

# Unvollständigkeit formaler Systeme

Prof. Dr. Thomas Streicher

SS 2000

# Inhaltsverzeichnis

1	Einleitung	2
2	Das Hilbertsche Programm	4
3	PRA, PA, $\Sigma_1$ und $\Pi_1$	7
4	Fixpunktsatz für arithmetische Prädikate	10
5	Der 1. Gödelsche Unvollständigkeitssatz	11
6	Der Satz von Löb und der 2. Gödelsche Unvollständigkeitssatz	14
7	Der Satz von Rosser oder die Unvollständigkeit beliebiger konsistenter r.e. Erweiterungen der PRA	17
8	Feferman's schwaches Reflexionsprinzip	20
9	Unentscheidbarkeit der Herleitbarkeit von $\Pi_1$ -Sätzen	21
10	Informatikrelevanz der Unvollständigkeitssätze	23
A	Beweisbarkeitslogik	25
B	Mathematische Unabhängigkeitsresultate	28
	B.1 Das Paris-Harrington Resultat . . . . .	29

# 1 Einleitung

Im Jahr 1930 hat Kurt Gödel den Vollständigkeitssatz für die Prädikatenlogik bewiesen, der besagt, daß ein Satz  $A$  aus einer Menge  $\mathcal{T}$  von Sätzen genau dann herleitbar ist, wenn  $A$  aus  $\mathcal{T}$  semantisch folgt, d.h. wenn  $A$  in jedem Modell gilt, in dem alle Sätze von  $\mathcal{T}$  gelten. Im Jahr 1931 hat Gödel seinen berühmten Unvollständigkeitssatz bewiesen, der besagt, daß es für jede hinreichend starke Axiomatisierung der Arithmetik einen wahren arithmetischen Satz  $G$  gibt, der aber aus dieser Axiomatisierung nicht herleitbar ist. Auf den ersten Blick scheinen dies widersprüchliche Aussagen zu sein. Dem ist aber definitiv nicht so. Betrachten wir z.B. die Peanoarithmetik PA. Der Vollständigkeitssatz besagt, daß ein arithmetischer Satz  $A$  in PA genau dann herleitbar ist, wenn er in *allen* Modellen von PA gilt. Der Gödelsche Unvollständigkeitssatz besagt hingegen, daß es einen arithmetischen Satz  $G$  gibt, der im “Standardmodell”  $\mathbb{N}$  wahr aber in PA nicht herleitbar ist.<sup>1</sup> Man kann die Aussage des Gödelschen Unvollständigkeitssatzes relativ einfach aus der Unentscheidbarkeit des Halteproblems herleiten, wie wir sogleich sehen werden.

Sei  $\mathcal{L}_{\text{pr}}$  die Sprache erster Stufe, die für jede (Definition einer) primitiv rekursiven Funktion eine Funktionskonstante bereitstellt und darüber hinaus bloß ein 2-stelliges Gleichheitsprädikat enthält. Sei  $\text{Ax}_{\text{pr}}$  die Menge aller Sätze in der Sprache  $\mathcal{L}_{\text{pr}}$ , die neben den üblichen Axiomen für die Gleichheit die definierenden Gleichungen für die primitiv rekursiven Funktion(sdefinition)en enthält. Offensichtlich gilt für jede  $k$ -stellige primitiv rekursive Funktion  $f$ , daß  $f(\vec{n}) = m$  genau dann, wenn  $\text{Ax}_{\text{pr}} \vdash f(\vec{n}) = \underline{m}$ <sup>2</sup>, da die Axiome in  $\text{Ax}_{\text{pr}}$  hinreichend stark sind, um geschlossene Terme “auszurechnen”. Infolgedessen gilt für alle primitiv rekursiven Funktionen  $f$ , daß  $\exists \vec{x}. f(\vec{x}) = 0$  genau dann in  $\mathbb{N}$  wahr ist, wenn  $\exists \vec{x}. f(\vec{x}) = \underline{0}$  aus  $\text{Ax}_{\text{pr}}$  herleitbar ist. Aus diesem Grund ist es unentscheidbar, ob eine Formel der Gestalt  $\exists \vec{x}. f(\vec{x}) = \underline{0}$  in  $\text{Ax}_{\text{pr}}$  herleitbar ist, da ja sonst das Halteproblem entscheidbar wäre, da  $n \in K$  genau dann, wenn  $H_n \equiv \exists k. T(\underline{n}, \underline{n}, k)$  aus  $\text{Ax}_{\text{pr}}$  herleitbar ist. Diese Reduktion auf das Halteproblem ist auch möglich für *r.e. konsistente* Erweiterungen von  $\text{Ax}_{\text{pr}}$ .

---

<sup>1</sup>In der Mathematik geht es um die Gültigkeit im Standardmodell  $\mathbb{N}$  und nicht um die Gültigkeit in allen (auch den Nichtstandard-)Modellen.

<sup>2</sup>Zur Erinnerung:  $\underline{n} \equiv \text{succ}^n(0)$ .

**Satz 1.1** Sei  $\mathcal{T}$  eine r.e. konsistente Erweiterung von  $\text{Ax}_{\text{pr}}$ . Dann gibt es ein  $n$  im Komplement von  $K$ , sodaß  $\forall k. \neg T(\underline{n}, \underline{n}, k)$  nicht aus  $\mathcal{T}$  herleitbar ist.

*Beweis:* Angenommen  $\mathcal{T}$  wäre eine r.e. konsistente Erweiterung von  $\text{Ax}_{\text{pr}}$ , aus der alle wahren Aussagen der Form  $\forall k. \neg T(\underline{n}, \underline{n}, k)$  herleitbar sind. Dann würde gelten

$$n \notin K \Leftrightarrow \mathcal{T} \vdash \forall k. \neg T(\underline{n}, \underline{n}, k)$$

da für  $n \in K$  die Aussage  $\exists k. T(\underline{n}, \underline{n}, k)$  bereits aus  $\text{Ax}_{\text{pr}}$  und somit aus  $\mathcal{T}$  herleitbar wäre im Widerspruch zur Konsistenz von  $\mathcal{T}$ .

Dieser Beweis der Existenz eines  $n \in \mathbb{C}K$  mit  $\mathcal{T} \not\vdash \forall k. \neg T(\underline{n}, \underline{n}, k)$  ist hochgradig *inkonstruktiv*. Ein alternatives konstruktives Argument sieht wie folgt aus. Sei  $W_e = \{n \in \mathbb{N} \mid \mathcal{T} \vdash \forall k. \neg T(\underline{n}, \underline{n}, k)\}$  (ein solches  $e$  existiert, da die Menge der Theoreme von  $\mathcal{T}$  ja als r.e. angenommen wurde). Es gilt  $e \notin W_e$ , da aus der Annahme  $e \in W_e$  folgt, daß  $\mathcal{T} \vdash \exists k. T(\underline{e}, \underline{e}, k)$  und  $\mathcal{T} \vdash \forall k. \neg T(\underline{e}, \underline{e}, k)$  im Widerspruch zur Konsistenz von  $\mathcal{T}$ . Also  $\mathcal{T} \not\vdash \forall k. \neg T(\underline{e}, \underline{e}, k)$ , obwohl  $\mathbb{N} \models \forall k. \neg T(\underline{e}, \underline{e}, k)$ , da ja  $\{e\}(e) \uparrow$ .  $\square$

Dies hat insbesondere folgende Konsequenz.

**Korollar 1.1** Eine r.e. konsistente Erweiterung  $\mathcal{T}$  von  $\text{Ax}_{\text{pr}}$  kann nicht alle wahren Aussagen der Form  $\forall x. f(x) \neq \underline{0}$  herleiten.

Wenn  $\forall x. f(x) \neq \underline{0}$  wahr, aber nicht aus  $\mathcal{T}$  herleitbar ist, dann kann man aufgrund des Vollständigkeitssatzes die falsche Aussage  $\exists x. f(x) = \underline{0}$  in konsistenter Weise zu  $\mathcal{T}$  hinzunehmen, Die so erhaltene Erweiterung von  $\mathcal{T}$  ist wiederum r.e. und konsistent, beweist jedoch falsche Sätze der Form  $\exists x. f(x) = \underline{0}$ . Für allgemeine r.e. konsistente Erweiterung  $\mathcal{T}$  von  $\text{Ax}_{\text{pr}}$  liegt somit folgende Situation vor.

- Alle wahren Aussagen der Form  $\exists x. f(x) = \underline{0}$  sind in  $\mathcal{T}$  herleitbar. Jedoch muß nicht jede in  $\mathcal{T}$  herleitbare Formel dieser Gestalt auch wahr sein.
- Alle in  $\mathcal{T}$  herleitbaren Aussagen der Form  $\forall x. \neg f(x) = \underline{0}$  sind wahr. Es gibt aber immer eine wahre Formel der Gestalt  $\forall x. \neg f(x) = \underline{0}$ , die in  $\mathcal{T}$  nicht herleitbar ist.

Konsistente r.e. Erweiterungen  $\mathcal{T}$  von  $\text{Ax}_{\text{pr}}$ , die nur wahre Sätze der Gestalt  $\exists x. f(x) = \underline{0}$  herzuleiten gestatten, heißen *1-konsistent* und sind im weiteren besonders wichtig.

## 2 Das Hilbertsche Programm

Ausgelöst durch verschiedene Paradoxien (Russel, Burali–Forti etc.) in Vorformen der heutigen (axiomatischen) Mengenlehre kam es um die Jahrhundertwende zu einer Grundlagenkrise der Mathematik. Es entstanden starke formale Systeme  $\mathcal{S}_{\text{ideal}}$  wie z.B. ZFC, die die Existenz “idealer Objekte” wie  $\mathcal{P}(\mathbb{N})$ , transfinite Ordinalzahlen etc. postulierten. Mithilfe dieser abstrakten Konzepte ließen sich viele Theorien und Beweise “schöner” darstellen, d.h. in der Form, wie sie uns heute geläufig sind. Die Gefahr dieser starken Systeme ist, daß sie haarscharf am “Abgrund der Inkonsistenz” vorbei(?) lavierten (vgl. z.B. die Ersetzung des inkonsistenten Komprehensionsschemas durch das (hoffentlich) konsistente Separationsschema in ZFC). David HILBERT, ein bekannter deutscher Mathematiker nicht nur seiner Zeit, wollte eine “Rettung” dieser abstrakten Methoden in die Wege leiten, da diese – hauptsächlich durch den Intuitionisten L.E.J. BROUWER – unter starken Beschuß geraten waren. Zumindest damals war man der Meinung, daß man sich von einem Gutteil der modernen Mathematik zu verabschieden hätte, wenn man gezwungen wäre, den Standpunkt der “Intuitionisten” (oder “Konstruktivisten”) in der Nachfolge Brouwer’s zu beziehen, die “nur ganz sichere Methoden” zuließen.

Die Idee von *Hilberts Programm* war nun, ein “ideales System”  $\mathcal{S}_{\text{ideal}}$  relativ zu einem “realen System”  $\mathcal{S}_{\text{real}}$ <sup>3</sup> durch ein “Konservativitätsresultat” abzusichern, d.h., daß

$$(\text{Refl}_A) \quad \mathcal{S}_{\text{ideal}} \vdash A \text{ impliziert } A$$

für alle “elementaren Aussagen”  $A$ , welche letztere Hilbert mit Aussagen der Form  $\forall \vec{x}. f(\vec{x}) = g(\vec{x})$  für primitiv rekursive  $f$  und  $g$  identifizierte (obwohl er sich da nie so genau festlegte). Das anvisierte Konservativitätsresultat besagt also, daß elementare Aussagen  $A$ , die in  $\mathcal{S}_{\text{ideal}}$  herleitbar sind, auch tatsächlich wahr sein sollen. Der Vorteil von  $\mathcal{S}_{\text{ideal}}$  wäre also ausschließlich “stilistischer” Natur in dem Sinne, daß es kürzere und/oder elegantere Beweise elementarer Aussagen gestattet. Das Konservativitätsresultat selbst sollte aber in  $\mathcal{S}_{\text{real}}$  bewiesen<sup>4</sup> werden, da es für alle (auch die Zweifler) einsehbar und akzeptabel sein sollte. Zu diesem Zweck muß man aber die Aussagen von  $\mathcal{S}_{\text{ideal}}$

---

<sup>3</sup>Im wesentlichen  $\text{Ax}_{\text{pr}}$  zusammen mit dem Axiom  $\text{succ}(0) \neq 0$  und dem Induktionsschema für *quantorenfreie* Prädikate, was auch unter dem Namen PRA (Primitiv Rekursive Arithmetik) bekannt ist.

<sup>4</sup>d.h. mit “finiten Methoden”, wie Hilbert es nannte

und “Herleitbarkeit in  $\mathcal{S}_{\text{ideal}}$ ” selber in  $\mathcal{S}_{\text{real}}$  ausdrücken können. Die Antwort auf dieses Problem heißt *Gödelisierung*, obwohl Hilbert selbst in  $\mathcal{S}_{\text{real}}$  über beliebige syntaktische Objekte zu reden gestattete (etwa Listen, Bäume etc. im Slang der heutigen Informatik). D.h. wir ordnen jeder Formel  $A$  von  $\mathcal{S}_{\text{ideal}}$  eine Zahl  $\ulcorner A \urcorner$ , den Code bzw. die Gödelnummer von  $A$ , zu und definieren ein primitiv rekursives Prädikat  $Bew$  mit  $Bew(n, \ulcorner A \urcorner)$  genau dann, wenn  $n$  eine Herleitung von  $A$  codiert, was wohl aufgrund der Ausdrucksstärke primitiv rekursiver Funktionen möglich sein sollte. Die formale Herleitbarkeit von  $A$  in  $\mathcal{S}_{\text{ideal}}$  kann nun in  $\mathcal{S}_{\text{real}}$  folgendermaßen ausgedrückt werden

$$Pr(\ulcorner A \urcorner) \equiv \exists n. Bew(n, \ulcorner A \urcorner) .$$

Das Konservativitätsprinzip für  $\mathcal{S}_{\text{ideal}}$  bzgl.  $\mathcal{S}_{\text{real}}$  läuft nun darauf hinaus, daß man fordert, daß

$$(Refl_A) \quad Pr(\ulcorner A \urcorner) \rightarrow A$$

bzw. die logisch äquivalente Aussage

$$Bew(n, \ulcorner A \urcorner) \rightarrow A$$

in  $\mathcal{S}_{\text{real}}$  herleitbar ist, und zwar für alle elementaren Sätze  $A$ . Da  $\perp \equiv 0 = 1$  zweifelsohne eine elementare Aussage ist, folgt aus dem Konservativitätsprinzip, daß  $\mathcal{S}_{\text{real}} \vdash Pr(\ulcorner \perp \urcorner) \rightarrow \perp$ , i.e.

$$\mathcal{S}_{\text{real}} \vdash \neg Pr(\ulcorner \perp \urcorner)$$

d.h. daß  $\mathcal{S}_{\text{real}}$  die Konsistenz von  $\mathcal{S}_{\text{ideal}}$  beweist. Da  $\mathcal{S}_{\text{real}}$  ein Subsystem von  $\mathcal{S}_{\text{ideal}}$  ist, hätte das insbesondere auch zur Folge, daß  $\mathcal{S}_{\text{real}}$  seine eigene Konsistenz beweist. Bereits dies ist jedoch aufgrund des 2. Gödelschen Unvollständigkeitssatzes nicht möglich, wie wir später sehen werden. Zuvor zeigen wir aber, daß aus der Herleitbarkeit von  $\neg Pr(\ulcorner \perp \urcorner)$  bereits die Herleitbarkeit des Konservativitätsprinzips folgt. Zu diesem Zwecke halten wir einige Eigenschaften des Beweisbarkeitsprädikats  $Pr$  fest, die sogenannten *Derivability Conditions*:

(D1) Wenn  $\mathcal{S}_{\text{ideal}} \vdash A$ , dann  $\mathcal{S}_{\text{real}} \vdash Pr(\ulcorner A \urcorner)$ .

(D2)  $\mathcal{S}_{\text{real}} \vdash Pr(\ulcorner A \rightarrow B \urcorner) \rightarrow Pr(\ulcorner A \urcorner) \rightarrow Pr(\ulcorner B \urcorner)$

(D3)  $\mathcal{S}_{\text{real}} \vdash \exists \vec{x}. f(\vec{x}) = 0 \rightarrow Pr(\ulcorner \exists \vec{x}. f(\vec{x}) = 0 \urcorner)$

Die Bedingungen (D1) und (D2) sind unmittelbar einsichtig. Die Bedingung (D3) zeigt man durch Metainduktion über den Aufbau von  $f$ , d.h. den Aufbau primitiv rekursiver Funktionsdefinitionen, und anschließender Induktion über natürliche Zahlen in  $\mathcal{S}_{\text{real}}$ . Letzteres geschieht durch die Angabe einer primitiv rekursiven Funktion  $\text{deriv}_f$ , für die man in  $\mathcal{S}_{\text{real}}$  beweisen kann, daß

$$\text{Bew}(\text{deriv}_f(\vec{x}), \ulcorner f(\vec{x}) = \underline{f(\vec{x})} \urcorner)$$

woraus unmittelbar die  $\mathcal{S}_{\text{real}}$ -Herleitbarkeit von

$$f(\vec{x}) = 0 \rightarrow \text{Pr}(\ulcorner f(\vec{x}) = 0 \urcorner)$$

und somit die  $\mathcal{S}_{\text{real}}$ -Herleitbarkeit von

$$f(\vec{x}) = 0 \rightarrow \text{Pr}(\ulcorner \exists \vec{x}. f(\vec{x}) = 0 \urcorner)$$

folgt.

**Satz 2.1** *Wenn  $\mathcal{S}_{\text{real}} \vdash \neg \text{Pr}(\ulcorner \perp \urcorner)$ , dann gilt für alle Formeln  $A$  der Gestalt  $\neg \exists n. f(n) = 0$ , daß  $\mathcal{S}_{\text{real}} \vdash \text{Pr}(\ulcorner A \urcorner) \rightarrow A$ .*

*Beweis:* Es gilt

$$(1) \mathcal{S}_{\text{real}} \vdash \exists n. f(n) = 0 \rightarrow \text{Pr}(\ulcorner \exists n. f(n) = 0 \urcorner)$$

aufgrund von (D3). Wegen (D2) gilt, daß

$$(2) \mathcal{S}_{\text{real}} \vdash \text{Pr}(\ulcorner A \urcorner) \rightarrow \text{Pr}(\ulcorner \exists n. f(n) = 0 \urcorner) \rightarrow \text{Pr}(\ulcorner \perp \urcorner).$$

Aufgrund der Annahme  $\mathcal{S}_{\text{real}} \vdash \neg \text{Pr}(\ulcorner \perp \urcorner)$  folgt aus (2), daß

$$(3) \mathcal{S}_{\text{real}} \vdash \text{Pr}(\ulcorner A \urcorner) \rightarrow \text{Pr}(\ulcorner \exists n. f(n) = 0 \urcorner) \rightarrow \perp.$$

Wegen (1) folgt dann

$$(4) \mathcal{S}_{\text{real}} \vdash \text{Pr}(\ulcorner A \urcorner) \rightarrow \exists n. f(n) = 0 \rightarrow \perp$$

und somit, daß

$$(5) \mathcal{S}_{\text{real}} \vdash \text{Pr}(\ulcorner A \urcorner) \rightarrow A$$

wie behauptet. □

Aus diesem Hilbert bereits bekannten Grund lief sein Programm darauf hinaus, in  $\mathcal{S}_{\text{real}}$  die Aussage  $\neg Pr(\ulcorner \perp \urcorner)$  herzuleiten, also mit “finiten Methoden” die Konsistenz von  $\mathcal{S}_{\text{ideal}}$  zu beweisen. Bedauerlicherweise(?) konnte Gödel 1931 zeigen, daß dies schon schief geht, wenn  $\mathcal{S}_{\text{ideal}}$  gleich  $\mathcal{S}_{\text{real}}$  ist. Somit war Hilberts Programm nachweisbar zum Scheitern verurteilt.<sup>5</sup>

Da

$$\mathcal{S}_{\text{real}} \vdash \neg Pr(\ulcorner \perp \urcorner) \leftrightarrow \forall n. \neg Bew(n, \ulcorner \perp \urcorner)$$

und  $\mathcal{S}_{\text{real}} \not\vdash \neg Pr(\ulcorner \perp \urcorner)$ , folgt unmittelbar, daß

$$\mathcal{S}_{\text{real}} \not\vdash \forall n. \neg Bew(n, \ulcorner \perp \urcorner)$$

und somit ist  $\forall n. \neg Bew(n, \ulcorner \perp \urcorner)$  ein elementarer Satz im Sinne Hilberts, der zwar wahr, aber in  $\mathcal{S}_{\text{real}}$  nicht beweisbar ist. Schlimmer noch läßt sich die Aussage  $\forall n. \neg Bew(n, \ulcorner \perp \urcorner)$  zwar in einem idealen System wie etwa ZFC herleiten (etwa durch Angabe eines Modells für  $\mathcal{S}_{\text{real}}$  in ZFC), jedoch nicht in  $\mathcal{S}_{\text{real}}$ . Also kann man ganz *explizit* einen “elementaren” Satz angeben, der in einem idealen System herleitbar ist, jedoch nicht in  $\mathcal{S}_{\text{real}}$ . Also sind abstrakte Konzepte/Methoden nicht konservativ bezüglich elementarer Sätze.

Offensichtlich ist  $\forall n. \neg Bew(n, \perp)$  in  $\mathcal{S}_{\text{real}}$  beweisbar äquivalent zur Divergenz des Programms, das nach der Herleitung von  $\perp$  in  $\mathcal{S}_{\text{real}}$  sucht. Somit ergibt sich aus Gödels 2. Unvollständigkeitssatz eine konkrete wahre Nichtterminationsaussage, die aber in  $\mathcal{S}_{\text{real}}$  nicht herleitbar ist. Obwohl im Beweis von Satz 1.1 ebenfalls eine wahre, aber nicht herleitbare Nichtterminationsaussage angegeben wurde, so erscheint die formalisierte Konsistenzaussage  $\neg Pr(\ulcorner \perp \urcorner)$  denn doch “philosophisch” ansprechender.

### 3 PRA, PA, $\Sigma_1$ und $\Pi_1$

In diesem Abschnitt fixieren wir die Definition zweier formaler Systeme der Arithmetik, nämlich PRA und PA, und definieren die Formelklassen  $\Sigma_1$  und  $\Pi_1$ .

---

<sup>5</sup>Aufgrund der überragenden Autorität Hilberts wurde dieses Scheitern aber nicht sogleich akzeptiert (*Weil nicht sein kann, was nicht sein darf!*). Dies war aber insofern nicht ganz absurd, weil Hilberts informeller Begriff “finiten Methoden” ziemlich dehnbar interpretiert werden konnte. Tatsächlich bewies wenige Jahre später Gerhard GENTZEN die Widerspruchsfreiheit der Arithmetik – allerdings unter Verwendung von Induktion über die Ordinalzahl  $\varepsilon_0$  für quantorenfreie Prädikate.

**Definition 3.1** Die Sprache (1. Stufe)  $\mathcal{L}_{\text{pr}}$  besteht aus Funktionssymbolen für jede primitiv rekursive Funktionsdefinition (inklusive der Konstante 0 und der Nachfolgeroperation  $\text{succ}$ ) und dem 2-stelligen Gleichheitsprädikat  $=$ .

Für eine Formel  $A$  in der Sprache  $\mathcal{L}_{\text{pr}}$  und eine Variable  $x$  bezeichnen wir mit  $\text{Ind}_{A,x}$  bzw.  $\text{Ind}_{A(x)}$  die Formel

$$A(0) \rightarrow \forall x.(A(x) \rightarrow A(\text{succ}(x))) \rightarrow \forall x.A(x)$$

genannt Induktionsprinzip für  $A(x)$ .

Die PRA (Primitiv Rekursive Arithmetik) besteht aus folgenden Axiomen

- (1) den Gleichheitsaxiomen und  $\text{Ax}_{\text{pr}}$ , der Menge aller definierenden Gleichungen für primitive rekursive Funktionen
- (2)  $\text{Ind}_{A,x}$  für alle **quantorenfreien** Formeln  $A$  und
- (3)  $\neg 0 = \text{succ}(0)$ .

Die PA (Peano Arithmetik) entsteht aus der PRA durch Hinzunahme von  $\text{Ind}_{A,x}$  für **beliebige** Formeln  $A$  in der Sprache  $\mathcal{L}_{\text{pr}}$ .  $\diamond$

**Bemerkung.**

- (1) Ohne dem Axiom  $\neg 0 = \text{succ}(0)$  hätte PA und somit auch PRA ein triviales Modell (dessen unterliegende Menge aus bloß einem Objekt besteht).
- (2) Obwohl die PRA bloß Induktion für quantorenfreie Formeln zuläßt, ist sie dennoch relativ ausdrucksstark (siehe etwa Troelstra&vanDalen *Constructivism in Mathematics*, Vol.1, 3.2). Beispielsweise lassen sich folgende Aussagen mühelos in der PRA herleiten

$$0 \neq \text{succ}(x)$$

$$x = 0 \vee x = \text{succ}(\text{pred}(x))$$

$$x \neq 0 \leftrightarrow \text{sig}(x) = \underline{1}$$

$$x = y \leftrightarrow \text{eq}(x, y) = 0$$

wobei  $\text{sig}$  die prim. rek. Signumsfunktion und  $\text{eq}$  die prim. rek. Gleichheitsfunktion  $\text{eq}(x, y) = (x \dot{-} y) + (y \dot{-} x)$  sind.

Schon die PRA wäre ein  $\mathcal{S}_{\text{real}}$  im Hilbertschen Sinn, wo das Gödelsche Unvollständigkeitsresultat zuschlägt, da das Prädikat  $\text{Bew}$  primitiv rekursiv ist und man die notwendigen Eigenschaften von  $\text{Bew}$  und  $\text{Pr}$  bereits in der PRA beweisen kann.

**Definition 3.2** ( $\Sigma_1$  und  $\Pi_1$  Formeln)

Eine  $\Sigma_1$  Formel ist eine Formel der Gestalt  $\exists \vec{x}.A$ , wobei  $A$  eine quantorenfreie Formel der Sprache  $\mathcal{L}_{\text{pr}}$  ist.

Eine  $\Pi_1$  Formel ist eine Formel der Gestalt  $\forall \vec{x}.A$ , wobei  $A$  eine quantorenfreie Formel der Sprache  $\mathcal{L}_{\text{pr}}$  ist.  $\diamond$

Da alle quantorenfreien Prädikate PRA-beweisbar primitiv rekursiv sind, gibt es für jede quantorenfreie Formel  $A(\vec{x})$  eine primitiv rekursive Funktion  $f$ , sodaß

$$A(\vec{x}) \leftrightarrow f(\vec{x}) = 0$$

in der PRA bewiesen werden kann.<sup>6</sup>

Also ist jede  $\Sigma_1$  Formel PRA-beweisbar äquivalent zu einer Formel der Gestalt  $\exists \vec{x}.f(\vec{x}, \vec{y}) = 0$  und jede  $\Pi_1$  Formel PRA-beweisbar äquivalent zu einer Formel der Gestalt  $\forall \vec{x}.f(\vec{x}, \vec{y}) = 0$ .

Da

$$\neg \forall \vec{x}.A(\vec{x}) \leftrightarrow \exists \vec{x}.\neg A(\vec{x}) \quad \neg \exists \vec{x}.A(\vec{x}) \leftrightarrow \forall \vec{x}.\neg A(\vec{x})$$

in der PRA herleitbar sind und quantorenfreie Formeln unter Negation abgeschlossen sind, ist die Negation einer  $\Sigma_1$  bzw.  $\Pi_1$  Formel PRA-beweisbar äquivalent zu einer  $\Pi_1$  bzw.  $\Sigma_1$  Formel. Aus der Rekursionstheorie wissen wir bereits, daß die durch  $\Sigma_1$  Formeln definierbaren Prädikate genau den r.e. Mengen entsprechen. Somit entsprechen die  $\Pi_1$ -definierbaren Prädikate genau den co-r.e. Mengen. Naheliegenderweise nennen wir die durch  $\Sigma_1$  bzw.  $\Pi_1$  Formeln definierbaren Prädikate bzw. Mengen auch  $\Sigma_1$  bzw.  $\Pi_1$  Prädikate bzw. Mengen. Mit  $\Delta_1$  wird gemeinhin der Schnitt von  $\Sigma_1$  und  $\Pi_1$  bezeichnet, d.h. die Prädikate bzw. Mengen, die sowohl durch eine  $\Sigma_1$  als auch durch eine  $\Pi_1$  Formel ausgedrückt werden können. Also entsprechen die  $\Delta_1$ -Mengen gerade den entscheidbaren Mengen.

Aus Abschnitt 1 wissen wir, daß

- PRA genau die wahren  $\Sigma_1$  Sätze beweist
- PRA nur wahre, aber nicht alle wahren  $\Pi_1$ -Sätze beweist.<sup>7</sup>

<sup>6</sup>Insbesondere entsprechen deshalb die  $\Pi_1$  Sätze den "elementaren Aussagen" im Sinne von Hilbert, d.h. den Sätzen der Gestalt  $\forall \vec{x}.f(\vec{x}) = g(\vec{x})$ , da letztere PRA-beweisbar äquivalent sind zu  $\forall \vec{x}.eq(f(\vec{x}), g(\vec{x})) = 0$ .

<sup>7</sup>Da wir allen Grund haben anzunehmen, daß  $\mathbb{N}$  ein Modell der PRA ist. Jedoch können wir dies nicht beweisen. Es ist bloß eine über die Jahrtausende gewachsene "Glaubensgewissheit". Man kann zwar z.B. in ZFC *beweisen*, daß  $\mathbb{N} \models \text{PRA}$ , jedoch wissen wir nicht, ob ZFC selbst 1-konsistent ist.

Dies gilt gleichermaßen für Erweiterungen der PRA, sofern diese im Sinne der folgenden Definition 1-konsistent sind.

**Definition 3.3** Eine (r.e.) Menge  $\mathcal{T}$  von Sätzen der Sprache  $\mathcal{L}_{\text{pr}}$  heißt 1-konsistent genau dann, wenn  $\mathcal{T}$  ausschließlich wahre  $\Sigma_1$  Sätze beweist.  $\diamond$

## 4 Fixpunktsatz für arithmetische Prädikate

Das grundlegende Hilfsmittel für den Beweis der Gödelschen Unvollständigkeitssätze (und verwandter Resultate) ist folgendes höchst verwunderliche Diagonalisierungs- oder Fixpunktlemma für arithmetische Prädikate.

**Satz 4.1** Für jedes arithmetische Prädikat  $P$  gibt es einen arithmetischen Satz  $A$ , sodaß

$$\text{PRA} \vdash A \leftrightarrow P(\ulcorner A \urcorner) .$$

*Beweis:* Sei  $\Theta$  eine Gödelisierung 1-stelliger arithmetischer Prädikate, d.h. von arithmetischen Formeln, die höchstens die ausgezeichnete Variable  $x$  frei enthalten. Es gibt dann eine 2-stellige primitiv rekursive Funktion  $sub$  mit

$$sub(n, m) = \ulcorner \Theta_n(\underline{m}) \urcorner$$

für alle  $n, m \in \mathbb{N}$ .

Für ein gegebenes arithmetisches Prädikat  $P$  sei

$$\Theta_{n_0}(x) \equiv P(sub(x, x))$$

und wir setzen

$$A \equiv \Theta_{n_0}(\underline{n_0}) \equiv P(sub(\underline{n_0}, \underline{n_0})) .$$

Da  $\text{PRA} \vdash sub(\underline{n_0}, \underline{n_0}) = \underline{sub(n_0, n_0)}$  gilt auch

$$\text{PRA} \vdash P(sub(\underline{n_0}, \underline{n_0})) \leftrightarrow P(\underline{sub(n_0, n_0)})$$

und somit

$$\text{PRA} \vdash A \leftrightarrow P(\ulcorner A \urcorner)$$

da  $sub(n_0, n_0) = \ulcorner \Theta_{n_0}(\underline{n_0}) \urcorner = \ulcorner P(sub(\underline{n_0}, \underline{n_0})) \urcorner = \ulcorner A \urcorner$ .  $\square$

Es sei bemerkt, daß kontextabhängig  $\ulcorner A \urcorner$  sowohl für die Gödelnummer von  $A$  als auch das entsprechende Numeral steht. D.h. ganz genau genommen steht  $P(\ulcorner A \urcorner)$  für  $P(\underline{n})$ , wobei  $n = \ulcorner A \urcorner$  die Gödelnummer des Satzes  $A$  ist.

Als relativ unmittelbare Konsequenz erhalten wir, daß  $Th(\mathbb{N})$ , die Menge der wahren arithmetischen Sätze, selbst nicht arithmetisch definierbar ist.

**Satz 4.2** *Es gibt kein arithmetisches Prädikat  $Tr$ , sodaß*

$$\mathbb{N} \models A \leftrightarrow Tr(\ulcorner A \urcorner)$$

*für alle arithmetischen Sätze  $A$ .*

*Beweis:* Wegen Satz 4.1 gibt es einen Satz  $A$ , sodaß  $A \leftrightarrow \neg Tr(\ulcorner A \urcorner)$  in PRA beweisbar ist. Somit würde gelten

$$\mathbb{N} \models Tr(\ulcorner A \urcorner) \leftrightarrow \neg Tr(\ulcorner A \urcorner)$$

was ja nicht möglich ist. □

Es sei bemerkt, daß dieser Satz auch richtig bleibt, wenn man das Standardmodell  $\mathbb{N}$  durch ein Nichtstandardmodell  $\mathcal{N}$  der PRA ersetzt (Übung!).

## 5 Der 1. Gödelsche Unvollständigkeitssatz

Wir beweisen nun, daß es für jede r.e. konsistente Erweiterung  $\mathcal{T}$  der PRA einen Satz  $G$  gibt, sodaß  $G$  in  $\mathcal{T}$  nicht bewiesen werden kann, obzwar  $G$  in  $\mathbb{N}$  wahr ist, und außerdem  $\neg G$  in  $\mathcal{T}$  auch nicht herleitbar ist, sofern  $\mathcal{T}$  1-konsistent ist (d.h. in diesem Fall wird  $G$  durch  $\mathcal{T}$  nicht entschieden). Dieser Satz  $G$  wurde erstmals in einer (bahnbrechenden) Arbeit von Kurt GÖDEL im Jahre 1931 angegeben und ist durch das sogenannte Lügnerparadoxon motiviert, da er durch die Eigenschaft

$$\mathcal{T} \vdash G \leftrightarrow \neg Pr(\ulcorner G \urcorner)$$

charakterisiert ist.

**Satz 5.1** *Sei  $\mathcal{T}$  eine r.e. konsistente Erweiterung der PRA. Dann gibt es aufgrund von Satz 4.1 einen arithmetischen Satz  $G$  mit*

$$\text{PRA} \vdash G \leftrightarrow \neg Pr_{\mathcal{T}}(\ulcorner G \urcorner)$$

*wobei  $Pr_{\mathcal{T}}$  die Beweisbarkeit in  $\mathcal{T}$  in der Sprache der (prim. rek.) Arithmetik ausdrückt.*

*Für  $G$  gelten nun folgende Eigenschaften*

- (1)  $\mathcal{T} \not\vdash G$

(2)  $\mathbb{N} \models G$

(3)  $\mathcal{T} \not\vdash \neg G$ , falls  $\mathcal{T}$  1-konsistent ist.

Es sei bemerkt, daß  $G$  in der PRA beweisbar äquivalent ist zu dem  $\Pi_1$  Satz  $\forall n. \neg \text{Bew}_{\mathcal{T}}(n, \ulcorner G \urcorner)$ .

*Beweis:* Zuallererst sei begemerkt, daß für  $Pr_{\mathcal{T}}$  die Derivability Conditions (D1)–(D3) aus Abschnitt 2 gelten, wobei  $\mathcal{S}_{\text{real}}$  die PRA und  $\mathcal{S}_{\text{ideal}}$  die Theorie  $\mathcal{T}$  ist.

Sei nun  $G$  ein arithmetischer Satz mit  $PRA \vdash G \leftrightarrow \neg Pr_{\mathcal{T}}(\ulcorner G \urcorner)$ , der aufgrund von Satz 4.1 auch tatsächlich existiert.

(1) Angenommen  $\mathcal{T} \vdash G$ , dann gilt wegen (D1), daß  $PRA \vdash Pr_{\mathcal{T}}(\ulcorner G \urcorner)$  und somit auch  $\mathcal{T} \vdash Pr_{\mathcal{T}}(\ulcorner G \urcorner)$ . Wegen  $\mathcal{T} \vdash G$  gilt aber auch  $\mathcal{T} \vdash \neg Pr_{\mathcal{T}}(\ulcorner G \urcorner)$  und somit  $\mathcal{T} \vdash \perp$  im Widerspruch zur Konsistenz von  $\mathcal{T}$ .

(2) Angenommen  $\mathbb{N} \not\models G$ . Dann  $\mathbb{N} \models \neg G$ , d.h.  $\mathbb{N} \models \exists n. \text{Bew}_{\mathcal{T}}(n, \ulcorner G \urcorner)$ , und somit  $\mathcal{T} \vdash G$  im Widerspruch zu (1).

(3) Angenommen  $\mathcal{T} \vdash \neg G$  und  $\mathcal{T}$  sei 1-konsistent. Dann  $\mathcal{T} \vdash \exists n. \text{Bew}_{\mathcal{T}}(n, \ulcorner G \urcorner)$  und somit  $\mathbb{N} \models \exists n. \text{Bew}_{\mathcal{T}}(n, \ulcorner G \urcorner)$  aufgrund der 1-Konsistenz von  $\mathcal{T}$ . Somit gilt aber  $\mathcal{T} \vdash G$  im Widerspruch zur Konsistenz von  $\mathcal{T}$ .

Da  $PRA \vdash \neg Pr_{\mathcal{T}}(\ulcorner G \urcorner) \leftrightarrow \forall n. \neg \text{Bew}_{\mathcal{T}}(n, \ulcorner G \urcorner)$  gilt auch

$$PRA \vdash G \leftrightarrow \forall n. \neg \text{Bew}_{\mathcal{T}}(n, \ulcorner G \urcorner)$$

wie behauptet. □

Als nächstes zeigen wir, daß der Gödelsatz  $G$  in der PRA beweisbar äquivalent ist zu  $\text{Con}_{\mathcal{T}} \equiv \neg Pr_{\mathcal{T}}(\ulcorner \perp \urcorner)$ , d.h. zur formalisierten Aussage der Konsistenz von  $\mathcal{T}$ .

**Konvention** Aus Gründen der Lesbarkeit schreiben wir in Zukunft

$$\Box A \quad \text{für} \quad Pr(\ulcorner A \urcorner)$$

was (durchaus nicht unabsichtlich!) an den Modaloperator “notwendigerweise” erinnern soll. Die “Derivability Conditions” lesen sich dann auf folgende suggestive Weise viel angenehmer:

$$(D1) \vdash A \Rightarrow PRA \vdash \Box A$$

$$(D2) PRA \vdash \Box(A \rightarrow B) \rightarrow \Box A \rightarrow \Box B$$

(D3)  $\text{PRA} \vdash \Box A \rightarrow \Box \Box A$

wobei bemerkt sei, daß man (D3) auch formulieren könnte als  $B \rightarrow \Box B$  für alle  $\Sigma_1$  Sätze  $B$ . Im weiteren ist jedoch bloß der Fall  $B \equiv \Box A \equiv \text{Pr}(\ulcorner A \urcorner) \equiv \exists n. \text{Bew}(n, \ulcorner A \urcorner)$  interessant.

**Satz 5.2** *Sei  $\mathcal{T}$  eine r.e. konsistente Erweiterung der PRA und  $G$  ein Gödel-satz für  $\mathcal{T}$ , d.h.  $\text{PRA} \vdash G \leftrightarrow \neg \text{Pr}_{\mathcal{T}}(\ulcorner G \urcorner)$ , dann gilt*

$$\text{PRA} \vdash G \leftrightarrow \neg \text{Pr}_{\mathcal{T}}(\ulcorner \perp \urcorner) .$$

*Beweis:* Wir schreiben abkürzend  $\vdash$  für  $\text{PRA} \vdash$ , d.h. Herleitbarkeit in PRA. Da  $G$  beweisbar äquivalent ist zu  $\neg \text{Pr}(\ulcorner G \urcorner)$  genügt es zu zeigen, daß

$$\vdash \text{Pr}(\ulcorner G \urcorner) \leftrightarrow \text{Pr}(\ulcorner \perp \urcorner) .$$

Da  $\vdash \perp \rightarrow G$ , gilt wegen (D1), daß  $\vdash \Box(\perp \rightarrow G)$  und somit wegen (D2), daß  $\vdash \Box \perp \rightarrow \Box G$ .

Weil  $\vdash G \rightarrow \neg \Box G$ , gilt wegen (D1), daß  $\vdash \Box(G \rightarrow \neg \Box G)$ , und somit wegen (D2), daß

$$(1) \vdash \Box G \rightarrow \Box \neg \Box G .$$

Wegen (D2) gilt aber

$$(2) \vdash \Box \neg \Box G \rightarrow \Box \Box G \rightarrow \Box \perp$$

da  $\vdash \neg \Box G \rightarrow \Box G \rightarrow \perp$ . Aus (1) und (2) folgt aber, daß

$$(3) \Box G \rightarrow \Box \Box G \rightarrow \Box \perp$$

und somit das erwünschte

$$\vdash \Box G \rightarrow \Box \perp$$

da ja aufgrund von (D3) gilt, daß  $\vdash \Box G \rightarrow \Box \Box G$ . □

Aus den beiden Sätzen 5.1 und 5.2 folgt nun für jede konsistente r.e. Erweiterungen  $\mathcal{T}$  der PRA, daß

$$\mathcal{T} \not\vdash \text{Con}_{\mathcal{T}}$$

obwohl ja  $\mathbb{N} \models \text{Con}_{\mathcal{T}}$ , da  $\mathcal{T}$  als konsistent vorausgesetzt wurde. Also ist  $\mathcal{T} + \neg \text{Con}_{\mathcal{T}}$  immer konsistent, allerdings **nicht** 1-konsistent.

Wenn  $\mathcal{T}$  überdies 1-konsistent ist, dann gilt auch

$$\mathcal{T} \not\vdash \neg \text{Con}_{\mathcal{T}}$$

und somit ist in diesem Fall auch  $\mathcal{T} + \text{Con}_{\mathcal{T}}$  konsistent<sup>8</sup>.

<sup>8</sup>Dies ist nicht trivial, da ja nicht notwendigerweise  $\mathbb{N} \models \mathcal{T}$ ! Zum Beispiel ist ja aufgrund obiger Betrachtungen die Theorie  $\text{PRA} + \text{Pr}_{\text{PRA}}(\ulcorner \perp \urcorner)$  konsistent. Somit ist aber auch in

## 6 Der Satz von Löb und der 2. Gödelsche Unvollständigkeitssatz

Wie im letzten Abschnitt dargelegt besagt der 1. Gödelsche Unvollständigkeitssatz, daß  $\mathcal{T} \not\vdash \text{Con}_{\mathcal{T}}$ , sofern  $\mathcal{T}$  eine konsistente r.e. Erweiterung der PRA ist. Diese metamathematische Aussage des 1. Unvollständigkeitssatzes läßt sich wiederum selbst folgendermaßen in der PRA formalisieren als

$$\text{Con}_{\mathcal{T}} \rightarrow \neg \Box \text{Con}_{\mathcal{T}}$$

und der 2. Gödelsche Unvollständigkeitssatz besagt nun, daß diese letztere Aussage selbst in der PRA beweisbar ist<sup>9</sup>.

Da  $\text{Con}_{\mathcal{T}} \rightarrow \neg \Box \text{Con}_{\mathcal{T}}$  logisch äquivalent ist zu  $\Box \text{Con}_{\mathcal{T}} \rightarrow \Box \perp$ , d.h.

$$\Box(\Box \perp \rightarrow \perp) \rightarrow \Box \perp$$

genügt es offensichtlich, diese letztere Aussage in der PRA herzuleiten. Allerdings ist dies bloß ein Spezialfall des Satzes von Löb, der besagt, daß

$$\text{PRA} \vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$$

bereits für alle arithmetischen Sätze  $A$  gilt.

### Satz 6.1 (Satz von Löb)

Sei  $\mathcal{T}$  eine konsistente r.e. Erweiterung der PRA, dann gilt für alle Sätze  $A$ , daß

- (1) Wenn  $\mathcal{T} \vdash \Box A \rightarrow A$ , dann bereits  $\mathcal{T} \vdash A$ .
- (2)  $\text{PRA} \vdash \Box(\Box A \rightarrow A) \rightarrow \Box A$  (“formalised Löb’s Theorem”)

wobei  $\Box A$  abkürzend für  $\text{Pr}_{\mathcal{T}}(\ulcorner A \urcorner)$  steht.

Da selbstverständlich auch  $\text{PRA} \vdash \Box A \rightarrow \Box(\Box A \rightarrow A)$  sind die Aussagen  $\Box A$  und  $\Box(\Box A \rightarrow A)$  in der PRA beweisbar äquivalent.

---

jeder r.e. konsistenten Erweiterung  $\mathcal{T}$  von  $\text{PRA} + \text{Pr}_{\text{PRA}}(\ulcorner \perp \urcorner)$  die Aussage  $\text{Pr}_{\mathcal{T}}(\ulcorner \perp \urcorner)$ , d.h.  $\neg \text{Con}_{\mathcal{T}}$  herleitbar, da ja ganz offensichtlich für beliebige Formeln  $A$  die Aussage  $\text{Pr}_{\text{PRA}}(\ulcorner A \urcorner) \rightarrow \text{Pr}_{\mathcal{T}}(\ulcorner A \urcorner)$  bereits in der PRA herleitbar ist.

<sup>9</sup>Der 2. Unvollständigkeitssatz wurde in der legendären Gödelschen Arbeit aus dem Jahre 1931 bloß angekündigt, aber nicht bewiesen. Der von Gödel in dieser Arbeit angekündigte Folgeartikel ist jedoch von ihm nie publiziert worden.

*Beweis:* Sei  $A$  ein beliebiger arithmetischer Satz. Aufgrund des Diagonalisierungslemmas 4.1 gibt es einen arithmetischen Satz  $L$  mit

$$(1) \text{ PRA} \vdash L \leftrightarrow \Box(L \rightarrow A) .$$

Wegen (D3) gilt auch, daß

$$(2) \text{ PRA} \vdash L \rightarrow \Box L$$

da  $L$  ja (zu) ein(em)  $\Sigma_1$  Satz (äquivalent) ist. Außerdem gilt wegen (D2), daß

$$(3) \text{ PRA} \vdash \Box(L \rightarrow A) \rightarrow \Box L \rightarrow \Box A .$$

Aus (1), (2) und (3) folgt nun, daß

$$(4) \text{ PRA} \vdash L \rightarrow \Box A$$

und somit auch

$$(5) \text{ PRA} \vdash (\Box A \rightarrow A) \rightarrow L \rightarrow A .$$

Aus (5) folgt mit (D1), daß

$$(6) \text{ PRA} \vdash \Box((\Box A \rightarrow A) \rightarrow L \rightarrow A)$$

und somit wegen (D2) auch

$$(7) \text{ PRA} \vdash \Box(\Box A \rightarrow A) \rightarrow \Box(L \rightarrow A) .$$

Aus (1) und (7) folgt nun, daß

$$(8) \text{ PRA} \vdash \Box(\Box A \rightarrow A) \rightarrow L .$$

Somit folgt aus (4) und (8), daß

$$\text{PRA} \vdash \Box(\Box A \rightarrow A) \rightarrow \Box A .$$

Also haben wir die zweite Behauptung gezeigt.

Für den Beweis der ersten Behauptung nehmen wir an, daß  $\mathcal{T} \vdash \Box A \rightarrow A$ . Dann folgt mit (D1), daß auch  $\text{PRA} \vdash \Box(\Box A \rightarrow A)$  und somit wegen der bereits bewiesenen 2. Behauptung des Satzes, daß  $\text{PRA} \vdash \Box A$ . Also insbesondere auch  $\mathcal{T} \vdash \Box A$ , woraus mit der Annahme  $\mathcal{T} \vdash \Box A \rightarrow A$  folgt, daß  $\mathcal{T} \vdash A$ , wie behauptet.  $\square$

**Korollar 6.1** (2. Gödelscher Unvollständigkeitssatz)

*Es gilt*

$$\text{PRA} \vdash \Box \text{Con}_{\mathcal{T}} \rightarrow \Box \perp$$

*und somit*

$$\text{PRA} \vdash \text{Con}_{\mathcal{T}} \rightarrow \neg \Box \text{Con}_{\mathcal{T}}$$

*Beweis:* (1) folgt aus der 2. Behauptung von Satz 6.1, wenn wir  $A$  durch  $\perp$  instanziiieren. Behauptung (2) folgt aus (1) rein logisch (mit Kontraposition).  $\square$

Wie schon in Abschnitt 2 erwähnt bezeichnen wir Sätze der Form  $\Box A \rightarrow A$  mit  $\text{Refl}_A$ , da es ein ‘‘Reflexionsprinzip für  $A$ ’’ in dem Sinne ausdrückt, daß ‘‘ $A$  wahr ist, falls  $A$  herleitbar ist’’. Leider läßt sich aufgrund des Satzes von Löb das Reflexionsprinzip für  $A$  in  $\mathcal{T}$  nur in den trivialen Fällen beweisen, wo  $A$  selbst schon in  $\mathcal{T}$  herleitbar ist. Anders ausgedrückt kann  $\mathcal{T}$  die Korrektheit von  $\mathcal{T}$  bloß für diejenigen Sätze beweisen, die in  $\mathcal{T}$  bereits herleitbar sind. Für alle anderen Sätze  $A$  kann man den Satz  $\text{Pr}_{\mathcal{T}}(\ulcorner A \urcorner) \wedge \neg A$  konsistenterweise zu  $\mathcal{T}$  hinzunehmen.

Wir können also festhalten, daß jedes konsistente formale System  $\mathcal{T}$  insofern offen ist, als die Hinzunahme von  $\text{Refl}_A$  für alle Sätze  $A$  eine echte Verstärkung des Systems zur Folge hat.

Wir haben bereits gesehen, daß der 2. Unvollständigkeitssatz aus Löbs Satz folgt. Es sei hier noch erwähnt, daß auch der 1. Unvollständigkeitssatz aus dem 2. unmittelbar folgt<sup>10</sup>. Also lassen sich alle (betrachteten) Unvollständigkeitsphänomene auf den Satz von Löb reduzieren.

---

<sup>10</sup>Sei  $\mathcal{T}$  eine konsistente r.e. Erweiterung der PRA. Angenommen  $\mathcal{T} \vdash \text{Con}_{\mathcal{T}}$ . Dann gilt aufgrund von (D1), daß  $\text{PRA} \vdash \Box \text{Con}_{\mathcal{T}}$ , und somit aufgrund des 2. Unvollständigkeitssatzes, daß auch  $\text{PRA} \vdash \Box \perp$ . Wegen der 1-Konsistenz der PRA gäbe es dann eine natürliche Zahl  $n$  mit  $\text{PRA} \vdash \text{Bew}_{\mathcal{T}}(\underline{n}, \ulcorner \perp \urcorner)$ , woraus folgte, daß  $\mathcal{T} \vdash \perp$  im Widerspruch zur angenommenen Konsistenz von  $\mathcal{T}$ .

Angenommen  $\mathcal{T}$  sei überdies 1-konsistent. Wenn  $\mathcal{T} \vdash \neg \text{Con}_{\mathcal{T}}$ , dann gilt aufgrund der 1-Konsistenz von  $\mathcal{T}$ , daß  $\mathbb{N} \models \exists n. \text{Bew}_{\mathcal{T}}(n, \ulcorner \perp \urcorner)$  und somit  $\mathcal{T} \vdash \perp$  im Widerspruch zur Konsistenz von  $\mathcal{T}$ .

Selbstverständlich kann man aber die erste Behauptung des 1. Gödelschen Unvollständigkeitssatzes auch direkt aus Satz 6.1 (1) folgern, indem man  $A$  durch  $\perp$  instanziiert.

## 7 Der Satz von Rosser oder die Unvollständigkeit beliebiger konsistenter r.e. Erweiterungen der PRA

Aus dem 1. Unvollständigkeitssatz von Gödel wissen wir bloß, daß für 1-konsistente r.e. Erweiterungen  $\mathcal{T}$  der PRA gilt, daß in  $\mathcal{T}$  die formalisierte Konsistenz von  $\mathcal{T}$  weder beweisbar noch widerlegbar ist. Daß für manche konsistente r.e. Erweiterungen  $\mathcal{T}$  der PRA die formalisierte Inkonsistenz von  $\mathcal{T}$  in  $\mathcal{T}$  selbst beweisbar ist, sieht man an folgendem Beispiel.

Sei  $\mathcal{T}$  eine konsistente r.e. Erweiterung der PRA, sodaß  $\mathcal{T} \vdash Pr_{PRA}(\ulcorner \perp \urcorner)$ , wie z.B.  $\mathcal{T} \equiv PRA + Pr_{PRA}(\ulcorner \perp \urcorner)$ . Weil ja offensichtlich

$$PRA \vdash Pr_{PRA}(\ulcorner \perp \urcorner) \rightarrow Pr_{\mathcal{T}}(\ulcorner \perp \urcorner)$$

folgt somit, daß

$$\mathcal{T} \vdash Pr_{\mathcal{T}}(\ulcorner \perp \urcorner)$$

d.h.

$$\mathcal{T} \vdash \neg Con_{\mathcal{T}} .$$

Also wird der Gödelsatz  $Con_{\mathcal{T}}$  in sehr vielen Fällen von  $\mathcal{T}$  (negativ) entschieden. Um diesen Defekt zu beheben, gibt es gottseidank den ‘‘Rosseratz’’ einer Theorie  $\mathcal{T}$ , der von  $\mathcal{T}$  *nie* entschieden wird.

Vorbereitend dazu benötigen wir folgendes Lemma.

**Lemma 7.1** *Für jedes  $n \in \mathbb{N}$  und jede Formel  $A(x)$  gilt*

- (1)  $PRA \vdash x < \underline{n} \leftrightarrow \bigvee_{k < n} x = \underline{k}$
- (2)  $PRA \vdash \exists x < \underline{n}. A(x) \leftrightarrow \bigvee_{k < n} A(\underline{k})$ .

*Beweis:* Behauptung (1) zeigt man leicht durch Metainduktion über  $n \in \mathbb{N}$ . Behauptung (2) folgt unmittelbar aus (1).  $\square$

**Satz 7.1** *Sei  $\mathcal{T}$  eine konsistente r.e. Erweiterung der PRA. Dann gibt es einen  $\Pi_1$  Satz  $R$ , sodaß*

- (i)  $\mathcal{T} \not\vdash R$
- (ii)  $\mathbb{N} \models R$
- (iii)  $\mathcal{T} \not\vdash \neg R$ .

*Beweis:* Anstelle von  $Bew_{\mathcal{T}}$  betrachten wir das etwas restriktivere Prädikat

$$Bew^*(n, \ulcorner A \urcorner) \equiv Bew_{\mathcal{T}}(n, \ulcorner A \urcorner) \wedge \forall k < n. \neg Bew_{\mathcal{T}}(k, \ulcorner \neg A \urcorner)$$

welches auch primitiv rekursiv ist, da die Allquantifikation beschränkt ist. Entsprechend betrachten wir auch folgende Verstärkung von  $Pr_{\mathcal{T}}$

$$Pr^*(\ulcorner A \urcorner) \equiv \exists n. Bew^*(n, \ulcorner A \urcorner)$$

wofür es aufgrund des Diagonalisierungslemmas einen Satz  $R$ , den sogenannten “Rossersatz von  $\mathcal{T}$ ”, gibt mit

$$(1) \text{ PRA} \vdash R \leftrightarrow \neg Pr^*(\ulcorner R \urcorner)$$

d.h.

$$(2) \text{ PRA} \vdash R \leftrightarrow \forall n. \neg Bew_{\mathcal{T}}(n, \ulcorner R \urcorner) \vee \exists k < n. Bew_{\mathcal{T}}(k, \ulcorner \neg R \urcorner) .$$

*ad (i):* Angenommen  $\mathcal{T} \vdash R$ . Dann gibt es eine natürliche Zahl  $n_0$  mit  $\text{PRA} \vdash Bew_{\mathcal{T}}(\underline{n_0}, \ulcorner R \urcorner)$ . Wegen (2) gilt aber

$$\mathcal{T} \vdash \neg Bew_{\mathcal{T}}(\underline{n_0}, \ulcorner R \urcorner) \vee \exists k < \underline{n_0}. Bew_{\mathcal{T}}(k, \ulcorner \neg R \urcorner)$$

woraus wegen  $\mathcal{T} \vdash Bew_{\mathcal{T}}(\underline{n_0}, \ulcorner R \urcorner)$  folgt, daß

$$\mathcal{T} \vdash \exists k < \underline{n_0}. Bew_{\mathcal{T}}(k, \ulcorner \neg R \urcorner)$$

und somit wegen Lemma 7.1(2), daß

$$\mathcal{T} \vdash \bigvee_{k < \underline{n_0}} Bew_{\mathcal{T}}(\underline{k}, \ulcorner \neg R \urcorner) .$$

Da  $\mathcal{T}$  (wie jede konsistente Erweiterung der PRA) genau die wahren, d.h. in  $\mathbb{N}$  gültigen,  $\Delta_1$  Sätze beweist, gibt es also ein  $k < n_0$  mit

$$\text{PRA} \vdash Bew_{\mathcal{T}}(\underline{k}, \ulcorner \neg R \urcorner)$$

woraus folgt, daß  $\mathcal{T} \vdash \neg R$  im Widerspruch zu  $\mathcal{T} \vdash R$  und der Konsistenz von  $\mathcal{T}$ .

*ad (ii):* Angenommen  $\mathbb{N} \not\models R$ . Dann

$$\mathbb{N} \models Pr^*(\ulcorner R \urcorner)$$

woraus folgt, daß es einen Code für einen Beweis von  $R$  in  $\mathcal{T}$  gibt. Dann gilt aber  $\mathcal{T} \vdash R$  im Widerspruch zu (i).

*ad* (iii): Angenommen  $\mathcal{T} \vdash \neg R$ . Dann gibt es eine natürliche Zahl  $n_0$  mit

$$\text{PRA} \vdash \text{Bew}_{\mathcal{T}}(\underline{n_0}, \ulcorner \neg R \urcorner)$$

und somit gilt auch

$$\text{PRA} \vdash (\forall k < n. \neg \text{Bew}_{\mathcal{T}}(k, \ulcorner \neg R \urcorner)) \rightarrow n \leq \underline{n_0} .$$

Also beweist die PRA auch den Satz

$$(\exists n. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner) \wedge \forall k < n. \neg \text{Bew}_{\mathcal{T}}(k, \ulcorner \neg R \urcorner)) \rightarrow \exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner)$$

und, da die Prämisse dieser Implikation PRA-beweisbar äquivalent zu  $\neg R$  ist, somit auch

$$\text{PRA} \vdash \neg R \rightarrow \exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner) .$$

Also gilt

$$\mathcal{T} \vdash \exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner) .$$

Wegen Lemma 7.1(2) gilt

$$\text{PRA} \vdash \exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner) \leftrightarrow \bigvee_{n \leq \underline{n_0}} \text{Bew}_{\mathcal{T}}(\underline{n}, \ulcorner R \urcorner)$$

und wegen der Konsistenz von  $\mathcal{T}$  gilt für jede natürliche Zahl  $n$ , daß

$$\text{PRA} \vdash \neg \text{Bew}_{\mathcal{T}}(\underline{n}, \ulcorner R \urcorner) .$$

Somit folgt, daß

$$\text{PRA} \vdash \neg \exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner) ,$$

und somit insbesondere, daß

$$\mathcal{T} \vdash \neg \exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner) .$$

Somit beweist  $\mathcal{T}$  sowohl  $\exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner)$  als auch seine Negation im Widerspruch zur Konsistenz von  $\mathcal{T}$ .  $\square$

Die Knackpunkt von Rossers Beweis ist, daß im Falle  $\mathcal{T} \vdash \neg R$  eine natürliche Zahl  $n_0$  existiert, sodaß nicht nur  $\text{PRA} \vdash \text{Bew}_{\mathcal{T}}(\underline{n_0}, \ulcorner \neg R \urcorner)$ , sondern auch  $\text{PRA} \vdash \neg R \rightarrow \exists n \leq \underline{n_0}. \text{Bew}_{\mathcal{T}}(n, \ulcorner R \urcorner)$ . Für  $G$  ist dies nicht möglich, weil man sonst beweisen könnte, daß  $\mathcal{T} \not\vdash \neg G$ , obwohl ja manche  $\mathcal{T}$  durchaus  $\neg G$  beweisen können, wie wir bereits gesehen haben.

Als Konsequenz des Satzes von Rosser ergibt sich folgendes Korollar.

**Korollar 7.1** Für jede konsistente r.e. Erweiterung  $\mathcal{T}$  der PRA gibt es eine konsistente r.e. Erweiterung  $\mathcal{T}'$ , die echt stärker ist als  $\mathcal{T}$ .

*Beweis:* Sei  $R$  der Rosser Satz von  $\mathcal{T}$ . Wir definieren  $\mathcal{T}' \equiv \mathcal{T} + R$ . Da  $\mathcal{T} \not\vdash \neg R$ , ist  $\mathcal{T}'$  konsistent. Außerdem ist  $\mathcal{T}'$  echt stärker als  $\mathcal{T}$ , weil  $\mathcal{T} \not\vdash R$ .  $\square$

## 8 Feferman's schwaches Reflexionsprinzip

Obzwar man aufgrund des Satzes von Löb das Reflexionsprinzip

$$(\text{Refl}_A) \quad \text{Bew}(n, \ulcorner A \urcorner) \rightarrow A$$

bloß für herleitbare  $A$  herleiten kann, gilt folgendes *abgeschwächte* Reflexionsprinzip

$$(\text{WRefl}_A) \quad \forall n. \Box(\text{Bew}(\underline{n}, \ulcorner A \urcorner) \rightarrow A)$$

für alle Sätze  $A$  der PRA, wie erstmals von S. FEFERMAN in den 60er Jahren des 20. Jahrhunderts gezeigt wurde.

**Satz 8.1** Für konsistente r.e. Erweiterungen  $\mathcal{T}$  der PRA gilt

$$\text{PRA} \vdash \forall n. \Box_{\mathcal{T}}(\text{Bew}_{\mathcal{T}}(\underline{n}, \ulcorner A \urcorner) \rightarrow A) .$$

*Beweis:* Aus Gründen der Lesbarkeit schreiben wir  $\text{Bew}$  bzw.  $\Box$  für  $\text{Bew}_{\mathcal{T}}$  bzw.  $\Box_{\mathcal{T}}$ .

Wir zeigen folgende beide Behauptungen

- (i)  $\text{PRA} \vdash \text{Bew}(n, \ulcorner A \urcorner) \rightarrow \Box(\text{Bew}(\underline{n}, \ulcorner A \urcorner) \rightarrow A)$
- (ii)  $\text{PRA} \vdash \neg \text{Bew}(n, \ulcorner A \urcorner) \rightarrow \Box(\text{Bew}(\underline{n}, \ulcorner A \urcorner) \rightarrow A)$

woraus die Behauptung unmittelbar durch Fallunterscheidung über  $\text{Bew}(n, \ulcorner A \urcorner)$  folgt (die auch konstruktiv zulässig ist, da es sich um ein primitiv rekursives Prädikat handelt).

*ad* (i): folgende Formeln sind PRA-beweisbar

$$(1) \quad A \rightarrow \text{Bew}(n, \ulcorner A \urcorner) \rightarrow A$$

(2)  $\Box A \rightarrow \Box(\text{Bew}(\underline{n}, \ulcorner A \urcorner) \rightarrow A)$  aus (1) mit (D1)<sup>11</sup> und (D2)

(3)  $\text{Bew}(n, \ulcorner A \urcorner) \rightarrow \Box A$

(4)  $\text{Bew}(n, \ulcorner A \urcorner) \rightarrow \Box(\text{Bew}(\underline{n}, \ulcorner A \urcorner) \rightarrow A)$  aus (2) und (3).

*ad* (ii): folgende Formeln sind PRA-beweisbar

(1)  $\neg \text{Bew}(n, \ulcorner A \urcorner) \rightarrow \Box(\neg \text{Bew}(\underline{n}, \ulcorner A \urcorner))$  wegen (D3)<sup>12</sup>

(2)  $\neg \text{Bew}(n, \ulcorner A \urcorner) \rightarrow \text{Bew}(n, \ulcorner A \urcorner) \rightarrow A$

(3)  $\Box(\neg \text{Bew}(\underline{n}, \ulcorner A \urcorner)) \rightarrow \Box(\text{Bew}(\underline{n}, \ulcorner A \urcorner) \rightarrow A)$  aus (2) mit (D1)<sup>12</sup> und (D2)

(4)  $\neg \text{Bew}(n, \ulcorner A \urcorner) \rightarrow \Box(\text{Bew}(\underline{n}, \ulcorner A \urcorner) \rightarrow A)$  aus (1) und (3).

□

## 9 Unentscheidbarkeit der Herleitbarkeit von $\Pi_1$ -Sätzen

Für den Rest dieses Abschnitts sei  $\mathcal{T}$  eine fest gewählte konsistente r.e. Erweiterung der PRA. Wir wissen, daß  $\mathcal{T}$  nur *wahre*  $\Pi_1$ -Sätze herleiten kann, jedoch auf keinen Fall alle solche, da sonst das Halteproblem entscheidbar wäre. Es stellt sich nun die Frage, ob die Menge der in  $\mathcal{T}$  herleitbaren  $\Pi_1$ -Sätze entscheidbar ist. Die in diesem Abschnitt gegebene Antwort ist negativ für *alle* konsistenten r.e. Erweiterung der PRA.

Als Vorbereitung benötigen wir folgenden Hilfssatz.

**Lemma 9.1** *Für  $n \in \mathbb{N}$  ist*

$$A_n \equiv \forall k. T(\underline{n}, \underline{n}, k) \rightarrow U(k) = 0$$

*ein  $\Pi_1$ -Satz. Wenn  $\{n\}(n) = 0$ , so gilt  $\text{PRA} \vdash A_n$ , und, wenn  $\{n\}(n) = 1$ , so gilt  $\text{PRA} \vdash \neg A_n$ .*

---

<sup>11</sup>Tatsächlich verwenden wir hier eine leichte Verallgemeinerung von (D1), die besagt, daß aus  $\mathcal{T} \vdash A(\vec{x})$  folgt, daß  $\text{PRA} \vdash \Box A(\vec{x})$ .

<sup>12</sup>Tatsächlich verwenden wir hier eine leichte Verallgemeinerung von (D3), die besagt, daß  $\text{PRA} \vdash A(\vec{x}) \rightarrow \Box A(\vec{x})$  für beliebige  $\Sigma_1$ -Formeln  $A(\vec{x})$  mit freien Variablen.

*Beweis:* In der PRA kann man beweisen, daß

$$(\dagger) \quad T(n, m, k_1) \wedge T(n, m, k_2) \rightarrow k_1 = k_2$$

da (Codes) terminierende(r) Berechnungssequenzen durch Programm und Eingabe eindeutig bestimmt sind!

Wenn  $\{n\}(n) = 0$ , so gibt es ein  $k \in \mathbb{N}$  mit  $T(n, n, k)$  und  $U(k) = 0$ , d.h.

$$\text{PRA} \vdash T(\underline{n}, \underline{n}, \underline{k}) \wedge U(\underline{k}) = 0 \quad .$$

Wegen  $(\dagger)$  gilt dann, daß  $\text{PRA} \vdash A_n$ .

Wenn  $\{n\}(n) = 1$ , so gibt es ein  $k \in \mathbb{N}$  mit  $T(n, n, k)$  und  $U(k) = 1$ , d.h.

$$\text{PRA} \vdash T(\underline{n}, \underline{n}, \underline{k}) \wedge U(\underline{k}) = \underline{1} \quad .$$

Wegen  $(\dagger)$  gilt dann, daß  $\text{PRA} \vdash \neg A_n$ . □

In unseren Ausführungen zur Rekursionstheorie haben wir bereits gesehen, daß die Mengen  $A_0 = \{n \in \mathbb{N} \mid \{n\}(n) = 0\}$  und  $A_1 = \{n \in \mathbb{N} \mid \{n\}(n) = 1\}$  nicht rekursiv trennbar sind. Dies benötigen wir im Beweis des folgenden Satzes, der eine negative Antwort auf die anfangs gestellte Frage liefert.

**Satz 9.1** *Die Menge der in  $\mathcal{T}$  herleitbaren  $\Pi_1$ -Sätze ist nicht entscheidbar.*

*Beweis:* Sei  $\Pi_1^{\mathcal{T}} := \{\ulcorner A \urcorner \mid A \text{ } \Pi_1\text{-Satz, } \mathcal{T} \vdash A\}$ . Wenn die total rekursive Funktion  $p$  die Menge  $\Pi_1^{\mathcal{T}}$  entscheidet, dann gälte für  $f := p \circ g$ , wobei  $g$  die rekursive Funktion mit  $g(n) = \ulcorner A_n \urcorner$  ist, daß

$$(i) \quad n \in A_0 \Rightarrow f(n) = 0$$

$$(ii) \quad n \in A_1 \Rightarrow f(n) = 1$$

da wegen Lemma 9.1  $n \in A_0 \Rightarrow \ulcorner A_n \urcorner \in \Pi_1^{\mathcal{T}}$  und  $n \in A_1 \Rightarrow \ulcorner A_n \urcorner \notin \Pi_1^{\mathcal{T}}$ . Weil  $f$  total rekursiv ist, wären somit  $A_0$  und  $A_1$  rekursiv trennbar im Widerspruch zu dem bekannten Resultat aus der Rekursionstheorie.

Ein alternativer, traditionellerer Beweis von Satz 9.1 basierend auf dem bewährten Diagonalisierungslemma sieht folgendermaßen aus.

Sei  $\{e\}$  eine total rekursive Funktion, die  $\Pi_1^{\mathcal{T}}$  entscheidet, d.h.

$$\{e\}(n) = 0 \quad \text{gdw} \quad n = \ulcorner A \urcorner \text{ für einen } \Pi_1\text{-Satz } A \text{ mit } \mathcal{T} \vdash A.$$

Das Prädikat  $P$  sei definiert als

$$P(n) \equiv \forall k. T(\underline{e}, n, k) \rightarrow U(k) \neq 0 .$$

Aufgrund des Diagonalisierungslemmas gibt es einen Satz  $A$  mit  $\text{PRA} \vdash A \leftrightarrow P(\ulcorner A \urcorner)$ . Durch genaue Inspektion des Beweises des Diagonalisierungslemmas ergibt sich, daß  $A$  *selbst ein*  $\Pi_1$ -Satz ist, nämlich

$$A \equiv \Theta_{n_0}(\underline{n_0}) \equiv P(\text{sub}(\underline{n_0}, \underline{n_0}))$$

wobei  $\Theta_{n_0}(x) \equiv P(\text{sub}(x, x))$ .

Sei nun  $k$  die kleinste natürliche Zahl, sodaß  $T(e, \ulcorner A \urcorner, k)$ . Es gilt dann, daß

- (i)  $U(k) = 0 \Rightarrow \mathcal{T} \vdash A$
- (ii)  $U(k) \neq 0 \Rightarrow \text{PRA} \vdash A \Rightarrow \mathcal{T} \vdash A$

und somit  $\mathcal{T} \vdash A$ . Dann gilt aber, daß  $U(k) = 0$ , weil  $\ulcorner A \urcorner \in \Pi_1^{\mathcal{T}}$ , und auch  $U(k) \neq 0$ , weil  $\mathcal{T} \vdash T(\underline{e}, \ulcorner A \urcorner, \underline{k}) \rightarrow U(\underline{k}) \neq 0$  und  $\mathcal{T} \vdash T(\underline{e}, \ulcorner A \urcorner, \underline{k})$ . Somit  $0 \neq 0$ . □

Da bereits die  $\mathcal{T}$ -Herleitbarkeit von  $\Pi_1$ -Sätzen unentscheidbar ist, ist auch ganz allgemein die  $\mathcal{T}$ -Herleitbarkeit von beliebigen arithmetischen Sätzen unentscheidbar. Außerdem folgt aus der Unentscheidbarkeit der in  $\mathcal{T}$  herleitbaren  $\Pi_1$ -Sätze, daß die Menge der mit  $\mathcal{T}$  konsistenten  $\Sigma_1$ -Sätze nicht rekursiv aufzählbar ist.

Wie auch immer, die Moral dieses Abschnitts ist, daß man für  $\Pi_1$ -Sätze nicht mechanisch vorab überprüfen kann, ob  $\mathcal{T}$  an ihrem Beweis scheitert.

## 10 Informatikrelevanz der Unvollständigkeitssätze

Bereits im ersten Teil der Vorlesung über Rekursionstheorie haben wir gesehen, daß wesentliche Eigenschaften von Programmen unentscheidbar, ja oft nicht einmal rekursiv aufzählbar sind. Beispielsweise ist das Halteproblem zwar rekursiv aufzählbar, aber nicht entscheidbar, wohingehend die Menge der (für jede Eingabe) terminierenden Algorithmen nicht rekursiv aufzählbar ist (noch ist es ihr Komplement!).

Mithilfe der Gödelsätze kann man für konsistente r.e. Erweiterungen der primitiv rekursiven Arithmetik auch *konkrete* Beispiele angeben, an denen  $\mathcal{T}$  scheitert:

1. Wenn man die Divergenz des Programms, das nach  $\mu n. Bew_{\mathcal{T}}(n, \ulcorner \perp \urcorner)$  sucht, in  $\mathcal{T}$  beweisen könnte, dann könnte  $\mathcal{T}$  seine Konsistenz beweisen.
2. Wenn man in  $\mathcal{T}$  die Totalität des Programms

$$f(n) \simeq \mathbf{if} \text{ } Bew_{\mathcal{T}}(n, \ulcorner \perp \urcorner) \mathbf{ then } \uparrow \mathbf{ else } 0$$

beweisen könnte, so könnte man in  $\mathcal{T}$  auch die Konsistenz von  $\mathcal{T}$  beweisen. Also ist  $f$  ein Programm, das für alle Eingaben terminiert (da ja  $\mathcal{T}$  als konsistent vorausgesetzt war), wofür man aber die Termination in  $\mathcal{T}$  selbst nicht beweisen kann.

3. Das obige Programm  $f$  ist *extensional* gleich dem Programm  $g(n) = 0$ . Man kann jedoch in  $\mathcal{T}$  die extensionale Gleichheit der Programme  $f$  und  $g$  nicht beweisen, weil sonst die Termination von  $f$  in  $\mathcal{T}$  beweisbar wäre. Selbstverständlich kann  $\mathcal{T}$  die Termination des Programms  $g$  beweisen, also sind die in  $\mathcal{T}$  beweisbar terminierenden Programme nicht unter extensionaler Gleichheit abgeschlossen!
4. Sei  $\mathcal{T}$  eine r.e. konsistente Erweiterung der PRA, die einen  $\Sigma_1$  Satz  $\exists k P(k)$  beweist, der in  $\mathbb{N}$  falsch ist. Dann gibt es zu jeder natürlichen Zahl  $e$  eine Zahl  $\tilde{e}$ , sodaß  $\phi_e = \phi_{\tilde{e}}$  aber  $\mathcal{T} \vdash \forall n \exists k T(\tilde{e}, n, k)$ . Man nehme für  $\tilde{e}$  die Gödelnummer des Algorithmus, der  $n$  abbildet auf  $f(\mu k. P(k) \vee T(e, n, k))$ , wobei  $f(k) \simeq \mathbf{if} P(k) \mathbf{ then } 0 \mathbf{ else } U(k)$ . Also sind nicht 1-konsistente  $\mathcal{T}$  höchst ungeeignet, um über Algorithmen zu rasonieren.

## Literatur

- [Sm] C. Smoryński *Self-Reference and Modal Logic* Springer Verlag, 1985.
- [Bo] G. Boolos *The Logic of Provability* Cambridge University Press, 1993.
- [Mat] Y. Matiyasevich *Hilbert's Thenth Problem* MIT Press, 1993.
- [GRS] R. Graham, B. Rothschild, J. Spencer *Ramsey Theory* Wiley 1980.
- [Mar] D. Marker *Model Theory: An Introduction* Springer 2002.

## A Beweisbarkeitslogik

Bereits im Verlauf der bisherigen Darstellung haben wir die modallogische Schreibweise verwendet, um die Derivability Conditions kompakt darzustellen und somit ihre Verwendung in Beweisen zu erleichtern. Tatsächlich steckt aber mehr dahinter als eine bloße Kurzschreibweise. G. Boolos, C. Smoryński (siehe [Bo, Sm]) und einige andere Logiker haben in den 1970er Jahren die Modallogik PRL (Provability Logik) vorgeschlagen, die wir in folgender Definition einführen wollen.

**Definition A.1** (Definition von PRL)

*Die Sprache der PRL ist die der üblichen Aussagenlogik angereichert um den 1-stelligen Operator  $\Box$ . Wir basieren den rein ausagenlogischen Teil auf die Junktoren  $\rightarrow$  (Implikation) und  $\perp$  (Konstante für die immer falsche Aussage). Wir schreiben  $\neg A$  als Abkürzung für  $A \rightarrow \perp$ . Wir verwenden  $p, q, r, \dots$  als Metavariablen für ausagenlogische Variable. Die Logik PRL wird durch folgende Axiome und Regeln axiomatisiert*

(A1) *Alle booleschen Tautologien sind Axiome.*

(A2)  $\Box(A \rightarrow B) \rightarrow \Box A \rightarrow \Box B$

(A3)  $\Box(\Box A \rightarrow A) \rightarrow \Box A$

(R1) *Aus  $A \rightarrow B$  und  $A$  schließe  $B$ .*

(R2) *Aus  $A$  schließe  $\Box A$ .*

Man beachte, daß für PRL nicht das Deduktionstheorem gilt, da  $\Box A$  aus  $A$  herleitbar ist, aber die Aussage nicht  $A \rightarrow \Box A$  im allgemeinen nicht herleitbar ist. Außerdem sieht man leicht, daß  $\Box A \rightarrow A$  nur dann herleitbar ist, wenn bereits  $A$  herleitbar ist.

Die Regel (R2) entspricht (D1) und Axiom (A2) entspricht (D2). Bei Axiom (A3) handelt es sich offenbar um den formalisierten Satz von Löb. Amüsanterweise ist das (D3) entsprechende Prinzip  $\Box A \rightarrow \Box \Box A$  bereits in PRL herleitbar.

**Lemma A.1** *Für alle Formeln  $A$  beweist PRL die Formel  $\Box A \rightarrow \Box \Box A$ .*

*Beweis:* Offenbar ist

$$(1) A \rightarrow (\Box\Box A \wedge \Box A) \rightarrow (\Box A \wedge A)$$

eine boolesche Tautologie, also auch

$$(2) A \rightarrow \Box(\Box A \wedge A) \rightarrow (\Box A \wedge A)$$

Somit gilt auch

$$(3) \Box A \rightarrow \Box(\Box(\Box A \wedge A) \rightarrow (\Box A \wedge A))$$

Als spezielle Instanz von (A3) erhalten wir

$$(4) \Box(\Box(\Box A \wedge A) \rightarrow (\Box A \wedge A)) \rightarrow \Box(\Box A \wedge A)$$

Aus (3) und (4) folgt nun

$$(5) \Box A \rightarrow \Box(\Box A \wedge A)$$

Da  $\Box(\Box A \wedge A) \rightarrow \Box\Box A$  folgt aus (5) wie gewünscht  $\Box A \rightarrow \Box\Box A$ .  $\square$

Der Witz von PRL ist, daß alle “arithmetischen Interpretation” von in PRL beweisbaren Aussagen in der PRA beweisbar sind. Eine “arithmetische Interpretation” ist eine Abbildung  $\phi$ , die Aussagenvariablen auf geschlossenen Formeln der PRA abbildet. Ein solches  $\phi$  induziert für jede modallogische Formel  $A$  eine geschlossene Formel der  $A^\phi$  der PRA wir folgt

$$p^\phi \equiv \phi(p) \quad \perp^\phi \equiv \perp \quad (A \rightarrow B)^\phi \equiv A^\phi \rightarrow B^\phi \quad (\Box A)^\phi \equiv Pr(\ulcorner A^\phi \urcorner)$$

Offenbar folgt aus  $\text{PRL} \vdash A$ , daß  $\text{PRA} \vdash A^\phi$  für alle arithmetischen Interpretationen  $\phi$ , da PRA ja den formalisierten Satz von Löb beweist und die Derivability Conditions (D1) und (D2) gelten. Um einiges schwieriger ist der Beweis des folgenden Satzes von R. Solovay (1976) (siehe [Sm, Bo] für detailliert Beweise).

**Satz A.1** (Arithmetische Vollständigkeit von PRL)

*Eine modallogische Formel  $A$  ist in PRL genau dann beweisbar, wenn PRA alle arithmetischen Interpretationen  $A^\phi$  beweist.*

Sei  $\text{PRL}_\omega$  diejenige Modallogik, die durch alle Theoreme von PRL, alle Formeln der Form  $\Box A \rightarrow A$  und den Modus Ponens axiomatisiert wird.<sup>13</sup> Da die PRA 1-konsistent ist und sogar  $\mathbb{N} \models \text{PRA}$ , folgt aus  $\text{PRL}_\omega \vdash A$ , daß in  $\mathbb{N}$  alle arithmetischen Interpretationen  $A^\phi$  gelten. R. Solovay (1976) hat auch bewiesen, daß

<sup>13</sup>Die Hinzunahme des Axiomenschemas  $\Box A \rightarrow A$  zu PRL ist ja inkonsistent!

**Satz A.2** (Arithmetische Vollständigkeit von  $\text{PRL}_\omega$ )

*Eine modallogische Formel  $A$  in  $\text{PRL}_\omega$  genau dann beweisbar ist, wenn alle arithmetischen Interpretationen  $A^\phi$  in  $\mathbb{N}$  gelten.*

Diese beiden Sätze sind von besonderem Interesse, da die Mengen der von  $\text{PRL}$  bzw.  $\text{PRL}_\omega$  bewiesenen Sätze entscheidbar sind. Hierbei verwendet man wesentlich, daß  $\text{PRL}$  vollständig ist bzgl. endlicher Kripke Modelle  $(W, R)$ , in denen  $R$  in dem Sinne wohlfundiert ist, daß es keine unendliche Kette

$$w_0 R w_1 R \dots w_n R w_{n+1} \dots$$

gibt. Dieses Vollständigkeitsresultat wird essentiell im Beweis der Sätze A.1 und A.2 verwendet. Im interessanten Teil des Beweises von A.1 geht man von einer in  $\text{PRL}$  nicht beweisbaren Formel aus, für die dann ein endliches wohlfundiertes Kripke Modell existiert, in dem  $A$  nicht gilt und mithilfe dessen man eine arithmetische Interpretation  $\phi$  konstruiert, sodaß  $A^\phi$  in der PRA nicht beweisbar ist.

Eine weitere wichtige Eigenschaft von  $\text{PRL}$  ist

**Satz A.3** *Für eine modallogische Formel  $A[p]$ , wo  $p$  nur innerhalb von  $\Box$  vorkommt, gibt es einen Satz  $B$ , dessen Aussagenvariablen alle von  $p$  verschieden sind und in  $A[p]$  vorkommen, sodaß*

(1)  $\text{PRL} \vdash B \leftrightarrow A[B]$  und

(1) für alle modallogischen Formeln  $C$  aus  $\text{PRL} \vdash C \leftrightarrow A[C]$  folgt, daß  $\text{PRL} \vdash C \leftrightarrow B$ .

D.h. “geboxte” Formeln  $A[p]$  haben in  $\text{PRL}$  Fixpunkte, die bis auf in  $\text{PRL}$  beweisbare Äquivalenz eindeutig sind.

## B Mathematische Unabhängigkeitsresultate

Wir haben gesehen, daß für alle r.e. konsistenten Erweiterungen  $T$  der PRA sich  $\Pi_1$  Sätze finden lassen, die in  $T$  nicht bewiesen werden können, typischerweise den  $\Pi_1$  Satz  $Con_T \equiv \forall n \neg Pr_T(\ulcorner \perp \urcorner)$ . Es käme jedoch kein Mathematiker, der nicht zufälligerweise Logiker ist, auf die Idee sich für einen Satz der Gestalt  $Con_T$  zu interessieren. Somit erhebt sich die Frage, ob es “mathematisch sinnvolle” Sätze gibt, die in PRA oder PA nicht beweisbar sind.<sup>14</sup> Natürlich ist “mathematisch sinnvoll” selbst kein mathematischer Begriff in dem Sinne, daß er präzise definiert werden könnte. Jedoch erkennt man mathematisch sinnvolle Aussagen, wenn man sie sieht. Im wesentlichen heißt das für uns, daß sie sich in ihrer Formulierung nicht auf logische Begriffe wie “beweisbar” oder “konsistent” beziehen. Außerdem sollten solche Aussagen sich halbwegs konzise formulieren lassen.

Letzteres trifft nicht ganz auf unser erstes Beispiel zu, nämlich Polynome mit ganzzahligen Koeffizienten, die keine ganzzahligen Nullstellen besitzen aber diese Aussage nicht in  $T$  beweisbar ist. Da sich die ganzen Zahlen leicht durch die natürlichen Zahlen kodieren lassen, ist die Aussage der Nullstellenfreiheit

---

<sup>14</sup>Selbstverständlich erhebt sich diese Frage auch für starke Systeme wie Z, ZF oder ZFC bzw. Erweiterungen dieser Systeme, z.B. durch die Forderung der Existenz großer Kardinalzahlen. Der Logiker Harvey Friedman hat sich mit dieser Frage intensiv beschäftigt (siehe etwa die Vorversion seines Buches *Boolean Relation Theory*, die man sich unter der Adresse [www.math.ohio-state.edu/~friedman/pdf/Whole103007.pdf](http://www.math.ohio-state.edu/~friedman/pdf/Whole103007.pdf) runterladen kann).

Ein erstes Beispiel in dieser Richtung wurde in den 1970er Jahren betrachtet und diente dazu, Z und ZF durch eine mathematisch natürliche Aussage zu trennen. Diese Aussage ist die sogenannte *Borel Determinacy*. Man betrachtet Spiele, wo in jedem Zug einer der beiden Spieler eine natürliche Zahl wählt. Als Resultat eines solchen unendlichen Spiels ergibt sich am Ende (aller Zeiten) eine unendliche Folge  $\alpha \in \mathbb{N}^{\mathbb{N}}$ , dem Baireschen Raum, den wir als mit der Produkttopologie ausgestattet denken. Für eine Teilmenge  $B \subseteq \mathbb{N}^{\mathbb{N}}$  erhebt sich nun die Frage, ob das Spiel für  $B$  “determiniert” ist, d.h. ob für den ersten oder den zweiten Spieler eine Gewinnstrategie existiert, wobei der erste Spieler genau dann gewinnt, wenn das Resultat des Spieles in  $B$  liegt. Die Borelmengen des Baireschen Raums sind die kleinste  $\sigma$ -Algebra, die die offenen Mengen des Baireschen Raums enthält. *Borel Determinacy* besagt nun, daß das Spiel für jede Borelmenge im Baireschen Raum determiniert ist. Der ursprünglich von D. Martin gegebene Beweis verwendet eine überabzählbare Iteration von  $\mathcal{P}$ , wofür man das Ersetzungsschema (*axiom of replacement*), das in ZF, aber nicht in Z postuliert wird, braucht. H. Friedman konnte zeigen, daß die *Borel Determinacy* in Z tatsächlich nicht beweisbar ist, d.h., daß ihr Beweis das mengentheoretische Ersetzungsschema benötigt. Dies ist bemerkenswert, da sich so gut wie alle der üblichen Aussagen der Mathematik ohne dem Ersetzungsschema beweisen lassen!

eines Polynoms  $P(\vec{x})$ , d.h.  $\forall \vec{x}. P(\vec{x}) \neq 0$ , auf recht einfache Art und Weise zu einer  $\Pi_1$  Aussage äquivalent. Für den Nachweis der Existenz eines solchen Polynoms verwenden wir den berühmten Satz von Y. Matiyasevich (siehe etwa [Mat]), der besagt, daß für jede r.e. Menge  $A$  ein Polynom  $P_A(x, \vec{y})$  mit ganzzahligen Koeffizienten existiert, sodaß  $x \in A$  genau dann, wenn  $x \geq 0$  und  $\exists \vec{y}. P_A(x, \vec{y}) = 0$ . Wenn nun  $A$  r.e. aber nicht rekursiv entscheidbar ist, dann kann  $T$  nicht für alle  $n \notin A$  die Aussage  $A_n \equiv \forall \vec{y}. P_A(n, \vec{y}) \neq 0$  beweisen, da man ja sonst ein Entscheidungsverfahren für  $A$  angeben könnte. Also gibt es eine natürliche Zahl  $n \in \mathbb{N} \setminus A$ , sodaß  $A_n \equiv \forall \vec{y}. P_A(n, \vec{y}) \neq 0$  gilt, aber in  $T$  nicht beweisbar ist. Es gibt also ein Polynom  $P$ , nämlich  $P(\vec{x}) = P_A(n, \vec{x})$ , das keine ganzzahligen Nullstellen hat, obwohl  $T$  diese Aussage nicht beweisen kann. Zugegebenermaßen ist solch ein Polynom nicht leicht anzugeben, hat aber zumindest weniger als 20 Variablen (siehe [Mat] für genauere Eingrenzungen).

Nebenbei bemerkt ist es noch nicht klar, ob der große Satz von Fermat (eine  $\Pi_1$  Aussage!) in PA beweisbar ist, obwohl manche Experten (A. MacIntyre) dieser Meinung sind. Ein Beweis dafür steht jedenfalls noch aus!

## B.1 Das Paris-Harrington Resultat

Das erste als solches akzeptierte “mathematische Unabhängigkeitsresultat” für die Peano Arithmetik wurde von J. Paris and L. Harrington Ende der 1970er Jahre bewiesen (siehe ihren Artikel *A Mathematical Incompleteness in Peano Arithmetic* im *Handbook of Mathematical Logic*). Unsere Präsentation des Resultats von Paris und Harrington folgt der Darstellung in Kapitel 5.4 von [Mar].

Bei dem Paris-Harringtonschen Unabhängigkeitsresultat handelt es sich um eine Verschärfung des endlichen Ramsey Satzes, die in der Arithmetik 2. Stufe herleitbar ist, aber nicht in der Peano Arithmetik, *die wir für diesen Abschnitt als in der Sprache  $\{+, \cdot, <\}$  (nebst Gleichheit) formuliert verstehen*.

Wir rufen zuerst den endlichen Ramsey Satz in Erinnerung und führen zu diesem Zweck etwas Notation ein. Wenn  $X$  eine Menge ist und  $k \in \mathbb{N}$ , dann bezeichne  $[X]^k$  die Menge der  $k$ -elementigen Teilmengen von  $X$ . Für eine natürliche Zahl  $m$  ist eine  $m$ -Färbung von  $[X]^k$  eine Abbildung  $f : [X]^k \rightarrow m$ , wobei  $m = \{0, 1, \dots, m-1\}$ .<sup>15</sup> Eine Teilmenge  $Y$  von  $X$  heißt *homogen* (bzgl.

<sup>15</sup>Im Falle  $k = m = 2$  läßt sich  $f : [X]^2 \rightarrow 2$  verstehen als *ungerichteter Graph* mit Knotenmenge  $X$ .

der Färbung  $f$ ), wenn  $f$  auf  $[Y]^k$  konstant ist. Wenn  $k, m, \ell, n \in \mathbb{N}$ , dann schreiben wir

$$n \longrightarrow (\ell)_m^k$$

als Abkürzung für die Aussage, daß für jede Färbung  $f : [n]^k \rightarrow m$  eine homogene  $\ell$ -elementige Teilmenge  $X \subseteq n$  existiert, sodaß die Einschränkung von  $f$  auf  $[X]^k$  konstant ist. In [GRS] z.B. findet man einen Beweis des folgenden Satzes

**Satz B.1** (endlicher Ramsey Satz)

$$\forall k, m, \ell \in \mathbb{N}. \exists n \in \mathbb{N}. n \longrightarrow (\ell)_m^k.$$

*Beweis:* Wir führen Induktion über  $k$ .

Für  $k = 1$  wähle  $n = 1 + m \cdot (\ell - 1)$ . Wenn  $f : n \rightarrow m$  und alle  $|f^{-1}(i)| \leq \ell - 1$  dann ist  $n \leq m \cdot (\ell - 1)$ , was unmöglich ist. Also gibt es ein  $i < m$  mit  $|f^{-1}(i)| \geq \ell$ .

Wir nehmen als Induktionshypothese an, daß  $\forall m, \ell \in \mathbb{N}. \exists n \in \mathbb{N}. n \longrightarrow (\ell)_m^k$ . Seien  $m, \ell \in \mathbb{N}$ . Aufgrund der Induktionshypothese gibt es ein  $t \in \mathbb{N}$ , sodaß  $t \longrightarrow (\ell)_m^k$ . Wir setzen  $n = 2m^c$ , wobei  $c = \sum_{i=k}^{t-1} \binom{i}{k}$ . Sei  $f : [n]^{k+1} \rightarrow m$ .

Wähle paarweise verschiedene  $a_1, \dots, a_k \in n$  und setze  $S_k = n \setminus \{a_1, \dots, a_k\}$ . Für  $i$  mit  $k < i \leq t$  definieren wir  $a_i$  und  $S_i$  folgendermaßen:

- wähle  $a_i \in S_{i-1}$  beliebig
- auf  $S_{i-1} \setminus \{a_i\}$  betrachte die Äquivalenzrelation  $x \equiv y$  genau dann, wenn für alle  $T \subseteq \{a_1, \dots, a_i\}$  mit  $|T| = k$  gilt  $f(T \cup \{x\}) = f(T \cup \{y\})$ ; wähle als  $S_i$  eine größtmögliche Äquivalenzklasse.

Es gibt höchstens  $m \binom{i}{k}$  Äquivalenzklassen. Also gilt  $|S_i| \geq (|S_{i-1}| - 1) \cdot m - \binom{i}{k}$ . Die Größe von  $n$  stellt sicher, daß wir  $a_1, \dots, a_t \in n$  und  $S_k, S_{k+1}, \dots, S_t$  geeignet wählen können. Angenommen  $i_1 < \dots < i_k < i_{k+1} \leq t$ . Dann gilt  $a_{i_{k+1}} \in S_{i_{k+1}-1} \subseteq S_{i_k}$  und somit gilt  $f(\{a_{i_1}, \dots, a_{i_k}, a_{i_{k+1}}\}) = f(\{a_{i_1}, \dots, a_{i_k}, x\})$  für alle  $x \in S_{i_k}$ . Wir betrachten eine  $m$ -Färbung  $f^*$  der  $k$ -elementigen Teilmengen von  $\{a_1, \dots, a_t\}$  mit

$$f^*(\{a_{i_1}, \dots, a_{i_k}\}) = f\{a_{i_1}, \dots, a_{i_k}, a_{i_{k+1}}\}$$

sofern  $i_k < t$  und beliebig sonst. Aufgrund der Induktionsannahme gibt es eine Teilfolge  $b_1, \dots, b_\ell$  von  $a_1, \dots, a_t$ , sodaß  $f^*$  auf  $[\{b_1, \dots, b_\ell\}]^k$  konstant ist. Sie  $r$  dieser konstante Wert. Für  $j_1 < \dots < j_k < j_{k+1} \leq \ell$  gilt dann

$$f(\{b_{j_1}, \dots, b_{j_k}, b_{j_{k+1}}\}) = f^*(\{b_{j_1}, \dots, b_{j_k}\}) = r$$

Also ist  $f$  auf  $[\{b_1, \dots, b_\ell\}]^{k+1}$  konstant. □

Mit etwas Codierung läßt sich dieser Satz in PA formulieren und auch beweisen, auch wenn  $n$  in Abhängigkeit von den Parametern  $k, m$  und  $\ell$  sehr schnell wächst. Um ein so großes Wachstum zu erzwingen, daß die Termination nicht mehr in PA bewiesen werden kann, muß man die Aussage von Satz B.1 geeignet verschärfen.

**Definition B.1**

Eine endliche Teilmenge  $X$  von  $\mathbb{N}$  heißt relativ groß (engl. “large”), wenn  $|M| \geq \min(M)$ . Wir schreiben

$$n \xrightarrow{*} (\ell)_m^k$$

als Abkürzung für die Aussage, daß für jede Färbung  $f : [n]^k \rightarrow m$  eine relativ große mindestens  $\ell$ -elementige homogene Teilmenge  $X$  von  $n$  existiert.

**Satz B.2** (J. Paris, L. Harrington)

Die Aussage

$$(PH) \quad \forall k, m, \ell \in \mathbb{N}. \exists n \in \mathbb{N}. n \xrightarrow{*} (\ell)_m^k$$

ist wahr, kann aber nicht in PA bewiesen werden.

“Wahr” heißt im vorliegenden Fall, daß die Aussage in der Arithmetik 2. Stufe bewiesen werden kann. Wir formulieren den Beweis von PH jedoch informell. Dazu benötigen wir die beiden folgenden Tatsachen aus der (unendlichen) Kombinatorik.

**Satz B.3** (Königs Lemma)

In jedem unendlichen, aber endlich verzweigende Baum gibt es einen unendlichen Pfad.

In [GRS] z.B. findet man einen Beweis des folgenden Satzes.

**Satz B.4** (unendlicher Ramsey Satz)

Wenn  $k$  und  $m$  natürliche Zahlen sind und  $f : [\mathbb{N}]^k \rightarrow m$ , dann gibt es eine unendliche Teilmenge  $X$  von  $\mathbb{N}$ , sodaß  $f$  eingeschränkt auf  $[X]^k$  konstant ist.

Wir weisen darauf hin, daß bereits die Formulierung der Sätze B.3 und B.4 die Bezugnahme auf Teilmengen von  $\mathbb{N}$  erfordert, d.h. die Sprache der Arithmetik 2. Stufe.

**Lemma B.1** *Es gilt PH.*

*Beweis:* Wir argumentieren mit Widerspruch, nehmen also an, es gebe natürliche Zahlen  $k, m, \ell$ , sodaß für jede natürliche Zahl  $n$  eine Funktion  $f : [n]^k \rightarrow m$  existiert, sodaß  $f$  auf jeder relativ großen mindestens  $\ell$ -elementigen Teilmenge von  $n$  verschiedene Werte annimmt. Die Menge dieser Gegenbeispiele ist bzgl.  $\subseteq$  ein endlich verzweigender unendlicher Baum. Wegen Satz B.3 existiert in diesem Baum ein unendlicher Pfad, d.h. eine Funktion  $f : [\mathbb{N}]^k \rightarrow m$ , sodaß für jedes  $n \in \mathbb{N}$  die Funktion  $f_n$ , die sich als Einschränkung von  $f$  auf  $[n]^k$  ergibt, ein Gegenbeispiel ist. Wegen Satz B.4 gibt es eine unendliche Teilmenge  $H$  von  $\mathbb{N}$ , sodaß die Einschränkung von  $f$  auf  $[H]^k$  konstant ist. Wähle eine natürliche Zahl  $t \geq \ell, \min(H)$ . Seien  $n_1 < \dots < n_t$  die  $t$  kleinsten Elemente von  $H$ . Wähle  $n > n_t$  und setze  $X = [n] \cap H$ . Es gilt nun  $|X| \geq t \geq \ell, \min(H)$ . Also hat  $X$  mindestens  $\ell$  Elemente und, da  $\min(H) = \min(X)$ , ist  $X$  auch relativ groß. Da aber  $f$  bereits auf  $[H]^k$  konstant ist, ist  $f$  umso mehr auf  $[X]^k$  konstant. Somit ist  $f_n$  kein Gegenbeispiel im Widerspruch zur Annahme.  $\square$

Um zu zeigen, daß PH in der PA nicht herleitbar ist, gehen wir folgendermaßen vor. Wir zeigen, daß aus PH ein kombinatorisches Prinzip (\*) folgt, das in einem geeignet konstruierten (Nichtstandard-)Modell der PA nicht gilt. Um (\*) zu formulieren, benötigen wir noch etwas Notation.

**Definition B.2** Sei  $X \subseteq \mathbb{N}$ . Eine Funktion  $f : [X]^k \rightarrow \mathbb{N}$  heißt regressiv, wenn  $f(A) < \min(A)$  für alle  $A \in [X]^k$ . Eine Teilmenge  $Y \subseteq X$  heißt min-homogen, wenn für alle  $A, B \in [Y]^k$  mit  $\min(A) = \min(B)$  bereits  $f(A) = f(B)$  gilt.

Für  $a, b \in \mathbb{N}$  schreiben wir  $(a, b)$  als Abkürzung für  $\{x \in \mathbb{N} \mid a < x < b\}$ .

**Lemma B.2** *In der PA kann man aus PH folgende Aussage herleiten.*

(\*) für alle  $c, k, \ell, n \in \mathbb{N}$  existiert ein  $d \in \mathbb{N}$ , sodaß für beliebige regressive  $f_1, \dots, f_n : [d]^k \rightarrow d$  ein  $Y \subseteq (c, d)$  existiert, sodaß  $|Y| \geq \ell$  und  $Y$  bzgl. aller  $f_i$  min-homogen ist

*Beweis:* Ist rein kombinatorisch (und etwas lästig) und somit in PA formalisierbar. Siehe Lemma 5.4.4 in [Mar].  $\square$

Um zu zeigen, daß (\*) in der PA nicht herleitbar ist, ist es nützlich folgenden Begriff aus der Modelltheorie heranzuziehen.

**Definition B.3**

Sei  $\mathcal{M}$  ein Modell der PA und  $\Gamma$  eine endliche Menge von Formeln in der Sprache der PA. Eine Teilmenge  $I$  von  $|\mathcal{M}|$  heißt sequence of diagonal indiscernibles (sdi) für  $\Gamma$ , wenn für jede Formel  $\phi(u_1, \dots, u_k, v_1, \dots, v_m)$  in  $\Gamma$  und  $a, x_1, \dots, x_n, y_1, \dots, y_n \in I$  mit  $a < x_1 < \dots < x_n$  und  $a < y_1 < \dots < y_n$  die Formel

$$\forall z_1, \dots, z_k < a (\phi(\vec{z}, \vec{x}) \leftrightarrow \phi(\vec{z}, \vec{y}))$$

in  $\mathcal{M}$  gilt.

Als nächstes beweist man unter Verwendung von (\*) folgende Eigenschaft des Standardmodells.

**Lemma B.3** Für  $c, \ell, m, n \in \mathbb{N}$  und Formeln  $\phi_1(u_1, \dots, u_k, v_1, \dots, v_n), \dots, \phi_\ell(u_1, \dots, u_k, v_1, \dots, v_n)$  gibt es eine endliche Menge  $I$  mit  $|I| \geq m$  und  $\min(I) > c$ , sodaß für  $a, x_1, \dots, x_n, y_1, \dots, y_n \in I$  mit  $a < x_1 < \dots < x_n$  und  $a < y_1 < \dots < y_n$  für alle  $i = 1, \dots, \ell$  die Formel

$$\forall z_1, \dots, z_k < a (\phi(\vec{z}, \vec{x}) \leftrightarrow \phi(\vec{z}, \vec{y}))$$

im Standardmodell  $\mathbb{N}$  gilt.

Angenommen  $m > 2n$ . Sei  $w$  die kleinste Zahl mit  $w \rightarrow (m+n)_{\ell+1}^{2n+1}$  und  $d$  die kleinste Zahl, sodaß ein  $Y \subseteq (c, d)$  existiert mit  $|Y| \geq w$  und  $Y$  min-homogen ist für alle  $f_j$ . Dann kann das obige  $I$  so gewählt werden, daß  $I \subseteq (c, d)$ .

*Beweis:* Siehe Beweis von Lemma 5.4.5 in [Mar].  $\square$

Sei  $\Delta_0$  die Menge der arithmetischen Formeln, wo alle Quantoren beschränkt sind. Man kann leicht zeigen, daß es ein arithmetisch definierbares Wahrheitsprädikat für geschlossene  $\Delta_0$  Formeln gibt und für solche Formeln Beweisbarkeit und Gültigkeit im Standardmodell koinzidieren. Wenn in Lemma B.3

die  $\phi_i$  alle  $\Delta_0$  Formeln sind, gilt deshalb dann für alle  $i = 1, \dots, \ell$ , daß

$$\forall z_1, \dots, z_k < a (\phi(\vec{z}, \vec{x}) \leftrightarrow \phi(\vec{z}, \vec{y}))$$

in der PA beweisbar ist. In diesem Fall läßt sich die Aussage auch – nach entsprechender Gödelisierung – in der PA beweisen.

**Lemma B.4** *Sei  $\mathcal{M}$  ein Modell der PA und  $\{c_n \mid n \in \mathbb{N}\}$  eine s.d.i. für die Menge der (geschlossenen)  $\Delta_0$  Sätze, wobei  $d_i < d_{i+1}$  für alle  $i \in \mathbb{N}$ . Dann ist  $|\mathcal{N}| = \{a \in |\mathcal{M}| \mid a < c_i \text{ for some } i\}$  unter Addition und Multiplikation abgeschlossen und die Unterstruktur  $\mathcal{N}$  von  $\mathcal{M}$  ist ein Modell der PA.*

*Beweis:* Siehe Beweis von Lemma 5.4.8 in [Mar]. □

Wir haben nun alles vorbereitet für den

*Beweis von Satz B.2 :*

Wir wissen bereits aus Lemma B.1, daß PH wahr ist. Um zu zeigen, daß PH in der PA nicht herleitbar ist, konstruieren wir ein Nichtstandardmodell der PA in dem (\*) nicht gilt und somit auch nicht PH.

Sei  $\mathcal{M}$  ein Nichtstandardmodell der PA, in dem (\*) gilt (etwa ein Nichtstandardmodell aller wahren arithmetischen Sätze). Sei  $c$  ein Nichtstandardelement von  $\mathcal{M}$ .

Da die PA den endlichen Ramsey Satz beweist, gibt es ein kleinstes  $w$  mit  $w \longrightarrow (3c + 1)_c^{2c+1}$ . Aufgrund der Annahme, daß (\*) in  $\mathcal{M}$  gilt, gibt es ein kleinstes  $d$ , sodaß für beliebige regressive  $f_1, \dots, f_c : [d]^{2c+1} \rightarrow d$  ein  $Y \subseteq (c, d)$  existiert mit  $|Y| \geq w$  und  $Y$  min-homogen bzgl. aller  $f_i$ .

Aus Lemma B.3 und nachfolgender Bemerkung folgt die Existenz einer endlichen Teilmenge  $I \subset (c, d)$  (im Sinne von  $\mathcal{M}$ !) mit  $|I| \geq c$ , sodaß  $I$  eine s.d.i. ist für alle  $\Delta_0$  Formeln  $\phi(u_1, \dots, u_c, v_1, \dots, v_c)$  mit Gödelnummer  $< c$ . Somit ist  $I$  insbesondere eine s.d.i. für alle standard  $\Delta_0$  Formeln.

Da  $|I| \geq c$  und  $c$  eine Nichtstandardzahl ist, ist  $I$  unendlich. Sei nun  $x_0 < x_1 < \dots < x_n < x_{n+1} < \dots$  eine aufsteigende Kette in  $I$ . Da s.d.i.s unter Teilmengen abgeschlossen sind, ist auch  $\{x_n \mid n \in \mathbb{N}\}$  eine s.d.i. für alle  $\Delta_0$  Formeln. Sei  $|\mathcal{N}| = \{y \in |\mathcal{M}| \mid y < x_i \text{ for some } i\}$ . Wegen Lemma B.4 ist die Unterstruktur  $\mathcal{N}$  mit Trägermenge  $|\mathcal{N}|$  ein Modell der PA.

Offensichtlich ist  $c \in |\mathcal{N}|$ . Da  $\mathcal{N}$  ein Modell der PA ist und man Satz B.1 in der PA beweisen kann, gibt es ein  $w'$  in  $\mathcal{N}$ , sodaß  $w' \longrightarrow (3c + 1)_c^{2c+1}$ . Da  $\mathcal{N}$  ein Anfangsstück von  $\mathcal{M}$  ist und  $w' \longrightarrow (3c + 1)_c^{2c+1}$  eine  $\Delta_0$  Formel ist,

gilt  $w' \rightarrow (3c+1)_c^{2c+1}$  eine  $\Delta_0$  auch in  $\mathcal{M}$ . Da  $w$  minimal gewählt war, folgt nun  $w \leq w'$  und somit  $w$  in  $\mathcal{N}$ .

Nehmen wir an, es gebe ein  $d'$  in  $\mathcal{N}$ , sodaß in  $\mathcal{N}$  gilt, daß für alle regressiven  $f_1, \dots, f_c : [d']^{2c+1} \rightarrow d'$  ein  $Y \subseteq (c, d')$  existiert mit  $|Y| \geq w$  und  $Y$  min-homogen bzgl. aller  $f_i$ . Mit einer ähnlichen Argumentation wie vorher zeigt man, daß die Aussage auch in  $\mathcal{M}$  gilt und somit  $d \leq d'$ , da  $d'$  minimal gewählt war. Dann wäre aber  $d$  in  $\mathcal{N}$ . Dies steht aber im Widerspruch dazu, daß  $d$  echt größer ist als alle Elemente von  $\mathcal{N}$ . Wir haben somit gezeigt, daß (eine Instanz) von (\*) in  $\mathcal{N}$  nicht gilt.