \subseteq [1]_{\Delta_0}$ $([0]_{\Delta_0} \setminus B \subseteq B' \subseteq [0]_{\Delta_0})$. Let $\xi(x)$ binumerate A' (B'), and let $\eta(x) := \theta(x) \wedge \neg\xi(x)$. Then $\eta(x)$ binumerates $[0]_{\Delta_0} \setminus B$ (A). □

3.4 Structural properties

Having studied the properties of individual sets of fixed points, we here turn to the study of collections of sets of fixed points, ordered to give rise to interesting structures. There are examples of such structures obtained from ordering equivalence classes of T under implication, e.g. (partial) Lindenbaums algebras and Magari algebras. As we are concerned with formulas with a free variable, it is not evident how an ordering under implication should be defined, so we have not pursued this course. Instead, we choose to order sets of fixed points under set inclusion.

We will briefly introduce some concepts from the field of lattice theory. For more details, see e.g. Davey & Priestley [8]. A *partially ordered set* is a set P together with a binary relation $\leq$ such that, for all $a, b, c \in P$:

1. $a \leq a$,
2. if $a \leq b$ and $b \leq a$, then $a = b$,
3. if $a \leq b$ and $b \leq c$, then $a \leq c$.

If P is an ordered set and $S \subseteq P$, then an element $a \in P$ is the *supremum* of S if a is the least element such that $s \leq a$ for all $s \in S$. We define the *infimum* dually.

P is an *upper semi-lattice* if, for each pair of elements $a, b \in P$, the set $\{a, b\}$ has a supremum in P. We will call such an element the *join* of a and b, and

denote this element by $a \oplus b$. Dually, P is a *lower semi-lattice* if, for each pair of elements $a, b \in P$, the set $\{a, b\}$ has an infimum, the *meet* of a and b ($a \otimes b$). A *lattice* is an ordered set that is both a lower and an upper semi-lattice.

A (semi-) lattice P has a greatest (top) element if there is an element $\top \in P$ such that $b \leq \top$ for all $b \in P$. A least (bottom, $\bot$) element is defined dually. A lattice having both a greatest and a least element is *bounded*. In a bounded lattice, we say that a is the complement of b ($a = b^{-1}$) if $a \otimes b = \bot$ and $a \oplus b = \top$. Note that we will consistently use $\otimes$, $\oplus$, $^{-1}$ as algebraic operations, and $\cap$, $\cup$ and ^{c} as set-theoretical operations.

Here we are interested in sets ordered under set inclusion, in which case we have the following situation. Let a, b be two elements of a set ordered under set inclusion. If $a \cup b$ is an element of the ordered set, then $a \cup b$ is indeed the join of a and b: It is evident that $a \cup b$ is greater than both a and b, so suppose that it is not the supremum. Then there is a $d \subset a \cup b$ such that $a \subseteq d$ and $b \subseteq d$. But since $a \cup b$ is the least set containing each element of both a and b, there can be no such d. Thus $a \cup b = a \oplus b$. A similar argument shows that if $a \cap b$ is an element of the ordered set, then $a \cap b = a \otimes b$. This means that in many cases, we can use the closureness under unions and intersections to show that ordered sets are (semi-) lattices.

Let $\mathscr{F}$ be the set of all sets of fixed points, ordered under set inclusion. We know little about how the elements of this sets are related to each other, e.g. whether the union of two arbitrary sets of fixed points is itself a set of fixed points or not. If we restrict ourselves to the set of all sets of Γ-fixed points, we are in a similar situation, and we fail to state interesting properties of these structures. Instead, we will consider subsets of the set of all sets of Γ-fixed points, based on our different sufficient conditions for r.e. sets to be sets of fixed points. As a first example, we define $\mathscr{F}_b$ to be the structure

$$\big(\{X : X \text{ is an r.e. subset of } \Gamma \text{ and } X^c \text{ has a lower bound}\}, \subseteq\big).$$

By Proposition 3.3.19, each set in $\mathscr{F}_b$ is the set of Γ-fixed points of some Γ-formula.

Proposition 3.4.1. *$\mathscr{F}_b$ is a distributive lattice with a greatest element.*

Proof. Let $a, b \in \mathscr{F}_b$, and let φ, ψ be some lower bounds of a^c and b^c, respectively. To show that $a \oplus b$ exists, it suffices to show that $a \cup b \in \mathscr{F}_b$, by the discussion above. It is clear that $a \cup b$ is an r.e. subset of Γ, so it remains to show that $(a \cup b)^c$ has a lower bound. But $(a \cup b)^c = a^c \cap b^c$, so for every element δ in this set, $T \vdash \varphi \vee \psi \to \delta$. Thus $\varphi \vee \psi$ is a lower bound of $(a \cup b)^c$, and it follows

that $a \cup b$ is an element of $\mathscr{F}_b$. A dual argument shows that $\varphi \wedge \psi$ is a lower bound of $(a \cap b)^c$, so $a \otimes b$ exists. Since join and meet is union and intersection, respectively, it follows that join and meet distribute over each other. Moreover, the set Γ is the greatest element as it is an r.e. subset of Γ, and every sentence is a lower bound for $X^c = \emptyset$. □

It is easy to see that $\mathscr{F}_b$ can have no least element. For suppose a is such an element. Then we can use Proposition 3.3.13 to remove any singleton set from a, obtaining an element that is in $\mathscr{F}_b$, but less than a. This argument also shows that the structure is not dense.

We now briefly discuss the structural properties of the set of recursive sets of fixed points. Here we use the sufficient conditions stated in propositions 3.3.15 and 3.3.16 to define our structures. As an introduction we present a somewhat neater example. Given $\theta(x)$, let $\mathscr{F}_\theta$ be the structure

$$\big(\{X : X \text{ is recursive and } \mathrm{Fix}_\Gamma(\theta) \subseteq X \subseteq \Gamma\}, \subseteq\big)$$

Proposition 3.4.2. *If $Fix_\Gamma(\theta)$ is non-recursive, $\mathscr{F}_\theta$ is a distributive lattice with greatest, but no least, element.*

Proof. Let $a, b \in \mathscr{F}_\theta$. To prove that $a \oplus b$ exists, it suffices to show that $a \cup b \in \mathscr{F}_\theta$. But $a \cup b$ is a recursive subset of Γ, containing $\mathrm{Fix}_\Gamma(\theta)$. Thus join is given by union. Similarily, since the intersection of a and b is a recursive set containing $\mathrm{Fix}_\Gamma(\theta)$, meet is given by intersection. This also shows that the structure is distributive.

The top element is Γ. Suppose that a is the bottom element of $\mathscr{F}_\theta$. Since $\mathrm{Fix}_\Gamma(\theta)$ is non-recursive, $a \setminus \mathrm{Fix}_\Gamma(\theta)$ is non-empty. Then we can remove a singleton set from $a \setminus \mathrm{Fix}_\Gamma(\theta)$, obtaining an element that is less than a, but still in the structure. Thus $\mathscr{F}_\theta$ can have no least element. □

In the case where $\mathrm{Fix}_\Gamma(\theta)$ is itself recursive, we obtain a countable Boolean algebra. By the same argument as before, join and meet are given by union and intersection, respectively. Γ is the top element, and $\mathrm{Fix}_\Gamma(\theta)$ is the bottom element. Finally, every element a has a complement a^{-1} which is given by $\Gamma \setminus (a \cup \mathrm{Fix}_\Gamma(\theta))$.

We now turn our attention to the conditions stated in propositions 3.3.15 and 3.3.16. Let, accordingly, $\mathscr{F}_\exists$ be the structure

$$\big(\{X : X \subseteq \Gamma \text{ is recursive, and there is a } \theta \in \Gamma \text{ s. t. } \mathrm{Fix}_\Gamma(\theta) \subseteq X\}, \subseteq\big),$$

and let $\mathscr{F}_R$ be the structure

$$\big(\{X : X \subseteq \Gamma \text{ is rec., and there is a } \psi \in \Gamma \text{ s. t. } [\psi] \cap X^c \text{ is rec.}\}, \subseteq\big).$$

Proposition 3.4.3. *$\mathscr{F}_\exists$ and $\mathscr{F}_R$ are upper semi-lattices with a greatest element.*

Proof. Let $a, b \in \mathscr{F}_\exists$. To prove that $a \oplus b$ exists, it suffices to show that $a \cup b \in \mathscr{F}_\exists$. But $a \cup b$ is a recursive subset of Γ, and since both sets contains a set of fixed points, it is clear that their union contains (at least) one set of fixed points. Thus join is given by union, and as before, Γ is the greatest element.

Suppose now that $a, b \in \mathscr{F}_R$. We want to show that $a \cup b \in \mathscr{F}_R$. This set is recursive, so it remains to show that there is a sentence ψ such that $(a \cup b)^c \cap [\psi]$ is recursive. $(a \cup b)^c \cap [\psi] = (a^c \cap b^c) \cap [\psi]$, and since intersection is commutative and associative, this set equals $(a^c \cap [\psi]) \cap b^c$. But by assumption, there is a sentence φ such that $a^c \cap [\varphi]$ is recursive. Since $b \in \mathscr{F}_R$, b^c is recursive, so $(a^c \cap [\varphi]) \cap b^c$ is a recursive set. Thus we have found a sentence ψ such that $(a \cup b)^c \cap [\psi]$ is recursive, and $a \cup b \in \mathscr{F}_R$. □

It is not clear if $a \otimes b$ exists for every $a, b \in \mathscr{F}_\exists$. Should a and b contain the very same set of fixed points, then their meet is the intersection of the two sets. A similar argument applies for $a, b \in \mathscr{F}_R$, for if both $a^c \cap [\psi]$ and $b^c \cap [\psi]$ are recursive, then the set $(a \cap b)^c \cap [\psi] = (a^c \cup b^c) \cap [\psi] = (a^c \cap [\psi]) \cup (b^c \cap [\psi])$ is recursive, and $a \cap b \in \mathscr{F}_R$.

By Proposition 3.3.14, any recursive subset of Γ in addition to an equivalence class is a set of Γ-fixed points of some Γ-formula. We briefly investigate the structure of these sets, so for each $\psi \in \Gamma$, let $\mathscr{F}_{[\psi]}$ be the structure

$$\big(\{X : \text{there is a recursive set } Y \subseteq \Gamma \text{ s.t. } Y \cup [\psi] = X\}, \subseteq\big).$$

The mapping from recursive subsets of Γ to elements of $\mathscr{F}_{[\psi]}$ is not injective, as for a given set X, there might be different recursive sets Y and Y' such that $X = Y \cup [\psi] = Y' \cup [\psi]$. Thus, for every element $a \in \mathscr{F}_{[\psi]}$ we let A be a representative of the equivalence class $\{A : A \cup [\psi] = a\}$.

Proposition 3.4.4. *The structure $\mathscr{F}_{[\psi]}$ is an atomic, bounded, distributive lattice.*

Proof. Suppose $a, b \in \mathscr{F}_{[\psi]}$. As before, it suffices to show that $a \cup b$ and $a \cap b$ are elements of $\mathscr{F}_{[\psi]}$. Let A and B be representatives of the equivalence classes corresponding to a and b, as above. It is clear that $a \cup b = A \cup B \cup [\psi]$, so $a \oplus b = a \cup b$. Similarily $a \otimes b = a \cap b$. This also shows that $\mathscr{F}_{[\psi]}$ is distributive.

Γ is the top element, and $[\psi]$ is the bottom element. Each element a of $\mathscr{F}_{[\psi]}$ is covered by an element constructed by adding any singleton set to a. □

Every recursive element of $\mathscr{F}_{[\psi]}$ has a complement. For let a be such an element, and let A' be the set $\Gamma \setminus (A \cup [\psi])$, where A is a representative of the equivalence class corresponding to a. It is clear that different choices of A does not affect the set A'. Now, let a^{-1} be the set $A' \cup [\psi]$. Since A' is recursive, this set is an element of $\mathscr{F}_{[\psi]}$. It follows that $a \oplus a^{-1} = \Gamma$ and $a \otimes a^{-1} = [\psi]$.

3.4.1 Finite differences

In each of the aforementioned structures, we can use the techniques of propositions 3.3.13 and 3.3.14 to add and remove singleton sets from any element. This shows that none of these structures are dense. We take this as a reason to introduce the order of set inclusion modulo finite sets, as is common in e.g. the study of the structure of r.e. sets. We will show that in some cases, the order modulo finite sets will give rise to dense structures. We will use the notation $a \subseteq^\star b$ to mean that $b \setminus a$ is a finite set. Thus $a \equiv^\star b$ iff $(a \setminus b) \cup (b \setminus a)$ is finite. We will also use the notation $a \subset^\star b$ whenever $a \subseteq^\star b$ and $a \not\equiv^\star b$. For each of the structures defined above, we use e.g. $\mathscr{F}^\star_\theta$, to indicate that we consider the structure $\mathscr{F}_\theta$, but with the new order $\subseteq^\star$. The elements of these structures are no longer sets, but rather equivalence classes $[a]$ under the relation $\equiv^\star$.

Proposition 3.4.5. *If $Fix_\Gamma(\theta)$ is non-recursive, the structure $\mathscr{F}^\star_\theta$ is a dense, distributive lattice with a greatest, but no least, element.*

Proof. Let a, b be representatives of the equivalence classes $[a], [b] \in \mathscr{F}^\star_\theta$. It is clear that the equivalence classes corresponding to the union and intersection, respectively, of these sets are elements of $\mathscr{F}^\star_\theta$. By the discussion in the introduction to this section, this suffices to show that $\mathscr{F}^\star_\theta$ is a distributive lattice.

Suppose, for a contradiction, that a is the least element of $\mathscr{F}^\star_\theta$. We can use Proposition 3.3.13 construct an element a' in $\mathscr{F}^\star_\theta$ such that $a' \subset^\star a$. Since a recursive set of fixed points intersects every equivalence class non-recursively, we can find an infinite recursive subset b of such an intersection, and let $a' = a \setminus b$.

Now, suppose $a \subset^\star b$. Since a and b are both recursive, the set $b \setminus a$ is recursive and infinite. This set can then be split in two infinite, recursive parts d, e, such that $a \subset^\star a \cup d = b \setminus e \subset^\star b$. □

As in the case of $\mathscr{F}_\theta$, if we choose $\theta(x)$ such that $\mathrm{Fix}_\Gamma(\theta)$ is recursive, then $\mathscr{F}^\star_\theta$ is a countable Boolean algebra, since $\mathrm{Fix}_\Gamma(\theta)$ is the least element of that structure. The proof above also suffices to show that $\mathscr{F}^\star_\theta$ is dense, so for each pair of Γ-formulas $\theta(x)$, $\chi(x)$ with recursive sets of Γ-fixed points, the structures $\mathscr{F}^\star_\theta$ and $\mathscr{F}^\star_\chi$ are isomorphic.

Using the argument from the proposition above, we can show that $\mathscr{F}^\star_R$ and $\mathscr{F}^\star_\exists$ are dense, upper semi-lattices with a greatest element. It is, on the other hand, unknown if $\mathscr{F}^\star_b$ and $\mathscr{F}_{[\psi]}$ are dense or not, as the aforementioned method does not suffice.

Another known method for proving denseness, successfully applied to partial Lindenbaum algebras, also fails.[6] When we try to adapt this proof, we run into the following problems: Let a and b be elements of $\mathscr{F}^\star_b$ such that $a \subset^\star b$, and let φ, ψ be some lower bounds to a^c and b^c, respectively. Let γ be a sentence that is undecidable in $T + \psi + \neg\varphi$. Then the set $d^c = \{\delta : T + \varphi \vee (\psi \wedge \gamma) \vdash \delta\}$ is a Σ_1-complete set with a lower bound. It follows that d is Π_1-complete, so $d \notin \mathscr{F}^\star_b$. If we could construct an r.e. extension of d, without making $d \equiv^\star b$, then we would have shown that $\mathscr{F}^\star_b$ is dense, but no such construction is known to us.

We would like to know more about these structures and the relation between them. E.g., if any pair of these structures are isomorphic, or whether they can be embedded into each other. As an example, it is clear that each $\mathscr{F}_\theta$ forms a proper filter in $\mathscr{F}_\exists$. It seems that the most interesting structures are $\mathscr{F}^\star_b$ and $\mathscr{F}^\star_{[\psi]}$. The former because it contains the sets of fixed points given by our only condition on possibly non-recursive, r.e. sets (Proposition 3.3.19), and the latter because it is the structure with the neatest structural properties. We wonder in particular how the structure $\mathscr{F}^\star_{[0]}$ is related to $\mathscr{F}^\star_b$. It is clear that each element of these structures is an r.e. extension of the equivalence class of the refutable sentences. However, $[0]$ is not an element of $\mathscr{F}^\star_b$, since $[0]^c$ can have no lower bound: Suppose that $[0]^c$ has a lower bound φ. Then we can always construct a sentence ψ such that $T \nvdash \psi$, $T \nvdash \varphi \to \psi$ and $T \vdash \psi \to \varphi$, contradicting that φ is a lower bound of $[0]^c$.

3.5 Connections

In this section, we show how our investigations relate to an problem on partially conservative sentences, formulated in Guaspari [18] and left open in Bennet [1]. Before we had acquired the results of Section 3.3.1, we had the following partial result. The fixed-point construction used in the proof of the proposition may also be of some interest on its own. Let the *kernel* of $\theta(x)$ be the set $\{k : T \vdash \theta(k)\}$, which we write as K_θ. Thus, K_θ is the set numerated by $\theta(x)$.

Proposition 3.5.1. *If K_θ is recursive, and $\theta \in \Gamma$, then $Fix_\Gamma(\theta)$ is Σ_1-complete.*

[6] Cf. Bennet [1].

Proof. Only the Π_1- (Σ_1-) case provides any difficulties as to complexity. We state the proof in the Π_1-case, and only comment on the Σ_1-case.

Let $\alpha(x)$ be a Σ_1-binumeration of K_θ, and let $X = \{k : \exists m T \vdash \rho(k,m)\}$, where $\rho(x,y)$ is a p.r. formula. Further, let $\xi_0(x)$ be a Π_1-numeration of X, and let $\xi_1(x)$ and $\delta(x)$, respectively, be such that, for all k,

$$\begin{aligned} T \vdash \xi_1(k) \leftrightarrow \exists z(\rho(k,z) \wedge \forall u{\leq}z \neg \mathrm{Prf}(\xi_1(k),u) \wedge \\ (\neg\alpha(\delta(k)) \rightarrow \forall u{\leq}z \neg \mathrm{Prf}_{T+\neg\theta(\delta(k))}(\xi_1(k),u)))) \\ T \vdash \delta(k) \leftrightarrow (\theta(\delta(k)) \wedge \xi_0(k)) \vee (\neg\alpha(\delta(k)) \wedge \neg\xi_1(k)). \end{aligned}$$

$\xi_0(x)$ and $\theta(x)$ are Π_1, $\xi_1(x)$ and $\alpha(x)$ are Σ_1, so $\delta(x)$ is Π_1.

It is easy to check that $\xi_1(x)$ is such that

1. if $k \in X$, then $T \vdash \xi_1(k)$
2. if $k \notin X$, then $T \nvdash \xi_1(k)$ and, if $T + \neg\theta(\delta(k))$ is consistent, $T + \neg\theta(\delta(k)) \nvdash \xi_1(k)$.

Suppose $k \in X$. Then, by 1, $T \vdash \xi_i(k)$, hence, by construction, $T \vdash \delta(k) \leftrightarrow \theta(\delta(k))$.

Now suppose, for a contradiction, that $k \notin X$ and $T \vdash \delta(k) \leftrightarrow \theta(\delta(k))$.

Suppose $\delta(k) \in K_\theta$, whence $T \vdash \theta(\delta(k))$. Now, $\alpha(x)$ binumerates K_θ, so $T \vdash \alpha(\delta(k))$ and thus $T \vdash \delta(k) \leftrightarrow \xi_0(k)$. By choice of $\xi_0(x)$, $T \nvdash \xi_0(k)$, so $T \nvdash \delta(k)$, and since $\delta(k)$ is a fixed point of θ, $T \nvdash \theta(\delta(k))$, a contradiction.

Thus $\delta(k) \notin K_\theta$. Then $T \nvdash \theta(\delta(k))$, so $T + \neg\theta(\delta(k))$ is consistent. But $T \vdash \neg\alpha(\delta(k))$, so $T + \neg\theta(\delta(k)) \vdash \delta(k) \leftrightarrow \neg\xi_1(k)$. Thus $T + \neg\theta(\delta(k)) \vdash \theta(\delta(k)) \leftrightarrow \neg\xi_1(k)$, whence $T + \neg\theta(\delta(k)) \vdash \xi_1(k)$.

Thus

$$k \notin X \Rightarrow T \nvdash \delta(k) \leftrightarrow \theta(\delta(k)).$$

The Σ_1-case goes through, *mutatis mutandis*, by letting $\xi_1(x)$ be such that, for all k:

$$\begin{aligned} T \vdash \xi_1(k) \leftrightarrow \forall z(\forall u{\leq}z \neg\rho(k,u) \rightarrow (\forall u{\leq}z \neg \mathrm{Prf}(\xi_1(k),u) \wedge \\ (\neg\alpha(\delta(k)) \rightarrow \forall u{\leq}z \neg \mathrm{Prf}_{T+\neg\theta(\delta(k))}(\xi_1(k),u))). \quad \square \end{aligned}$$

Another criterion for Σ_1-completeness that we used in an earlier stage of research was the existence of certain kinds of numerations $\alpha(x)$ of r.e. sets.

Definition 3.5.2. *Given $\theta(x) \in \Gamma$, let θ^* be the following statement:*
There is an $\alpha(x) \in \Gamma^d$ such that, for all k,

1. *If $T \vdash \theta(k)$, then $T \vdash \alpha(k)$,*
2. *if $T \nvdash \theta(k)$, then $\neg\alpha(k) \in Cons(\Gamma^d, T + \neg\theta(k))$.*

Let θ^*_H be θ^* with Cons replaced by HCons. By definition, θ^*_H implies θ^*.

Proposition 3.5.3 (Bennet)**.** *If $\theta(x) \in \Gamma$ and θ^* holds, then $Fix_\Gamma(\theta)$ is Σ_1-complete.*

Proof. The proof is almost the same as the proof of Theorem 3.5.1. Let X, $\rho(x,y)$, $\xi_0(x)$, $\xi_1(x)$, and $\delta(x)$ be as in the statement of that theorem, and let $\alpha(x)$ be as regulated by θ^*. It is important to note that if $\theta(x) \in \Gamma$, we can choose $\xi_1(x)$ to be in Γ^d, and to be such that if $T + \neg\theta(\delta(k))$ is consistent, then $T + \neg\theta(\delta(k)) \nvdash \xi_1(k)$.

If $k \in X$, then $T \vdash \xi_i(k)$, so $T \vdash \delta(k) \leftrightarrow \theta(\delta(k))$.

Now, suppose that $k \notin X$, and for a contradiction, that $T \vdash \delta(k) \leftrightarrow \theta(\delta(k))$. If $T \vdash \theta(\delta(k))$, then $T \vdash \alpha(\delta(k))$, whence $T \vdash \delta(k) \rightarrow \xi_0(k)$. But $\delta(k)$ is a fixed point of $\theta(x)$, so $T \vdash \xi_0(k)$, a contradiction. It follows that $T \nvdash \theta(\delta(k))$. Thus $T \nvdash \theta(\delta(k))$. But then $\neg\alpha(\delta(k)) \in \mathrm{Cons}(\Gamma^d, T + \neg\theta(\delta(k))$, whence $T + \neg\theta(\delta(k)) + \neg\alpha(\delta(k))$ is consistent, and proves $\delta(k) \leftrightarrow \neg\xi_1(k)$. But $\delta(k)$ is a fixed point of $\theta(x)$, so $T + \neg\theta(\delta(k)) + \neg\alpha(\delta(k)) \vdash \xi_1(k)$. By the conservativity of $\neg\alpha(\delta(k))$, $T + \neg\theta(\delta(k)) \vdash \xi_1(k)$, a contradiction.

Thus

$$k \notin X \Rightarrow T \nvdash \delta(k) \leftrightarrow \theta(\delta(k)).$$

□

While, in fact, this proposition yields nothing beyond Proposition 3.3.11, there is an interesting relationship between the existence of these numerations and of the set K_θ being recursive. First, let us show that θ^* implies that there is a set of fixed points disjoint from $\mathrm{Fix}_\Gamma(\theta)$.

Proposition 3.5.4. *If θ^* holds, then $Fix_\Gamma(\theta)$ is disjoint from $Fix_\Gamma(\neg\alpha)$.*

Proof. Suppose, for a contradiction, that θ^* holds, and that there is a δ such that $T \vdash \delta \leftrightarrow \theta(\delta) \leftrightarrow \neg\alpha(\delta)$.

Suppose first that $T \vdash \delta$. Then $T \vdash \theta(\delta)$, so, by θ^*, $T \vdash \alpha(\delta)$. But by supposition it follows that $T \vdash \neg\alpha(\delta)$. Thus $T \nvdash \delta$. But then $T \nvdash \theta(\delta)$, whence $\neg\alpha(k) \in \mathrm{Cons}(\Gamma^d, T + \neg\theta(k))$. Now $T \vdash \delta \leftrightarrow \neg\alpha(\delta)$, so $T + \neg\theta(\delta) + \delta$ is consistent. But this theory proves both $\theta(\delta)$ and $\neg\theta(\delta)$, a contradiction. □

We next invoke a theorem of Bennet to shed light on the relationship between the existence of the numerations described by θ^*_H and recursivity of the set K_θ.

Theorem 3.5.5 (Bennet [1])**.** *Given two theories T_0, T_1, the following are equivalent:*

1. $\Gamma \cap \mathit{HCons}(\Gamma^d, T_0) \setminus \mathit{Th}(T_1) = \emptyset$,
2. $\Gamma \cap \mathit{HCons}(\Gamma^d, T_0) \setminus (\mathit{Th}(T_0) \cup \mathit{Th}(T_1)) = \emptyset$,
3. $\mathit{Th}_\Gamma(T_1)$ *is inconsistent with* T_0.

Proposition 3.5.6 (Bennet)**.** *Given $\theta(x) \in \Gamma$, K_θ is recursive iff θ^*_H holds.*

Proof. Let $\theta(x)$ be any Γ-formula such that K_θ is recursive. Let $\alpha(x)$ be a Δ_1-binumeration of K_θ. Then:

1. If $T \vdash \theta(k)$, then $T \vdash \alpha(k)$, since $\alpha(x)$ binumerates K_θ.
2. If $T \nvdash \theta(k)$, then $T \vdash \neg\alpha(k)$. Any provable sentence is trivially conservative, and since $T + \neg\theta(k)$ is consistent, it follows that $\neg\alpha(k) \in \mathrm{HCons}(\Gamma^d, T + \neg\theta(k))$.

Thus, if K_θ is recursive, θ^*_H holds. For the other direction, let $\theta(x)$ be any Γ-formula such that θ^*_H holds. For any k, Let $T_0 := T + \neg\theta(k)$ and $T_1 := T + \theta(k)$. Then $\mathrm{Th}_\Gamma(T_1)$ is inconsistent with T_0, so by Theorem 3.5.5, $\Gamma \cap \mathrm{HCons}(\Gamma^d, T_0) \setminus \mathrm{Th}(T_1) = \emptyset$. Thus we have:

1. If $T \vdash \theta(k)$, then $T \vdash \alpha(k)$, and thus $T + \theta(k) \nvdash \neg\alpha(k)$.
2. If $T \nvdash \theta(k)$, then $T + \theta(k) \vdash \neg\alpha(k)$ and $T \nvdash \alpha(k)$.

These clauses are mutually exclusive, and we have r.e. methods for testing both provability and non-provability of $\theta(k)$. It follows that K_θ is recursive. □

By inspection of the definition of θ^*_H and θ^*, it is easy to confirm the following two facts.

1. If $\theta(x) \in \Delta^T_1$, then θ^*_H,
2. if $\theta(x) \in \Delta^T_n$ for some n, then θ^*.

By the theorem above, we also see that if θ^*_H holds, then K_θ is recursive, so there is a Δ_1-numeration of K_θ. Thus a set is recursive iff it is numerated by a Δ_1-formula iff it is numerated by a formula such that θ^*_H holds. We would like to know if there is a similar relationship when it comes to the statement θ^* and partial (as opposed to hereditary partial) conservativity instead. The solution is not as easily found, and again we invoke a theorem of Bennet.

Theorem 3.5.7 (Bennet [1])**.** *Given two theories T_0, T_1, the following are equivalent:*

1. $\Pi_n \cap Cons(\Sigma_n, T_0) \setminus Th(T_1) = \emptyset$,
2. $\Pi_n \cap Cons(\Sigma_n, T_0) \setminus (Th(T_0) \cup Th(T_1)) = \emptyset$,
3. $Th_{\Pi_n}(T_0) \subseteq Th_{\Pi_n}(T_1)$ *and* $Th_{\Pi_n}(T_1)$ *is inconsistent with* T_0.

Proposition 3.5.8 (Bennet)**.** *For any r.e. set X, there is a Π_n-numeration $\theta(x)$ of X such that K_θ is recursive iff θ^* holds.*

Proof. Let X be any r.e. set, and let, by Lemma 3.2.8 $\theta(x)$ be a Π_n-numeration of X such that, for all k, if $k \in X$, then $T \vdash \theta(k)$, and if $k \notin X$, then $\neg\theta(k) \in \mathrm{Cons}(\Pi_n, T)$. By construction, $K_\theta = X$, and if K_θ is recursive, then θ^* holds by definition.

So suppose θ^* holds. For any k, let $T_0 := T + \neg\theta(k)$ and let $T_1 = T + \theta(k)$. $\mathrm{Th}_{\Pi_n}(T_1)$ is inconsistent with T_0, and since $\neg\theta(k)$ is Π_n-conservative over T, $\mathrm{Th}_{\Pi_n}(T_0) \subseteq \mathrm{Th}_{\Pi_n}(T_1)$. By Theorem 3.5.7, $\Pi_n \cap \mathrm{Cons}(\Sigma_n, T_0) \setminus \mathrm{Th}(T_1) = \emptyset$. Thus:

1. If $T \vdash \theta(k)$, then $T \vdash \alpha(k)$, and $T + \theta(k) \nvdash \neg\alpha(k)$.
2. If $T \nvdash \theta(k)$, then $T + \theta(k) \vdash \neg\alpha(k)$ and $T \nvdash \alpha(k)$.

Again, it follows that K_θ is recursive. □

We know nothing on these lines when it comes to Σ_n-numerations. If we had a result dual to Theorem 3.5.7, we could prove that $\theta(x) \in \Delta^T_n$ iff θ^* holds. But for each $\theta(x) \in \Delta^T_n$, $\mathrm{Fix}_\Gamma(\theta)$ is Σ_1-complete, so such a proof would show that the method of constructing numerations $\alpha(x)$ according to θ^* yields nothing new on the recursion theoretic complexity of sets of fixed points. Now, we have no such proof, and instead we are left with the open question from Bennet [1] of whether one may have

$$\Sigma_n \cap \mathrm{Cons}(\Pi_n, T_0) \setminus \mathrm{Th}(T_1) = \emptyset.$$

Chapter 4

Conclusion and further work

After an introduction, we gave an historical survey of the field of metamathematics. Starting in the early 1900s, with the foundational struggle of Hilbert and others, and the impact of Gödel's incompleteness theorems on these projects, we gave proofs of Rosser's incompleteness theorem, Tarski's theorem of the undefinability of truth, the general Fixed point theorem as stated by Carnap, and Löb's solution to Henkin's problem on provability. Further, we discussed different strengthenings of the Fixed point theorem, due to Ehrenfucht, Feferman and Montague. We studied in some detail the connections between interpretability, partial conservativity and relative consistency. Using examples from Bennet, Bernardi, Guaspari, Hájek, Lindström, Shavrukov, and Solovay, we showed, in the final section of Chapter 2, how more elaborate fixed-point constructions were used in the 70s and 80s.

In Chapter 3, we introduced the notion of a set of fixed points, a concept which we studied in technical detail. We here briefly sum up our knowledge of the recursion theoretic complexity of sets of fixed points.

1. Every set $\mathrm{Fix}(\theta)$ is Σ_1-complete. (Theorem 3.3.3.)

2. Every r.e. Σ_n- (or Π_n-) set is a set of fixed points of a B_n-formula. (Theorem 3.3.6.)

3. Every set of Γ-fixed points whose intersection with another set of Γ-fixed points is recursive and disjoint from some set of Γ-fixed points, is Σ_1-complete. This includes any set of Γ-fixed points of an extensional formula, except for formulas provably equivalent to $\mathrm{Tr}_\Gamma(x)$. (Theorem 3.3.11.)

We also gave examples of Σ_1-complete sets of (Γ-) sentences that are not sets of fixed points of any (Γ-) formula. Moreover, we showed that there are non-extensional formulas with recursive sets of fixed points. The following are sufficient conditions for a set X to be a set of fixed points.

1. X is a recursive subset of Γ, and there is a formula $\theta(x) \in \Gamma$ such that $\mathrm{Fix}_\Gamma(\theta) \cap X^c = \emptyset$. (Proposition 3.3.15.)

2. X is a recursive subset of Γ, and there is a sentence $\psi \in \Gamma$ such that $[\psi] \cap X^c$ is recursive. (Proposition 3.3.16.)

3. X is an r.e. subset of Γ, and there is a sentence $\psi \in \Gamma$ such that $T \vdash \psi \to \varphi$, for all $\varphi \in X^c$. (Proposition 3.3.19.)

The first condition can be changed to the condition that $\mathrm{Fix}_\Gamma(\theta) \cap X^c$ is recursive and disjoint from some set of fixed points, which implies that there is a formula $\chi(x)$ such that $\mathrm{Fix}_\Gamma(\chi) \cap X^c = \emptyset$.

Further on, we investigated the structural properties of sets of fixed points, ordered under set inclusion. As we have no characterisation of sets of fixed points, we could obtain no results on the set of all sets of fixed points. Instead, we had to rely on the three sufficient conditions stated above to define structures with neater properties. A more thorough investigation of this type would probably demand better characterisations.

We were able to relate our knowledge on sets of fixed points to older results, e.g. Löb's theorem and Bernardi's results on inseparability of equivalence classes. The latter connection is due to the fact that every equivalence class is a set of fixed points. Also related to this is the observation that the set of Γ-self-provers of T is a Σ_1-complete set of fixed points.

There are also some relation to open problems in metamathematics, as seen in Section 3.5. There we saw, as in the section on partial Lindenbaum algebras at the end of Chapter 2, that some results may depend on whether we are considering Σ_n- or Π_n-formulas. A possibility is that future refinements of our present results may have to take this situation into account.

Another problem relating to ours is that of Guaspari & Solovay [17]. They ask whether all Rosser sentences are equivalent, i.e. how many equivalence classes the set of Rosser sentences intersects. In their paper, they prove that the answer to this question depends on the choice of proof predicate. By Proposition 3.3.1 it is clear, however, that the Σ_1-completeness of the set of Rosser sentences is independent of this choice. This follows directly from the observation that no Rosser sentence can be provable nor refutable in T, so the set of

Rosser sentences is disjoint from both [0] and [1]. It seems here that we have to take into account the actual syntactical properties of formulas to settle the question, an approach we have not taken in this project.[1]

We intend to take the study of sets of fixed points further, and conclude by listing some open questions that are implicit or explicit in the text.

1. Is there a set of Γ-fixed points that is neither recursive nor Σ_1-complete?
2. Is Y is a recursive set of Γ-fixed points iff Y is a recursive subset of Γ such that $\mathrm{Fix}_\Gamma(\theta) \cap Y$ is non-recursive for all $\theta(x) \in \Gamma$?
3. Is there a Σ_1-complete set of Γ-fixed points that intersects every other set of Γ-fixed points non-recursively?
4. Which conditions are necessary for an r.e. set of (Γ-) sentences to be a set of (Γ-) fixed points?
5. What is the relation between the structure of sets of fixed points ordered under inclusion (or implication) to other known structures, as Boolean algebras and partial Lindenbaum algebras?

[1] See also Blanck [5].

Bibliography

[1] Bennet, C. (1986). **On some orderings of extensions of arithmetic**, Ph. D. thesis, Department of Philosophy, University of Göteborg.

[2] Bennet, C. (1986) *Lindenbaum algebras and partial conservativity* in **Proceedings of the AMS**, Vol. 97, No. 2, pp. 323-327.

[3] Bernardi, C. (1981). *On the Relation Provable Equivalence and on Partitions in Effectively inseparable sets* in **Studia Logica**, Vol. 40, No. 1, pp. 29-37.

[4] Bernardi, C. (1984). *A shorter proof of a recent result by R. Di Paola* in **Notre Dame Journal of Formal Logic**, Vol. 25, No. 4, pp. 390-393.

[5] Blanck, R. (2006). **On Rosser sentences and proof predicates**, Master's thesis, Department of Philosophy, University of Göteborg.

[6] Carnap, R. (1937). **The Logical Syntax of Language**, Routledge & Kegan Paul, London.

[7] Craig, W. (1953). *On axiomatizability within a system* in **The Journal of Symbolic Logic**, Vol. 18, No. 1, pp. 30-32.

[8] Davey, B. A. & Priestley, H. A. (2002). **Introduction to Lattices and Order**, 2nd edition, Cambridge University Press.

[9] Davis, M. (1973) *Hilbert's tenth problem is unsolvable* in **The American Mathematical Monthly**, Vol. 80, No. 3, pp. 233-269.

[10] Ehrenfeucht, A. & Feferman, S. (1960). *Representability of recursively enumerable sets in formal theories* in **Archive for Mathematical Logic**, Vol. 5, No. 1-2, pp. 37-41.

[11] Feferman, S. (1960). *Arithmetization of metamathematics in a general setting* in **Fundamenta Mathematicae**, Vol. 49, pp. 35-92.

[12] Feferman, S., Kreisel, G., Orey, S. (1960). *1-consistency and faithful interpretations* in **Archive for Mathematical Logic**, Vol. 6, pp. 52-63.

[13] Feferman, S. (1988). *Hilbert's program relativized* in **The Journal of Symbolic Logic**, Vol. 53, No. 2, pp. 364-384.

[14] Feferman, S. (1997). *My route to arithmetization* in **Theoria**, Vol. 63, pp. 168-181.

[15] Franzén, T. (2005). **Gödel's Theorem**, A. K. Peters, Wellesley.

[16] Frege, G. (1879). **Begriffsschrift**, translated in [26].

[17] Guaspari, D. & Solovay, R. M. (1979). *Rosser sentences* in **Annals of Mathematical Logic**, Vol. 16, No. 1, pp. 81–99.

[18] Guaspari, D. (1979). *Partially conservative extensions of arithmetic* in **Transactions of the AMS**, Vol. 254, pp. 47-68.

[19] Gödel, K. (1931). *Über Formal Unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme, I.* in **Monatshefte für Math. u. Physik**, Vol. 38, pp. 173-198, translated in [21].

[20] Gödel, K. (1934). *On undecidable propositions of formal mathematical systems*, lecture notes, printed in [21].

[21] Gödel, K. (1986). **Collected works**, eds. Feferman et al., Oxford University Press.

[22] Hájek, P. (1984). *On a new notion of partial conservativity* in **Computation and proof theory**, Springer lecture notes in mathematics, Vol. 1104, pp. 217-232.

[23] Hájek, P. & Pudlák, P. (1998). **Metamathematics of First-Order Arithmetic**, 2nd edition, Springer-Verlag, Berlin.

[24] Hájková, M. (1971). *The lattice of bi-numerations of arithmetic I* in **Commentationes Mathematicae Universitatis Carolinae**, Vol. 12, No. 1, pp. 81-104.

[25] Hájková, M. (1971). *The lattice of bi-numerations of arithmetic II* in **Commentationes Mathematicae Universitatis Carolinae**, Vol. 12, No. 2, pp. 281-306.

[26] van Heijenoort, J. (1999). **From Frege to Gödel**, 2nd edition, toExcel.

[27] Henkin, L. (1952). *A problem concerning provability* in **The Journal of Symbolic Logic**, Vol. 17, No. 2, p. 160.

[28] Hilbert, D. (1904). *On the foundations of logic and arithmetic* in [26], pp. 129-138.

[29] Hilbert, D. (1922). *Neubegründung der Mathematik. Erste Mitteilung*, translated in [45], pp. 198-214.

[30] Hilbert, D. (1925). *On the infinite* in [26], pp. 367-392.

[31] Hilbert, D. (1927). *The foundations of mathematics* in [26], pp. 464-479.

[32] Hilbert, D. & Bernays, P. (1934-1939) **Grundlagen der Mathematik**, Vol. 1-2, Springer-Verlag, Berlin.

[33] Kitcher, P. (1976). *Hilbert's epistemology* in **Philosophy of Science**, Vol. 43, No. 1, pp. 99-115.

[34] Kent, C. F. (1973). *The relation of A to* $\mathrm{Prov}^{\ulcorner} A^{\urcorner}$ *in the Lindenbaum sentence algebra* in **The Journal of Symbolic Logic**, Vol. 38, No. 2, pp. 295-298.

[35] Kleene, S. C. (1938). *On notation for ordinal numbers* in **The Journal of Symbolic Logic**, Vol. 3, No. 4, pp. 150-155.

[36] Kleene, S. C. (1952). **Introduction to metamathematics**, North-Holland.

[37] Lessan, H. (1978). **Models of arithmetic**, Ph. D. thesis, University of Manchester.

[38] Lindström, P. (1979). *Some results on interpretability* in **Proceedings of the 5th Scandinavian Logic Symposium 1979**, Aalborg University Press, pp. 329-361.

[39] Lindström, P. (1984). *On partially conservative sentences and interpretability* in **Proceedings of the AMS**, Vol. 91, pp. 436-443.

[40] Lindström, P. (1984). *On certain lattices of degrees of interpretability* in **Notre Dame Journal of Formal Logic**, Vol. 25, pp. 127-140.

[41] Lindström, P. (1984). *On faithful interpretability* in **Computation and proof theory**, Springer lecture notes in mathematics, Vol. 1104, pp. 279-288.

[42] Lindström, P. (1988). *Partially generic formulas in arithmetic* in **Notre Dame Journal of Formal Logic**, Vol. 29, No. 2, pp. 185-192.

[43] Lindström, P. (2003). **Aspects of Incompleteness**, 2nd edition, A. K. Peters, Natick.

[44] Löb, M. H. (1955). *Solution to a problem of Leon Henkin* in **The Journal of Symbolic Logic**, Vol. 20, No. 2, pp. 115-118.

[45] Mancosu, P. (1998). **From Brouwer to Hilbert**, Oxford University Press.

[46] Montague, R. (1962). *Theories incomparable with respect to relative interpretability* in **The Journal of Symbolic Logic**, Vol. 27, No. 2, pp. 195-211.

[47] von Neumann, J. (2005) **John von Neumann: Selected letters**, (History of Mathematics, Volume 27), ed. Miklós Rédei, Providence, RI: American Mathematical Society.

[48] Orey, S. (1961). *Relative interpretations* in **Zeitschrift f. math. Logik u. Grundlagen d. Mathematik**, Vol. 7, pp. 146-153.

[49] di Paola, R. A. (1984). *A uniformely, extremely nonextensional formula of arithmetic with many undecidable fixed points in many theories* in **Proceedings of the AMS**, Vol. 91, No. 2, pp. 291-297.

[50] Putnam, H. & Smullyan, R. M. (1960). *Exact separation of recursively enumerable sets within theories* in **Proceedings of the AMS**, Vol. 11, pp. 574-577.

[51] Quine, W. V. O. (1962). *The ways of paradox* in [52], pp. 3-20.

[52] Quine, W. V. O. (1966). **The ways of paradox and other essays**, Random House, New York.

[53] Rogers, H. Jr (1967). **Theory of Recursive Functions and Effective Computability**, McGraw-Hill.

[54] Rosser, J. B. (1936). *Extensions of some theorems of Gödel and Church* in **The Journal of Symbolic Logic**, Vol. 1, No. 3, pp. 87-91.

[55] Rosser, J. B. (1939). *An informal exposition of Proofs of Gödel's theorems and Church's theorem* in **The Journal of Symbolic Logic**, Vol. 4, No. 2, pp. 53-60.

[56] Russell, B. & Whitehead, A. N. (1910-1913). **Principia mathematica**, Vols. 1-3, Cambridge University Press.

[57] Shavrukov, V. Yu. (1991). *The Lindenbaum fixed point algebra is undecidable* in **Studia Logica**, Vol. 50, No. 1, pp. 143-148.

[58] Simpson, S. G. (1986). *Partial Realizations of Hilbert's Program* in **The Journal of Symbolic Logic**, Vol. 53, No. 2, pp. 349-363.

[59] Sinaceur, H. (2001). *Alfred Tarski: Semantic Shift, Heuristic Shift In Metamathematics* in **Synthese**, Vol. 126, No. 1-2, pp. 49-65.

[60] Smoryński, C. (1981). *Fifty Years of Self-Reference in Arithmetic* in **Notre Dame Journal of Formal Logic**, Vol. 22, No. 4, pp. 357-375.

[61] Smoryński, C. (1981). *Calculating self-referential statements: Guaspari sentences of the first kind* in **The Journal of Symbolic Logic**, Vol. 46, No. 2, pp. 329-344.

[62] Smoryński, C. (1985). **Self-reference and Modal Logic**, Springer-Verlag, Berlin.

[63] Smullyan, R. M. (1957). *Languages in which self reference is possible* in **The Journal of Symbolic Logic**, Vol. 22, No. 1, pp. 55-67.

[64] Smullyan, R. M. (1961). **Theory of Formal Systems**, Annals of Mathematics Studies, No. 47, Princeton University Press.

[65] Soare, R. I. (1987). **Recursively enumerable sets and degrees**, Springer-Verlag, Berlin.

[66] Solovay, R. M. (1985). *Explicit Henkin Sentences* in **The Journal of Symbolic Logic**, Vol. 50, No. 1, pp. 91-93.

[67] Švejdar, V. (1978). *Degrees of interpretability* in **Comment. Math. Univ. Carol.**, Vol. 19, pp. 789-813.

[68] Tait, W. W. (1981). *Finitism* in **The Journal of Philosophy**, Vol. 78, No. 9, pp. 524-546.

[69] Tarski, A. (1933). *Pojęcie prawdy w językach nauk dedukcyjnych*, translated in [70], pp. 152-278.

[70] Tarski, A. (1983). **Logic, semantics, metamathematics**, ed. J. Corcoran, Hackett Publishing Company.

[71] Visser, A. (1990). *Interpretability logic* in **Mathematical Logic**, ed. P. P. Petkov, Plenum Press, pp. 195-209.

[72] Yanofsky, N. S. (2003). *A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points* in **The Bulletin of Symbolic Logic**, Vol. 9, No. 3, pp. 362-386.