Unvollständigkeit formaler Systeme

Prof. Dr. Thomas Streicher

WS 2014/15

Inhaltsverzeichnis

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 Unvollständigleitssatz besagt hingegen, daß es einen arithmetischen Satz G gibt, der im "Standardmodell" $\mathbb{N}$ wahr aber in PA nicht herleitbar ist.[1]
Man kann die Aussage des Gödelschen Unvollständigkeitssatzes relativ einfach aus der Unentscheidbarkeit des Halteproblems herleiten, wie wir sogleich sehen werden.
Sei $\mathcal{L}_{\rm pr}$ 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 $\mathrm{Ax}_{\rm pr}$ die Menge aller definierenden Gleichungen der primitiv rekursiven Funktionsdefinitionen. Offensichtlich gilt für jede k–stellige primitiv rekursive Funktion f, daß $f(\vec{n}) = m$ genau dann, wenn $\mathrm{Ax}_{\rm pr} \vdash f(\underline{\vec{n}}) = \underline{m}$ [2], da die Axiome in $\mathrm{Ax}_{\rm pr}$ 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 $\mathrm{Ax}_{\rm pr}$ herleitbar ist. Aus diesem Grund ist es unentscheidbar, ob eine Formel der Gestalt $\exists\vec{x}.f(\vec{x}) = \underline{0}$ in $\mathrm{Ax}_{\rm pr}$ 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 $\mathrm{Ax}_{\rm pr}$ herleitbar ist. Diese Reduktion auf das Halteproblem ist auch möglich für *r.e. konsistente* Erweiterungen von $\mathrm{Ax}_{\rm pr}$, die das Axiom $0 \neq succ(x)$ enthalten.

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

[1] In der Mathematik geht es um die Gültigkeit im Standardmodell $\mathbb{N}$ und nicht um die Gültigkeit in allen (auch den Nichtstandard–)Modellen.

[2] Zur Erinnerung: $\underline{n} \equiv succ^n(0)$.

ist.

Beweis: Angenommen $\mathcal{T}$ wäre eine r.e. konsistente Erweiterung von $\mathrm{Ax}_{\mathrm{pr}}$, 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 $\mathrm{Ax}_{\mathrm{pr}}$ und somit aus $\mathcal{T}$ herleitbar wäre im Widerspruch zur Konsistenz von $\mathcal{T}$.

Dieser Beweis der Existenz eines $n \in \complement K$ mit $\mathcal{T} \nvdash \forall k.\neg T(\underline{n}, \underline{n}, k)$ 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 $\mathcal{T} \nvdash \forall k.\neg T(\underline{e}, \underline{e}, k)$, 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* $\mathrm{Ax}_{\mathrm{pr}}$ *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 $\mathrm{Ax}_{\mathrm{pr}}$ 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 $\mathrm{Ax}_{\mathrm{pr}}$, 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(?) lavieren (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" $\mathcal{S}_{\text{real}}$ [3] durch ein "Konservativitätsresultat" abzusichern, d.h., daß

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

für alle "elementaren Aussagen" A, welch 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[4] werden, da es für alle (auch die Zweifler) einsehbar und akzeptabel sein sollte. Zu diesem Zweck muß man aber die Aussagen von $\mathcal{S}_{\text{ideal}}$

[3] Im wesentlichen Ax_{pr} zusammen mit dem Axiom $succ(0) \neq 0$ und dem Induktionschema für *quantorenfreie* Prädikate, was auch unter dem Namen PRA (Primitiv Rekursive Arithmetik) bekannt ist.

[4] 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 $\mathit{Bew}(n, \ulcorner A \urcorner)$ 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

$$\mathit{Pr}\big(\ulcorner A \urcorner\big) \equiv \exists n.\, \mathit{Bew}(n, \ulcorner A \urcorner) \ .$$

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

$$(\mathrm{Refl}_A) \qquad \mathrm{Pr}\big(\ulcorner A \urcorner\big) \to A$$

bzw. die logisch äquivalente Aussage

$$\mathit{Bew}(n, \ulcorner A \urcorner) \to A$$

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

$$\mathcal{S}_{\text{real}} \vdash \neg\, \mathrm{Pr}\big(\ulcorner \bot \urcorner\big)$$

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 $\neg \mathit{Pr}\big(\ulcorner \bot \urcorner\big)$ 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 $\mathcal{S}_{\text{real}} \vdash \mathrm{Pr}\big(\ulcorner A \urcorner\big)$.

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

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

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 $\mathcal{S}_{\mathrm{real}}$. Letzteres geschieht durch die Angabe einer primitiv rekursiven Funktion deriv_f, für die man in $\mathcal{S}_{\mathrm{real}}$ beweisen kann, daß

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

woraus unmittelbar die $\mathcal{S}_{\mathrm{real}}$–Herleitbarkeit von

$$f(\vec{x}) = 0 \to \mathit{Pr}\big(\ulcorner f(\vec{x}) = 0\urcorner\big)$$

und somit die $\mathcal{S}_{\mathrm{real}}$–Herleitbarkeit von

$$f(\vec{x}) = 0 \to \mathit{Pr}\big(\ulcorner \exists \vec{x}.\, f(\vec{x}) = 0\urcorner\big)$$

folgt.

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

Beweis: Es gilt

(1) $\mathcal{S}_{\mathrm{real}} \vdash \exists n.\, f(n) = 0 \to \mathit{Pr}\big(\ulcorner\exists n.\, f(n) = 0\urcorner\big)$

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

(2) $\mathcal{S}_{\mathrm{real}} \vdash \mathit{Pr}\big(\ulcorner A\urcorner\big) \to \mathit{Pr}\big(\ulcorner\exists n.\, f(n) = 0\urcorner\big) \to \mathit{Pr}\big(\ulcorner\bot\urcorner\big)$.

Aufgrund der Annahme $\mathcal{S}_{\mathrm{real}} \vdash \neg\mathit{Pr}\big(\ulcorner\bot\urcorner\big)$ folgt aus (2), daß

(3) $\mathcal{S}_{\mathrm{real}} \vdash \mathit{Pr}\big(\ulcorner A\urcorner\big) \to \mathit{Pr}\big(\ulcorner\exists n.\, f(n) = 0\urcorner\big) \to \bot$.

Wegen (1) folgt dann

(4) $\mathcal{S}_{\mathrm{real}} \vdash \mathrm{Pr}\big(\ulcorner A\urcorner\big) \to \exists n.\, f(n) = 0 \to \bot$

und somit, daß

(5) $\mathcal{S}_{\mathrm{real}} \vdash \mathrm{Pr}\big(\ulcorner A\urcorner\big) \to A$

wie behauptet. $\square$

Aus diesem Hilbert bereits bekannten Grund lief sein Programm darauf hinaus, in $\mathcal{S}_{\text{real}}$ die Aussage $\neg Pr(\ulcorner\bot\urcorner)$ 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.[5]
Da

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

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

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

und somit ist $\forall n.\, \neg Bew(n, \ulcorner\bot\urcorner)$ ein elementarer Satz im Sinne Hilberts, der zwar wahr, aber in $\mathcal{S}_{\text{real}}$ nicht beweisbar ist. Schlimmer noch läßt sich die Aussage $\forall n.\, \neg Bew(n, \ulcorner\bot\urcorner)$ 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 $\forall n.\, \neg Bew(n, \bot)$ in $\mathcal{S}_{\text{real}}$ beweisbar äquivalent zur Divergenz des Programms, das nach der Herleitung von $\bot$ 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 $\neg Pr(\ulcorner\bot\urcorner)$ denn doch “philosophisch” ansprechender.

3 PRA, PA, Σ_1 und Π_1

In diesem Abschnitt fixieren wir die Definition zweier formaler Systeme der Arithmetik, nämlich PRA und PA, und definieren die Formelklassen Σ_1 und Π_1.

[5] 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 “finiter 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 ε_0 für quantorenfreie Prädikate.

Definition 3.1 *Die Sprache (1. Stufe) $\mathcal{L}_{\mathrm{pr}}$ besteht aus Funktionssymbolen für jede primitiv rekursive Funktionsdefinition (inklusive der Konstante* 0 *und der Nachfolgeroperation succ) und dem* 2*-stelligen Gleichheitsprädikat* $=$.
Für eine Formel A in der Sprache $\mathcal{L}_{\mathrm{pr}}$ und eine Variable x bezeichnen wir mit $\mathrm{Ind}_{A,x}$ *bzw.* $\mathrm{Ind}_{A(x)}$ *die Formel*

$$A(0) \to \forall x.(A(x) \to A(succ(x))) \to \forall x.A(x)$$

genannt Induktionsprinzip für $A(x)$.
Die PRA (Primitiv Rekursive Arithmetik) *besteht aus folgenden Axiomen*

(1) *den Gleichheitsaxiomen und* $\mathrm{Ax}_{\mathrm{pr}}$*, der Menge aller definierenden Gleichungen für primitive rekursive Funktionen*

(2) $\mathrm{Ind}_{A,x}$ *für alle* **quantorenfreien** *Formeln A und*

(3) $\neg\, 0 = succ(0)$.

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

Bemerkung.
(1) Ohne dem Axiom $\neg\, 0 = succ(0)$ hätte PA und somit auch PRA ein triviales Modell (dessen unterliegende Menge aus bloß einem Objekt besteht).
(2) Obzwar 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

$$0 \neq succ(x)$$

$$x = 0 \lor x = succ(pred(x))$$

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

$$x = y \leftrightarrow eq(x, y) = 0$$

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

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

Definition 3.2 (Σ_1 und Π_1 Formeln)
Eine Σ_1 Formel ist eine Formel der Gestalt $\exists\vec{x}.A$, wobei A eine quantorenfreie Formel der Sprache $\mathcal{L}_{\mathrm{pr}}$ ist.
Eine Π_1 Formel ist eine Formel der Gestalt $\forall\vec{x}.A$, wobei A eine quantorenfreie Formel der Sprache $\mathcal{L}_{\mathrm{pr}}$ ist. $\diamondsuit$

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.[6]
Also ist jede Σ_1 Formel PRA–beweisbar äquivalent zu einer Formel der Gestalt $\exists\vec{x}.f(\vec{x},\vec{y}) = 0$ und jede Π_1 Formel PRA–beweisbar äquivalent zu einer Formel der Gestalt $\forall\vec{x}.f(\vec{x},\vec{y}) = 0$.
Da

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

in der PRA herleitbar sind und quantorenfreie Formeln unter Negation abgeschlossen sind, ist die Negation einer Σ_1 bzw. Π_1 Formel PRA–beweisbar äquivalent zu einer Π_1 bzw. Σ_1 Formel. Aus der Rekursionstheorie wissen wir bereits, daß die durch Σ_1 Formeln definierbaren Prädikate genau den r.e. Mengen entsprechen. Somit entsprechen die Π_1–definierbaren Prädikate genau den co-r.e. Mengen. Naheliegenderweise nennen wir die durch Σ_1 bzw. Π_1 Formeln definierbaren Prädikate bzw. Mengen auch Σ_1 bzw. Π_1 Prädikate bzw. Mengen. Mit Δ_1 wird gemeinhin der Schnitt von Σ_1 und Π_1 bezeichnet, d.h. die Prädikate bzw. Mengen, die sowohl durch eine Σ_1 als auch durch eine Π_1 Formel ausgedrückt werden können. Also entsprechen die Δ_1–Mengen gerade den entscheidbaren Mengen.
Aus Abschnitt 1 wissen wir, daß

- PRA genau die wahren Σ_1 Sätze beweist
- PRA nur wahre, aber nicht alle wahren Π_1–Sätze beweist.[7]

[6] Insbesondere entsprechen deshalb die Π_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 $\forall\vec{x}.\,eq(f(\vec{x}), g(\vec{x})) = 0$.

[7] 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 "Glaubengewissheit". Man kann zwar z.B. in ZFC *beweisen*, daß $\mathbb{N} \models$ 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 $\mathcal{L}_{\mathrm{pr}}$ heißt* 1–konsistent *genau dann, wenn $\mathcal{T}$ ausschließlich wahre Σ_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*

$$\mathrm{PRA} \vdash A \leftrightarrow P\left(\ulcorner A \urcorner\right) \quad .$$

Beweis: Sei Θ 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

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

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

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

und somit

$$\mathrm{PRA} \vdash A \leftrightarrow P\left(\ulcorner A \urcorner\right)$$

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 $P\left(\ulcorner A \urcorner\right)$ 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 Nichstandardmodell $\mathcal{N}$ 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*

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

wobei $Pr_{\mathcal{T}}$ die Beweisbarkeit in $\mathcal{T}$ in der Sprache der (prim. rek.) Arithmetik ausdrückt.
Für G gelten nun folgende Eigenschaften

(1) $\mathcal{T} \nvdash G$

(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* Π_1 *Satz* $\forall n. \neg Bew_{\mathcal{T}}(n, \ulcorner G \urcorner)$.

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 $\vdash G \leftrightarrow \neg Pr_{\mathcal{T}}\big(\ulcorner G \urcorner\big)$, der aufgrund von Satz 4.1 auch tatsächlich existiert.
(1) Angenommen $\mathcal{T} \vdash G$, dann gilt wegen (D1), daß PRA $\vdash Pr_{\mathcal{T}}\big(\ulcorner G \urcorner\big)$ und somit auch $\mathcal{T} \vdash Pr_{\mathcal{T}}\big(\ulcorner G \urcorner\big)$. Wegen $\mathcal{T} \vdash G$ gilt aber auch $\mathcal{T} \vdash \neg Pr_{\mathcal{T}}\big(\ulcorner G \urcorner\big)$ und somit $\mathcal{T} \vdash \bot$ im Widerspruch zur Konsistenz von $\mathcal{T}$.
(2) Angenommen $\mathbb{N} \not\models G$. Dann $\mathbb{N} \models \neg G$, d.h. $\mathbb{N} \models \exists n.\, Bew_{\mathcal{T}}(n, \ulcorner G \urcorner)$, und somit $\mathcal{T} \vdash G$ im Widerspruch zu (1).
(3) Angenommen $\mathcal{T} \vdash \neg G$ und $\mathcal{T}$ sei 1–konsistent. Dann $\mathcal{T} \vdash \exists n.\, Bew_{\mathcal{T}}(n, \ulcorner G \urcorner)$ und somit $\mathbb{N} \models \exists n.\, Bew_{\mathcal{T}}(n, \ulcorner G \urcorner)$ aufgrund der 1–Konsistenz von $\mathcal{T}$. Somit gilt aber $\mathcal{T} \vdash G$ im Widerspruch zur Konsistenz von $\mathcal{T}$.
Da PRA $\vdash \neg Pr_{\mathcal{T}}\big(\ulcorner G \urcorner\big) \leftrightarrow \forall n. \neg Bew_{\mathcal{T}}(n, \ulcorner G \urcorner)$ gilt auch

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

wie behauptet. □

Für jede r.e. konsistente Erweiterung $\mathcal{T}$ der PRA kann man aber auch einen Gödelsatz basteln, ohne die Syntax von $\mathcal{T}$ zu gödelisieren und ohne den arithmetischen Fixpunktsatz 4.1 zu verwenden. Das Argument ist wohl altbekannt, aber in der Literatur nicht leicht zu finden, weswegen wir es hier ausformulieren.[8]
Für jede natürliche Zahl n sei

$$W_n = \{m \in \mathbb{N} \mid \{n\}(m)\!\downarrow\}$$

d.h. $m \in W_n$ gdw $\exists k. T(n, m, k)$. Bekanntlich stellt dies eine effektive Numerierung der r.e. Mengen natürlicher Zahlen bereit. Sei nun $W_e = \{n \in \mathbb{N} \mid \mathcal{T} \vdash n \notin W_n\}$. Angenommen $e \in W_e$. Dann $\mathcal{T} \vdash e \in W_e$ und somit

[8]Wir folgen hier der Darstellung in [Sti].

$\mathcal{T} \not\vdash e \in W_e$, da ja $\mathcal{T}$ als konsistent angenommen wurde. Also ist $e \notin W_e$ wahr und somit $\mathcal{T} \not\vdash e \notin W_e$. Also ist $e \notin W_e$ ein Gödelsatz. Wenn $\mathcal{T}$ überdies 1-konsistent ist, gilt auch $\mathcal{T} \not\vdash e \in W_e$.
Ein Vorteil des Gödelsatzes G aus Satz 5.1 ist aber, daß er in der PRA beweisbar äquivalent ist zu $\mathrm{Con}_\mathcal{T} \equiv \neg Pr_\mathcal{T}(\ulcorner\bot\urcorner)$, d.h. zur formalisierten Aussage der Konsistenz von $\mathcal{T}$, wie wir in Satz 6.1 beweisen werden. Daraus folgt dann sehr schnell der 2. Gödelsche Unvollständigkeitssatz.

Konvention Aus Gründen der Lesbarkeit schreiben wir in Zukunft

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

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

(D1) $\vdash A \quad \Rightarrow \quad \mathrm{PRA} \vdash \Box A$

(D2) $\mathrm{PRA} \vdash \Box(A \to B) \to \Box A \to \Box B$

(D3) $\mathrm{PRA} \vdash \Box A \to \Box\Box A$

wobei bemerkt sei, daß man (D3) auch formulieren könnte als $B \to \Box B$ für alle Σ_1 Sätze B. Im weiteren ist jedoch bloß der Fall $B \equiv \Box A \equiv \mathrm{Pr}(\ulcorner A\urcorner) \equiv \exists n.Bew(n, \ulcorner A\urcorner)$ interessant.

6 Der 2. Gödelsche Unvollständigkeitssatz

Wir beweisen zuerst, daß jeder Gödelsatz G für eine Theorie $\mathcal{T}$ schon in der PRA beweisbar äquivalent zu $\mathrm{Con}_\mathcal{T} \equiv \neg Pr_\mathcal{T}(\ulcorner\bot\urcorner)$ ist.

Satz 6.1 *Sei $\mathcal{T}$ eine r.e. konsistente Erweiterung der* PRA *und G ein Gödelsatz für $\mathcal{T}$, d.h.* $\mathrm{PRA} \vdash G \leftrightarrow \neg Pr_\mathcal{T}(\ulcorner G\urcorner)$, *dann gilt*

$$\mathrm{PRA} \vdash G \leftrightarrow \neg Pr_\mathcal{T}(\ulcorner\bot\urcorner) \quad .$$

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

$$\vdash Pr(\ulcorner G\urcorner) \leftrightarrow Pr(\ulcorner\bot\urcorner) \quad .$$

Da $\vdash \bot \to G$, gilt wegen (D1), daß $\vdash \Box(\bot \to G)$ und somit wegen (D2), daß $\vdash \Box\bot \to \Box G$.
Weil $\vdash G \to \neg\Box G$, gilt wegen (D1), daß $\vdash \Box(G \to \neg\Box G)$, und somit wegen (D2), daß

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

Wegen (D2) gilt aber

(2) $\vdash \Box\neg\Box G \to \Box\Box G \to \Box\bot$

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

(3) $\Box G \to \Box\Box G \to \Box\bot$

und somit das erwünschte

$$\vdash \Box G \to \Box\bot$$

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

Aus den beiden Sätzen 5.1 und 6.1 folgt nun für jede konsistente r.e. Erweiterungen $\mathcal{T}$ der PRA, daß

$$\mathcal{T} \nvdash \mathrm{Con}_{\mathcal{T}}$$

obwohl ja $\mathbb{N} \models \mathrm{Con}_{\mathcal{T}}$, da $\mathcal{T}$ als konsistent vorausgesetzt wurde. Also ist $\mathcal{T} + \neg\mathrm{Con}_{\mathcal{T}}$ immer konsistent, allerdings **nicht** 1–konsistent.
Wenn $\mathcal{T}$ überdies 1–konsistent ist, dann gilt auch

$$\mathcal{T} \nvdash \neg\mathrm{Con}_{\mathcal{T}}$$

und somit ist in diesem Fall auch $\mathcal{T} + \mathrm{Con}_{\mathcal{T}}$ konsistent[9].
Oft wird die Aussage, daß $\mathcal{T} \nvdash \mathrm{Con}_{\mathcal{T}}$ für konsistente r.e. Erweiterungen $\mathcal{T}$ der PRA, als 2. Gödelscher Unvollständigkeitssatz bezeichnet.[10] Dies läßt sich jedoch wie folgt internaliseren.

[9]Dies ist nicht trivial, da ja nicht notwendigerweise $\mathbb{N} \models \mathcal{T}$! Zum Beispiel ist ja aufgrund obiger Betrachtungen die Theorie PRA + $Pr_{\mathrm{PRA}}(\ulcorner\bot\urcorner)$ konsistent. Somit ist aber auch in jeder r.e. konsistenten Erweiterung $\mathcal{T}$ von PRA + $Pr_{\mathrm{PRA}}(\ulcorner\bot\urcorner)$ die Aussage $Pr_{\mathcal{T}}(\ulcorner\bot\urcorner)$, d.h. $\neg\mathrm{Con}_{\mathcal{T}}$ herleitbar, da ja ganz offensichtlich für beliebige Formeln A die Aussage $\mathrm{Pr}_{\mathrm{PRA}}(\ulcorner A\urcorner) \to \mathrm{Pr}_{\mathcal{T}}(\ulcorner A\urcorner)$ bereits in der PRA herleitbar ist.

[10]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.

Satz 6.2 (2. Gödelscher Unvollständigkeitsatz)
Für jede konsistente r.e. Erweiterung $\mathcal{T}$ *der* PRA *gilt* PRA $\vdash Con_{\mathcal{T}} \leftrightarrow \neg\Box Con_{\mathcal{T}}$.

Beweis: Aufgrund von Satz 6.1 beweist bereits die PRA, daß jeder Gödelsatz G für $\mathcal{T}$ zu $Con_{\mathcal{T}} \equiv \neg\,\mathrm{Pr}_{\mathcal{T}}(\ulcorner\bot\urcorner)$ äquivalent ist. Daraus folgt dann mit (D1) und (D2), daß PRA $\vdash \Box_{\mathcal{T}} G \leftrightarrow \Box_{\mathcal{T}} Con_{\mathcal{T}}$ und somit PRA $\vdash \neg\Box_{\mathcal{T}} G \leftrightarrow \neg\Box_{\mathcal{T}} Con_{\mathcal{T}}$. Da PRA $\vdash G \leftrightarrow \neg\Box_{\mathcal{T}} G$, gilt auch PRA $\vdash G \leftrightarrow \neg\Box_{\mathcal{T}} Con_{\mathcal{T}}$ und somit PRA $\vdash Con_{\mathcal{T}} \leftrightarrow \neg\Box_{\mathcal{T}} Con_{\mathcal{T}}$. $\Box$

7 Der Satz von Löb

Die nichttriviale Richtung des 2. Gödelschen Unvollstïdigkeitssatzes besagt, daß $\Box(\Box\bot \to \bot) \to \Box\bot$ bereits in der PRA beweisbar ist. Dies ist ein Spezialfall des viel allgemeineren Satzes von H. Löb.

Satz 7.1 (Satz von Löb)
Sei $\mathcal{T}$ *eine konsistente r.e. Erweiterung der* PRA*, dann gilt für alle Sätze* A*, daß*

(1) *Wenn* $\mathcal{T} \vdash \Box A \to A$*, dann bereits* $\mathcal{T} \vdash A$.

(2) PRA $\vdash \Box(\Box A \to A) \to \Box A$ (“formalised Löb's Theorem”)

wobei $\Box A$ *abkürzend für* $Pr_{\mathcal{T}}(\ulcorner A\urcorner)$ *steht.*

Da selbstverständlich auch PRA $\vdash \Box A \to \Box(\Box A \to A)$ *sind die Aussagen* $\Box A$ *und* $\Box(\Box A \to A)$ *in der* PRA *beweisbar äquivalent.*

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

(1) PRA $\vdash L \leftrightarrow \Box(L \to A)$.

Wegen (D3) gilt auch, daß

(2) PRA $\vdash L \to \Box L$

da L ja (zu) ein(em) Σ_1 Satz (äquivalent) ist. Außerdem gilt wegen (D2), daß

(3) $\mathrm{PRA} \vdash \Box(L \to A) \to \Box L \to \Box A$.

Aus (1), (2) und (3) folgt nun, daß

(4) $\mathrm{PRA} \vdash L \to \Box A$

und somit auch

(5) $\mathrm{PRA} \vdash (\Box A \to A) \to L \to A$.

Aus (5) folgt mit (D1), daß

(6) $\mathrm{PRA} \vdash \Box((\Box A \to A) \to L \to A)$

und somit wegen (D2) auch

(7) $\mathrm{PRA} \vdash \Box(\Box A \to A) \to \Box(L \to A)$.

Aus (1) und (7) folgt nun, daß

(8) $\mathrm{PRA} \vdash \Box(\Box A \to A) \to L$.

Somit folgt aus (4) und (8), daß

$$\mathrm{PRA} \vdash \Box(\Box A \to A) \to \Box A \ .$$

Also haben wir die zweite Behauptung gezeigt.[11]

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

Wie schon in Abschnitt 2 erwähnt bezeichnen wir Sätze der Form $\Box A \to A$ mit 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

[11] Man kann die zweite Behauptung des Satzes alternativ auch folgendermaßen beweisen. Aufgrund des arithmetischen Fixpunktsatzes gibt es einen Satz L, sodaß $\mathsf{PRA} \vdash L \leftrightarrow (\Box L \to A)$. Aufgrund der Derivabilty Conditions folgt dann $\mathsf{PRA} \vdash \Box L \to \Box\Box L \to \Box A$ und somit auch $\mathsf{PRA} \vdash \Box L \to \Box A$. Also $\mathsf{PRA} \vdash (\Box A \to A) \to \Box L \to A$ und somit auch $\mathsf{PRA} \vdash (\Box A \to A) \to L$. Also gilt auch $\mathsf{PRA} \vdash \Box(\Box A \to A) \to \Box L$ und somit $\mathsf{PRA} \vdash \Box(\Box A \to A) \to \Box A$.

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 $\mathrm{Pr}_{\mathcal{T}}(\ulcorner A \urcorner) \wedge \neg A$ konsistenterweise zu $\mathcal{T}$ hinzunehmen.
Offensichtlich ist der 2. Gödelsche Unvollständigkeitssatz eine unmittelbare Instanz des Satzes von Löb. Es sei hier noch erwähnt, daß auch der 1. Unvollständigkeitssatz aus dem 2. unmittelbar folgt[12]. Also lassen sich alle (betrachteten) Unvollständigkeitsphänomene auf den Satz von Löb reduzieren.

8 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 $\mathcal{T} \vdash \mathit{Pr}_{\mathrm{PRA}}(\ulcorner \bot \urcorner)$, wie z.B. $\mathcal{T} \equiv \mathrm{PRA} + \mathrm{Pr}_{\mathrm{PRA}}(\ulcorner \bot \urcorner)$. Weil ja offensichtlich

$$\mathrm{PRA} \vdash \mathit{Pr}_{\mathrm{PRA}}(\ulcorner \bot \urcorner) \to \mathit{Pr}_{\mathcal{T}}(\ulcorner \bot \urcorner)$$

folgt somit, daß

$$\mathcal{T} \vdash \mathit{Pr}_{\mathcal{T}}(\ulcorner \bot \urcorner)$$

[12]Sei $\mathcal{T}$ eine konsistente r.e. Erweiterung der PRA. Angenommen $\mathcal{T} \vdash \mathrm{Con}_{\mathcal{T}}$. Dann gilt aufgrund von (D1), daß $\mathrm{PRA} \vdash \Box\,\mathrm{Con}_{\mathcal{T}}$, und somit aufgrund des 2. Unvollständigkeitssatzes, daß auch $\mathrm{PRA} \vdash \Box\bot$. Wegen der 1–Konsistenz der PRA gäbe es dann eine natürliche Zahl n mit $\mathrm{PRA} \vdash \mathit{Bew}_{\mathcal{T}}(\underline{n}, \ulcorner \bot \urcorner)$, woraus folgte, daß $\mathcal{T} \vdash \bot$ im Widerspruch zur angenommenen Konsistenz von $\mathcal{T}$.
Angenommen $\mathcal{T}$ sei überdies 1–konsistent. Wenn $\mathcal{T} \vdash \neg\,\mathrm{Con}_{\mathcal{T}}$, dann gilt aufgrund der 1–Konsistenz von $\mathcal{T}$, daß $\mathbb{N} \models \exists n.\, \mathit{Bew}_{\mathcal{T}}(n, \ulcorner \bot \urcorner)$ und somit $\mathcal{T} \vdash \bot$ im Widerspruch zur Konsistenz von $\mathcal{T}$.
Selbstverständlich kann man aber die erste Behauptung des 1. Gödelschen Unvollständigkeitssatz auch direkt aus Satz 7.1 (1) folgern, indem man A durch $\bot$ instanziiert.

d.h.

$$\mathcal{T} \vdash \neg\, \mathrm{Con}_{\mathcal{T}} \ \ .$$

Also wird der Gödelsatz $\mathrm{Con}_{\mathcal{T}}$ in sehr vielen Fällen von $\mathcal{T}$ (negativ) entschieden. Um diesen Defekt zu beheben, gibt es gottseidank den "Rossersatz" einer Theorie $\mathcal{T}$, der von $\mathcal{T}$ *nie* entschieden wird.
Vorbereitend dazu benötigen wir folgendes Lemma.

Lemma 8.1 *Für jedes $n \in \mathbb{N}$ und jede Formel $A(x)$ gilt*

(1) $\mathrm{PRA} \vdash x < \underline{n} \leftrightarrow \bigvee_{k<n} x = \underline{k}$

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

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

Satz 8.1 *Sei $\mathcal{T}$ eine konsistente r.e. Erweiterung der* PRA. *Dann gibt es einen Π_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 $Pr_{\mathcal{T}}$

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

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

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

d.h.

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

ad (i): Angenommen $\mathcal{T} \vdash R$. Dann gibt es eine natürliche Zahl n_0 mit 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 8.1(2), daß

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

Da $\mathcal{T}$ (wie jede konsistente Erweiterung der PRA) genau die wahren, d.h. in $\mathbb{N}$ gültigen, Δ_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

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

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

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

und somit gilt auch

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

Also beweist die PRA auch den Satz

$$(\exists n.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner) \wedge \forall k < n.\, \neg Bew_{\mathcal{T}}(k, \ulcorner \neg R \urcorner)) \to \exists n \leq \underline{n_0}.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner)$$

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

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

Also gilt
$$\mathcal{T} \vdash \exists n \le \underline{n_0}.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner) \ .$$
Wegen Lemma 8.1(2) gilt
$$\mathrm{PRA} \vdash \exists n \le \underline{n_0}.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner) \leftrightarrow \bigvee_{n \le n_0} Bew_{\mathcal{T}}(\underline{n}, \ulcorner R \urcorner)$$
und wegen der Konsistenz von $\mathcal{T}$ gilt für jede natürliche Zahl n, daß
$$\mathrm{PRA} \vdash \neg Bew_{\mathcal{T}}(\underline{n}, \ulcorner R \urcorner) \ .$$
Somit folgt, daß
$$\mathrm{PRA} \vdash \neg\, \exists n \le \underline{n_0}.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner) \ ,$$
und somit insbesondere, daß
$$\mathcal{T} \vdash \neg\, \exists n \le \underline{n_0}.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner) \ .$$
Somit beweist $\mathcal{T}$ sowohl $\exists n \le \underline{n_0}.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner)$ als auch seine Negation im Widerspruch zur Konsistenz von $\mathcal{T}$. ☐

Der Knackpunkt von Rossers Beweis ist, daß im Falle $\mathcal{T} \vdash \neg R$ eine natürliche Zahl n_0 existiert, sodaß nicht nur $\mathrm{PRA} \vdash Bew_{\mathcal{T}}(\underline{n_0}, \ulcorner \neg R \urcorner)$, sondern auch $\mathrm{PRA} \vdash \neg R \to \exists\, n \le \underline{n_0}.\, Bew_{\mathcal{T}}(n, \ulcorner R \urcorner)$. Für G ist dies nicht möglich, weil man sonst beweisen könnte, daß $\mathcal{T} \nvdash \neg G$, 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 8.1 *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 $\mathcal{T} \nvdash \neg R$, ist $\mathcal{T}'$ konsistent. Außerdem ist $\mathcal{T}'$ echt stärker als $\mathcal{T}$, weil $\mathcal{T} \nvdash R$. ☐

9 Feferman's schwaches Reflexionsprinzip

Obzwar man aufgrund des Satzes von Löb das Reflexionsprinzip
$$(\mathrm{Refl}_A) \qquad Bew(n, \ulcorner A \urcorner) \to A$$

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

$$(\mathrm{WRefl}_A) \qquad \forall n.\,\Box(Bew(\underline{n},\ulcorner A\urcorner) \to A)$$

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

Satz 9.1 *Für konsistente r.e. Erweiterungen $\mathcal{T}$ der* PRA *gilt*

$$\mathrm{PRA} \vdash \forall n.\,\Box_{\mathcal{T}}(Bew_{\mathcal{T}}(\underline{n},\ulcorner A\urcorner) \to A) \ .$$

Beweis: Aus Gründen der Lesbarkeit schreiben wir Bew bzw. $\Box$ für $Bew_{\mathcal{T}}$ bzw. $\Box_{\mathcal{T}}$.
Wir zeigen folgende beide Behauptungen

(i) $\mathrm{PRA} \vdash Bew(n,\ulcorner A\urcorner) \to \Box(Bew(\underline{n},\ulcorner A\urcorner) \to A)$

(ii) $\mathrm{PRA} \vdash \neg Bew(n,\ulcorner A\urcorner) \to \Box(Bew(\underline{n},\ulcorner A\urcorner) \to A)$

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

ad (i): folgende Formeln sind PRA–beweisbar

(1) $A \to Bew(n,\ulcorner A\urcorner) \to A$

(2) $\Box A \to \Box(Bew(\underline{n},\ulcorner A\urcorner) \to A)$ aus (1) mit (D1)[13] und (D2)

(3) $Bew(n,\ulcorner A\urcorner) \to \Box A$

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

ad (ii): folgende Formeln sind PRA–beweisbar

(1) $\neg Bew(n,\ulcorner A\urcorner) \to \Box(\neg Bew(\underline{n},\ulcorner A\urcorner))$ wegen (D3)[14]

(2) $\neg Bew(n,\ulcorner A\urcorner) \to Bew(n,\ulcorner A\urcorner) \to A$

(3) $\Box(\neg Bew(\underline{n},\ulcorner A\urcorner)) \to \Box(Bew(\underline{n},\ulcorner A\urcorner) \to A)$ aus (2) mit (D1)[12] und (D2)

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

$\Box$

[13] Tatsächlich verwenden wir hier eine leichte Verallgemeinerung von (D1), die besagt, daß aus $\mathcal{T} \vdash A(\vec{x})$ folgt, daß $\mathrm{PRA} \vdash \Box A(\underline{\vec{x}})$.

[14] Tatsächlich verwenden wir hier eine leichte Verallgemeinerung von (D3), die besagt, daß $\mathrm{PRA} \vdash A(\vec{x}) \to \Box A(\underline{\vec{x}})$ für beliebige Σ_1–Formeln $A(\vec{x})$ mit freien Variablen.

10 Unentscheidbarkeit der Herleitbarkeit von Π_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* Π_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 Π_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 10.1 *Für $n \in \mathbb{N}$ ist*

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

ein Π_1–Satz. Wenn $\{n\}(n) = 0$, so gilt PRA $\vdash A_n$, *und, wenn $\{n\}(n) = 1$, so gilt* PRA $\vdash \neg A_n$.

Beweis: In der PRA kann man beweisen, daß

$$(\dagger) \qquad T(n, m, k_1) \wedge T(n, m, k_2) \to k_1 = k_2$$

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.

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

Wegen $(\dagger)$ gilt dann, daß 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.

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

Wegen $(\dagger)$ gilt dann, daß 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 10.1 *Die Menge der in $\mathcal{T}$ herleitbaren Π_1–Sätze ist nicht entscheidbar.*

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

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

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

da wegen Lemma 10.1 $n \in A_0 \Rightarrow \ulcorner A_n \urcorner \in \Pi_1^{\mathcal{T}}$ 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 10.1 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

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

Aufgrund des Diagonalisierungslemmas gibt es einen Satz A mit $\mathrm{PRA} \vdash A \leftrightarrow P(\ulcorner A \urcorner)$. Durch genaue Inspektion des Beweises des Diagonalisierungslemmas ergibt sich, daß A *selbst ein* Π_1*–Satz ist*, nämlich

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

wobei $\Theta_{n_0}(x) \equiv P(\mathrm{sub}(x, x))$.
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) $U(k) \neq 0 \Rightarrow \mathrm{PRA} \vdash A \Rightarrow \mathcal{T} \vdash A$

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 $\mathcal{T} \vdash T(\underline{e}, \ulcorner A \urcorner, \underline{k}) \to U(\underline{k}) \neq 0)$ und $\mathcal{T} \vdash T(\underline{e}, \ulcorner A \urcorner, \underline{k})$. Somit $0 \neq 0$. □

Da bereits die $\mathcal{T}$–Herleitbarkeit von Π_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 Π_1–Sätze, daß die Menge der mit $\mathcal{T}$ konsistenten Σ_1–Sätze nicht rekursiv aufzählbar ist.
Wie auch immer, die Moral dieses Abschnitts ist, daß man für Π_1–Sätze nicht mechanisch vorab überprüfen kann, ob $\mathcal{T}$ an ihrem Beweis scheitert.

11 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

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

   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 Σ_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 $f(k) \simeq \textbf{if}\ P(k)\ \textbf{then}\ 0\ \textbf{else}\ U(k)$. Also sind nicht 1-konsistente $\mathcal{T}$ höchst ungeeignet, um über Algorithmen zu räsonieren.

12 Beweisbarkeitslogik

Bereits im Verlauf der bisherigen Darstellung haben wir die modallogische Schreibweise verwendet, um die Derivability Conditions kompakt formulieren zu können 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 12.1 (Definition von PRL)
Die Sprache der PRL *ist die der üblichen Aussagenlogik angereichert um den 1-stelligen Operator* $\Box$*. Wir basieren den rein aussagenlogischen 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* $p, q, r, \ldots$ *als Metavariablen für aussagenlogische Konstante. 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 $A \rightarrow \Box A$ 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) entspicht (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 12.1 *Für alle Formeln* A *beweist* PRL *die Formel* $\Box A \rightarrow \Box\Box A$*.*

Beweis: Offenbar ist

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

eine boolesche Tautologie, also auch

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

Somit gilt auch

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

Als spezielle Instanz von (A3) erhalten wir

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

Aus (3) und (4) folgt nun

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

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

Der Witz von PRL liegt darin, daß alle "arithmetischen Interpretationen" von in PRL beweisbaren Aussagen in der PRA beweisbar sind. Eine "arithmetische Interpretation" ist eine Abbildung ϕ, die aussagenlogische Konstante auf geschlossen Formeln der PRA abbildet. Ein solches ϕ induziert für jede modallogische Formel A eine geschlossene Formel A^ϕ der PRA wie folgt

$$p^\phi \equiv \phi(p) \qquad \bot^\phi \equiv \bot \qquad (A \to B)^\phi \equiv A^\phi \to B^\phi \qquad (\Box A)^\phi \equiv Pr(\ulcorner A^\phi \urcorner)$$

Offenbar folgt aus $\mathsf{PRL} \vdash A$, daß PRA $\vdash A^\phi$ für alle arithmetischen Interpretationen ϕ, 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] oder Appendix B für detaillierte Beweise).

Satz 12.1 (Arithmetische Vollständigkeit von PRL)
Eine modallogische Formel A ist in PRL *genau dann beweisbar, wenn* PRA *alle arithmetischen Interpretationen A^ϕ beweist.*

Sei PRL_ω diejenige Modallogik, die durch alle Theoreme von PRL, alle Formeln der Form $\Box A \to A$ und den Modus Ponens axiomatisiert wird.[15] Da die PRA 1-konsistent ist und sogar $\mathbb{N} \models$ PRA, folgt aus $\mathsf{PRL}_\omega \vdash A$, daß in $\mathbb{N}$ alle arithmetischen Interpretationen A^ϕ gelten. R. Solovay (1976) hat auch bewiesen, daß

[15] Die Hinzunahme des Axiomenschemas $\Box A \to A$ zu PRL ist ja inkonsistent!

Satz 12.2 (Arithmetische Vollständigkeit von PRL_ω)
Eine modallogische Formel A in PRL_ω genau dann beweisbar ist, wenn alle arithmetischen Interpretationen A^ϕ in $\mathbb{N}$ gelten.

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

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

gibt, d.h. R ist transitiv und irreflexiv. Dieses Vollständigkeitsresultat wird essentiell im Beweis der Sätze 12.1 und 12.2 verwendet. Im interessanten Teil des Beweises von 12.1 geht man von einer in 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 ϕ konstruieren kann, sodaß A^ϕ in der PRA nicht beweisbar ist. Die Logik PRL_ω ist deswegen entscheidbar, weil man im Beweis von Satz 12.2 auch zeigt, daß $\mathsf{PRL}_\omega \vdash A$ genau dann, wenn $\mathsf{PRL} \vdash A^*$ mit

$$A^* \equiv (\Box C_1 \to C_1) \to \ldots \to (\Box C_k \to C_k) \to A$$

wobei $C_1, \ldots, C_k$ die Liste derjenigen Unterformeln C von A ist, sodaß $\Box C$ eine Unterformel von A ist. Also ist $\mathsf{PRL}_\omega \vdash A$ äquivalent zu $\mathsf{PRL} \vdash A^*$, womit sich die Entscheidbarkeit von PRL_ω auf die Entscheidbarkeit von PRL zurückführen läßt.
Wir nennen modallogische Formeln, in denen keine Aussagenkonstanten vorkommen, "gödelsch". Für solche Formeln A ist A^ϕ unabhängig von ϕ und wird mit $\widehat{A}$ bezeichnet. Wir nennen $\widehat{A}$ die "arithmetische Lesart von A". Aufgrund des 2. Solovayschen Vollständigkeitssatzes und der Entscheidbarkeit von PRL_ω, können wir

> für "gödelsche" Formeln entscheiden, ob ihre arithmetische Lesart im Standardmodell $\mathbb{N}$ wahr ist

was bemerkenswert ist, da diese Formeln über hochgradig unentscheidbare formale Systeme sprechen. Es sind also bestimmte metamathematische Eigenschaften solcher Systeme entscheidbar. Aber, wie am Ende von [Bo] bewiesen, ist dies nicht mehr der Fall für Provability Logic mit Quantoren.

Aber auch der erste Vollständigkkeitssatz von Solovay hat Anwendungen. Man kann sich fragen, ob die PRA für alle arithmetischen Sätze S beweist, daß $\mathrm{Pr}(\ulcorner \mathrm{Pr}(\ulcorner S \urcorner) \vee \mathrm{Pr}(\ulcorner \neg S \urcorner) \urcorner) \to \mathrm{Pr}(\ulcorner S \urcorner) \vee \mathrm{Pr}(\ulcorner \neg S \urcorner)$, d.h., daß das Reflexionsprinzip für Aussagen der Form "S wird durch das formale System entschieden" gilt. Aufgrund von Satz 12.1 ist diese Frage äquivalent zur Herleitbarkeit von $\Box(\Box p \vee \Box \neg p) \to \Box p \vee \Box \neg p$ in PRL. Diese modallogische Formel wird aber durch ein Kripke Modell (mit 3 Welten) refutiert (Übung!), also ist die Antwort auf die Frage negativ.

Eine weitere wichtige Eigenschaft von PRL ist folgende.

Satz 12.3 *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) $\mathsf{PRL} \vdash B \leftrightarrow A[B]$ *und*

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

d.h. "geboxte" Formeln $A[p]$ haben in PRL Fixpunkte, die bis auf in PRL beweisbare Äquivalenz eindeutig sind.

Für ausführliche Beweise der in diesem Abschnitt vorgestellten Resultate konsultiere man die Bücher [Sm] und [Bo] bzw. die Appendices A und B.

13 Mathematische Unabhängigkeitsresultate

Wir haben gesehen, daß für alle r.e. konsistenten Erweiterungen T der PRA sich Π_1 Sätze finden lassen, die in T nicht bewiesen werden können, typischerweise den Π_1 Satz $Con_T \equiv \forall n \neg Bew_T(\ulcorner \bot \urcorner)$. 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.[16] Natürlich ist "mathematisch sinnvoll" selbst kein mathematischer Begriff in dem Sinne, daß er präzise definiert werden könnte. Jedoch erkennt

[16]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 die-

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 vielleicht 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 eines Polynoms $P(\vec{x})$, d.h. $\forall\vec{x}.P(\vec{x}) \neq 0$, auf recht einfache Art und Weise zu einer Π_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 Koeffizient 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 ganzahligen Nullstellen hat, obwohl T diese Aussage nicht beweisen kann. Zugegebenermaßen ist solch ein

---

ser Frage intensiv beschäftigt, siehe etwa die Vorversion seines Buches *Boolean Relation Theory and Incompleteness*, die man sich unter der Adresse `u.osu.edu/friedman.8/foundational-adventures/booleanrelation-theory-book/` 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) einen 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 zweite 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 σ-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 gegeben 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 das Ersetzungsschema beweisen lassen!

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 Π_1 Aussage!) in PA beweisbar ist, obwohl manche Experten (A. MacIntyre) dieser Meinung sind. Ein Beweis dafür steht jedenfalls noch aus!

Alle bisher in diesem Abschnitt betrachteten Beispiele sind wahre Π_1-Sätze, die aber in der Peano Arithmetik nicht beweisbar sind. Es gibt aber auch "mathematische" Π_2-Sätze, die in der Peano Arithmetik nicht beweisbar sind. Das erste solche geht auf J. Paris and L. Harrington zurück und wird in Appendix D beschrieben (siehe [Mar] für eine ausführlichere Darstellung). Ein weiteres solches findet sich in [KP], wo gezeigt wird, daß die wahre, von R. Goodstein in den 40er Jahren bewiesene Aussage, daß alle *Goodstein sequences* schlußendlich konstant 0 werden, nicht in der Peano Arithmetik bewiesen werden kann. In beiden Fällen handelt es sich um total rekursive Funktionen, die so stark wachsen, daß ihre Termination nicht in der PA bewiesen werden kann.

Ein weiteres Unabhängigkeitsresulat ist im Zusammenhang mit der Cantorschen *Kontinuumshypothese* (CH) zu erwähnen, die besagt, daß jede unendliche Teilmenge von $\mathcal{P}(\mathbb{N})$ zu $\mathbb{N}$ oder $\mathcal{P}(\mathbb{N})$ gleichmächtig ist. Gödel hat 1938 gezeigt, daß CH mit ZF in dem Sinne kompatibel ist, dass ZFC genau dann konsistent ist, wenn $\mathsf{ZFC}+$ CH konsistent ist. Paul Cohen hat 25 Jahre später gezeigt (1963), daß CH in ZFC nicht bewiesen werden kann, da man mithife der sogenannten "forcing" Methode ein Modell von ZFC konstruieren kann, in dem CH nicht gilt. Also ist die Kontinuumshypothese von ZFC unabhängig. Dies ist bemerkenswert, da sie von Cantor im 19. Jahrhundert lange vor dem Gödelschen Unvollständigkeitssatz postuliert wurde.

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.

[Sti] J. Stillwell *Reverse Mathematics: Proofs from the Inside Out* Princeton University Press, 2018.

[GRS] R. Graham, B. Rothschild, J. Spencer *Ramsey Theory* Wiley 1980.

[Mar] D. Marker *Model Theory: An Introduction* Springer 2002.

[KP] L. Kirby, J. Paris *Accessible independence results for Peano arithmetic* Bulletin of the London Mathematical Society no. 14, pp. 285-293, 1982.

A Das allgemeine Schema von Diagonalisierungsbeweisen

In der mathematischen Logik erweisen sich sogenannte Diagonalisierungsbeweise als höchst nützliche Methode. Erstmals wurde diese von Cantor verwendet, um nachzuweisen daß es keine surjektive Abbildung von einer Menge X auf ihre Potenzmenge $\mathcal{P}(X)$ gibt. Russell verwendete diese Methode, um die Inkonsistenz des uneingeschränkten Komprehensionsprinzip (in der Mengenlehre) nachzuweisen. In der Berechenbarkeitstheorie wird diese Methode verwendet, um etwa die Unentscheidbarkeit des Halteproblems zu beweisen. Wir geben nun eine allgemeine prädikatenlogische Antinomie an, die solchen Diagonaliserungsbeweisen zugrunde liegt. Für beliebige zweistellige Prädikate $A(x,y)$ ist

$$(D_A) \qquad \exists x \forall y \big(A(x,y) \leftrightarrow \neg A(y,y)\big)$$

eine prädikatenlogische Antinomie, da aus $\forall y\big(A(x,y) \leftrightarrow \neg A(y,y)\big)$ unmittelbar $A(x,x) \leftrightarrow \neg A(x,x)$ folgt, was schon rein aussagenlogisch eine Antinomie darstellt.[17] Ein Diagonalisierungsbeweis von $\neg B$ ist nun ein Beweis von $B \to D_A$ für eine geeignet gewählte Aussage A.

Wir präsentieren nun eine auf Lawvere zurückgehende kategorielle Sichtweise von Diagonalisierungsbeweisen.

Satz A.1 (Fixpunktsatz von Lawvere)
Sei $f : A \times A \to B$, sodaß $e : A \to B^A : a \mapsto \lambda x{:}A.f(a,x)$ surjektiv ist. Dann hat jede Abbildung $g : B \to B$ einen Fixpunkt.

Beweis: Sei $h = g \circ f \circ \delta_A : A \to B$. Dann existiert ein $a \in A$ mit $e(a) = h$ und somit $f(a,a) = e(a)(a) = h(a) = g(f(a,a))$. $\square$

Mit Kontraposition erhalten wir folgende äquivalente Aussage.

Satz A.2 *Wenn $g : B \to B$ keinen Fixpunkt hat, dann gibt es keine surjektive Abbildung $e : A \to B^A$.*

Wir geben nun eine kategorielle Formulierung des arithmetischen Fixpunktsatzes an, aus dem dann die gödelschen Sätze und der Satz von Löb folgen.

[17]Insbesondere auch schon in der intuitionistischen Aussagenlogik, wie man folgendermaßen sieht. Auch intuitionistisch gilt $\neg(A \wedge \neg A)$. Aus der Annahme $A \leftrightarrow \neg A$ folgen dann damit sowohl $\neg A$ als auch $\neg\neg A$, womit wir $\neg(A \leftrightarrow \neg A)$ gezeigt haben.

Sei $\mathcal{T}$ eine 1-konsistente Theorie in der Sprache der Arithmetik. Sei $\mathcal{C}$ die zugehörige Lindenbaum-Tarski Kategorie, welche ein Objekt N der natürlichen Zahlen und ein Wahrheitswertobjekt Ω enthält. Alle Objekte dieser Kategorie sind endliche Produkte dieser beiden Objekte.
Es gibt nun in $\mathcal{C}$ Morphismen $\Theta : N \times N \to \Omega$ und $\mathsf{subst} : N \times N \to N$, sodaß

(1) für alle $c : 1 \times N \to \Omega$ ein $\ulcorner c \urcorner : 1 \to N$ existiert, sodaß

$$\begin{tikzcd} N \times N \arrow[r, "\Theta"] & \Omega \\ 1 \times N \arrow[u, "\ulcorner c \urcorner \times \mathsf{id}_N"] \arrow[ur, "c"'] & \end{tikzcd}$$

in $\mathcal{C}$ kommutiert, und

(2) für jedes Numeral $n : 1 \to N$ in $\mathcal{C}$ gilt, daß $\Theta(\mathsf{subst}(n, m), 0) = \Theta(n, m)$.

Also ist $\mathsf{dec} = \Theta(-, 0) : N \to \Omega$ eine "Decodierungsfunktion", die Gödelnummern in Propositionen umwandelt. Für $p \in \Omega$ sei $c_p : N \to \Omega$ die konstante Funktion mit Wert p. Wegen (1) gilt dann $p = \Theta(\ulcorner c_p \urcorner, 0) = \mathsf{dec}\big(\ulcorner c_p \urcorner\big)$. Also hat jede Proposition eine Gödelnummer.
Mithilfe dieser Postulate läßt sich der Beweis des arithmetischen Fixpunktsatzes, d.h. für Satz 4.1, wie folgt durchführen.
Sei $p : N \to \Omega$. Wir betrachten $c = p \circ \mathsf{subst} \circ \delta_N : N \to \Omega$. Wegen (1) gibt es ein $\ulcorner c \urcorner : 1 \to N$, sodaß $\Theta(\ulcorner c \urcorner, -) = c(-)$ und somit $c \circ \ulcorner c \urcorner = p \circ \mathsf{subst} \circ \delta_N \circ \ulcorner c \urcorner = p(\mathsf{subst}(\ulcorner c \urcorner, \ulcorner c \urcorner))$, und wegen (2), daß $\Theta(\mathsf{subst}(\ulcorner c \urcorner, \ulcorner c \urcorner), 0) = \Theta(\ulcorner c \urcorner, \ulcorner c \urcorner) = c \circ \ulcorner c \urcorner = p(\mathsf{subst}(\ulcorner c \urcorner, \ulcorner c \urcorner)$. Für $a = \mathsf{subst}(\ulcorner c \urcorner, \ulcorner c \urcorner) : 1 \to N$ gilt nun $\mathsf{dec}(a) = \Theta(\mathsf{subst}(\ulcorner c \urcorner, \ulcorner c \urcorner), 0) = p(\mathsf{subst}(\ulcorner c \urcorner, \ulcorner c \urcorner) = p(a)$, also ist a die Gödelnummer einer Aussage, die zu $p(a)$ äquivalent ist.

B Vollständigkeit von PRL bezüglich endlicher wohlfundierter Modelle

Ein *Kripke Modell* ist ein Tripel $(W, R, \Vdash)$, sodaß W eine nichtleere Menge, R eine zweistellige Relation auf W und $\Vdash$ eine Relation zwischen Elementen aus W und Aussagenkonstanten (von PRL) ist. Die Elemente von W stellt man sich als "mögliche Welten" vor und R als sog. "accessibility relation", die angibt, welche Welten von einer gegebenen als "möglich" oder "erreichbar" erscheinen. Die "forcing" Relation $\Vdash$ kann auf folgende Weise zu einer Relation zwischen W und Formeln von PRL erweitert werden

(1) für kein $w \in W$ gilt $w \Vdash \bot$

(2) $w \Vdash A \to B$ genau dann, wenn aus $w \Vdash A$ folgt, daß $w \Vdash B$

(3) $w \Vdash \Box A$ genau dann, wenn $w' \Vdash A$ für alle w' mit wRw'.

Eine Aussage A gilt in $(W, R, \Vdash)$, wenn $w \Vdash A$ für alle $w \in W$, wofür wir $\Vdash A$ schreiben. Man zeigt leicht, daß in einem Kripke Modell die Axiome (A1) und (A2) und die Regeln (R1) und (R2) in dem Sinne gelten, daß die Konklusion im Modell gilt, falls alle Prämissen im Modell gelten.
Im allgemeinen gilt (A3) nicht in jedem Kripke Modell. Aber man kann leicht zeigen (Übung!), daß (A3) in jedem Kripke Modell gilt, in dem die Relation R transitiv und in dem Sinne wohlfundiert ist, daß es keine Folge $w : \mathbb{N} \to W$ gibt mit $w_n R w_{n+1}$ für alle $n \in \mathbb{N}$. Falls W endlich ist, ist eine transitive Relation R auf W genau dann wohlfundiert, wenn sie irreflexiv ist, d.h. wRw für kein $w \in W$ gilt. Solche Modelle nennen wir *endlich wohlfundiert* und sie validieren alle Axiome und Regeln von PRL.
Im Rest dieses Abschnitts zeigen wir die Umkehrung dieser Aussage, d.h., dass $\mathsf{PRL} \vdash D$, wenn D in allen endlichen wohlfundierten Modellen gilt. Tatsächlich zeigen wir die Kontraposition dieser Aussage, nämlich daß für jede in PRL nicht beweisbare Aussage D ein endlich wohlfundiertes Modell existiert, in dem D nicht gilt. Unser Beweis dieses Sachverhalts folgt im wesentlichen der Darstellung in [Bo].
Für den Rest dieses Abschnitts sei D eine fixe, aber beliebige Formel, die in PRL nicht beweisbar ist. Mit $\mathcal{S}(D)$ bezeichnen wir die (endliche) Menge der Teilformeln von D und mit $\mathcal{C}(D)$ die (endliche) Menge $\mathcal{S}(D) \cup \{\neg A \mid A \in \mathcal{S}(D)\}$ der negierten und unnegierten Teilformeln von D.

Wir basteln nun ein endlich wohlfundiertes Kripke Modell $(W, R, \vdash)$, in dem D nicht gilt.
Die Elemente von W sind maximal konsistente Teilmengen von $\mathcal{C}(D)$, d.h. Teilmengen w von $\mathcal{C}(D)$, sodaß $\mathsf{PRL} \nvdash \neg \bigwedge w$ und für jede Formel $A \in \mathcal{S}(D)$ entweder $A \in w$ oder $\neg A \in w$. Auf W definieren wir eine accessibility Relation R wie folgt: wRw' genau dann, wenn folgende beide Bedingungen erfüllt sind

(R1) wenn $\Box A \in w$, dann $A, \Box A \in w'$

(R2) es gibt ein $\Box B \in w'$, sodaß $\neg \Box B \in w$.

Schlussendlich definieren wir $w \Vdash p$ als $p \in w$.

Lemma B.1 *Das eben definierte Modell* $(W, R, \vDash)$ *ist endlich wohlfundiert.*

Beweis: Offensichtlich ist W endlich, da $\mathcal{C}(D)$ endlich ist.
Wir zeigen nun, daß R transitiv ist. Sei wRw' und $w'Rw''$. Wir zeigen nun, daß wRw''. Sei $\Box A \in w$. Dann ist wegen (R1) auch $\Box A \in w'$ und somit wieder wegen (R1) auch $A, \Box A \in w''$. Wegen (R2) gibt es ein $\Box B \in w'$, sodaß $\neg \Box B \in w$. Wegen (R1) ist auch $\Box B \in w''$.
Die Relation R ist irreflexiv, da im Falle wRw, wegen (R2) ein $B \in \mathcal{S}(D)$ existiert, mit $\Box B \in w$ und $\neg \Box B \in w$, was wegen (R1) unmöglich ist. $\Box$

Um zu zeigen, daß D in $(W, R, \Vdash)$ nicht gilt, benötigen wir eine Reihe von Lemmata.

Lemma B.2 *Sei* $X \subseteq \mathcal{C}(D)$ *konsistent, d.h.* $\mathsf{PRL} \nvdash \neg \bigwedge X$. *Dann existiert ein* $w \in W$ *mit* $X \subseteq w$.

Beweis: Sei $X \subseteq \mathcal{C}(D)$ konsistent und $A \in \mathcal{S}(D)$. Wenn $X \cup \{A\}$ und $X \cup \{\neg A\}$ beide inkonsistent wären, d.h. $\mathsf{PRL} \vdash \neg(A \wedge \bigwedge X)$ und $\mathsf{PRL} \vdash \neg(\neg A \wedge \bigwedge X)$, dann $\mathsf{PRL} \vdash \neg \bigwedge X$, da $(A \wedge \bigwedge X) \wedge (\neg A \wedge \bigwedge X)$ rein aussagenlogisch zu $\bigwedge X$ äquivalent ist. Da aber X nach Voraussetzung konsistent ist, d.h. $\mathsf{PRL} \nvdash \neg \bigwedge X$, ist also $X \cup \{A\}$ oder $X \cup \{\neg A\}$ konsistent.
Durch wiederholte Anwendung des eben bewiesenen Sachverhalts läßt sich jedes konsistente $X \subseteq \mathcal{C}(D)$ zu einer maximal konsistenten Teilmenge w mit $X \subseteq w$ ergänzen. $\Box$

Lemma B.3 *Wenn $w \in W$ und $A \in \mathcal{S}(D)$ mit $\mathsf{PRL} \Vdash \bigwedge w \to A$, dann ist auch $A \in w$.*

Beweis: Wir beweisen die Behauptung erst mal für den Fall, daß $A \in \mathcal{S}(D)$. Da $w \in W$, gilt $A \in w$ oder $\neg A \in w$. Da $\mathsf{PRL} \vdash \bigwedge w \to A$ und w konsistent ist, folgt $\neg A \notin w$ also $A \in w$.
Andernfalls ist $A \equiv \neg B$ für ein $B \in \mathcal{S}(D)$. Da $w \in W$, gilt $B \in w$ oder $\neg B \in w$. Da nach Voraussetzung $\mathsf{PRL} \vdash \bigwedge w \to \neg B$, folgt $B \notin w$. Also ist $A \equiv \not B \in w$. $\square$

Lemma B.4 *Für $\Box B \in \mathcal{C}(D)$ und $w \in W$ gilt $\Box B \in w$ genau dann, wenn $B \in w'$ für all $w' \in W$ mit wRw'.*

Beweis: Wenn $\Box B \in w$ und wRw', dann ist $B \in w'$.
Die Rückrichtung beweisen wir mit Kontraposition. Vorerst stellen wir aber erst mal folgende Hilfsüberlegung an. Sei $X = \{\neg B, \Box B\} \cup \{C, \Box C \mid \Box C \in w\}$ und $\{C_1, \ldots, C_n\} = \{C \mid \Box C \in w\}$. Wenn X inkonsistent ist, dann $\mathsf{PRL} \vdash \neg(\neg B \wedge \Box B \wedge C_1 \wedge \Box C_1 \wedge \ldots C_n \wedge \Box C_n)$ also auch $\mathsf{PRL} \vdash C_1 \wedge \Box C_1 \wedge \ldots C_n \wedge \Box C_n \to \Box B \to B$, also auch $\mathsf{PRL} \vdash \Box C_1 \wedge \Box\Box C_1 \wedge \ldots \Box C_n \wedge \Box\Box C_n \to \Box(\Box B \to B)$, also auch $\Box C_1 \wedge \ldots \Box C_n \to \Box B$.
Angenommen $\Box B \notin w$. Also ist X konsistent, da andernfalls $\Box B \in w$ wegen Lemma B.3, da ja $\Box C_1, \ldots, \Box C_n \in w$. Also gibt es wegen Lemma B.2 ein $w' \in W$ mit $X \subseteq w'$. Da $\neg B \in X \subseteq w' \in W$, gilt $B \notin w'$. Wenn $\Box C \in w$, dann $C, \Box C \in X \subseteq w'$. Da $\Box B \notin w$, ist $\neg\Box B \in w$. Also wRw'. $\square$

Lemma B.5 *Für $A \in \mathcal{S}(D)$ und $w \in W$ gilt $A \in w$ genau dann, wenn $w \Vdash A$.*

Beweis: Wir beweisen die Aussage mit Induktion über den Aufbau von A.
Falls A eine Aussagenvariable p ist, dann gilt die Behauptung aufgrund der Definition von $\Vdash$.
Sei $A \equiv B \to C$ und wir nehmen als Induktionshypothese an es gelte die Behauptung für B und C. Sei $B \to C \notin w$. Dies ist äquivalent zu $\neg(B \to C) \in w$. Wegen Lemma B.3 ist dies äquivalent zu $B \in w$ und $\neg C \in w$, d.h. zu $B \in w$ und $C \notin w$. Aufgrund der Induktionshypothese ist dies äquivalent zu $w \Vdash B$ and $w \nVdash C$, was hinwiederum äquivalent ist zu $w \nVdash B \to C$.
Sei $A \equiv \Box B$ und wir nehmen als Induktionshypothese an, die Behauptung gelte für B. Angenommen $\Box B \in w$. Wegen Lemma B.4 ist dann $B \in w'$ für

alle w' mit wRw'. Aufgrund der Induktionshypothese gilt dann $w' \Vdash B$ für alle w' mit wRw' und somit $w \Vdash \Box B$. Angenommen $w \Vdash \Box B$. Dann gilt $w' \Vdash B$ für alle w' mit wRw'. Dann gilt aufgrund der Induktionshypothese, daß $B \in w'$ für alle w' mit wRw', woraus mit Lemma B.4 $\Box B \in w$ folgt. $\Box$

Satz B.1 *Die Aussage D gilt nicht im oben definierten Modell $(W, R, \Vdash)$.*

Beweis: Da $\mathsf{PRL} \nvdash D$, gibt es aufgrund von Lemma B.2 ein $w \in W$ mit $\neg D \in w$ und somit $D \notin w$. Aufgrund von Lemma B.5 folgt nun $w \nVdash D$ und somit gilt D nicht in $(W, R, \Vdash)$. $\Box$

Für das Modell $(W, R, \Vdash)$, das in Abhängigkeit von D konstruiert wird, gilt $|W| \leq 2^{|\mathcal{C}(D)|}$. Also erhalten wir folgendes

Korollar B.1 *Eine Formel D ist in* PRL *genau dann herleitbar, wenn D in allen Modellen $(W, R, \Vdash)$ gilt, für die $|W| \leq 2^{|\mathcal{C}(D)|}$.*

Daraus folgt folgender

Satz B.2 *Die Logik* PRL *ist entscheidbar.*

C Beweis der Solovayschen Vollständigkeitssätze

Wenn $\mathsf{PRL} \vdash A$, dann beweist PRA jede arithmetische Interpretation A^ϕ von A, da PRA ja das formalisierte Löbsche Theorem beweist.
Nehmen wie hingegen an, daß $\mathsf{PRL} \nvdash A$. Wir werden eine arithmetische Interpretation ϕ konstruieren, sodaß A^ϕ nicht in PRA bewiesen werden kann. Aufgrund des Vollständigkeitsresultats für PRL gibt es ein endlich wohlfundiertes Modell $(W, R, \Vdash)$, in dem A nicht gilt. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß $w \nVdash A$ und wRw' für alle von w verschiedenen w' und außerdem $W = \{1, \ldots, n\}$ mit $w = 1$. Wir schreiben W' für $W \cup \{0\}$ und R' für $R \cup \{\langle 0, w\rangle \mid w \in W\}$ und erweitern $\Vdash$ durch die Setzung $0 \Vdash p$ genau dann, wenn $1 \Vdash p$ für aussagenlogische Konstanten p.
Aufgrund der 2. Kleeneschen Rekursionstheorems gibt es eine total rekursive Funktion $h : \mathbb{N} \to \{0, 1, \ldots, n\}$, sodaß

- $h(0) = 0$
- wenn x Gödelnummer eines PRA Beweises von $\neg \mathsf{Lim}_w$ für eine w mit $h(x)R'w$ ist, dann $h(x+1) = w$, und andernfalls $h(x+1) = h(x)$

wobei Lim_w für $\exists m \forall k \geq m \ h(k) = w$ steht. Wir definieren die gesuchte arithmetische Interpretation ϕ als $\phi(p) \equiv \bigvee\{\mathsf{Lim}_w \mid w \in W' \text{ und } w \Vdash p\}$.

Lemma C.1 *Es gelten folgende Aussagen*

(1) *für* $0 \leq i < j \leq n$ *beweist* PRA*, daß* $\neg(\mathsf{Lim}_i \wedge \mathsf{Lim}_j)$

(2) PRA *beweist* $h(a) = \underline{i} \to \mathsf{Lim}_i \vee \bigvee_{iR'j} \mathsf{Lim}_j$

(3) PRA *beweist* $\bigvee_{0 \leq i \leq n} \mathsf{Lim}_i$

(4) *für* $iR'j$ *beweist* PRA *die Aussage* $\mathsf{Lim}_i \to \neg \Pr\left(\ulcorner \neg \mathsf{Lim}_j \urcorner\right)$

(5) *für* $1 \leq i \leq n$ *beweist* PRA *die Aussage* $\mathsf{Lim}_i \to \Pr\left(\ulcorner \neg \mathsf{Lim}_i \urcorner\right)$

(6) *für* $1 \leq i \leq n$ *beweist* PRA *die Aussage* $\mathsf{Lim}_i \to \Pr\left(\ulcorner \bigvee_{iRj} \mathsf{Lim}_j \urcorner\right)$

Beweis: Die Funktion h kann nicht gleichzeitig gegen verschiedene Werte konvergieren und diese Argument läßt sich in der PRA formalisieren, also gilt (1).
Für alle k gilt $h(k) = h(k+1)$ oder $h(k)R'h(k+1)$. Außerdem ist $h(k) \leq n$ für alle k. Wenn deshalb h irgendwann den Wert i annimmt und irgenwannmal konstant wird, dann liegt dieser konstante Wert notwendigerweise in $\{i\} \cup \{j \leq n \mid iR'j\}$. Dieses Argument läßt sich in der PRA formalisieren, also gilt (2).
Indem man a in (2) gleich 0 setzt, folgt (3).
Behauptung (4) folgt daraus, daß man folgendes informelle Argument in der PRA formalisieren kann. Angenommen $iR'j$, Lim_i und PRA $\vdash \mathsf{Lim}_j$. Dann gibt es eine natürliche Zahl x, sodaß $h(y) = i$ für alle $y \geq x$. Wenn Lim_j in der PRA beweisbar ist, dann gibt es Beweise von Lim_j mit beliebig großen Gödelnummern. Sei also $y \geq x$ Gödelnummer eines PRA-Beweises von Lim_j. Aufgrund der Konstruktion von h ist dann $h(y) = j$. Da aber auch $h(y) = i$, folgt $i = j$ im Widerspruch zu $iR'j$.
Behauptung (5) zeigt man folgendermaßen. Sei $i \geq 1$. Die PRA beweist $h(x) = \underline{i} \to \exists y\, Bew(y, \ulcorner\neg\mathsf{Lim}_i\urcorner)$. Da die PRA auch beweist, daß $\mathsf{Lim}_i \to \exists x\, h(x) = \underline{i}$, beweist die PRA auch $\mathsf{Lim}_i \to \mathrm{Pr}\big(\ulcorner\neg S_i\urcorner\big)$.
Wir beweisen nun Behauptung (6). Sei i eine natürliche Zahl zwischen 1 und n. Wegen Lemma C.1(2) gilt

$$\mathrm{PRA} \;\vdash \exists x\, h(x) = \underline{i} \to \mathsf{Lim}_i \vee \bigvee_{iR'j} \mathsf{Lim}_j$$

und somit auch

$$\mathrm{PRA} \;\vdash \mathrm{Pr}\big(\ulcorner\exists x\, h(x) = \underline{i}\urcorner\big) \to \mathrm{Pr}\big(\ulcorner\mathsf{Lim}_i \vee \bigvee_{iR'j} \mathsf{Lim}_j\urcorner\big)$$

und somit auch

$$\mathrm{PRA} \;\vdash \exists x\, h(x) = \underline{i} \to \mathrm{Pr}\big(\ulcorner\mathsf{Lim}_i \vee \bigvee_{iR'j} \mathsf{Lim}_j\urcorner\big)$$

da ja PRA $\vdash \exists x\, h(x) = \underline{i} \to \mathrm{Pr}\big(\ulcorner\exists x\, h(x) = \underline{i}\urcorner\big)$, weil $\exists x\, h(x) = \underline{i}$ eine Σ_1 Formel ist. Da PRA $\vdash \mathsf{Lim}_i \to \exists x\, h(x) = \underline{i}$, gilt auch

$$\mathrm{PRA} \;\vdash \mathsf{Lim}_i \to \mathrm{Pr}\big(\ulcorner\mathsf{Lim}_i \vee \bigvee_{iR'j} \mathsf{Lim}_j\urcorner\big)$$

Wegen Lemma C.1(5) gilt

$$\mathrm{PRA} \vdash \mathsf{Lim}_i \to \mathrm{Pr}\big(\ulcorner \neg\mathsf{Lim}_i \urcorner\big)$$

Da PRA $\vdash \mathrm{Pr}\big(\ulcorner \neg\mathsf{Lim}_i \urcorner\big) \wedge \mathrm{Pr}\big(\ulcorner \mathsf{Lim}_i \vee \bigvee_{iR'j} \mathsf{Lim}_j \urcorner\big) \to \mathrm{Pr}\big(\ulcorner \bigvee_{iR'j} \mathsf{Lim}_j \urcorner\big)$, folgt nun PRA $\vdash \mathsf{Lim}_i \to \mathrm{Pr}\big(\ulcorner \bigvee_{iR'j} \mathsf{Lim}_j \urcorner\big)$ wie behauptet. $\square$

Lemma C.2 *Für alle* $1 \leq i \leq n$ *und Unterformeln* B *von* A *gilt*

(1) *wenn* $i \Vdash B$*, dann* PRA $\vdash \mathsf{Lim}_i \to B^\phi$

(2) *wenn* $i \nVdash B$*, dann* PRA $\vdash \mathsf{Lim}_i \to \neg B^\phi$

Beweis: Wir beweisen die Behauptung durch Induktion über den Aufbau von B.
Sei B eine Aussagenkonstante p. Wenn $i \Vdash p$, dann PRA $\vdash \mathsf{Lim}_i \to p^\phi$. Wenn $i \nVdash p$, dann ist jedes Glied Lim_j der Disjunktion p^ϕ verschieden von Lim_i. Wegen Lemma C.1(1) gilt aber PRA $\vdash \mathsf{Lim}_i \to \neg\mathsf{Lim}_j$, und deshalb PRA $\vdash \mathsf{Lim}_i \to \neg p^\phi$.
Sei $B \equiv B_1 \to B_2$ und wir nehmen als Induktionshypothese an, die Behauptung gelte für B_1 und B_2. Wenn $i \Vdash B_1 \to B_2$, dann $i \nVdash B_1$ oder $i \Vdash B_2$. Wenn $i \nVdash B_1$, dann gilt aufgrund der Induktionshypothese, daß PRA $\vdash \mathsf{Lim}_i \to \neg B_1^\phi$ und somit PRA $\vdash \mathsf{Lim}_i \to B_1^\phi \to B_2^\phi$. Wenn $i \Vdash B_2$, dann gilt aufgrund der Induktionshypothese, daß PRA $\vdash \mathsf{Lim}_i \to B_2^\phi$ und somit PRA $\vdash \mathsf{Lim}_i \to B_1^\phi \to B_2^\phi$. Wenn $i \nVdash B_1 \to B_2$, dann $i \Vdash B_1$ und $i \nVdash B_2$. Aufgrund der Induktionshypothese beweist PRA sowohl $\mathsf{Lim}_i \to B_1^\phi$ als auch $\mathsf{Lim}_i \to \neg B_2^\phi$ und somit PRA $\vdash \mathsf{Lim}_i \to \neg(B_1^\phi \to B_2^\phi)$.
Sei $B \equiv \Box C$ und wir nehmen als Induktionshypothese an, die Behauptung gelte für C. Wir erinnern, daß $B^\phi \equiv \mathrm{Pr}\big(\ulcorner C^\phi \urcorner\big)$.
Angenommen $i \Vdash \Box C$. Dann $j \Vdash C$ für alle j mit iRj. Also gilt aufgrund der Induktionshypothese, daß PRA $\vdash \mathsf{Lim}_j \to C^\phi$ für alle j mit iRj. Also auch PRA $\vdash \bigvee_{iRj} \mathsf{Lim}_j \to C^\phi$ und somit PRA $\vdash \mathrm{Pr}\big(\ulcorner \bigvee_{iRj} \mathsf{Lim}_j \urcorner\big) \to \mathrm{Pr}\big(\ulcorner C^\phi \urcorner\big)$, d.h. PRA $\vdash \mathrm{Pr}\big(\ulcorner \bigvee_{iRj} \mathsf{Lim}_j \urcorner\big) \to B^\phi$. Somit folgt mit Lemma C.1(6), daß PRA $\vdash \mathsf{Lim}_i \to B^\phi$ wie behauptet.
Angenommen $i \nVdash \Box C$. Dann gibt es ein j mit iRj, sodaß $j \nVdash C$. Aufgrund der Induktionshypothese gilt dann PRA $\vdash \mathsf{Lim}_j \to \neg C^\phi$, also PRA $\vdash$

$C^\phi \to \neg\mathsf{Lim}_j$, also PRA $\vdash \mathrm{Pr}(\ulcorner C^\phi \urcorner) \to \mathrm{Pr}(\ulcorner \neg\mathsf{Lim}_j \urcorner)$ und somit PRA $\vdash \neg\,\mathrm{Pr}(\ulcorner \neg\mathsf{Lim}_j \urcorner) \to \neg B^\phi$. Da iRj, gilt wegen Lemma C.1(4), daß PRA $\vdash \mathsf{Lim}_i \to \neg\,\mathrm{Pr}(\ulcorner \mathsf{Lim}_j \urcorner)$ und somit PRA $\vdash \mathsf{Lim}_i \to \neg B^\phi$ wie behauptet. $\square$

Lemma C.3 *Die* PRA *beweist* $\mathsf{Lim}_0 \to \neg\,\mathrm{Pr}(\ulcorner A^\phi \urcorner)$.

Beweis: Da ja $1 \not\Vdash A$, folgt mit Lemma C.2, daß PRA $\vdash \mathsf{Lim}_1 \to \neg A^\phi$, also auch PRA $\vdash A^\phi \to \neg\mathsf{Lim}_1$, also auch PRA $\vdash \mathrm{Pr}(\ulcorner A^\phi \urcorner) \to \mathrm{Pr}(\ulcorner \neg\mathsf{Lim}_1 \urcorner)$, also auch PRA $\vdash \neg\,\mathrm{Pr}(\ulcorner \neg\mathsf{Lim}_1 \urcorner) \to \neg\,\mathrm{Pr}(\ulcorner A^\phi \urcorner)$. Da $0R'1$, gilt wegen Lemma C.1(4), daß PRA $\vdash \mathsf{Lim}_0 \to \neg\,\mathrm{Pr}(\ulcorner \neg\mathsf{Lim}_1 \urcorner)$, und somit PRA $\vdash \mathsf{Lim}_0 \to \neg\,\mathrm{Pr}(\ulcorner A^\phi \urcorner)$ wie behauptet. $\square$

Diese Lemmata erlauben uns nun das erste Solovaysche Vollständigkeitsresultat zu beweisen.

Beweis von Satz 12.1 : Alle in PRA beweisbaren Aussagen sind in $\mathbb{N}$ wahr. Deshalb gilt wegen Lemma C.1(5) für $1 \le i \le n$, daß $\mathbb{N} \models \mathsf{Lim}_i \to \mathrm{Pr}(\ulcorner \neg\mathsf{Lim}_i \urcorner)$. Angenommen $\mathbb{N} \models \mathsf{Lim}_i$. Dann auch $\mathbb{N} \models \mathrm{Pr}(\ulcorner \neg\mathsf{Lim}_i \urcorner)$ und somit PRA $\vdash \neg\mathsf{Lim}_i$. Also $\mathbb{N} \models \neg\mathsf{Lim}_i$ im Widerspruch zur Annahme. Wir haben also gezeigt, daß $\mathbb{N} \not\models \mathsf{Lim}_i$ für $1 \le i \le n$. Andererseits gilt wegen Lemma C.1(3), daß $\bigvee\limits_{0 \le i \le n} \mathsf{Lim}_i$. Also $\mathbb{N} \models \mathsf{Lim}_0$. Wegen Lemma C.3 folgt nun $\mathbb{N} \models \neg\,\mathrm{Pr}(\ulcorner A^\phi \urcorner)$ und somit beweist PRA nicht A^ϕ. $\square$

Unter Verwendung der bisherigen Resultate schreiten wir num zum

Beweis von Satz 12.2 : Wir zeigen, daß für jede modallogische Formel A folgende drei Aussagen äquivalent sind

i) $\mathsf{PRL} \vdash A^*$

ii) $\mathsf{PRL}_\omega \vdash A$

iii) $\mathbb{N} \models A^\phi$ für alle arithmetischen Interpretation ϕ.

Klarerweise folgt ii) aus i). Da $\mathbb{N}$ alle arithmetischen Interpretationen der Axiome von PRL_ω validiert, folgt iii) aus ii). Daß i) aus iii) folgt, beweisen wir mit Kontraposition. Nehmen wir also an, daß $\mathsf{PRL} \not\vdash A^*$. Wir müssen zeigen, daß es eine arithmetische Interpretation ϕ gibt, sodaß $\mathbb{N} \not\models A^\phi$.

Wegen (dem Beweis von) Satz 12.1 gibt es ein endlich wohlfundiertes Modell $(W, R, \Vdash)$ mit $w = \{1, \ldots, n\}$ und $1Ri$ für alle $1 < i \leq n$, sodaß $1 \nVdash A^*$. Seien W' und R' und ϕ wie im Beweis für Satz 12.1. Es gilt nun Lemma C.2 für alle Unterformeln von A^* und somit insbesondere für alle Unterformeln von A.
Da $1 \nVdash A^*$, gilt $1 \nVdash A$ aber $1 \Vdash \Box C_i \to C_i$ für $i = 1, \ldots, k$, wobei $C_1, \ldots, C_k$ eine Liste aller Unterformeln C von A ist, sodaß $\Box C$ eine Unterformel von A ist.
Wir zeigen jetzt durch Induktion über alle Unterformeln B von A, daß

(1) wenn $1 \Vdash B$, dann PRA $\vdash \mathsf{Lim}_0 \to B^\phi$

(2) wenn $1 \nVdash B$, dann PRA $\vdash \mathsf{Lim}_0 \to \neg B^\phi$.

Angenommen B ist eine aussagenlogische Konstante p. Wenn $1 \Vdash p$, dann auch $0 \Vdash p$ und somit ist Lim_0 eines der Glieder der Disjunktion p^ϕ, also PRA $\vdash \mathsf{Lim}_0 \to p^\phi$. Wenn $1 \nVdash p$, dann $0 \nVdash p$ und somit ist Lim_0 keines der Glieder der Disjunktion p^ϕ, woraus wegen Lemma C.1(1) folgt, daß PRA $\vdash \mathsf{Lim}_0 \to \neg p^\phi$.
Angenommen $B \equiv B_1 \to B_2$ und es gelte als Induktionshypothese die Behauptung für B_1 und B_2. Wenn $1 \Vdash B_1 \to B_2$, dann $1 \nVdash B_1$ oder $1 \Vdash B_2$. Wenn $1 \nVdash B_1$, dann gilt aufgrund der Induktionshypothese, daß PRA $\vdash \mathsf{Lim}_0 \to \neg B_1^\phi$ und somit PRA $\vdash \mathsf{Lim}_0 \to B_1^\phi \to B_2^\phi$, d.h. PRA $\vdash \mathsf{Lim}_0 \to (B_1 \to B_2)^\phi$. Wenn $1 \Vdash B_2$, dann gilt aufgrund der Induktionshypothese, daß PRA $\vdash \mathsf{Lim}_0 \to B_2^\phi$ und somit PRA $\vdash \mathsf{Lim}_0 \to B_1^\phi \to B_2^\phi$, d.h. PRA $\vdash \mathsf{Lim}_0 \to (B_1 \to B_2)^\phi$. Wenn $1 \nVdash B_1 \to B_2$, dann $1 \Vdash B_1^\phi$ und $1 \nVdash B_2^\phi$. Aufgrund der Induktionshypothese gilt dann PRA $\vdash \mathsf{Lim}_0 \to B_1^\phi$ und $\vdash \mathsf{Lim}_0 \to \neg B_2^\phi$, also PRA $\vdash \mathsf{Lim}_0 \to \neg(B_1^\phi \to B_2^\phi)$, also PRA $\vdash \mathsf{Lim}_0 \to \neg(B_1 \to B_2)^\phi$.
Angenommen $B \equiv \Box C$ und es gelte als Induktionshypothese die Behauptung für C.
Wenn $1 \Vdash \Box C$, dann $i \Vdash C$ für $1 < i \leq n$. Wegen Lemma C.2 gilt dann PRA $\vdash \mathsf{Lim}_i \to C^\phi$ für $1 < i \leq n$. Da $1 \Vdash \Box C \to C$, gilt auch $1 \Vdash C$. Wegen Lemma C.2 gilt dann PRA $\vdash \mathsf{Lim}_1 \to C^\phi$ und wegen der Induktionshypothese gilt auch PRA $\vdash \mathsf{Lim}_0 \to C^\phi$. Wegen Lemma C.1(3) gilt aber PRA $\vdash \bigvee_{0 \leq i \leq n} \mathsf{Lim}_i$ und somit PRA $\vdash C^\phi$, also auch PRA $\vdash \mathrm{Pr}(\ulcorner C^\phi \urcorner)$, d.h. PRA $\vdash (\Box C)^\phi$ und deshalb auch PRA $\vdash \mathsf{Lim}_0 \to (\Box C)^\phi$.
Wenn $1 \nVdash \Box C$, dann $i \nVdash C$ für ein i mit $1 < i \leq n$. Wegen Lemma C.2 gilt dann PRA $\vdash \mathsf{Lim}_i \to \neg C^\phi$. Also auch PRA $\vdash C^\phi \to \neg \mathsf{Lim}_i$, also PRA $\vdash$

$\mathrm{Pr}(\ulcorner C^\phi \urcorner) \to \mathrm{Pr}(\ulcorner \neg \mathsf{Lim}_i \urcorner)$ und somit PRA $\vdash \neg \mathrm{Pr}(\ulcorner \neg \mathsf{Lim}_i \urcorner) \to \neg \mathrm{Pr}(\ulcorner C^\phi \urcorner)$, d.h. PRA $\vdash \neg \mathrm{Pr}(\ulcorner \neg \mathsf{Lim}_i \urcorner) \to \neg(\Box C)^\phi$. Da $0R'i$, gilt wegen Lemma C.1(4), daß PRA $\vdash \mathsf{Lim}_0 \to \neg \mathrm{Pr}(\ulcorner \neg \mathsf{Lim}_i \urcorner)$ und somit PRA $\vdash \mathsf{Lim}_0 \to \neg(\Box C)^\phi$.
Da $1 \not\Vdash A$, gilt PRA $\vdash \mathsf{Lim}_0 \to \neg A^\phi$ und somit $\mathbb{N} \models \mathsf{Lim}_0 \to \neg A^\phi$. Aus dem Beweis von Satz 12.1 wissen wir, daß $\mathbb{N} \models \mathsf{Lim}_0$, und somit $\mathbb{N} \models \neg A^\phi$. Also $\mathbb{N} \not\models A^\phi$ wie behauptet. $\Box$

Als wichtige Konsequenz erhalten wir folgenden

Satz C.1 *Die Logik* PRL_ω *ist entscheidbar.*

Beweis: Die Behauptung folgt aus der Entscheidbarkeit von PRL, da $\mathsf{PRL}_\omega \vdash A$ genau dann, wenn $\mathsf{PRL} \vdash A^*$. $\Box$

Also ist es entscheidbar, ob eine modallogische Formel unter allen arithmetischen Interpretationen gilt.

D 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 $f : [X]^k \to m$, wobei $m = \{0, 1, \ldots, m-1\}$.[18] Eine Teilmenge Y von X heißt *homogen* (bzgl. der Färbung f), wenn f auf $[Y]^k$ konstant ist. Wenn $k, m, \ell, n \in \mathbb{N}$, dann schreiben wir

$$n \longrightarrow (\ell)^k_m$$

als Abkürzung für die Aussage, daß für jede Färbung $f : [n]^k \to m$ eine bzgl. f homogene ℓ-elementige Teilmenge $X \subseteq n$ existiert.[19] In [GRS] z.B. findet man einen Beweis des folgenden Satzes

Satz D.1 (endlicher Ramsey Satz)
$\forall k, m, \ell. \in \mathbb{N}.\ \exists n \in \mathbb{N}.\ n \longrightarrow (\ell)^k_m$.

Beweis: Wir führen Induktion über k.
Für $k = 1$ wähle $n = 1 + m \cdot (\ell - 1)$. Wenn $f : n \to m$ 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 $\forall m, \ell \in \mathbb{N}.\ \exists n \in \mathbb{N}.\ n \longrightarrow (\ell)^k_m$.
Seien $m, \ell \in \mathbb{N}$. Aufgrund der Induktionshypothese gibt es ein $t \in \mathbb{N}$, sodaß $t \longrightarrow (\ell)^k_m$. Wir setzen $n = 2m^c$, wobei $c = \sum_{i=k}^{t-1} \binom{i}{k}$. Sei $f : [n]^{k+1} \to m$.

[18] Im Falle $k = m = 2$ läßt sich $f : [X]^2 \to 2$ verstehen als *ungerichteter* Graph mit Knotenmenge X.

[19] also ist insbesondere $n \geq \ell$

Wähle paarweise verschiedene $a_1, \ldots, a_k \in n$ und setze $S_k = n \setminus \{a_1, \ldots, a_k\}$. 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 $T \subseteq \{a_1, \ldots, a_i\}$ 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 $a_1, \ldots, a_t \in n$ und $S_k, S_{k+1}, \ldots, S_t$ geeignet wählen können. Angenommen $i_1 < \cdots < i_k < i_{k+1} \leq t$. Dann gilt $a_{i_{k+1}} \in S_{i_{k+1}-1} \subseteq S_{i_k}$ und somit gilt $f(\{a_{i_1}, \ldots, a_{i_k}, a_{i_{k+1}}\}) = f(\{a_{i_1}, \ldots, a_{i_k}, x\})$ für alle $x \in S_{i_k}$. Wir betrachten eine m-Färbung f^* der k-elementigen Teilmengen von $\{a_1, \ldots, a_t\}$ mit

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

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

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

Also ist f auf $\left[\{b_1, \ldots, b_\ell\}\right]^{k+1}$ 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 ℓ 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 D.1 geeignet verschärfen.

Definition D.1
Eine endliche Teilmenge X von $\mathbb{N}$ heißt relativ groß *(engl.* "large"*), wenn $|M| \geq \min(M)$. Wir schreiben*

$$n \xrightarrow{*} (\ell)^k_m$$

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

Satz D.2 (J. Paris, L. Harrington)
Die Aussage

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

ist wahr, kann aber nicht in PA *bewiesen werden.*

“Wahr” heißt im vorliegendem 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 D.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 D.4 (unendlicher Ramsey Satz)
Wenn k und m natürliche Zahlen sind und $f : [\mathbb{N}]^k \to m$, 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 D.3 und D.4 die Bezugnahme auf Teilmengen von $\mathbb{N}$ erfordert, d.h. die Sprache der Arithmetik 2. Stufe.

Lemma D.1 *Es gilt* PH*.*

Beweis: Wir argumentieren mit Widerspruch, nehmen also an, es gebe natürliche Zahlen k, m, ℓ, sodaß für jede natürliche Zahl n eine Funktion $f : [n]^k \to m$ existiert, sodaß f auf jeder relativ großen mindestens ℓ-elementigen Teilmenge von n verschiedene Werte annimmt. Die Menge dieser Gegenbeispiele ist bzgl. $\subseteq$ ein endlich verzweigender unendlicher Baum. Wegen Satz D.3 existiert in diesem Baum ein unendlicher Pfad, d.h. eine Funktion $f : [\mathbb{N}]^k \to m$, 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 D.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 $n_1 < \cdots < n_t$ 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 ℓ 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. □

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 D.2 *Sei $X \subseteq \mathbb{N}$. Eine Funktion $f : [X]^k \to \mathbb{N}$ 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 D.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 $f_1, \ldots, f_n : [d]^k \to d$ 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]. □

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

Definition D.3
Sei $\mathcal{M}$ ein Modell der PA *und Γ eine endliche Menge von Formeln in der Sprache der* PA*. Eine Teilmenge I von $|\mathcal{M}|$ heißt* sequence of diagonal indiscernibles (sdi) *für Γ, wenn für jede Formel $\phi(u_1, \ldots, u_k, v_1, \ldots, v_m)$ in Γ und $a, x_1, \ldots, x_n, y_1, \ldots, y_n \in I$ mit $a < x_1 < \ldots x_n$ und $a < y_1 < \ldots y_n$ die Formel*

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

in $\mathcal{M}$ gilt.

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

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

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

im Standardmodell $\mathbb{N}$ *gilt.*
Angenommen $m > 2n$*. Sei* w *die kleinste Zahl mit* $w \longrightarrow (m+n)^{2n+1}_{\ell+1}$ *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]. □

Sei Δ_0 die Menge der arithmetischen Formeln, wo alle Quantoren beschränkt sind. Man kann leicht zeigen, daß es ein arithmetisch definierbares Wahrheitsprädikat für geschlossene Δ_0 Formeln gibt und für solche Formeln Beweisbarkeit und Gültigkeit im Standardmodell koinzidieren. Wenn in Lemma D.3 die ϕ_i alle Δ_0 Formeln sind, gilt deshalb dann für alle $i = 1, \ldots, \ell$, daß

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

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

Lemma D.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)* Δ_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 D.2 :
Wir wissen bereits aus Lemma D.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 $f_1, \ldots, f_c : [d]^{2c+1} \to d$ ein $Y \subseteq (c, d)$ existiert mit $|Y| \geq w$ und Y min-homogen bzgl. aller f_i.
Aus Lemma D.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 Δ_0 Formeln $\phi(u_1, \ldots, u_c, v_1, \ldots, v_c)$ mit Gödelnummer $< c$. Somit ist I insbesondere eine s.d.i. für alle standard Δ_0 Formeln.
Da $|I| \geq c$ und c eine Nichtstandardzahl ist, ist I unendlich. Sei nun $x_0 < x_1 < \cdots < x_n < x_{n+1} < \ldots$ 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 Δ_0 Formeln. Sei $|\mathcal{N}| = \{y \in |\mathcal{M}| \mid y < x_i \text{ for some } i\}$. Wegen Lemma D.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 D.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 Δ_0 Formel ist, gilt $w' \longrightarrow (3c+1)_c^{2c+1}$ eine Δ_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 $f_1, \ldots, f_c : [d']^{2c+1} \to d'$ 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. $\square$