



universität  
wien

# MASTERARBEIT

Titel der Masterarbeit

“Deflationismus und formale Wahrheitstheorien  
nach Tarski und Kripke  
(unter besonderer Berücksichtigung von Kripkes Fixpunktsemantik)”

Verfasser

Leo Stadlmüller, BA

angestrebter akademischer Grad

Master of Arts (MA)

Wien, 2012

Studienkennzahl lt. Studienblatt: A 066 941

Studienrichtung lt. Studienblatt: Philosophie

Betreuerin: MMag. DDr. Esther Ramharter

# Inhaltsverzeichnis

|          |                                                                                 |           |
|----------|---------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Einleitung</b>                                                               | <b>3</b>  |
| <b>2</b> | <b>Tarskis Erbe</b>                                                             | <b>4</b>  |
| 2.1      | Technische Vorbemerkungen . . . . .                                             | 4         |
| 2.2      | Tarskis Theorem . . . . .                                                       | 6         |
| 2.3      | TB . . . . .                                                                    | 7         |
| 2.4      | Kompositionale Wahrheit . . . . .                                               | 12        |
| 2.4.1    | CT $\uparrow$ . . . . .                                                         | 12        |
| 2.4.2    | CT . . . . .                                                                    | 13        |
| <b>3</b> | <b>Deflationismus und Konservativität</b>                                       | <b>15</b> |
| 3.1      | Allgemeines zum Deflationismus . . . . .                                        | 15        |
| 3.2      | Das Generalisierungsproblem . . . . .                                           | 18        |
| 3.3      | Konservativität . . . . .                                                       | 19        |
| 3.4      | Deflationismus ohne Konservativität . . . . .                                   | 21        |
| <b>4</b> | <b>Kripke</b>                                                                   | <b>22</b> |
| 4.1      | Kritik an Tarski . . . . .                                                      | 22        |
| 4.2      | Induktive Definitionen, monotone Operatoren und der minimale Fixpunkt . . . . . | 25        |
| 4.3      | Die Semantik von Kripkes Theorie . . . . .                                      | 28        |
| 4.4      | KF . . . . .                                                                    | 35        |
| 4.5      | Wahrheitswert-Überschneidungen und Supervaluierung . . . . .                    | 35        |
| 4.6      | Revenge . . . . .                                                               | 37        |
| <b>5</b> | <b>Deflationismus nach Kripke: Die zeitgenössische Debatte</b>                  | <b>40</b> |
| 5.1      | Eine neue Form des Deflationismus . . . . .                                     | 40        |
| 5.2      | Field . . . . .                                                                 | 42        |
| 5.3      | PKF und inferentieller Deflationismus . . . . .                                 | 46        |
| 5.4      | Zusammenfassung . . . . .                                                       | 51        |
| <b>6</b> | <b>Anhang</b>                                                                   | <b>53</b> |
| 6.1      | Literaturverzeichnis . . . . .                                                  | 53        |
| 6.2      | Abstract . . . . .                                                              | 56        |

## 1 Einleitung

Seit den Anfängen der Philosophie ist Wahrheit ein zentraler Begriff philosophischer Reflexion. Dennoch ist es angemessen zu behaupten, dass dieser Begriff erstmals im letzten Jahrhundert formal präzisiert und dadurch exakten und strengen formalen Regeln unterworfen wurde. Diese Arbeit verdanken wir Alfred Tarski, welcher als Gründervater jeglicher formaler Wahrheitstheorien gilt. Durch seine und darauf folgende Arbeit wurde es möglich, umgangssprachlich formulierte philosophische Intuitionen mathematisch zu modellieren und rigoros Konsequenzen daraus abzuleiten. Nichtsdestotrotz ist das Verhältnis von philosophischer und formaler Wahrheit dadurch nicht geklärt; oder gar der philosophische Wahrheitsbegriff durch einen formal-mathematischen ersetzt worden. Denn einerseits liegen vielen formalen Theorien philosophische Intuitionen zugrunde, andererseits lassen sich aus formalen Ergebnissen erneut Rückschlüsse auf die Philosophie ziehen. Was diese Verbindungen sind und wie sie hergestellt werden können, ist neben der Präsentation und Diskussion von Tarskis und Kripkes formalen Wahrheitstheorien ein Hauptanliegen der vorliegenden Arbeit. Ein zweites zentrales Anliegen dieser Arbeit ist eine detaillierte Ausarbeitung von Kripkes Fixpunktsemantik. Dies scheint mir aus mehreren Gründen sehr wichtig zu sein. Denn obwohl das zentrale Lemma (das Kripke Fixpunktlemma) eigentlich sehr einfach zu beweisen ist, benötigt man ein Verständnis der Feinstruktur dahinter um weitere wichtige Theoreme daraus abzuleiten zu können, um die Unterschiede zu anderen Theorien zu verstehen und um die Entwicklung nach Kripke nachvollziehen zu können.

Abgesehen von dieser Einleitung und dem Anhang besteht diese Arbeit aus vier Kapiteln. Die ersten beiden stellen eine Präsentation von Tarskis Wahrheitstheorien in ihrer modernen Formulierung (Kap. 2) und der daran anschließenden philosophischen Diskussion über den Deflationismus (Kap.3) dar. Die beiden anderen bestehen aus einer Auseinandersetzung mit dem formalen Rahmen von Kripkes Theorie (Kap. 4) und der daran anschließenden formalen Weiterentwicklung seiner Theorie im Zusammenhang mit der modernen Deflationismusdebatte (Kap.5). Insofern verläuft die Arbeit (abgesehen von dem erwähnten Schwerpunkt auf Kripkes Fixpunktsemantik) symmetrisch und erlaubt es die zentralen Unterschiede und Problemstellungen der Theorien und philosophischen Debatten klar herauszuarbeiten. Da es sehr gut argumentierbar ist, dass Tarskis und Kripkes Theorien die zentralen Grundlagen der modernen Forschung darstellen, wird dem Leser damit eine solide Grundlage für weitere Arbeit am Thema geschaffen.

Ich hatte das Glück ungewöhnlich viel Unterstützung bei der Erarbeitung der für diese Arbeit relevanten Inhalte zu haben. Ich bedanke mich diesbezüglich bei allen Mit-

gliedern des ‘Wiener Forum für Analytische Philosophie‘ für zahlreiche Diskussionen in den Kolloquien, bei allen Teilnehmern der Konferenz ‘Nothing but the Truth‘ im März 2012 und bei allen Teilnehmern des Seminars ‘Philosophische und formale Wahrheitstheorien‘, das Georg Schiemer und ich im WS 2011 gehalten haben. Für all dies bin ich darüber hinaus Prof. Richard Heinrich, Prof. Elisabeth Nemeth, Prof. Esther Ramharter und dem Dekanat des Instituts für Philosophie und Bildungswissenschaften für überaus freundliche Unterstützung zu Dank verpflichtet. Außerdem bedanke ich mich bei Prof. Hannes Leitgeb für die freundliche Aufnahme an seinem ‘Munich Center for Mathematical Philosophy‘ während meines Erasmusaufenthalts in München im SS 2011.

Außerdem bin ich den folgenden Personen für sehr aufschlussreiche Diskussionen zu Dank verpflichtet: Martin Fischer, Ole Hjortland, Leon Horsten, Paul Horwich, Jeffrey Ketland, Naomi Osorio-Kupferblum, Hannes Leitgeb, Julien Murzi, Thomas Schindler, Katharina Sodoma, Michael Toppel und Maximilian Wieländer.

Ganz besonders bedanke ich mich bei Günther Eder und Georg Schiemer für zahlreiche Diskussionen und Unterstützung bei allen formalen und philosophischen Angelegenheiten dieser Arbeit.

Nicht zuletzt bedanke ich mich bei meiner Betreuerin MMag. DDr. Esther Ramharter für ihr Engagement und ihre konstruktive Kritik bei allen mathematischen und philosophischen Belangen der Arbeit.

Nichtsdestotrotz bin ich für den hier präsentierten Inhalt selbst verantwortlich und hoffe sehr, dass er in Anbetracht der gerade aufgelisteten Hilfestellungen nicht allzu bescheiden ausgefallen ist.

## 2 Tarskis Erbe

In diesem Abschnitt werden die wichtigsten Wahrheitstheorien von Alfred Tarski in ihrer modernen Formulierung präsentiert. Im Zuge dessen werden einige zentrale Probleme ausgearbeitet, welche im weiteren Verlauf der Arbeit wichtig sein werden. Da Tarskis Wahrheitstheorien eine wichtige historische Grundlage bilden und die sich daraus ergebenden Problemstellungen für alle weiteren Diskussion zentral sind, ist deren Darstellung sowohl ein natürlicher Ausgangspunkt, als auch eine wichtige Einleitung.

### 2.1 Technische Vorbemerkungen

Um unserer Diskussion einen formalen Boden, auf dem wir gehen können, zu geben, müssen einige technische Vorbedingungen geklärt werden. Da unsere Basissprache - diejenige Sprache über die wir sprechen - im Zuge aller technischen Überlegungen Peano Arithmetik (= PA) sein wird, muss diese dargestellt werden. Außerdem benötigen wir

eine passende Kodierung dieser Sprache. Abgesehen davon nehme ich einfach an, dass der Leser mit den grundlegenden Eigenschaften der natürlichen Zahlen vertraut ist. Die Menge der natürlichen Zahlen bezeichnen wir als  $\omega$ .

Die Sprache der Peano Arithmetik  $\mathcal{L}_{PA}$  besteht aus einem logischen und einem mathematischen Vokabular, das gemeinsam die Grundlage für die Axiome derselben Theorie bildet. *Das logische Vokabular von  $\mathcal{L}_{PA}$*  besteht aus den logischen Konnektiven  $\wedge, \neg, \forall$  ( $\vee, \rightarrow, \leftrightarrow, \exists$  werden durch diese 3 definiert), sowie aus unendlich viele Variablen  $x_i$  für  $i \in \mathbb{N}$  und dem Identitätssymbol  $=$ . *Das mathematische Vokabular von  $\mathcal{L}_{PA}$*  besteht aus  $s$  (Nachfolger),  $0$  (Nullkonstante) und den Funktionssymbolen  $+$  und  $\times$  für Addition und Multiplikation.

In der Sprache  $\mathcal{L}_{PA}$  kann nun die Theorie PA formuliert werden. Diese besteht aus folgenden nichtlogischen Axiomen:

1.  $\neg \exists (s(x) = 0)$
2.  $\forall x \exists y (s(x) = y)$
3.  $\forall x \forall y (s(x) = s(y) \rightarrow x = y)$
4.  $\forall x (x + 0) = x$
5.  $\forall x \forall y (s(x + y) = x + s(y))$
6.  $\forall x (x \times 0) = 0$
7.  $\forall x \forall y (x \times s(y) = (x \times y) + x)$
8.  $(\phi(0) \wedge \forall x (\phi(x) \rightarrow \phi(s(x)))) \rightarrow \forall x (\phi(x))$  für alle  $\phi(x) \in \mathcal{L}_{PA}$

Weiters sei die Sprache  $\mathcal{L}_T$  die Sprache  $\mathcal{L}_{PA}$  mit Wahrheitsprädikat und die Theorie  $PA_T$  die Theorie PA formuliert in  $\mathcal{L}_T$  (vgl. [37], S. 28f).

Die Ausdrücke der Sprache  $\mathcal{L}_T$  können außerdem durch natürliche Zahlen kodiert werden.<sup>1</sup> Das Numeral einer Zahl  $n \in \omega$  ist das Symbol  $0$  mit  $n$  Nachfolgersymbolen vorangestellt. Das Numeral der Zahl 3 ist z.B. SSS0 und wir bezeichnen dieses als  $\bar{3}$ . Für eine beliebige Zahl  $n$  bezeichnet  $\bar{n}$  das Numeral von  $n$ .

Da wir alle Ausdrücke  $\phi \in \mathcal{L}_T$  kodieren können, sei  $\ulcorner \phi \urcorner$  das Numeral des Kodes von  $\phi$ . Sätze der Form  $\phi(x)$  stehen für Formeln der Sprache  $\mathcal{L}$  der Arithmetik.

Ein Punkt über den logischen Konnektiven ( $\neg, \wedge, \forall$ ) steht für das ihnen korrespondierende Funktionssymbol in  $\mathcal{L}$ . So steht z.B.  $\dot{\neg}$  für das Funktionssymbol in  $\mathcal{L}$ , welches, wenn es auf eine Formel  $\phi$  angewendet wird, deren Negation  $\neg\phi$  liefert. Analoges gilt für  $\dot{\wedge}$  und  $\dot{\forall}$ .

Weiters benötigen wir eine Substitutionsfunktion  $x(t/s)$ , die für eine Formel  $x$ , einen Term  $s$  und einen Term  $t$ , das Resultat der Substitution von  $s$  mit  $t$  liefert. Ähnlich

<sup>1</sup>Diese Kodierung nennt man Gödel-Kodierung. Details dazu finden sich in z.B. in [7], Kap.14 und [59], Kap 15.

liefert  $x(t/v)$  das Resultat der Substitution von  $t$  mit der Variable  $v$  in einer Formel. Das Symbol  $^\circ$  ist die Evaluierungsfunktion, die für einen geschlossenen Term von  $\mathcal{L}$  dessen Wert liefert. Z.B. drückt  $x^\circ = y$  aus, dass  $y$  der Wert von  $x$  ist; und  $x^\circ = y^\circ$  drückt aus, dass der Wert von  $x$  gleich dem Wert von  $y$  ist.

Die Menge aller Sätze von  $\mathcal{L}$  wird durch  $Sent(x)$  bezeichnet, die Menge aller Sätze von  $\mathcal{L}_T$  durch  $Sent_T(x)$ . (vgl. [35], Kap. 5)

Grundsätzlich sind damit alle Anliegen der Notation geklärt. Für weitere Details sei jedoch auf [35], [7] und [59] verwiesen. Da diese Notation sehr kompliziert sein kann, werde ich aus Gründen besserer Anschaulichkeit öfter zu einer einfacheren (obwohl nicht ganz korrekten) Notation wechseln. Dies ist sowohl zweckdienlich als auch weit verbreitet (siehe [37], [24]). Das heißt wir schreiben manchmal Sätze der Form

$$\forall \phi \in \mathcal{L}_{PA} : T(\phi \wedge \psi) \leftrightarrow T\phi \wedge T\psi,$$

anstatt dem viel schwieriger zu lesenden Satz

$$\forall x \forall y (Sent(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow T(x) \wedge T(y))),$$

den der Leser gerne nach den gerade gegebenen Definitionen entschlüsseln darf.

## 2.2 Tarskis Theorem

Das erste limitative Resultat aus dem Bereich der mathematischen Logik ist Tarskis Theorem. Dieses zeigt, dass keine Theorie, die minimalen formalen Anforderungen genügt, ihr eigenes (unrestriktiertes) Wahrheitsprädikat ausdrücken kann; d.i. alle Sätze der Form  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$ , für alle  $\phi \in \mathcal{L}_T$  beweisen kann. Intuitiv bedeutet das, dass ‘es keine Wahrheitsdefinition für die Arithmetik in der Arithmetik gibt’ ([16], S.195). Es gilt nämlich für jede Theorie mit Wahrheitsprädikat, die stark genug ist, um ihre eigene Syntax auszudrücken und Diagonalisierung erlaubt<sup>2</sup>, dass es einen Satz der Form  $\lambda \leftrightarrow \neg T(\ulcorner \lambda \urcorner)$  gibt. Dieser Satz  $\lambda$  ist der Lügnersatz und sagt intuitiv von sich selbst aus, dass er falsch ist.

**Theorem 2.1** *Eine Theorie, die den Lügnersatz beweist, kann nicht alle T-Bikonditionale der Form  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$ , für alle  $\phi \in \mathcal{L}_T$  ableiten.*

*Beweis:* Angenommen es gilt  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$ , für alle  $\phi \in \mathcal{L}_T$ . Dann gilt für den Lügnersatz  $T(\ulcorner \lambda \urcorner) \leftrightarrow \lambda$ . Da  $\lambda$  equivalent ist zu  $\neg T(\ulcorner \lambda \urcorner)$ , folgt daraus  $T(\ulcorner \lambda \urcorner) \leftrightarrow \neg T(\ulcorner \lambda \urcorner)$ .  $\square$

Tarski hat aufgrund von diesem Problem darauf geschlossen, dass die Anforderung einer Wahrheitsdefinition in einer unrestrictierten (universalen) Sprache, für eben diese

<sup>2</sup>siehe [16], [7], [18] [59] für eine Erklärung dieser Begriffe

Sprache selbst, unmöglich ist. Diese Tatsache veranlasste ihn auch dazu, skeptisch bezüglich der Möglichkeit einer Wahrheitsdefinition für die Umgangssprache zu sein, da diese derart universal ist:

Der Versuch, eine strukturelle Definition des Terminus ‘wahre Aussage’ aufzubauen, stösst - auf die Umgangssprache angewendet - auf Schwierigkeiten, die wir nicht überwinden können. ([62], S. 277f)

Sind obige Bemerkungen richtig, so scheint selbst die Möglichkeit eines konsequenten und dabei mit den Grundsätzen der Logik und dem Geiste der Umgangssprache übereinstimmenden Gebrauchs des Ausdrucks ‘wahre Aussage’ und, was daraus folgt, die Möglichkeit des Aufbaus irgend welcher korrekter Definitionen dieses Ausdrucks sehr in Frage gestellt. ([62], S. 279)

Wäre diese Diagnose zutreffend, könnte man zurecht daran zweifeln, ob eine formale Wahrheitstheorie außerhalb der mathematischen Logik etwas zu sagen hat. Die Tradition nach Tarski ist ihm bei diesem Zweifel jedoch nicht gefolgt. Einerseits könnte man nämlich gegen Tarski einfach den Allgemeinheitsanspruch fallen lassen und sich passenden Fragmenten der Umgangssprache zuwenden; andererseits könnte man versuchen eine Wahrheitstheorie aufzustellen, welche der ‘universalen’ Umgangssprache gerecht wird. Letzteres steht, wie wir noch sehen werden, spätestens seit Saul Kripikes Arbeit ([43]) im Raum und wird heute als legitime Herangehensweise an das von Tarski gestellte Problem betrachtet.

## 2.3 TB

Angesichts der Probleme, die der Lügnersatz verursacht, war Tarskis Lösung eine Typisierung des Wahrheitsprädikats. Informell bedeutet das, dass das Wahrheitsprädikat nicht unrestrictiert auf sich selbst angewendet werden darf. D.h; dass wir keine Sätze der Form  $T...TT(\ulcorner 2 + 2 = 4 \urcorner)$  bilden können, sondern nur Sätze der Form  $T_n...T_1T(\ulcorner 2 + 2 = 4 \urcorner)$  und, dass auf der untersten Stufe die Applikation des Wahrheitsprädikats auf Sätze von  $\mathcal{L}_{PA}$  (PA ohne Wahrheitsprädikat) eingeschränkt wird. Diese restriktierte Form erlaubt es nun, eine widerspruchsfreie Theorie aufzustellen.

Zusätzlich zu ihrer Widerspruchsfreiheit, muss die Theorie jedoch auch material adäquat sein. Damit meint Tarski, dass alle Äquivalenzen der Form

(T) X ist wahr genau dann, wenn p,

wobei p eine Aussage der Basissprache und X der Name von p ist, ableitbar sein sollen. In Tarskis Worten:

Nun sind wir endlich imstande, die Bedingungen, unter denen wir den Gebrauch und die Definition des Terms ‘wahr’ vom sachlichen Standpunkt aus als angemessen berachten, in eine präzise Form zu bringen: wir wollen den Term ‘wahr’ so gebrauchen, daß alle Äquivalenzen der Form (T) behauptet werden können, und *wir wollen eine Definition der Wahrheit ‘angemessen’ nennen, wenn alle diese Äquivalenzen aus ihr folgen.* ([63], S. 145)

Da ich hier eine andere und modernere Terminologie als Tarski verwende, wird diese Anforderung in folgende Definition übersetzt:

**Definition 2.2** ([37], S. 49) *Eine Wahrheitsdefinition für die Sprache  $\mathcal{L}_{PA}$  ist material adäquat, genau dann, wenn  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  für jeden Satz  $\phi \in \mathcal{L}_{PA}$  in der Metatheorie beweisbar ist.*

Die einfachste Theorie, die diese Anforderung erfüllt, ist die Theorie TB (Tarski Bikonditionale), welche sich aus  $PA_T$  und den restriktierten Tarski-Bikonditionalen ergibt.<sup>3</sup>

**Definition 2.3** (vgl. [35], S. 53, [37], S. 50) *Die Theorie TB besteht aus dem Axiom:*

$$T(\ulcorner \phi \urcorner) \leftrightarrow \phi \text{ für alle } \phi \in \mathcal{L}_{PA}$$

TB hat folgende nette<sup>4</sup> Eigenschaften (vgl. [37], S. 51f):

- Die Modelle von TB basieren auf den natürlichen Zahlen. Die Extension des Wahrheitsprädikats sind daher alle arithmetischen Wahrheiten.
- TB ist konsistent.

Obwohl TB eine konsistente und material adäquate Theorie ist, erfüllt sie insbesondere zwei intuitiv wünschenswerte Anforderungen nicht:

- TB beweist keine infiniten Generalisierungen.
- TB ist konservativ über PA.

Da die Beweise dieser beiden Fakten sehr ähnlich sind, behandeln wir sie gemeinsam. Eine allgemeine Diskussion des Problems der Konservativität wird auf das nächste Kapitel verschoben. Ich beginne damit zu zeigen, dass TB keine infiniten Generalisierungen beweist. Zuerst wird der abstrakte Beweis präsentiert, dann gebe ich Beispiele. Die Anforderung infinite Generalisierungen zu beweisen, bedeutet informell, dass eine

<sup>3</sup>Ich folge nicht Tarskis ursprünglicher Präsentation, sondern orientiere mich an einer modernen axiomatischen Darstellung, wie sie in [35] und [37] gegeben wird.

<sup>4</sup>Ich verwende ‘nett’ im technischen Sinn und meine damit ‘von einem formalen Standpunkt aus wünschenswert’.

Wahrheitstheorie Sätze wie z.B.

Alle Sätze der Form ‘Wenn  $p$ , dann  $p$ ‘ sind wahr

beweisen soll. Es kann jedoch gezeigt werden, dass TB diese Form der Generalisierung nicht beweisen kann. Mit der Phrase ‘alle Sätze‘ verlangen wir nämlich eine unendliche Anzahl von Sätzen, TB kann jedoch nur eine endliche Anzahl dieser beweisen. Daher lässt sich im Fall von TB auch nur von finiten Generalisierungen sprechen (vgl. [35], S. 57).

**Theorem 2.4** ([35], S. 57f, [37], S. 69f) *Wenn  $TB \vdash \forall x(\phi(x) \rightarrow Tx)$  und  $\phi(x)$  ist eine  $\mathcal{L}$  - Formel, dann gibt es ein  $n$ , sodass PA beweist, dass es höchstens  $n$   $\phi$ ’s gibt; bzw, sodass gilt:  $PA \vdash \exists_{\leq n} x \phi(x)$ .*

Wobei  $\exists_{\leq n} x \phi(x)$  eine Abkürzung der Formel ist, die ausdrückt, dass es höchstens  $n$  Objekte gibt, die  $\phi(x)$  in der Sprache der Logik erster Stufe mit Identität erfüllen. Ungekürzt hat sie folgende Form:

$$\forall x_1 \dots \forall x_{n+1} (\phi(x_1) \wedge \dots \wedge \phi(x_{n+1}) \rightarrow x_1 = x_2 \vee \dots \vee x_1 = x_{n+1} \vee \dots \vee x_n = x_{n+1})$$

Daher ist jede beweisbare Generalisierung, die durch das Wahrheitsprädikat in TB ausgedrückt wird, nur eine finite Generalisierung. Der Beweis dafür basiert auf der Möglichkeit, jeden Beweis einer arithmetischen Formel mit Wahrheitsprädikat (aus TB) in einen Beweis in PA derselben Formel umzuwandeln.

*Beweis: Sei ein Beweis von  $\forall x(\phi(x) \rightarrow Tx)$  gegeben. Da dieser Beweis endlich ist, kann er auch nur endlich viele Sätze der Form  $T^\top \psi^\top \leftrightarrow \psi$  beinhalten. D.h. wir können diese Sätze auflisten*

- $T^\top \psi_1^\top \leftrightarrow \psi_1$
- $T^\top \psi_2^\top \leftrightarrow \psi_2$
- ...
- 
- 
- 
- $T^\top \psi_n^\top \leftrightarrow \psi_n$

*Sei  $\tau(x)$  die Formel*

$$(x = \ulcorner \psi_1 \urcorner \wedge \psi_1) \vee \dots \vee (x = \ulcorner \psi_n \urcorner \wedge \psi_n).$$

Ersetze das Wahrheitsprädikat überall im Beweis durch  $\tau$  und füge für alle  $i \leq n$  die Beweise für  $\tau(\ulcorner \psi_i \urcorner) \leftrightarrow \psi_i$  hinzu, wo diese Sätze die Disquotationssätze ersetzt haben. Dies führt zu einem Beweis von  $\forall x(\phi(x) \rightarrow \tau(x))$ . Gemeinsam mit dem logischen Theorem  $\exists_{\leq n} x \tau(x)$  liefert dies das erwünschten Resultat.  $\square$

Abhängig davon, welchen Satz wir für  $\phi(x)$  einsetzen, können wir Beispiele für dieses Theorem anführen. Bemerke zuerst, dass obwohl TB keine infiniten Generalisierungen beweist, sehr wohl z.B. jede Instanz von Kompositionalität beweist.

**Korollar 2.5** (vgl. [35], S. 57, [37], S. 70)  $\forall \phi \in \mathcal{L}_{PA} \text{ TB} \vdash T(\phi \wedge \psi) \leftrightarrow T(\phi) \wedge T(\psi)$ .

*Beweis:* das folgt aus Theorem 2.4, da jeder Beweis in TB in einen PA Beweis verwandelt werden kann und umgekehrt.  $\square$

Dennoch kann TB diese Instanzen nicht zu einem Theorem zusammenfassen.

**Korollar 2.6** ([37], S. 70)  $\text{TB} \not\vdash \forall \phi, \psi \in \mathcal{L}_{PA} : (T(\phi) \wedge T(\psi)) \leftrightarrow T(\phi \wedge \psi)$

*Beweis:* Das folgt direkt aus Theorem 2.4.  $\square$

Aus denselben Gründen beweist TB diese infiniten Generalisierungen auch nicht für andere logische Prinzipien. Als Beispiel ziehe ich das Gesetz des ausgeschlossenen Dritten heran.

**Korollar 2.7** ([37], S. 70)  $\text{TB} \not\vdash \forall \phi \in \mathcal{L}_{PA} : T(\phi) \vee T(\neg \phi)$

*Beweis:* Analog zu Korollar 2.6.  $\square$

Ähnliches gilt natürlich für logische Gesetze bezüglich der Negation und der Quantoren (vgl. [37], S. 70). Dies ist ein erster wichtiger Grund dafür, dass TB keine gute Wahrheitstheorie ist. Bezüglich des Mangels an Kompositionalität schreiben Leon Horsten:

In summary, (TB) fails to fully validate our intuitions concerning the *compositional nature* of the notion of truth (i.e. the fact the property of truth ‘distributes’ over the logical connectives). In fact, our intuition that truth is compositional is (...) just as basic as our intuition that truth is a disquotational device. Our truth theory has an obligation to either do justice to it or explain what is wrong with it. ([37], S. 70)

Bezüglich der mangelnden infiniten Generalisierungen schreibt Volker Halbach:

The main criticism that deflationist theories based on the disquotation sentences or similar axioms have to meet was raised by Tarski: the disquotation

sentences to not allow one to *prove* generalizations (...). ([35], S.61f)

Jeffrey Ketland bringt das Problem auf den Punkt und nennt TB aufgrund dieser Probleme  $\omega$  - unvollständig.

Compare this with the notion of  $\omega$ - *incompleteness* from mathematical logic.

A theory  $T$  in the language of arithmetic  $\mathcal{L}$  is  $\omega$  -incomplete if and only if there is a formula  $P(x)$  such that

1. For all natural numbers  $n$ ,  $T \vdash P(\bar{n})$
2.  $T$  does not imply  $\forall x P(x)$

([40], S. 84)

Ein weiteres Problem von TB ist, dass sie konservativ über PA ist.<sup>5</sup>

**Definition 2.8** ([35], S. 44, [37], S. 75) *Eine Wahrheitstheorie  $\mathcal{T}$  ist konservativ über einer Theorie  $S$ , formuliert in der Sprache  $\mathcal{L}$  ohne Wahrheitsprädikat, genau dann, wenn alle  $\mathcal{T}$  - Theoreme  $\phi$  in der Sprache  $\mathcal{L}$  auch Theoreme von  $S$  sind.*

Im Folgenden sei unsere Theorie  $\mathcal{S} = PA$ .

**Theorem 2.9** ([35], S. 55, [37], S. 75) *TB ist konservativ über PA.*

*Beweisskizze: Die Beweisidee ist dieselbe wie in Theorem 2.4. In einem Beweis von TB können nur endlich viele Disquotationsaxiome  $T \ulcorner \phi_i \urcorner \leftrightarrow \phi_i (i \leq n)$  vorkommen. In dem Beweis werden atomare Formeln der Form  $Ts$ , für einen Term  $s$ , mit folgender Formel substituiert:*

$$(s = \ulcorner \phi_1 \urcorner \wedge \phi_1) \vee \dots \vee (s = \ulcorner \phi_n \urcorner \wedge \phi_n),$$

*wobei  $\phi_1$  bis  $\phi_n$  sämtliche Sätze  $\phi_i$  der Disquotationsaxiome im Beweis sind. Um den Beweis zu vervollständigen, muss gezeigt werden, dass jeder TB-Beweis einer arithmetischen Formel in einen PA-Beweis der selben Formel transformiert werden kann.*

□

TB unterscheidet sich daher in seiner beweistheoretischen Stärke nicht von PA und fügt deshalb letzterer Theorie auch keine neuen Theoreme hinzu. Dass dies nicht auf jede Wahrheitstheorie zutrifft, wird sich im nächsten Abschnitt zeigen.

---

<sup>5</sup>dass dies ein Problem für TB ist, ist a priori nicht klar. Dies ist Gegenstand einer eigenen Debatte, die im nächsten Abschnitt behandelt wird.

## 2.4 Kompositionale Wahrheit

Eine Theorie, welche die oben besprochenen Probleme von TB überwindet, ist die Theorie CT (= Compositional Truth). Diese geht wie TB auf Tarski zurück und wird hier als Axiomatisierung seiner sogenannten induktiven Klauseln präsentiert. Diese stellen eine Wahrheitsdefinition für die Sprache  $\mathcal{L}$  der Arithmetik dar. Da wir Sätze mit Zahlen identifizieren, wird die Menge  $\mathcal{S}$  der wahren  $\mathcal{L}$  - Sätze eine Menge von Zahlen sein. Ich beginne mit Halbachs Version von Tarskis induktiven Klauseln und gebe dann zwei Definitionen der Theorie ‘Compositional Truth’ an.

**Definition 2.10** ([35], S. 64) (*Wahrheitsdefinition für  $\mathcal{L}$* ) Die Menge der wahren  $\mathcal{L}$  - Sätze ist die Menge  $\mathcal{S}$  von  $\mathcal{L}$  - Sätzen, welche folgende Bedingungen, für alle geschlossenen Terme  $s, t$  und  $\mathcal{L}$  - Sätze  $\phi, \psi, \forall v\chi$ , und  $\exists v\chi$ , erfüllt:

- (i)  $s = t \in \mathcal{S}$  gdw der Wert von  $s$  identisch ist mit dem Wert von  $t$ .
- (ii)  $\neg\phi \in \mathcal{S}$  gdw  $\phi \notin \mathcal{S}$ .
- (iii)  $\phi \wedge \psi$  gdw ( $\phi \in \mathcal{S}$  und  $\psi \in \mathcal{S}$ ).
- (iv)  $\phi \vee \psi$  gdw ( $\phi \in \mathcal{S}$  oder  $\psi \in \mathcal{S}$ ).
- (v)  $\forall v\chi \in \mathcal{S}$  gdw für alle geschlossenen Terme  $t$ ,  $\chi(t/v) \in \mathcal{S}$ .
- (vi)  $\exists v\chi \in \mathcal{S}$  gdw für einen geschlossenen Term  $t$ ,  $\chi(t/v) \in \mathcal{S}$ .

### 2.4.1 CT $\uparrow$

Im Folgende unterscheide ich die axiomatische Theorie CT $\uparrow$ , in welcher das Wahrheitsprädikat nicht im Induktionsaxiom vorkommen darf, von der Theorie CT, in welcher dies der Fall sein darf. Es wird sich zeigen, dass diese Unterscheidung einen sehr feinen, jedoch wichtigen Punkt impliziert, der in der Literatur kaum explizit gemacht wurde.<sup>6</sup>

**Definition 2.11** ([35], S. 65) (*CT $\uparrow$* ) Das System CT $\uparrow$  besteht aus allen Axiomen von PA und den folgenden Axiomen:

- CT1  $\forall s \forall t (T(s \doteq t) \leftrightarrow s^\circ = t^\circ)$
- CT2  $\forall x (Sent(x) \rightarrow (T(\neg x) \leftrightarrow \neg T x))$
- CT3  $\forall x \forall y (Sent(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow T(x) \wedge T(y)))$
- CT4  $\forall x \forall y (Sent(x \dot{\vee} y) \rightarrow (T(x \dot{\vee} y) \leftrightarrow T(x) \vee T(y)))$
- CT5  $\forall v \forall x (Sent(\dot{\forall} vx) \rightarrow (T(\dot{\forall} vx) \leftrightarrow \forall t T(x(t/v))))$
- CT6  $\forall v \forall x (Sent(\dot{\exists} vx) \rightarrow (T(\dot{\exists} vx) \leftrightarrow \exists t T(x(t/v))))$

<sup>6</sup>Eine Ausnahme ist natürlich [35], wo dies vollkommen klar gemacht wurde. Dennoch ist der Punkt aufgrund des Umfangs der Untersuchung nicht leicht zu erkennen.

Wie leicht ersichtlich ist, beweist  $CT \upharpoonright$  die geforderten Generalisierungen aus Abschnitt 2.3. So ist zum Beispiel die Anforderung der Kompositionalität schon in den Axiomen enthalten. Das Gesetz des ausgeschlossenen Dritten ist ebenfalls leicht aus den Axiomen zu beweisen. Dennoch gibt es Generalisierungen, welche in  $CT \upharpoonright$  nicht beweisbar sind. Zum Beispiel beweist  $CT \upharpoonright$  nicht, dass alle Sätze der Form  $0 = 0 \wedge \dots \wedge 0 = 0$  wahr sind ([35], S.102) und auch nicht, dass die Substitution von identischen Termen Wahrheit beibehält. D.i.:

**Theorem 2.12** ([35], S. 102)  $CT \upharpoonright \not\vdash \forall x \forall s \forall t (Sent(\forall v x \wedge s^\circ = t^\circ \rightarrow (Tx(s/v) \leftrightarrow Tx(t/v)))$

*Beweis:* siehe ([35], S. 102).  $\square$

Weiters - und das ist etwas überraschend - gilt für  $CT \upharpoonright$  folgendes:

**Theorem 2.13** (Halbach)  $CT \upharpoonright$  ist konservativ über PA.

*Beweis:* siehe ([35], Kap. 8.1).  $\square$

Dies ist sehr bemerkenswert, da man erwarten könnte, dass die geforderten Generalisierungen aus Kap. 2.3 einen beweistheoretischen Unterschied machen. Die Konservativität von  $CT \upharpoonright$  zeigt jedoch, dass kein direkter Zusammenhang zwischen den geforderten Verallgemeinerungen und Nichtkonservativität besteht. Den zentralen Unterschied macht daher, wie wir gleich sehen werden, offensichtlich das Induktionsaxiom aus.

#### 2.4.2 CT

**Definition 2.14** ([35], S. 102) (CT) Das System CT resultiert aus  $CT \upharpoonright$ , indem alle Instanzen des Induktionsaxioms zu der Sprache  $\mathcal{L}_T$  hinzugefügt werden.

Die formalen Eigenschaften der Theorie CT unterscheiden sich in zentraler Hinsicht von denen ihrer Vorgänger TB und  $CT \upharpoonright$ . Insbesondere ist CT nicht konservativ über PA. Dies folgt aus einigen Lemmata, welche im Folgenden dargestellt werden. Sei  $Bew_{PA}$  das Beweisbarkeitsprädikat.

**Lemma 2.15** ([35], S. 104, [37], S. 76) Die Theorie CT beweist das globale Reflexionsprinzip für PA; d.i.:

$$CT \vdash \forall x (Sent(x) \wedge Bew_{PA}(x) \rightarrow Tx)$$

*Beweisidee:* Der Beweis ist via Induktion über die Länge der Beweise in PA. Es muss

gezeigt werden, dass die logischen Axiomenschemata wahr sind und, dass die arithmetischen Axiome von PA inklusive dem Induktionsaxiom wahr sind.  $\square$

Wie angekündigt ist es für dieses Theorem zentral, dass das Wahrheitsrädikat im Induktionsaxiom vorkommt.

(...) allowing T to occur in the induction scheme is essential for the proof of the uniform reflection theorem to work. (...) the truth truth predicate (...) occurs in the instance of mathematical induction that is used to show that all instances of the induction axiom of PA are true. ([37], S. 78)

Da CT das globale Reflexionsprinzip beweist, gilt außerdem

**Lemma 2.16**  $CT \vdash \neg Bew_{PA}(\ulcorner 0 = 1 \urcorner)$

*Beweis:* Angenommen es gilt  $CT \vdash Bew_{PA}(\ulcorner 0 = 1 \urcorner)$ . Dann würde gemäß 2.15 auch  $T(\ulcorner 0 = 1 \urcorner)$  gelten. Dies ist offensichtlich falsch.  $\square$

Sei  $Con(PA)$  definiert als  $\neg Bew_{PA}(\ulcorner \perp \urcorner)$ , wobei  $\perp$  das Zeichen für eine Absurdität ist. Dann folgt aus dem gerade Gesagten:

**Lemma 2.17**  $CT \vdash Con(PA)$

*Beweis:*  $\neg Bew_{PA}(\ulcorner 0 = 1 \urcorner)$  ist per definitionem äquivalent zu  $Con(PA)$ .  $\square$

Dies ergibt weiters:

**Theorem 2.18** *CT ist nicht konservativ über PA.*

*Beweis.* Es gilt  $CT \vdash Con(PA)$ . Gödels zweiter Unvollständigkeitssatz besagt, dass  $PA \not\vdash Con(PA)$  gilt. Daraus folgt nach der oben gegebenen Definition von Konservativität, dass CT nicht konservativ über PA ist.  $\square$

Zu bemerken ist außerdem, dass CT Generalisierungen beweist, welche in  $CT \upharpoonright$  nicht beweisbar sind. So beweist CT, dass die Substitution von identischen Termen Wahrheit beibehält.

**Lemma 2.19** ([35], S. 102)  $CT \vdash \forall x \forall s \forall t (Sent(\ulcorner \forall x \wedge s^\circ = t^\circ \urcorner \rightarrow (Tx(s/v) \leftrightarrow Tx(t/v))))$

*Beweis:* siehe ([35], S. 102).  $\square$

Letztlich sei darauf hingewiesen, dass weder TB, noch  $CT \upharpoonright$  oder CT Wahrheitsiterationen der Form  $T...T(\ulcorner \phi \urcorner)$  beweisen. Die Axiome wurden auf die Basistheorie PA ohne Wahrheitsprädikat eingeschränkt und verhindern damit genau das, da sonst Tarskis

Theorem zu einem Widerspruch führen würde. Da eine Theorie wie CT, aus denselben Gründen wie im Fall von PA, seine eigene Konsistenz nicht beweisen kann, ist es naheliegend, erneut zu einer stärkeren Theorie aufzusteigen um die Konsistenz der vorhergehenden zu beweisen. Die stärkere Theorie darf natürlich ebenfalls nicht ihr eigenes Wahrheitsprädikat enthalten. Deshalb zieht man getypte Wahrheitsprädikate  $T_0, T_1, T_2, T_3 \dots T_n$ , heran, wobei jedes  $T_n$  nur auf ein Wahrheitsprädikat  $T_i, (i < n)$  niedrigerer Stufe angewendet werden darf. Axiomatisiert man diesen Prozess, erhält man eine Theorie für ramifizierte Wahrheit (siehe [35], S.127f).

### 3 Deflationismus und Konservativität

Im Anschluss an die ursprünglich von Tarski vorgebrachten Theorien ergeben sich einige sehr interessante philosophische Diskussionen, welche insbesondere für die Deflationismusdebatte zentrale Bedeutung haben. Im Folgenden werden die wichtigsten Probleme diesbezüglich dargestellt.

#### 3.1 Allgemeines zum Deflationismus

Deflationismus ist eine These mit vielen Gesichtern und historisch betrachtet ist es schwierig einen einheitlichen Kern zu erkennen. Die Auseinandersetzung damit ist vor allem in der klassischen Sprachphilosophie zentral und eng mit Problemen der Bedeutung, Referenz und ‘Inhalt‘ verbunden.<sup>7</sup> Da dies jedoch nicht Thema der vorliegenden Arbeit ist, werde ich nur diejenigen Thesen des Deflationismus heranziehen, welche ich für common-sense in der Forschungslandschaft halte. Diese werden daher auch in Bezug auf Autoritäten der Geschichte und Gegenwart der analytischen Philosophie gerechtfertigt und synthetisiert. Ich werde zuerst die wichtigsten Thesen des Deflationismus präsentieren und diese mit einigen Zitaten belegen. Zu beachten ist hierbei, dass sich die Thesen gegenseitig bedingen und daher nur gemeinsam eine adäquate Darstellung ergeben. Die vier zentralen Thesen des Deflationismus sind:

1. Wahrheit ist ein metaphysisch leichter Begriff ohne Substanz.
2. Wahrheit beinhaltet nicht Repräsentation.
3. Wahrheit ist nur ein logischer Begriff.
4. Das Bikonditional  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  ist die zentrale (wenn nicht einzige) Rolle des Wahrheitsprädikats.

Die erste These ist höchstwahrscheinlich die zentrale philosophische Leitidee des Deflationismus. Diese Behauptung wird von Leon Horsten explizit ausgesprochen:

---

<sup>7</sup>siehe z.B.: [9], [14], [20], [21], [39], [53]

The core idea of deflationism is that truth is an insubstantial, light notion. It is a notion that does not carry a heavy philosophical burden or commitment of any kind (metaphysical, epistemological, moral, etc.) ([36], S. 557)

Natürlich ist dies sehr intuitiv und ungenau. A priori ist nämlich nicht klar, was es bedeuten soll, dass Wahrheit leicht und insubstantiell ist. Dies kann jedoch veranschaulicht werden, wenn man beachtet, dass eine substantielle Wahrheitstheorie oft mit dem Begriff der Repräsentation verknüpft wird. So geht die Idee des Deflationismus, dass Wahrheit metaphysisch wenig Gewicht hat, oft mit einem Skeptizismus bezüglich Repräsentation einher; d.h. mit These 2. Dies dürfte z.B. die Idee von von Jeffrey Ketlands Charakterisierung der deflationistischen Position sein.<sup>8</sup>

(...) the idea (of deflationism; LS) perhaps is that we need to ‘deflate’ the correspondence notion that truth expresses a substantial or theoretically significant language-world relation. ([40], S. 69)

Um ein Beispiel anzuführen, wogegen sich dies wendet, kann man (ironischer weise) Tarskis eigene Interpretation seiner Theorie heranziehen, da dieser selbst seine Wahrheitstheorie als eine Korrespondenztheorie interpretiert hat, wie folgendes Zitat veranschaulichen soll:

Wir möchten, daß unsere Definition den Intuitionen der *klassischen aristotelischen Konzeption der Wahrheit* gerecht wird - die in ihren Ausdruck in den wohlbekannten Worten der *Metaphysik* des Aristoteles finden: *Von etwas, das ist, zu sagen, daß es nicht ist, oder von etwas, das nicht ist, daß es ist, ist falsch, während von etwas, das ist, zu sagen, daß es ist, oder von etwas, das nicht ist, daß es nicht ist, wahr ist.*

Wenn wir uns der modernen philosophischen Terminologie anpassen wollen, könnten wir diese Konzeption vielleicht durch die bekannte Formulierung ausdrücken: *Die Wahrheit einer Aussage besteht in ihrer Übereinstimmung (oder Korrespondenz) mit der Wirklichkeit.*

(...) so können wir zum selben Zweck auch sagen: *Eine Aussage ist wahr, wenn sie einen existierenden Sachverhalt bezeichnet.* (...) Dergestalt muß die Definition der Wahrheit, wenn sie unserer Konzeption entspricht, die folgende Äquivalenz implizieren: *Die Aussage ‘Schnee ist weiß’ ist wahr genau dann, wenn Schnee weiß ist.* ([63], S. 142)

Da wir nun wissen, wogegen sich der Deflationismus wendet, benötigen wir im nächsten Schritt eine positive Charakterisierung dieser Position. Dies führt uns zuerst zu These 3, welche besagt, dass Wahrheit ein logischer Begriff ist und deshalb genauso unwichtig, bzw insubstantiell ist, wie die logischen Konnektive.

<sup>8</sup>siehe auch [61]: ‘Philosophers often make suggestions like the following: truth consists in correspondence to the facts (...). According to the deflationist, however, such suggestions are mistaken (...).’

At a first approximation, deflationism claims that the notion of truth is akin to logical notions like ‘and,’ and ‘not’. Logical notions are usually not regarded as deep philosophical notions. In a similar fashion, truth should be regarded as a superficial notion. This is what is meant when it is said that the notion of truth should be *deflated*. ([37], S. 60)

Das Wahrheitsprädikat hat demnach zwar durchaus einen logischen Mehrwert. Wie wir schon gesehen haben, können wir durch das Hinzufügen eines Wahrheitsprädikats zu einer Sprache gewisse logische Tatsachen beweisen. Dennoch bleibt die zentrale These bestehen, dass Wahrheit abgesehen von dieser Funktion, welche als insubstantiell betrachtet wird, eigentlich entbehrlich ist.

Eine weiterer Aspekt der Behauptung, dass das Wahrheitsprädikat insubstantiell ist, besteht nun aus These 4. Diese drückt nämlich formal aus, dass das Wahrheitsprädikat eigentlich entbehrlich und redundant ist.

According to the doctrine known as *deflationism about truth*, the concept of truth (...) is redundant and dispensable (...). ([40]. S. 69)

Diese These findet sich historisch betrachtet in mehreren Ausformungen. Die zwei zentralen davon sind die Redundanz-Theorie und der ‘Disquotationalismus’.<sup>9</sup> Erstere Theorie geht auf Ramsey zurück und besagt, dass das Wahrheitsprädikat redundant ist. D.h; alles was sich über Wahrheit sagen lässt, wird durch das Schema

die Proposition, ‘dass p’ ist wahr genau dann, wenn p.

ausgedrückt, welches Ramsey folgendermaßen interpretiert.

Truth and falsity are ascribed primarily to propositions. (...) it is evident that ‘It is true that Caesar was murdered’ means no more than that Caesar was murdered, and ‘It is false that Caesar was murdered’ means that Caesar was not murdered. They are phrases which we sometimes use for emphasis or for stylistic reasons. ([54], S. 142)

Die zweite Theorie geht auf W.V.O Quine zurück und besagt, dass die Funktion des Wahrheitsprädikat gänzlich durch das Disquotationsschema der Form

Der Satz ‘xyz’ ist wahr genau dann, wenn xyz.

ausgedrückt wird. Dieses interpretiert Quine wie folgt:

---

<sup>9</sup>Der zentrale Unterschied dieser beiden Positionen ist, dass gemäß der Redundanz-Theorie Propositionen die Träger von Wahrheit sind (anders formuliert bedeutet dies, dass das Wahrheitsprädikat nur Propositionen zugeschrieben wird), wobei der Disquotationalismus behauptet, dass Sätze die genuinen Träger von Wahrheit sind. Diesbezüglich gibt es zahlreiche Diskussionen. Siehe z.B. [11], [39]

To say that the statement ‘Brutus killed Caesar’ is true, or that ‘The atomic weight of sodium is 23’ is true, is in effect simply to say that Brutus killed Caesar, or that the atomic weight of sodium is 23. ([52], S. 24)

Zu sagen, dass ein Satz wahr ist, ist daher (nach Quine) dasselbe, wie den entsprechenden Satz selbst zu äußern und dies wird formal durch das Tarski-Bikonditional  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  eingefangen.

Obwohl diese historischen Beispiele die entsprechende Problematik sehr schön veranschaulichen, muss dennoch darauf hingewiesen werden, dass es formallogisch wenig Unterschied macht, ob das Wahrheitsprädikat eine Eigenschaft von Sätzen oder Propositionen ist. Dies ist deshalb so, weil hier das Wahrheitsprädikat eine Eigenschaft von Gödel - Codes von Sätzen ist.

Da somit der Hintergrund zum Deflationismus zumindest allgemein geklärt ist, gehe ich im Folgenden auf zwei zentrale Kritikpunkte diesbezüglich ein. Dabei handelt es sich um Kritiken aus dem Bereich der mathematischen Logik, welche die formale Adäquatheit der soeben vorgebrachten Thesen betreffen.

### 3.2 Das Generalisierungsproblem

Eine weitere zentrale Behauptung des zeitgenössischen Deflationismus, welche in enger Verbindung mit den oben besprochenen Thesen 1-4 steht, ist, dass der einzige Zweck des Wahrheitsprädikats derjenige ist, infinite Generalisierungen auszudrücken, die anders nicht ausgedrückt werden können. Dies ist offensichtlich eine Spielart von These 3 und findet sich erstmals bei Quine [53] und wurde später von prominenten Autoren wie Horwich (vgl. [38], [39]) und Field (siehe[21]) übernommen. Quine schreibt:

The truth predicate is a device of disquotation. We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about sentences, then the truth predicate has its use. ([53], S. 12)

Horwich schreibt:

Thus our concept of truth is a device of generalization. That is its function. And it is able to perform that function in virtue of its role within the so-called *equivalence schema*, ‘ $\langle p \rangle$  is true  $\leftrightarrow p$ ’. Nothing more is needed; and nothing less will do. ([39], S. 5)

Würde man nun die Thesen 1-4, gemeinsam mit der Behauptung, dass das Wahrheitsprädikat einzig ein Instrument zur Generalisierung ist, in einen formalen Rahmen gießen, ergibt dies die Theorie TB, die wir in Abschnitt 2.3 kennengelernt haben.

Genau dies wurde von vielen Autoren getan und wird heute weitgehend als korrekte mathematische Präzisierung der gerade dargestellten Behauptungen des Deflationismus angesehen (vgl. [21], [35], [37], [40], [56]). Daraus ergibt sich die These des frühen Deflationismus (kurz FD):

(FD) Die Theorie TB drückt formal alles aus, was über Wahrheit zu sagen ist.

Wie jedoch schon in Abschnitt 2.3 gezeigt wurde, entstehen hier die ersten Probleme. Das erste ist, dass TB keine infiniten Generalisierungen beweisen kann, wie z.B. Quine dies behauptet hat. Diese geforderten Generalisierungen (wie z.B. Kompositionalität) sind jedoch grundlegende Intuitionen, welche von einer Wahrheitstheorie auch erfüllt werden sollten. Der Deflationist muss diesbezüglich daher entweder seine Theorie modifizieren oder seine philosophischen Intuitionen ändern.

Obwohl daher die Theorie TB, eigentlich schon aus diesem Grund keine gute Wahrheitstheorie ist, wurde dieses Problem nicht als bezwingend betrachtet.<sup>10</sup> Aus diesem Grund wurde ein stärkeres Argument gegen TB und der mit ihr verbundenen philosophischen Thesen vorgebracht. Der zentrale Kritikpunkt betrifft hier ebenfalls die These (FD), jedoch in Zusammenhang mit der Leitidee des Deflationismus, dass Wahrheit ein metaphysisch leichter und unwichtiger Begriff ist. Um diese Problematik angemessen darzustellen, muss jedoch weiter ausgeholt werden.

### 3.3 Konservativität

Zuerst sei als Wiederholung angemerkt, dass die Theorie TB konservativ über Peano Arithmetik ist und daher keine Theoreme beweisen kann, welche nicht schon von PA bewiesen werden können (Theorem 2.9). Insofern resultiert durch das Hinzufügen der Tarski-Bikonditionale nichts Wichtiges oder Neues. Dies ist, man möchte meinen, für den Deflationismus, genau so wie es sein soll. Zumindest scheint es These 1 gerecht zu werden, welche besagt, dass Wahrheit ein metaphysisch leichter und unwichtiger Begriff ist. Wären aus dem Hinzufügen einer Wahrheitstheorie zu einer gegebenen Basistheorie neue Theoreme beweisbar - so die Idee - , könnte man nicht mehr behaupten, dass Wahrheit ein unwichtiger Begriff ist. Aus diesem Grund haben Jeffrey Ketland und Stewart Shapiro dem Deflationisten unterstellt, dass er auf eine konservative Wahrheitstheorie festgelegt ist.<sup>11</sup>

Indeed, one might go further: *if* truth is non-substantial - as deflationists claim - *then* the theory of truth *should* be conservative. Roughly: *non-substantiality*  $\equiv$  *conservativeness*. ([40], S. 79)

---

<sup>10</sup>siehe dazu [38], [39]

<sup>11</sup>Eigentlich geht dieses Argument auf Leon Horsten zurück. Leider ist es jedoch sehr schwierig auszumachen, wo und wie er dies vorgebracht hat.

I submit that in one form or another, conservativeness is essential to deflationism. ([56], S. 497)

Die Tatsache jedoch, dass TB konservativ über PA ist, stellt nach Ketland und Shapiro ein zentrales Problem dar. Denn wie wir bereits wissen, kann eine konservative Wahrheitstheorie, aufgrund von Gödels zweiten Unvollständigkeitssatzes, nicht die Konsistenz von PA beweisen. Ketland und Shapiro argumentieren jedoch, dass dies eine adäquate Wahrheitstheorie unbedingt leisten sollte.

Suppose we accept a standard axiomatization of arithmetic (PA, say). It seems correct to say that we also (implicitly) accept its truth, and thus we surely think that it is consistent. (...) So, we have *accepted* PA, we think it's true, and we seem to be committed to thinking it consistent. ([40], S. 91)

Formal gesprochen erwarten wir daher, dass eine Wahrheitstheorie alle Theoreme von PA als wahr (kurz:  $T(\text{PA})$ ) beweist. Dies wird am besten durch das globale Reflexionsprinzip der Form

$$\forall x(\text{Sent}(x) \wedge \text{Bew}_{\text{PA}}(x) \rightarrow T(x))$$

ausgedrückt. Deshalb schreibt Jeffrey Ketland:

Any adequate theory of truth should be able to prove the 'equivalence' of a (...) *theory* T and its 'truth'  $\text{True}(\text{T})$ <sup>12</sup> (that is, the metalanguage formula  $\forall x(\text{Sent}(x) \wedge \text{Bew}_{\text{PA}}(x) \rightarrow T(x))$ ).<sup>13</sup> ([40], S. 90)

Wie wir bereits wissen, ist das globale Reflexionsprinzip beweistheoretisch äquivalent zu der Aussage  $\text{Con}(\text{PA})$ . Das Argument von Ketland und Shapiro lautet nun kurz gesagt, dass die Theorie TB aufgrund ihrer Konservativität keine gute Wahrheitstheorie ist.<sup>14</sup> Eine Theorie wie TC jedoch - so geht das Argument weiter -, welche die Konsistenz von PA und daher auch  $T(\text{PA})$  beweist, ist für einen Deflationisten aus zwei Gründen nicht akzeptabel (vgl. [40], [56]):

1. Sie bringt Konsequenzen mit sich, welche durch das Hinzufügen einer Wahrheitstheorie (zu einer Basistheorie) entstehen. D.h. das Wahrheitsprädikat hat nicht nur eine disquotationale Funktion. In der Tat erlaubt es uns Theoreme zu beweisen, welche die Basistheorie alleine nicht beweisen kann.
2. Aussagen wie  $T(\text{PA})$  und  $\text{Con}(\text{PA})$  sind substantiell. Eine Wahrheitstheorie, die sie beweist, kann nicht unwichtig und metaphysisch leicht sein.

---

<sup>12</sup>hier PA

<sup>13</sup>Notation angepasst

<sup>14</sup>Aus denselben Gründen ist daher  $\text{CT}^\dagger$  ebenfalls keine gute Wahrheitstheorie.

Viele Autoren haben daraus geschlossen, dass Deflationismus in der oben angegebenen Form nicht aufrecht zu erhalten ist. Obwohl (meines Wissens nach) die meisten nachfolgenden Autoren diesbezüglich mit Ketland und Shapiro übereinstimmen, werden wir jedoch im nächsten Abschnitt Gründe dafür angeben, dass der Deflationismus von der Konservativitätsanforderung befreit werden sollte.

### 3.4 Deflationismus ohne Konservativität

Zuerst sei bemerkt, dass die Behauptung, dass der Deflationismus auf eine konservative Wahrheitstheorie festgelegt ist, sehr stark ist. Wäre dies die richtige Charakterisierung der deflationistischen Position, müsste sie aufgrund der beschriebenen Mängel von TB ein für alle Mal aufgegeben werden. Dass dies problematisch ist, wurde im weiteren Verlauf dieser Diskussion auch eingebracht.

Voker Halbach (in [33]) hat als erster angemerkt, dass die Anforderung der Konservativität, wie sie von Ketland und Shapiro eingebracht wurde, vage ist, da nicht klar ist, worüber eine Wahrheitstheorie konservativ sein soll. Zu diesem Zweck unterscheidet er die beiden Anforderungen ‘konservativ über PA’ und ‘konservativ über Logik’ und hat gezeigt, dass letztere Anforderung unmöglich zu erfüllen ist. Zuerst jedoch zu einer Charakterisierung dessen, was ‘konservativ über Logik’ bedeutet. Dazu schreibt Leon Horsten:

A truth theory  $T$  is conservative over logic if and only if all sentences that are provable from  $T$  but that do not contain the truth predicate are provable from the laws of logic alone. ([37], S. 81)

Diese Anforderung kann jedoch nicht einmal von der schwächsten Version von TB erfüllt werden. Diese beweist nämlich schon, dass es zumindest zwei Objekte gibt und das ist mehr, als aus reiner Logik folgt. Um dies zu veranschaulichen, nehmen wir die folgenden restriktierten Bikonditionale zur Hand:

$$\begin{aligned} T(\forall x : x = x) &\leftrightarrow \forall x : x = x \\ T(\forall x : x \neq x) &\leftrightarrow \forall x : x \neq x \end{aligned}$$

Die Logik erster Stufe impliziert nun, dass die rechte Seite des ersten Bikonditionals wahr ist und die rechte Seite des zweiten Bikonditionals falsch ist. Da identische Objekte dieselben Eigenschaften haben, muss der Satz  $\forall x : x = x$  ein anderes Objekt sein als der Satz  $\forall x : x \neq x$ . Das heißt, dass die Theorie TB, auch ohne eine gegebene Basisstheorie, beweist, dass es zumindest zwei Objekte gibt. Reine Logik kann dies jedoch nicht tun.

Ausgehend davon, sind die meisten modernen Autoren der Ansicht, dass der Deflatio-

nismus von der Konservativitätsanforderung befreit werden sollte. Die Analogie dürfte hier sein, dass ähnlich, wie keine Wahrheitstheorie konservativ über Logik sein kann, keine *gute* Wahrheitstheorie konservativ über PA sein kann. So schreibt Horsten:

The nonconservativeness of truth over mathematics is a *phenomenon*.

So Halbach is right when he states that if deflationism about truth is to be given a chance at all, it must be dissociated from mathematical conservativeness claims (Halbach, 2001). Conservativeness over mathematics should not be regarded as an essential component of deflationism about truth. ([37], S. 92)

Das Argument gegen die Konservativitätsanforderung lautet demnach, dass Nichtkonservativität eine Tatsache ist, die man hinnehmen muss. Würde man weiterhin behaupten, dass Konservativität essentiell für den Deflationismus ist, müsste man ihn für immer verabschieden, da keine gute Wahrheitstheorie diese Anforderung erfüllt. Dennoch braucht man Kriterien dafür, welche verfügbaren Wahrheitstheorien deflationistisch sind und welche nicht. Könnte man diese Unterscheidung nicht anführen, wäre die Debatte überflüssig, da ansonsten jede gute Wahrheitstheorie schlicht inflationistisch wäre. Daraus folgt, dass die Frage, welche Elemente zentral für den Deflationismus sind, erneut gestellt werden muss. Darauf werden wir in Kapitel 5 im Kontext moderner Variationen von Kripkes Wahrheitstheorie zurückkommen.

## 4 Kripke

Neben der Arbeit von Tarski, wurde im zwanzigsten Jahrhundert nur eine ähnlich einflussreiche Wahrheitstheorie hervorgebracht; und diese wurde von Saul Kripke in seinem ‘Outline of a Theory of Truth’ [43] dargelegt. In diesem Abschnitt wird zuerst kurz Kripkes Kritik an Tarski präsentiert, danach werden die mathematischen Aspekte seiner Theorie detailliert aufgearbeitet; wobei ein Schwerpunkt auf seine semantische Theorie und dem Beweis des Fixpunktlemmas gelegt wird.

### 4.1 Kritik an Tarski

Wie wir bereits wissen, konnte Tarski, durch die Einführung eines getypten Wahrheitsprädikats, eine konsistente Wahrheitstheorie formulieren und so Widersprüche vermeiden. Kripke setzt nun genau an diesem Punkt an und behauptet, dass ein typisiertes Wahrheitsprädikat unnatürlich und unintuitiv ist. Seine Gegenthese lautet schlicht, dass unsere Umgangssprache eigentlich ein ungetyptes Wahrheitprädikat enthält.

Surely our language contains just one word ‘true’, not a sequence of distinct

phrases  $\ulcorner true_n \urcorner$ , applying to sentences of higher and higher levels. ([43], S. 695)

Aber warum ist dies so? In Tarskis Fall hatten wir eine syntaktische Markierung, welche dem Wahrheitsprädikat die relevanten Levels zuschreibt. Ist dies nicht möglich, ist der Satz in Tarskis Theorie syntaktisch nicht wohlgeformt. Dennoch äußern wir in unserer Umgangssprache wohlgeformte Sätze, die diese Anforderung nicht erfüllen. So können wir z.B. ohne Bedenken folgenden Satz formulieren, ohne die relevanten Levels im voraus zu kennen.

(1) Alles, was Sokrates sagt ist wahr.

Diesbezüglich schreibt Kripke:

In general, if a sentence such as (1) asserts (all, some, most, etc.) of the sentences of a certain class C are true, its truth value can be ascertained if the truth values of the sentences in the class C can be ascertained. If some of these sentences themselves involve the notion of truth, their truth value in turn must be ascertained by looking at other sentences, and so on. If ultimately this process terminates in sentences not mentioning the concept of truth, we call the original sentence grounded; otherwise ungrounded. ([43], S. 693f)

Ein Sprecher, der eine Aussage macht, kann jedoch nie sicher sein, ob sein Satz grounded ist oder nicht und kann daher seinen Aussagen nicht immer die relevanten Levels zuschreiben.

Ordinarily, however, a speaker *has no way of knowing the ‘levels’ of (the) relevant utterances*. ([43], S. 695)

Die Tradition nach Kripke hat ihm diesbezüglich zugestimmt. Dennoch müssen wir nun versuchen, diese Intuitionen formal zu modellieren. Da wir von Tarskis Theorem wissen, dass ein ungetyptes Wahrheitsprädikat, in klassischer Logik formuliert, zu einem Widerspruch führt, verwendet Kripke eine nichtklassische Logik um seine Intuitionen formal umzusetzen. Die grundlegende Idee fasst er wie folgt zusammen:

There is to be only one truth predicate, applicable to sentences containing the predicate itself; but paradox is to be avoided by allowing truth-value gaps and by declaring that paradoxical sentences in particular suffer from such a gap. ([43], S. 698)

Die nichtklassische Logik, von welcher Kripke spricht, wird im Folgenden (in erster Linie) eine Starke-Kleene Logik mit Wahrheitswertlücken ( $K_3$ ) sein.<sup>15</sup> Anders als in einer zweiwertigen Logik, in welcher für jeden Satz gilt, dass er entweder wahr oder falsch ist, erlaubt diese Logik drei Werte: wahr, falsch und unbestimmt. Damit ändert sich einiges. So können wir insbesondere Tarskis induktive Klauseln (Definition 2.10) nicht mehr verwenden. Dies ist deshalb so, weil in einer klassischen Logik gilt, dass die Vereinigung der Extension des Wahrheitsprädikates  $T$ , mit deren Komplement (die Antiextension von  $T$ ), die Menge der natürlichen Zahlen  $\omega$  ist. Verwenden wir dagegen eine mehrwertige Logik mit Wahrheitswert-Lücken, bildet die Vereinigung von Extension und Antiextension nicht  $\omega$ . Deshalb verwenden wir das Starke-Kleene Evaluierungsschema, welches nicht-klassische Modelle mit Extension und Antiextension liefert. Aus diesen Definitionen folgen die wichtigsten formalen Eigenschaften von Kripkes Theorie, welche im Folgenden expliziert werden.

Bevor wir beginnen, wird auf informeller Ebene der grundlegende Unterschied zwischen Ordinalzahlen und Kardinalzahlen erklärt. Dies wird für das nächste Kapitel wichtig sein. Kardinalzahlen messen die Mächtigkeit oder Größe einer Menge, Ordinalzahlen repräsentieren den Zählprozess, durch welchen eine gegebene Menge generiert wird. Zwei Kardinalzahlen  $\alpha, \beta$  sind gleich groß, gdw es eine Bijektion von  $\alpha$  auf  $\beta$  gibt; zwei Ordinalzahlen sind gleich groß, wenn sie bezüglich der Ordnung, die durch den Zählprozess generiert wird, gleich groß sind. Im Finiten gibt es keinen Unterschied zwischen Ordinalzahlen und Kardinalzahlen. Z.B. ist der Kardinalitätsaspekt der Menge  $\{1, 2, 3\}$  gleich ihrem ordinalen Aspekt, nämlich 3. Im Transfiniten macht dies jedoch einen Unterschied. Sei  $\omega$  die Menge der natürlichen Zahlen. Wir generieren die Menge  $\omega$ , indem wir mit 0 beginnen und einfach ins Transfinite 0, 1, 2, 3, ... weiterzählen. Ändern wir nun den Zählprozess und beginnen mit 1, 2, 3, ... und nehmen 0 als letztes Element dieser Menge, erreichen wir eine neue Ordinalzahl, da die Menge unter diesem letzten Element ordnungs-isomorph zu der Menge  $\omega$  ist und daher ein Element mehr besitzt. Wir nennen diese Menge  $\omega + 1$ . Diesen ‘Trick‘ können wir nun unendlich oft wiederholen; dh. wir können mit 2, 3, 4, ... zu zählen beginnen und 0, 1 als letzte Elemente anhängen und so die Menge  $\omega + 2$  erzeugen. Ganz allgemein können wir mit irgendeiner Zahl  $n$  beginnen und alle  $n-1$  hinten anhängen: also  $n, n+1, \dots, 0, 1, \dots, n-1$  und so eine Menge  $\omega + n$  erzeugen. Die so erzeugten Mengen sind jedoch alle gleich groß und haben daher dieselbe Kardinalität. Diese Mengen nennt man abzählbar unendlich und ihre Kardinalzahl nennt man  $\aleph_0$ . Die Menge  $\omega$  der natürlichen Zahlen ist die kleinste abzählbar unendliche Ordinalzahl und hat daher die Kardinalität  $\aleph_0$ . Wichtig für die vorliegenden Zwecke ist außerdem, dass es keine Menge aller Ordinalzahlen gibt. Es gibt überabzählbar viele abzählbare Ordinalzahlen. Außerdem sei angemerkt, dass es

---

<sup>15</sup>Wir werden diese Logik erst später formal explizieren (siehe S. 30).

auch viel größere Kardinalzahlen als  $\aleph_0$  gibt. Z.B. sind die reellen Zahlen überabzählbar unendlich. Für weitere Details diesbezüglich siehe [16] und [50].

## 4.2 Induktive Definitionen, monotone Operatoren und der minimale Fixpunkt

Das starke Kleene Evaluierungsschema (K3) mit Wahrheitswertlücken erlaubt es uns einen monotonen Operator auf Teilmengen der Extension und Antiextension zu definieren, was eine Vorbedingung für Kripkes Fixpunktlemma ist. Diese Tatsache folgt aus abstrakten Überlegungen aus der Rekursions- und Beweistheorie, welche im Folgenden besprochen werden.

Zuerst werden wir einen monotonen Operator definieren und im allgemeinen Fall das Fixpunktlemma beweisen. Um maximale Klarheit zu erreichen, wird weiters gezeigt, dass der allgemeine Fall auf Kripkes Wahrheitstheorie übertragen werden kann. Um ein Gefühl dafür zu bekommen, wie monotone Operatoren eingeführt werden, werfen wir zuerst einen Blick auf induktive Definitionen.<sup>16</sup> Wir werden zeigen, dass nur positive induktive Definitionen monoton sind.

**Definition 4.1** (vgl.[35], S. 116) *Sei  $S \subseteq \omega$ . Eine positive induktive Definition hat die Form:*

$n \in S$  genau dann wenn  $\theta(n, S)$ ,

wobei  $\in S$  nur positiv in  $\theta(n, S)$  vorkommt; d.i.: nur innerhalb einer geraden Anzahl von Negationssymbolen. Insbesondere darf  $\notin S$  nicht in  $\theta(n, S)$  vorkommen. Abgesehen von  $\in S$ , ist  $\theta(n, S)$  ein arithmetisches Prädikat.

Zu beachten ist, dass in Tarskis Theorie die Existenz einer Menge  $S$ , welche die Bedingungen (i) - (vi) (siehe Definition 2.10) erfüllt, durch Induktion über den Formelaufbau bewiesen wird. Hat man jedoch positive induktive Definitionen, folgt die Existenz einer Menge  $S$ , die gewisse Bedingungen erfüllt, aus allgemeinen Fakten über eben diese Definitionen. (vgl. [35], S. 117).

**Lemma 4.2** *Positive induktive Definitionen sind monoton; d.i.: gegeben zwei Mengen  $S_1$  und  $S_2$  sodass,  $S_1 \subseteq S_2$ , dann gilt für alle  $n \in \omega$ :*

$$(\theta(n, S_1) \rightarrow \theta(n, S_2))$$

*Beweis: Der Beweis wird durch Induktion nach den Rang der Formeln geführt (siehe [50], S. 88).  $\square$*

---

<sup>16</sup>Ein Beispiel für induktive Definitionen sind zum Beispiel Tarski's induktive Klauseln in Definition 2.10.

**Lemma 4.3** *Tarskis Definition ist aufgrund des Vorkommens von  $\notin S$  in Klausel (ii) von Definition 2.10 nicht monoton.*

*Beweis:* Sei  $S_1 \subseteq S$  und sei das Prädikat  $\notin S$  in  $\theta(n, S)$  enthalten. Angenommen  $n$  ist  $\phi$  und  $\phi \notin S_1$ . Dann gibt es eine Menge  $S_2$  sodass  $S_1 \subseteq S_2$  und  $\phi \in S_2$ . Dh.: es gilt nicht  $\theta(n, S_1) \rightarrow \theta(n, S_2)$  für  $S_1 \subseteq S_2$ .  $\square$

Auf der Basis dieser positiven induktiven Definition können wir nun einen monotonen Operator einführen. Im Folgenden sei  $Pot(x)$  die Potenzmengenoperation.<sup>17</sup>

**Definition 4.4** (vgl.[35], S. 117f, [50], S. 84f) *Ein monotoner Operator für eine Menge  $S \subseteq \omega$  ist die Operation:*

$$\Gamma_\theta : Pot(S) \rightarrow Pot(S), \text{ sodass}$$

$$\Gamma_\theta(S) := \{n : \theta(n, S)\}$$

Die Idee dieser Definition ist, dass, ausgehend von der leeren Menge, durch die wiederholte Anwendung des Operators  $\Gamma_\theta$  immer neue Mengen erzeugt werden, für welche  $\theta(n, S)$  gilt. Weiters nennen wir eine Menge  $S$  *korrekt*, wenn gilt, dass  $S \subseteq \Gamma_\theta(S)$ .

**Definition 4.5** ([68], S. 119) *Wenn  $S$  korrekt ist, sei die Sequenz  $\langle \Gamma_\theta^\alpha(S) \mid \alpha \text{ eine Ordinalzahl} \rangle$  über transfinite Rekursion<sup>18</sup> wie folgt definiert:*

$$\Gamma_\theta^0(S) = S \text{ und } \Gamma_\theta^\alpha(S) = \Gamma_\theta(\Gamma_\theta^{\alpha-1}(S)) \text{ für alle } \alpha > 0,$$

*wobei, wenn  $\alpha$  eine Limes-Ordinalzahl ist<sup>19</sup> ist, dann gilt:*

$$\Gamma_\theta^{\alpha-1}(S) = \bigcup_{\beta < \alpha} \Gamma_\theta^\beta(S).$$

Außerdem sollte aus dem gerade gesagten klar sein, dass insbesondere für eine Limes-Ordinalzahl  $\alpha$  gilt, dass  $\Gamma_\theta^\alpha(S) = \Gamma_\theta(\bigcup_{\beta < \alpha} \Gamma_\theta^\beta(S))$ .

**Definition 4.6** *Sei  $I_\theta^\alpha := \Gamma_\theta^\alpha(S)$ . Ausgehend von Definition 4.5 ist  $I_\theta^\alpha$  die unter  $\Gamma_\theta$  abgeschlossene Menge.*

Oben wurde gezeigt, dass positive induktive Definitionen monoton sind. Daher ist auch der soeben eingeführte Operator  $\Gamma_\theta$  monoton. D.i; es gilt:

<sup>17</sup>Siehe [16], S. 35 f für weitere Details.

<sup>18</sup>transfinite Rekursion generalisiert die primitive Rekursion ins Transfinite für alle Ordinalzahlen. Für Details, siehe [50] S. 23f

<sup>19</sup>Eine Ordinalzahl  $\alpha$  ist eine Limes-Ordinalzahl, wenn  $\forall \beta (\beta < \alpha \rightarrow \beta' < \alpha)$  gilt. Für weitere Details, siehe [50], S. 22.

**Lemma 4.7** ([68], S. 116) *Die Sequenz  $\langle \Gamma_\theta^\alpha(S) \rangle$  ist nichtabsteigend. D.h. Für alle Ordinalzahlen  $\alpha, \beta$ , wenn  $\alpha < \beta$ , dann  $\Gamma_\theta^\alpha(S) \subseteq \Gamma_\theta^\beta(S)$ .*

*Beweis: Es ist zu zeigen, dass  $\Gamma_\theta^{\alpha-1}(S) \subseteq \Gamma_\theta^\alpha(S)$ , für jede Ordinalzahl  $\alpha > 0$ , gilt. Der Beweis ist über transfinite Induktion.<sup>20</sup>*

*Wenn  $\alpha = 1$ , dann ist  $\Gamma_\theta^{\alpha-1}(S) \subseteq \Gamma_\theta^\alpha(S)$  äquivalent zu der Aussage, dass  $S$  korrekt ist.*

*Wenn  $\alpha$  eine Nachfolgerordinalzahl ist, dann gilt aufgrund der Induktionshypothese  $\Gamma_\theta^{\alpha-2}(S) \subseteq \Gamma_\theta^{\alpha-1}(S)$ . Aus der Monotonie von  $\Gamma_\theta$  folgt  $\Gamma_\theta^{\alpha-1}(S) \subseteq \Gamma_\theta^\alpha(S)$ .*

*Wenn  $\alpha$  eine Limes-Ordinalzahl ist, dann gilt  $\Gamma_\theta^{\alpha-1}(S) = \bigcup_{\beta < \alpha} \Gamma_\theta^\beta(S)$ , sodass für alle Ordinalzahlen  $\beta < \alpha$ ,  $\Gamma_\theta^\beta(S) \subseteq \Gamma_\theta^{\alpha-1}(S)$ . Eine Anwendung von  $\Gamma_\theta$  auf beide Seiten ergibt für alle  $\beta < \alpha$ ,  $\Gamma_\theta^{\beta+1}(S) \subseteq \Gamma_\theta^\alpha(S)$ . Die Vereinigung aller  $\beta < \alpha$  ergibt das erwünschte Resultat  $\Gamma_\theta^{\alpha-1}(S) \subseteq \Gamma_\theta^\alpha(S)$ .  $\square$*

Aus Kardinalitätsgründen können wir nun zeigen, dass der Prozess, in welchem durch den Operator  $\Gamma_\theta$  immer größere Mengen generiert werden, an einem gewissen Punkt endet. Dieser Punkt ist der minimale Fixpunkt des Operators  $\Gamma_\theta$ . Dieser wird ausgehend von der leeren Menge generiert und ist folgendermaßen definiert:

**Definition 4.8** (vgl. [35], S. 118, [50], S. 86) *Die Menge  $I_\theta^\xi$  ist der minimale Fixpunkt der Operation  $\Gamma_\theta$ , wenn es eine Ordinalzahl  $\xi < \omega^+$  gibt, für die gilt:*

$$\Gamma_\theta^\xi(S) = \Gamma_\theta^{\xi-1}(S).$$

Um aufwändige technische Details zu vermeiden, gebe ich einfach nur eine Skizze des Kardinalitätsarguments (Details, siehe [50], S. 85f).

Sei  $|X|$  die Kardinalität einer Menge  $X$  und sei  $|(X)|^+$  das kleinste Kardinal größer als  $|X|$ . Eine Menge  $a$  ist abzählbar, wenn  $|a| \leq \omega$ . Insbesondere ist die Menge  $S$ , welche oben definiert wurde, abzählbar. Sei  $\omega^+ := \omega_1$  das kleinste überabzählbare Kardinal größer  $\omega$ . Dann gibt es  $|(X)|^+$  viele Ordinalzahlen der Kardinalität kleiner gleich  $|X|$ . Insbesondere gibt es  $\omega^+$  viele abzählbare Ordinalzahlen. Für unsere Zwecke ist außerdem wichtig, dass es  $\omega^+$  viele abzählbare Ordinalzahlen  $\leq \omega$  gibt. Das ist deshalb so, weil  $\sup\{\alpha : \alpha < \omega^+\} = \omega^+ = |\{\alpha : |\alpha| \leq \omega\}|$ .<sup>21</sup> Anschaulich gesprochen bedeutet das, dass wir mehr Ordinalzahlen zur Verfügung haben, als wir benötigen um eine abzählbare Menge zu erzeugen. In unserem Fall wollen wir durch den monotonen

<sup>20</sup>Transfinite Induktion ist die Verallgemeinerung der Induktion ins Transfinite für alle Ordinalzahlen. Für Details siehe [50], S. 22f.

<sup>21</sup> $\sup$  steht für das Supremum einer Menge. Ein Ordinal  $\alpha$  ist das Supremum einer Menge  $M$ , wenn  $\alpha$  das minimale Ordinal ist, für das gilt,  $\forall \beta \in M (\beta \leq \alpha)$ . Siehe [50], S. 22.

Operator  $\Gamma_\theta$  die Menge  $\omega$  der natürlichen Zahlen erzeugen, haben jedoch  $\omega^+$  viele Ordinalzahlen zur Hand um diese Menge zu erzeugen. Daraus folgt, dass unsere Menge an einem gewissen Punkt voll ist und das Hinzufügen einer weiteren Operation  $\Gamma_\theta^\xi$  keine neue Menge mehr erzeugt.

**Theorem 4.9** ([50], S. 86, [49], S. 118, [35], S. 118) (*Fixpunkt*) *Es gibt eine Ordinalzahl  $\xi < \omega^+$ , sodass  $\Gamma_\theta^\xi(S) = \Gamma_\theta^{\xi-1}(S)$ . Die so generierte Menge  $I_\theta^\xi$  ist der minimale Fixpunkt der Operation  $\Gamma_\theta$ .*

*Beweis:* Für alle durch den Operator  $\Gamma_\theta$  generierten Mengen  $S \subseteq \omega$  gilt  $|S| \leq \omega < \omega^+$ . Es gibt  $\omega^+$  viele Ordinalzahlen  $< \omega^+$ . Insbesondere gibt es  $\omega^+$  viele Ordinalzahlen  $\leq \omega$ . Deshalb gibt es eine Ordinalzahl  $\xi \leq \omega < \omega^+$  für die  $\Gamma_\theta^\xi(S) = \Gamma_\theta^{\xi-1}(S)$  gilt. Die so erzeugte Menge  $I_\theta^\xi$  ist der minimalen Fixpunkt der Operation  $\Gamma_\theta$ .  $\square$

Außerdem gilt natürlich  $I_\theta^\xi = I_\theta^{\xi-1}$ . Man nennt  $\xi$  die Abschlussordinalzahl von  $\Gamma_\theta$ , wenn  $\xi$  die minimale Ordinalzahl ist, für die  $\Gamma_\theta^\xi(S) = \Gamma_\theta^{\xi-1}(S)$  gilt. Für diese Ordinalzahl gilt außerdem, dass  $I_\theta^\xi = \bigcup_{\alpha \in On} I_\theta^\alpha$ , wobei  $On$  die Klasse aller Ordinalzahlen ist.

### 4.3 Die Semantik von Kripkes Theorie

Ausgehend von diesen Überlegungen, können wir nun eine Kripke-Wahrheitsmenge definieren. Diese ist die Kripke-Variante von Definition 2.10 und bildet die Grundlage für das Kripke-Feferman Axiomensystem, welches später besprochen wird.

**Definition 4.10** ([35], S. 204) *Eine Menge  $S \subseteq \omega$  ist eine Kripke-Wahrheitsmenge genau dann, wenn sie nur  $\mathcal{L}_T$  - Sätze enthält und, für geschlossene Terme  $s$  und  $t$ , sowie für die Sätze  $\phi, \psi$ , und  $\forall v\chi$  aus  $\mathcal{L}_T$ , die folgenden Bedingungen erfüllt:*

- (i)  $s = t \in S$  gdw  $s$  und  $t$  denselben Wert haben.
- (ii)  $s \neq t \in S$  gdw  $s$  und  $t$  nicht denselben Wert haben.
- (iii)  $\neg\neg\phi \in S$  gdw  $\phi \in S$ .
- (iv)  $\phi \wedge \psi \in S$  gdw  $\phi \in S$  und  $\psi \in S$ .
- (v)  $\neg(\phi \wedge \psi) \in S$  gdw  $\neg\phi \in S$  oder  $\neg\psi \in S$ .
- (vi)  $\phi \vee \psi \in S$  gdw  $\phi \in S$  oder  $\psi \in S$ .
- (vii)  $\neg(\phi \vee \psi) \in S$  gdw  $\neg\phi \in S$  und  $\neg\psi \in S$ .
- (viii)  $\forall x\chi \in S$  gdw für alle  $n$ ,  $\chi(\bar{n}) \in S$
- (ix)  $\neg\forall x\chi \in S$  gdw es ein  $n$  gibt mit  $\neg\chi(\bar{n}) \in S$ .
- (x)  $\exists x\chi \in S$  gdw es gibt ein  $n$  mit  $\chi(\bar{n}) \in S$ .
- (xi)  $\neg\exists x\chi \in S$  gdw für alle  $n$ ,  $\neg\chi(\bar{n}) \in S$ .
- (xii)  $Tt \in S$  gdw der Wert von  $t$  ein  $\mathcal{L}_T$  Satz in  $S$  ist.
- (xiii)  $\neg Tt \in S$  gdw der Wert von  $t$  kein  $\mathcal{L}_T$  Satz ist oder ein  $\mathcal{L}_T$  ist, dessen Negation in  $S$  ist.

Bemerke, dass in dieser Definition das Prädikat  $\in S$  nur positiv vorkommt und daher, wie oben beschrieben, Anlass zu einem monotonen Operator  $\Gamma$  gibt. Daher gilt, dass  $S$  eine Kripke Wahrheitsmenge ist gdw sie ein Fixpunkt dieses Operators ist; d.i.: gdw  $\Gamma(S) = S$ . Weiters ist der minimale Fixpunkt  $\Gamma(S) = S$  konsistent; d.i. es gibt keinen  $\mathcal{L}_T$  - Satz, sodass sowohl der Satz als auch seine Negation in  $\Gamma(S) = S$  ist. Außerdem sei bemerkt, dass wenn  $\Gamma(S) = S$  gilt, dann gilt auch  $T(\neg\phi) \in S$  gdw  $\phi \in S$ . All dies folgt aus den in Abschnitt 4.2 präsentierten Eigenschaften von induktiven Definitionen und monotonen Operatoren. Was sich geändert hat ist einzig, dass in Definition 4.10 ein Wahrheitsprädikat vorkommt. Da Definition 4.10 jedoch noch nicht in 3-wertiger Logik formuliert ist und daher von Kripkes Ansatz abweicht, sparen wir uns die Beweise für eine Sprache mit Wahrheitsprädikat für später auf. (vgl. [50], S. 204ff)<sup>22</sup>

Definition 4.10 bildet die Grundlage für das Axiomensystem Kripke-Feferman, in dem das Starke Kleene Evaluierungsschema nur implizit ist. Im Folgenden werden wir daher dieses Evaluierungsschema explizieren und seine formalen Eigenschaften beschreiben. Dabei wählen wir in erster Linie eine dreiwertige Logik, welche Wahrheitswert-Lücken erlaubt. Aufgrund dieser Tatsache müssen wir auch unsere Modelle ändern. Klassische Modelle sind Paare  $(\mathfrak{M}, S)$ , für die gilt, dass die Antiextension  $S^-$  von  $S^+$  schlicht das Komplement von  $S^+$  ist. Da wir jedoch Wahrheitswert-Lücken erlauben, gilt dies nicht: die Vereinigung von Extension und Antiextension bildet nicht die gesamte Menge der natürlichen Zahlen. Formal:  $S^+ \cup S^- \neq \omega$ . Weil letztere Aussage äquivalent zu der Aussage ist, dass die Theorie vollständig ist, spricht man auch von einer paravollständigen Theorie. Starke-Kleene Modelle haben daher die Form  $(\mathfrak{M}, S^+, S^-)$ , wobei  $\mathfrak{M}$  ein Modell von  $\mathcal{L}$  ist und  $S^+$  und  $S^-$  Teilmengen der Domäne von  $\mathfrak{M}$  sind. Da hier das arithmetische Vokabular standardmäßig interpretiert ist, definieren wir das Starke-Kleene Evaluierungsschema  $\models_{SK}$  für  $\mathfrak{M} = \mathbb{N}$ . Unsere Modelle sind also Tripel der Form  $(\mathbb{N}, S^+, S^-)$ .

**Definition 4.11** ([35], S. 205) (*Das Starke Kleene Evaluierungsschema.*) Für die Mengen  $S^+, S^- \subseteq \omega$ , ist  $(\mathbb{N}, S^+, S^-) \models_{SK}$ , für alle geschlossenen Terme  $s, t$ ,  $\mathcal{L}_T$  - Sätze  $\phi, \psi$  und  $\mathcal{L}_T$  - Formeln  $\chi(x)$ , mit genau  $x$  frei, folgendermaßen definiert:

- (i)  $(\mathbb{N}, S^+, S^-) \models_{SK} s = t$  gdw  $s$  und  $t$  denselben Wert haben.
- (ii)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg s = t$  gdw  $s$  und  $t$  nicht denselben Wert haben.
- (iii)  $(\mathbb{N}, S^+, S^-) \models_{SK} Tt$ , gdw  $t$  ein geschlossener Term ist, dessen Wert ein  $\mathcal{L}_T$  - Satz in  $S^+$  ist.
- (iv)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg Tt$  gdw  $t$  ein geschlossener Term ist, dessen Wert ein  $\mathcal{L}_T$  in  $S^-$  ist oder dessen Wert kein  $\mathcal{L}_T$  - Satz ist.
- (v)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\neg\phi$  gdw  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$ .
- (vi)  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi \wedge \psi$  gdw  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$  und  $(\mathbb{N}, S^+, S^-) \models_{SK} \psi$ .

<sup>22</sup>Ich folge hier [35]. Obwohl es an dieser Stelle vielleicht ungewöhnlich erscheint, die Konsequenzen von Definition 4.10 einfach kurz aufzulisten, ersparen wir uns dadurch einige Wiederholungen und somit auch viel Arbeit. Im Prinzip wurde damit nur etwas vorgegriffen.

- (vii)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg(\phi \wedge \psi)$  gdw  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi$ , oder  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\psi$ .  
(viii)  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi \vee \psi$  gdw  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$  oder  $(\mathbb{N}, S^+, S^-) \models_{SK} \psi$ .  
(ix)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg(\phi \vee \psi)$  gdw  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi$  und  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\psi$ .  
(x)  $(\mathbb{N}, S^+, S^-) \models_{SK} \forall x\phi(x)$  gdw für alle  $n \in \omega$   $(\mathbb{N}, S^+, S^-) \models_{SK} \phi(\bar{n})$ .  
(xi)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\forall x\phi(x)$  gdw für zumindest ein  $n \in \omega$   $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi(\bar{n})$ .  
(xii)  $(\mathbb{N}, S^+, S^-) \models_{SK} \exists x\phi(x)$  gdw für zumindest ein  $n \in \omega$   $(\mathbb{N}, S^+, S^-) \models_{SK} \phi(\bar{n})$ .  
(xiii)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\exists x\phi(x)$  gdw für alle  $n \in \omega$   $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi(\bar{n})$ .  
 $(\mathbb{N}, S^+, S^-) \models_{SK} n$  gilt für kein  $n$  das nicht ein Kode eines  $\mathcal{L}_T$  - Satzes ist.

Offensichtlich erfüllt Definition 4.11 die Anforderungen einer positiven induktiven Definition. Die folgenden Beispiele veranschaulichen dies und sind direkte Konsequenzen der Definition:

- ...  
(iii)'  $(\mathbb{N}, S^+, S^-) \models_{SK} T\bar{n}$  gdw  $n \in S^+$   
(iv)'  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg T\bar{n}$  gdw  $n \in S^-$   
...  
(vi)'  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi \wedge \psi$  gdw  $\phi \in S^+$  und  $\psi \in S^+$   
(vii)'  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg(\phi \wedge \psi)$  gdw  $\neg\phi \in S^-$  oder  $\neg\psi \in S^-$ .  
...

Da wir Wahrheitswert-Lücken erlauben, ist ein defekter, bzw. paradoxer Satz weder wahr noch falsch. Sei  $n$  so ein Satz. Dann gilt  $n \notin S^+ \cup S^-$  und weder  $T\bar{n}$ , noch  $\neg T\bar{n}$  werden vom Modell  $(\mathbb{N}, S^+, S^-)$  erfüllt (vgl. [35], S. 206).

**Definition 4.12 (Konsistenz)** ([35], S. 207) *Ein SK Modell ist konsistent gdw es keinen  $\mathcal{L}_T$  - Satz  $\phi$  gibt, für den gilt  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$  und  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi$ .*

Um zu veranschaulichen wie wir diese konsistenten Modelle erhalten, werden im Folgenden die semantischen Eigenschaften einer starken-Kleene Logik ( $K_3$ ) mit Wahrheitswert-Lücken dargestellt. Die semantischen Eigenschaften der Konnektive sind in den folgenden drei Tabellen dargestellt (vgl. [51] S. 122f, [35], S. 207):

| $\phi$ | $\neg\phi$ |
|--------|------------|
| w      | f          |
| f      | w          |
| -      | -          |

| $\phi$ | $\psi$ | $\phi \wedge \psi$ |
|--------|--------|--------------------|
| w      | w      | w                  |
| w      | f      | f                  |
| f      | w      | f                  |
| f      | f      | f                  |
| w      | -      | -                  |
| f      | -      | f                  |
| -      | w      | -                  |
| -      | f      | f                  |
| -      | -      | -                  |

| $\phi$ | $\psi$ | $\phi \vee \psi$ |
|--------|--------|------------------|
| w      | w      | w                |
| w      | f      | w                |
| f      | w      | w                |
| f      | f      | f                |
| w      | -      | w                |
| f      | -      | -                |
| -      | w      | w                |
| -      | f      | -                |
| -      | -      | -                |

Aufgrund dieser Definition ist mehr oder weniger klar, dass, wenn  $\lambda$  der Lügner Satz ist, er den Wahrheitswert ‘-‘ (unbestimmt) erhält. Wie wir bereits wissen, gibt es aufgrund des Diagonallemmas einen Satz  $\lambda \leftrightarrow \neg T(\ulcorner \lambda \urcorner)$ . Würden wir in klassischer Logik arbeiten, würde aufgrund der Definition 4.11 sowohl  $\lambda \in S^+$  ( $\lambda$  und Definition 4.11 (iii)) als auch  $\lambda \in S^-$  gelten ( $\neg T(\ulcorner \lambda \urcorner)$  und Definition 4.11 (iv)). Somit wäre  $\lambda \in S^+ \cap S^-$  und Definition 4.11 wäre inkonsistent (nach Definition 4.12). Da wir jedoch mit der dreiwertigen Logik  $K_3$  arbeiten, ist  $\lambda$  unbestimmt. Diese Tatsache, gemeinsam mit der offensichtlichen Definition von  $\leftrightarrow$  aus  $\wedge, \vee$  und  $\neg$  aus der Semantik für  $K_3$ , ergibt, dass  $\lambda \leftrightarrow \neg T(\ulcorner \lambda \urcorner)$  unbestimmt ist und  $\lambda$  weder in  $S^+$  noch in  $S^-$  ist und deshalb wie erwünscht  $S^+ \cap S^- = \emptyset$  gilt.

Daher erhalten wir folgendes Lemma:

**Lemma 4.13** ([35], S. 207) *Wenn ein SK-Modell  $(\mathbb{N}, S^+, S^-)$  konsistent ist, dann gibt es keinen Satz  $\phi$  für den gilt  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$  und  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi$ .*

*Beweis: Induktion über die Länge von  $\phi$ . Wenn  $\phi$  atomar ist, folgt die Behauptung direkt aus der Konsistenz des Modells (siehe Definition 4.12). Für komplexe Formeln nehmen wir die Konjunktion als Beispiel und machen die folgenden zwei Annahmen um einen Widerspruch zu erzeugen:*

- (i)  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi \wedge \psi$
- (ii)  $(\mathbb{N}, S^+, S^-) \models_{SK} \neg(\phi \wedge \psi)$

*Aus (i) und Definition 4.10 folgt*

$$(\mathbb{N}, S^+, S^-) \models_{SK} \phi \text{ und } (\mathbb{N}, S^+, S^-) \models_{SK} \psi.$$

*Annahme (ii) besagt jedoch, dass entweder*

$$(\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi \text{ oder } (\mathbb{N}, S^+, S^-) \models_{SK} \neg\psi \text{ gelten muss.}$$

*Dies widerspricht jedoch der Induktionsvoraussetzung (Definition 4.12) die besagt, dass ein SK-Modell nicht gleichzeitig  $\phi$  und  $\neg\phi$  oder  $\psi$  und  $\neg\psi$  enthalten kann. Wäre dem so, dann wäre, aufgrund der Definition von  $(\mathbb{N}, S^+, S^-)$ ,  $\phi$  sowohl in  $S^+$  als auch in  $S^-$ . Dies widerspricht der Induktionsvoraussetzung direkt. Analoges gilt für die anderen Fälle.  $\square$*

Nun können wir das Fixpunktlemma für das starke Kleene Evaluierungsschema beweisen. Wie oben bemerkt, kann Definition 4.11 leicht in eine positive induktive Definition umgewandelt werden. Die positive induktive Definition über  $\theta(n, S)$ , aus dem

vorhergehenden Abschnitt, muss daher auf diese Definition mit Extension  $S^+$  und Antiextension  $S^-$  verallgemeinert werden. Das ist kein Problem, da hier ausgehend von Definition 4.11 und Definition 4.1 eine positive induktive Definition die Form  $n \in (S^+, S^-)$  gdw  $\sigma(n, S^+, S^-)$  hat. Außerdem ist klar, dass als entsprechende Variation davon  $\sigma(n, S^+, S^-)$  gdw  $(\mathbb{N}, S^+, S^-) \models_{SK} n$  gilt. Dies vorausgesetzt, können wir nun einen monotonen Operator einführen.

**Definition 4.14** ([35], S. 208, [37], S. 118) *Ein monotoner Operator für die Mengen  $(S^+, S^-) \subseteq \omega$  ist die Operation:*

$$\Delta_\sigma : Pot(S^+, S^-) \rightarrow Pot(S^+, S^-), \text{ sodass}$$

$$\Delta_\sigma(S^+, S^-) := (\{\phi \in \mathcal{L}_T : (\mathbb{N}, S^+, S^-) \models_{SK} \phi\}, NSatz \cup \{\phi \in \mathcal{L}_T : (\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi\}),$$

wobei  $NSatz$  die Menge der Zahlen ist, die keine Codes von  $\mathcal{L}_T$  - Sätzen sind.

Intuitiv geschieht in dieser Konstruktion Folgendes: wir beginnen mit einem Tupel von zwei leeren Mengen. Die Extension des Wahrheitsprädikats ist daher  $\mathfrak{M}_0 = (\mathbb{N}, \emptyset, \emptyset)$ . D.h. an dieser Stelle sind alle Sätze, auf die das Wahrheitsprädikat zutrifft, undefiniert. Dann wenden wir den monotonen Operator auf diese Menge an und bestimmen so die Menge der Sätze  $S_1^+$ , die von  $\mathfrak{M}_0$  wahr gemacht wurden und die Menge der Sätze  $S_1^-$ , deren Negation von  $\mathfrak{M}_0$  wahr gemacht wurden. So erhalten wir ein neues Modell  $\mathfrak{M}_1 = (\mathbb{N}, S_1^+, S_1^-)$ , welches dazu verwendet wird, wieder eine neue neue Extension  $S_2^+$  und Antiextension  $S_2^-$  zu bestimmen, usw. Der monotone Operator  $\Delta_\sigma$  liefert also durch wiederholte Anwendung immer die Menge der Sätze, welche vom Vorgängermodell wahr gemacht werden (ad infinitum) (vgl. [37], S. 118, [43], S. 703). Formal wird  $\Delta_\sigma$  analog zu Definition 4.5 wie folgt definiert.

**Definition 4.15** *Für die Operation  $\Delta_\sigma$  sei die Sequenz  $\langle \Delta_\sigma^\alpha(S^+, S^-) \mid \alpha \text{ eine Ordinalzahl} \rangle$  über transfinite Rekursion wie folgt definiert:*

$$\Delta_\sigma^0(S^+, S^-) = (S^+, S^-) \text{ und } \Delta_\sigma^\alpha(S^+, S^-) = \Delta_\sigma(\Delta_\sigma^{\alpha-1}(S^+, S^-)), \text{ für alle } \alpha > 0,$$

wobei, wenn  $\alpha$  eine Limes-Ordinalzahl ist, dann gilt:

$$\Delta_\sigma^{\alpha-1}(S^+, S^-) = \bigcup_{\beta < \alpha} \Delta_\sigma^\beta(S^+, S^-).$$

Insbesondere gilt für eine Limes-Ordinalzahl, dass  $\Delta_\sigma^\alpha(S^+, S^-) = \Delta_\sigma(\bigcup_{\beta < \alpha} \Delta_\sigma^\beta(S^+, S^-))$ .

Offensichtlich ist der Operator  $\Delta_\sigma$  monoton. Daraus ergibt sich:

**Lemma 4.16** *Wenn  $S_1^+ \subseteq S_2^+$ , und  $S_1^- \subseteq S_2^-$ , dann  $\Delta_\sigma(S_1^+, S_1^-) \subseteq \Delta_\sigma(S_2^+, S_2^-)$ .*

*Beweis: Dies folgt aus der Monotonie der positiven induktiven Definition.  $\square$*

sowie

**Lemma 4.17** *Für alle Ordinalzahlen  $\alpha, \beta$ , wenn  $\alpha < \beta$ , dann  $\Delta_\sigma^\alpha(S^+, S^-) \subseteq \Delta_\sigma^\beta(S^+, S^-)$ .*

*Beweis: Analog zu Lemma 4.7.  $\square$*

Wie wir oben gesehen haben, haben diese Lemmata die folgende Konsequenz:

**Theorem 4.18** *Es gibt ein Ordinal  $\xi$ , für das gilt:  $\Delta_\sigma^\xi(S^+, S^-) = \Delta_\sigma^{\xi-1}(S^+, S^-)$ .*

*Beweis: Analog zu Theorem 4.9.  $\square$*

Die durch  $\Delta_\sigma^\xi$  generierte Menge ist der minimale Fixpunkt und enthält alle  $\phi \in \mathcal{L}_T$ , für die  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$  und  $NSatz \cup (\mathbb{N}, S^+, S^-) \models_{SK} \neg\phi$  gilt.

Dieses Theorem liefert uns nun die minimale Extension für das Wahrheitsprädikat. Zu bemerken ist hier, dass der Fixpunkt stufenweise, ausgehend von der leeren Menge, aufgebaut wurde. Eine ähnliche Konstruktion findet sich bei Tarski für mehrere getypte Wahrheitsprädikate  $T_0 \dots T_n$ . Kripke merkt dazu an:

The minimal fixed point (...) avoids explicit subscripts but nevertheless has a notion of level; in this respect it can be compared with standard set theory as opposed to the theory of types. ([43], Fußnote 10)

Dies ist deshalb bemerkenswert, weil hier eine historische Analogie anklingt: Ähnlich wie man im Fall der Metamathematik im frühen zwanzigsten Jahrhundert von einer Typentheorie zur Mengenlehre übergegangen ist, wird hier von einer getypten zu einer ungetypten Wahrheitstheorie gewechselt, welche als natürlicher angesehen wird als ihr Vorgänger.

Das nächste Theorem zeigt eine weitere wesentliche Eigenschaft von Kripkes Theorie, welche - wie wir sehen werden, eine zentrale Rolle in der modernen Deflationismusdebatte spielt. Es etabliert die Intersubstituierbarkeit von  $T(\ulcorner \phi \urcorner)$  und  $\phi$  für alle  $\phi \in \mathcal{L}_T$ .

**Theorem 4.19** ([35], S. 209, [37], S. 121) *Wenn  $\Delta_\sigma^\xi(S^+, S^-)$  ein Fixpunkt ist, dann gilt  $(\mathbb{N}, S^+, S^-) \models_{SK} T^\ulcorner \phi \urcorner$  gdw  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$ .*

*Beweis: Definition 4.10 liefert  $(\mathbb{N}, S^+, S^-) \models_{SK} T^\ulcorner \phi \urcorner$  gdw  $\phi \in S^+$ . Im minimalen Fixpunkt ist  $S^+$  die Menge aller  $\mathcal{L}_T$  - Sätze  $\phi$ , für die gilt, dass  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$ . Deshalb gilt, wenn  $\phi \in S^+$ , ebenfalls  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$ .  $\square$*

Bezüglich dem minimalen Fixpunkt verdanken wir Kripke außerdem folgende Definition (vgl. [43], S. 706f):

**Definition 4.20** (*Grounding*) Ein  $\mathcal{L}_T$  - Satz  $\phi$  ist *grounded* gdw er im minimalen Fixpunkt enthalten ist.

Ein Satz  $\phi$  ist *ungrounded* gdw er nicht im minimalen Fixpunkt enthalten ist.

Intuitiv besagt Grounding, dass ein Satz von nicht-formalsemantischen Fakten abhängt. Da wir bei der Konstruktion des minimalen Fixpunkts von der leeren Menge ausgegangen sind, dann Sätze ohne Wahrheitsprädikat betrachtet haben und so immer weiter gegangen sind, hängt jeder Satz im minimalen Fixpunkt von Sätzen ohne Wahrheitsprädikat ab.

(...), the least fixed point model is special. In the first stage, the extension and the anti-extension of the truth predicate were taken to be empty. Then we looked at the world (the world of natural numbers, in this case) to include sentences in the extension and in the anti-extension. Then we built on that. As a result, the sentences in the extension and anti-extension of the least fixed point are all *grounded* in the world. Their truth value is ultimately determined by facts in the (mathematical) world. For other fixed points, this is not always the case. ([35], S. 122)

Zu bemerken ist hierbei, dass es mehrere Fixpunkte neben dem minimalen gibt. Obwohl mehrere Autoren dem minimalen Fixpunkt einen speziellen Status einräumen (siehe Zitat Horsten oben, [45]) leisten die nichtminimalen Fixpunkte wichtige Beiträge zur Behandlung verschiedener Paradoxien und sind daher auf keinen Fall abzutun.

Arguably, Kripkes main contribution was not so much the construction of the smallest fixed point of  $\Delta$  but rather his classification of the different consistent fixed points and the discussion of their use for discriminating between ungrounded sentences, paradoxical sentences, and so on. ([35], S. 210).

Dieser Ansicht war auch Kripke selbst.

Nor is it even my present purpose to make any firm recommendation between the minimal fixed point of a particular valuation scheme and the various other fixed points.\* Indeed, without the nonminimal fixed points we could not have defined the intuitive difference between ‘grounded’ and ‘paradoxical’. My purpose is rather to provide a family of flexible instruments which can be explored simultaneously and whose fertility and consonance with intuition can be checked. ([43], S. 712)

\*Though the minimal fixed point certainly is singled out as natural in many respects. (Fußnote 29, p.712)

Da die Klassifizierung diverser Fixpunkte eine eigene Arbeit erfordern würde, sei hier nur auf [27], [9], [43] verwiesen.

#### 4.4 KF

Die Theorie KF ist eine Axiomatisierung von Kripkes semantischer Theorie, die auf Solomon Feferman zurückgeht. Die Basis für KF ist Definition 4.10.

**Definition 4.21** (KF) ([35], S. 201) *Das System KF besteht aus allen Axiomen von  $PA_T$  und den folgenden Axiomen:*

- KF1  $\forall s \forall t (T(s \doteq t) \leftrightarrow s^\circ = t^\circ)$
- KF2  $\forall s \forall t (T(\dot{\neg}s \doteq t) \leftrightarrow s^\circ \neq t^\circ)$
- KF3  $\forall x (Sent_T(x) \rightarrow (T(\dot{\neg}\dot{\neg}x) \leftrightarrow Tx))$
- KF4  $\forall x \forall y ((Sent_T(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow Tx \wedge Ty))$
- KF5  $\forall x \forall y ((Sent_T(x \dot{\wedge} y) \rightarrow (T(\dot{\neg}(x \dot{\wedge} y)) \leftrightarrow T\dot{\neg}x \text{ oder } T\dot{\neg}y))$
- KF6  $\forall x \forall y ((Sent_T(x \dot{\vee} y) \rightarrow (T(x \dot{\vee} y) \leftrightarrow Tx \vee Ty))$
- KF7  $\forall x \forall y ((Sent_T(x \dot{\vee} y) \rightarrow (T(\dot{\neg}(x \dot{\vee} y)) \leftrightarrow T\dot{\neg}x \wedge T\dot{\neg}y))$
- KF8  $\forall v \forall x ((Sent_T(\dot{\forall}vx) \rightarrow (T(\dot{\forall}vx) \leftrightarrow \forall t T(x(t/v))))$
- KF9  $\forall v \forall x ((Sent_T(\dot{\forall}vx) \rightarrow (T(\dot{\neg}\dot{\forall}vx) \leftrightarrow \exists t T(\dot{\neg}x(t/v))))$
- KF10  $\forall v \forall x ((Sent_T(\dot{\exists}vx) \rightarrow (T(\dot{\exists}vx) \leftrightarrow \exists t T(x(t/v))))$
- KF11  $\forall v \forall x ((Sent_T(\dot{\exists}vx) \rightarrow (T(\dot{\neg}\dot{\exists}vx) \leftrightarrow \forall t T(\dot{\neg}x(t/v))))$
- KF12  $\forall t (T(\dot{T}t) \leftrightarrow Tt^\circ)$
- KF13  $\forall t (T\dot{\neg}\dot{T}t \leftrightarrow (T\dot{\neg}t^\circ \vee \neg Satz_T(t^\circ)))$

Da wir schon bei der Besprechung der Semantik von KF viele wichtige Eigenschaften der Theorie kennengelernt haben und die wichtigsten philosophischen Überlegungen im Zuge dieses Kapitels geklärt wurden, begnüge ich mich hier nur mit drei zentralen Konsequenzen von KF:

- KF ist  $\omega$  - konsistent (siehe [37], S.126)
- KF ist arithmetisch korrekt. (siehe [37], S.126)
- KF beweist alle Disquotations-Sätze der Form  $\forall t_1, \dots, \forall t_n (T^\top \phi(\dot{t}_1, \dots, \dot{t}_n)^\top \leftrightarrow \phi(t_1^\circ, \dots, t_n^\circ))$ , für alle T-positiven Formeln  $\phi(x_1, \dots, x_n)$ . (siehe [35], S.201)

#### 4.5 Wahrheitswert-Überschneidungen und Supervaluierung

Bisher haben wir in unserer Evaluierung eine dreiwertige Logik verwendet, die Wahrheitswert-Lücken erlaubt. Diese ist jedoch nicht exklusiv. Ebenso können wir eine Logik verwenden, die Wahrheitswert-Überschneidungen erlaubt. Galt für eine Logik mit Wahrheitswert Lücken, dass die Schnittmenge von Extension und Antiextension leer ist (siehe

auch Definition 4.12), die Vereinigung von Extension und Antiextension jedoch nicht die Menge der natürlichen Zahlen  $\omega$  ist (da nicht jeder Satz einen Wahrheitswert erhält), gilt für eine Logik mit Wahrheitswert-Überschneidungen genau das Gegenteil: die Vereinigung von Extension und Antiextension ist  $\omega$  (d.h. sie ist vollständig), die Schnittmenge von Extension und Antiextension ist jedoch nicht leer. Formal ([35], S.208):

- (a)  $S^+ \cup S^- = \omega$  (Vollständigkeit)  
 (b)  $S^+ \cap S^- = \emptyset$  (Konsistenz)

Gemeinhin wird eine Logik mit Wahrheitswert-Überschneidungen LP (= Logik der Paradoxien) genannt. Die Semantik dieser Logik wird in folgender Tabelle veranschaulicht ([35], S. 207, [51], S. 124):

| $\phi$ | $\neg\phi$ |
|--------|------------|
| w      | f          |
| f      | w          |
| w/f    | w/f        |

| $\phi$ | $\psi$ | $\phi \wedge \psi$ |
|--------|--------|--------------------|
| w      | w      | w                  |
| w      | f      | f                  |
| f      | w      | f                  |
| f      | f      | f                  |
| w      | w/f    | w/f                |
| f      | w/f    | f                  |
| w/f    | w      | w/f                |
| w/f    | f      | f                  |
| w/f    | w/f    | w/f                |

| $\phi$ | $\psi$ | $\phi \vee \psi$ |
|--------|--------|------------------|
| w      | w      | w                |
| w      | f      | w                |
| f      | w      | w                |
| f      | f      | f                |
| w      | w/f    | w                |
| f      | w/f    | w/f              |
| w/f    | w      | w                |
| w/f    | f      | w/f              |
| w/f    | w/f    | w/f              |

Ein zentraler Unterschied zwischen den beiden Logiken  $K_3$  und LP ist, dass in  $K_3$  das Gesetz des ausgeschlossenen Dritten nicht gilt; also  $K_3 \not\models \phi \vee \neg\phi$ , die Logik LP es jedoch erfüllt; also  $LP \models \phi \vee \neg\phi$ . (vgl. [51], S. 124f)

Will man in einem axiomatischen System wie KF Wahrheitswert-Überschneidungen ausschließen, kann man ihm folgendes Konsistenz Axiom hinzufügen, welches besagt, dass kein Satz sowohl wahr als auch falsch ist (vgl. [35], S. 212).

$$(\text{Konsistenz}) \quad \forall x (Sent_T(x) \rightarrow \neg(Tx \wedge T\dot{\neg}x))$$

Will man Wahrheitswert-Lücken ausschließen, kann man folgendes Vollständigkeits Axiom hinzufügen, welches besagt, dass jeder Satz entweder wahr oder falsch ist (oder beides) (vgl. [35], S. 212).

$$(\text{Vollständigkeit}) \quad \forall x (Sent_T(x) \rightarrow (Tx \vee T\dot{\neg}x))$$

Sowohl die Logik  $K_3$  als auch die Logik LP basieren auf dem starken Kleene Evaluierungsschema. Ein alternatives Evaluierungsschema, welches hier nur angemerkt wird, ist das Supervaluierungsschema. In diesem Schema ist ein Satz  $\phi \in \mathcal{L}_T$  gültig in einem partiellen Modell  $(\mathbb{N}, S^+, S^-)$  gdw er gültig in jedem klassischen Modell  $(\mathbb{N}, S)$  ist. Der minimale Fixpunkt dieses Schemas lässt den Lügner Satz  $\lambda$  unbestimmt. Da jedoch das Gesetz des ausgeschlossenen Dritten gültig ist, gilt ebenfalls  $\lambda \vee \neg\lambda$ . Das führt zwar nicht zu einem direkten Widerspruch, hat jedoch die unangenehme Konsequenz, dass die Supervaluations-Variante von Kripkes Theorie nicht kompositional ist. Da dies jedoch eine der wenigen intuitiven Anforderungen ist, an die wir auch fest glauben, werten wir diese Tatsache als unakzeptablen Nachteil dieser Theorie. Eine Axiomatisierung dieser Theorie, genannt VF (nach van Fraassen), wurde von Cantini vorgebracht. (vgl. [37], S. 123, [35], S. 264f, [51], S.134f)

## 4.6 Revenge

Kripkes Theorie hat viele nette Eigenschaften, jedoch ein zentrales Problem: sie ist nicht revenge-immun. Dies ergibt sich aus der Tatsache, dass, obwohl wir eine mehrwertige innere Logik verwendet haben, die Metatheorie (die Axiome KF) in klassischer Logik formuliert sind. So ist zum Beispiel aufgrund unserer mehrwertigen Logik und der Tatsache, dass weder der Lügnersatz  $\lambda$ , noch seine Negation, im minimalen Fixpunkt enthalten sind, der Lügnersatz  $\lambda$  weder wahr noch falsch. Demnach ist  $T^\top \lambda \vee \neg\lambda^\top$  kein Theorem von KF. Dennoch ist  $\lambda \vee \neg\lambda$  eine klassische Tautologie und daher ein Theorem von KF und wir würden gerne sagen, dass der Lügnersatz nicht wahr ist, können es aber nicht. (vgl. [35], S. 206, [37], S. 124, [24], S.73)

For Kripke's theory, the problem is simply that he has presented his theory as a piece of orthodox mathematics, in a classical metalanguage, and has allowed himself to say, for instance, that if  $\lambda$  is the liar sentence, then  $\lambda$  is not true. But according to his own theory, if  $\lambda$  is the liar sentence, then not  $T(\lambda)$  is itself not true. So in expounding his theory, he seems to be saying things that, according to that very theory are not true. ([10] S. 13)

Formal ist die geradlinigste mir bekannte Art und Weise dies auszudrücken, die Behauptung, dass Kripkes Theorie nicht symmetrisch ist. Um dies formulieren zu können, benötigen wir jedoch noch drei Definitionen ([35], S. 152f).

**Definition 4.22** (*NEC*) Die Regel NEC ist:

$$\frac{\phi}{T(\top\phi^\top)}$$

Diese Regel korrespondiert dem Schema  $\phi \rightarrow T(\ulcorner \phi \urcorner)$  und erlaubt es von  $\phi$  auf  $T(\ulcorner \phi \urcorner)$  zu schließen. NEC darf nicht auf Annahmen in einem Kalkül angewendet werden und ist insofern schwächer als das Schema.

**Definition 4.23** (*CONEC*) Die Regel CONEC ist:

$$\frac{T(\ulcorner \phi \urcorner)}{\phi}$$

Diese Regel korrespondiert dem Schema  $T(\ulcorner \phi \urcorner) \rightarrow \phi$  und erlaubt es von  $T(\ulcorner \phi \urcorner)$  auf  $\phi$  zu schließen. CONEC darf ebenfalls nicht auf Annahmen in einem Kalkül angewendet werden und ist insofern schwächer als das Schema.

**Definition 4.24** (*Symmetrie*) Eine Wahrheitstheorie ist symmetrisch gdw sie abgeschlossen ist unter NEC und CONEC.

Ist eine Wahrheitstheorie nicht symmetrisch, muss sie entweder NEC oder CONEC einschränken. Da die beiden Regeln NEC und CONEC schwächer sind als die ihnen korrespondierenden Axiome ( $T(\ulcorner \phi \urcorner) \rightarrow \phi$  und  $\phi \rightarrow T(\ulcorner \phi \urcorner)$ ), die gemeinsam das Bikonditional ergeben, ist es außerdem klar, dass eine asymmetrische Theorie nicht unrestriktiert alle Bikonditionale ableiten kann.

Die folgenden Lemmata zeigen, dass  $KF + NEC$  inkonsistent und  $KF$  daher asymmetrisch ist.

**Lemma 4.25** ([37], S. 127) Für alle  $\phi \in \mathcal{L}_T$ ,  $KF \vdash T(\ulcorner \phi \urcorner) \rightarrow \phi$ .

*Beweis:* Für jeden Satz  $T(\ulcorner \phi \urcorner)$ , der im minimalen Fixpunkt enthalten ist, gilt, dass auch  $\phi$  im minimalen Fixpunkt enthalten ist, da ein Satz mit Wahrheitsprädikat nur dann einen Wert erhält, wenn er auch ohne Wahrheitsprädikat einen Wahrheitswert hat.  $\square$

**Lemma 4.26** ([37], S. 127)  $KF \vdash \lambda \wedge \neg T(\ulcorner \lambda \urcorner)$ .

*Beweis:* Es gibt einen Satz  $\lambda$ , sodass  $KF \vdash \lambda \leftrightarrow \neg T(\ulcorner \lambda \urcorner)$ .

- (i)  $\neg \lambda$  (Annahme)
- (ii)  $T(\ulcorner \lambda \urcorner)$  (Def. von  $\lambda$ )
- (iii)  $\lambda$  (Lemma 4.20)
- (iv)  $\lambda$  (von (iii) und (i))
- (v)  $\neg T(\ulcorner \lambda \urcorner)$  ((iv) und Def.  $\lambda$ )
- (vi)  $\lambda \wedge \neg T(\ulcorner \lambda \urcorner)$  ((iv) und (v)).  $\square$

**Lemma 4.27** ([37], S. 128) *KF + NEC ist inkonsistent.*

*Beweis:* Dies folgt aus Lemma 4.26. Eine Anwendung von NEC auf (iv) ergibt (\*)  $T(\ulcorner \lambda \urcorner)$ . (\*) und (v) ergeben  $T(\ulcorner \lambda \urcorner) \wedge \neg T(\ulcorner \lambda \urcorner)$ .  $\square$

**Korollar 4.28** *KF ist nicht symmetrisch. Das folgt aus Lemma 4.27 und der Definition von Symmetrie.*

Natürlich hat Kripke auf dieses Problem selbst aufmerksam gemacht. Da Kripkes Formulierungen präzise, wichtig und einzigartig sind, lohnt es sich, diese hier ausführlich zu wiederholen.

It seems likely that many who have worked on the truth-gap approach to the semantic paradoxes have hoped for a universal language, one in which everything that can be stated at all can be expressed. (...) Now the languages of the present approach contain their own truth predicate and even their own satisfaction predicates, and thus to this extent the hope has been realized. Nevertheless the present approach does not claim to give a universal language, and I doubt that such a goal can be achieved. First, the induction defining the minimal fixed point is carried out in a set-theoretic metalanguage, not in the object language itself. Second, there are assertions we can make about the object language which we cannot make in the object language. For example, Liar sentences are *not true* in the object language, in the sense that the induction process never makes them true; but we are precluded from saying this in the object language by our interpretation of negation and the truth predicate. (...) The necessity to ascend to a metalanguage may be one of the weaknesses of the present theory. The ghost of the Tarski hierarchy is still with us. ([43], S. 714)

Kripke selbst hat den positiven Wert seiner Arbeit jedoch vor allem darin gesehen, eine Theorie vorgebracht zu haben, die ein ungetyptes Wahrheitsprädikat definiert und daher unserer Umgangssprache näher kommt.

On the basis of the fact that the goal of a universal language seems elusive, some have concluded that truth-gap approaches, or any approaches that attempt to come closer to natural language than does the orthodox approach, are fruitless. I hope that the fertility of the present approach, and its agreement with intuitions about natural language in a large number of instances, cast doubt upon such negative attitudes. ([43], S. 715)

Zusammenfassend sind wir daher einerseits aufgrund des Revenge-Problems und der Tatsache, dass wir auf mengentheoretische Ressourcen weitgehend nicht verzichten

können, berechtigt zu bezweifeln, dass wir eine universale Sprache aufstellen können. Andererseits enthält die Sprache von Kripkes Theorie ihr eigenes ungetyptes Wahrheits- und Erfüllungsprädikat und kommt so unserem intuitiven Sprachverständnis näher als Tarskis Theorie. Im nächsten Kapitel werden wir sehen, wie die Tradition nach Kripke diese Problemstellungen beurteilt hat.

## 5 Deflationismus nach Kripke: Die zeitgenössische Debatte

In Kapitel 3 wurde der frühe Deflationismus im Anschluss an Tarskis Wahrheitstheorien besprochen. Wir haben damit geschlossen, dass die These des frühen Deflationismus nicht haltbar ist und wir uns deshalb einer anderen Form des Deflationismus zuwenden sollten. Genau dies wurden im Anschluss an Kripke sowohl in formaler als auch in philosophischer Hinsicht getan. Im Folgenden werden daher die zentralen Entwicklungen der letzten Jahre nach Kripke präsentiert und kritisch begutachtet.

### 5.1 Eine neue Form des Deflationismus

In Abschnitt 3 wurde behauptet, dass ein zentrales Anliegen des Deflationismus die Ableitung aller Tarski-Bikonditionale ist. Im Kontext von Tarskis Theorien ist dies die Forderung,  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  für alle  $\phi \in \mathcal{L}_{\mathcal{PA}}$  abzuleiten. Die zwei zentralen Probleme in diesem Zusammenhang sind: (1) Eine Theorie mit dem Tarski - Bikonditional als einzigem Axiom (TB), ist formal nicht adäquat, da sie keine infiniten Generalisierungen beweist und konservativ ist (siehe Kapitel 3 und Abschnitt 2.3). In diesem Zusammenhang wurde dann auch die Idee verabschiedet, dass Deflationismus auf Konservativität festgelegt ist. (2) Ein getyptes Wahrheitsprädikat entspricht nicht unseren Intuitionen und ist daher aus philosophischen Gründen zu verabschieden.

Daraus ergibt sich natürlich die Frage, worauf der Deflationismus im Kontext einer ungetypten Wahrheitstheorie festgelegt ist und welche philosophischen Ideen damit verbunden sind. Wie zuvor ist diese Frage nicht ganz eindeutig zu beantworten. Hinzu kommt an diesem Punkt außerdem, dass es sich dabei um sehr junge Probleme handelt und die verfügbare Literatur zum Thema sowohl rar als auch skizzenhaft ist. Nichtsdestotrotz werde ich versuchen, die Position so klar wie möglich, mit den mir zur Verfügung stehenden Mitteln, darzustellen.

Oben wurde gezeigt, dass Kripke eine konsistente Wahrheitstheorie aufgestellt hat, in welcher es möglich ist, ein ungetyptes Wahrheitsprädikat für die Sprache  $\mathcal{L}_{\mathcal{T}}$  aufzustellen. Seine Theorie hat außerdem die nette Eigenschaft, dass in jedem Fixpunkt  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  für alle  $\phi \in \mathcal{L}_{\mathcal{T}}$  gilt. Dennoch ist seine Theorie (wie in Abschnitt 4.6 gezeigt wurde) nicht revenge-immun; bzw nicht symmetrisch:  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  für alle  $\phi \in \mathcal{L}_{\mathcal{T}}$  hält in seiner Theorie nicht *im Allgemeinen*.

Weiters wurde schon besprochen, dass die Forderung der Intersubstituierbarkeit von  $T(\ulcorner \phi \urcorner)$  und  $\phi$  eng mit der Ansicht verbunden ist, dass Wahrheit metaphysisch unwichtig, bzw. insubstantiell ist und daher sehr gut zum Projekt des Deflationismus passt (vgl. Abschnitt 3.1). Obwohl sowohl Tarskis als auch Kripkes Theorien dies in der einen oder anderen Form leisten, ist dies bei ihnen nicht *im Allgemeinen* (respektive für die Sprachen  $\mathcal{L}_{\mathcal{PA}}$  und  $\mathcal{L}_{\mathcal{T}}$ ) möglich, da sie jeweils weder revenge-immun sind, noch ihre eigene Metasprache enthalten können. Nimmt man die Forderung der Intersubstituierbarkeit von  $T(\ulcorner \phi \urcorner)$  und  $\phi$  daher ernst, sind für einen Deflationisten weder Tarskis noch Kripkes Theorie akzeptabel und es ergibt sich eine neue These des Deflationismus (TD). Diese ist, dass in einer deflationistische Theorie alle Sätze der Form

(TD)  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  für alle  $\phi \in \mathcal{L}_{\mathcal{T}}$  *im Allgemeinen*

ableitbar sein müssen. Eine zentrale Konsequenz von (TD) ist, dass ihre Umsetzung eine durchgängige Revision klassischer Logik erfordert. Wie bereits gezeigt wurde, wurde in Kripkes Theorie das Revenge - Problem und die Asymmetrie durch eine Diskrepanz von innerer und äußerer Logik verursacht. D.i; während die innere Logik (auf welcher die Konstruktion des Fixpunktes basiert) nichtklassisch ist, wurde die Metatheorie - insbesondere das Axiomensystem KF - in klassischer Logik formuliert. Um daher eine revenge-immune Theorie, die (TD) erfüllt, aufstellen zu können, muss dieser Fehler ausgemärzt werden und die Metatheorie ebenfalls in nichtklassischer Logik formuliert werden.

Da Kripkes Theorie und ihr zugehöriges Axiomensystem, von einem formalen Standpunkt aus betrachtet, natürlich konsistent ist, gibt (TD) auch ein Abgrenzungskriterium zu nicht-deflationistischen Theorien an die Hand. Ein Inflationist kann nämlich, wenn er sich nicht auf (TD) festlegt, zB Kripkes Theorie durchaus akzeptieren.

Um zu sehen, warum dies so ist, bemerke zuerst, dass abgesehen von der These (TD) - welche hier die These (FD) und die Konservativitätsanforderung ersetzt - und der damit einhergehenden Forderung einer durchhängigen Revision klassischer Logik, der Deflationismus (und daher auch der Inflationismus) immer noch so definiert ist wie zuvor. Bezüglich (TD) und der Revision klassischer Logik scheiden sich jedoch erneut die Geister. Gegen den Deflationismus behauptet nämlich der Inflationismus, dass (TD) unmotiviert ist, da erstens die Forderung einer vollkommen abgeschlossenen Sprache eine Illusion ist; und zweitens, dass eine hierarchische Wahrheitstheorie durchaus natürlich ist (vgl. [28], [29]).<sup>23</sup> Aus ähnlichen Gründen sieht der Inflationist auch keinen Anlass dazu, von der klassischen Logik abzuweichen (vgl. [28], [29]).<sup>24</sup>

Diese Punkte sind weitgehend skizzenhaft und bedürften tiefergehender Diskussion.

<sup>23</sup>Bemerke, dass es eine wesentliche Motivation von (TD) ist, eine hierarchische Wahrheitstheorie zu vermeiden.

<sup>24</sup>An diesem Punkt bin ich Jeffrey Ketland Hannes Leitgeb zu Dank verpflichtet.

Leider ist es jedoch so, dass sich diese meines besten Wissens nach in der derzeitigen Forschungslandschaft nicht findet. Da es deshalb auch kein Anliegen der vorliegenden Arbeit ist, diese zu klären, hoffe ich, dass sich der Leser damit zufrieden gibt. Aufgrund des Mangels an Forschungsliteratur wird diesbezüglich im Folgenden auch kein Standpunkt eingenommen.

## 5.2 Field

Der erste Philosoph, welcher das Projekt, eine durchgängig nichtklassische Wahrheitstheorie zu formulieren, die (TD) erfüllt, ernsthaft verfolgt hat, ist Hartry Field. Allgemein formuliert handelt es sich hierbei um eine nichtklassische Erweiterung von Kripkes Theorie.<sup>25</sup> Field formuliert sein Projekt folgendermaßen:

The book ([24], LS)(...) offers a truth theory that includes all Tarski biconditionals (all sentences of the form  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$ ), and allows the general intersubstitutivity of  $T(\ulcorner \phi \urcorner)$  with  $\phi$  (except in opaque contexts). It does this by restricting excluded middle (...); but classical reasoning emerges unscathed except in contexts where ‘True’ or related predicates have ‘ungrounded’ occurrences. ([25])

Bemerke, dass die Formulierung ‘all sentences of the form  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$ , and allows the general intersubstitutivity of  $T(\ulcorner \phi \urcorner)$  with  $\phi$  (except in opaque contexts)’ äquivalent ist zu (TD). Field nennt dies die *naive Wahrheitstheorie* ([23], S. 139). Außerdem sollte klar sein, dass die Qualifikation ‘except in opaque contexts’ - ähnlich wie bei Kripkes Theorie - keine Einschränkung der Bikonditionale bedeutet, sondern vielmehr deren unbeschränkte Ableitung erst ermöglichen soll.

Im Folgenden sein die grundlegendsten Ideen und formalen Eigenschaften von Fields Theorie dargestellt.<sup>26</sup> Dabei werde ich mich auf diejenigen Aspekte seiner Arbeit konzentrieren, welche ausgehend von Kripkes Theorie verständlich gemacht werden können. Field beginnt seine Kritik an Kripke mit der Bemerkung, dass seine Theorie kein Konditional enthält.

(...) An unpleasant feature of the theory is that there is no biconditional  $\leftrightarrow$  for which  $\phi \leftrightarrow \phi$  is a logical law, or (equivalently, given the intersubstitutivity property) for which  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  holds in general. Indeed, there is no conditional  $\rightarrow$  for which  $\phi \rightarrow \phi$  or its converse hold generally. ([23], S. 139)

Dies ist ein sehr guter Punkt, welcher auf einfache Art und Weise veranschaulicht werden kann.<sup>27</sup> Zuerst sei bemerkt, dass das gewöhnliche materiale Konditional  $\phi \supset \psi$

<sup>25</sup>‘I will propose a natural extension of Kripke’s theory (...)’ [23].

<sup>26</sup>Da Fields Theorie verhältnismäßig kompliziert ist und ihr ein eigener formaler Apparat zu Grunde liegt, würde eine genaue Darstellung den Rahmen dieser Arbeit sprengen.

<sup>27</sup>Ich halte es für einen großen Vorteil dieser Argumentation, dass sie die Probleme ohne aufwendige Details derart abstrakt auf den Punkt bringt. Da in dieser Arbeit schon jedes diesbezüglich wichtige Theorem und Lemma bewiesen

definiert ist als  $\neg\phi \vee \psi$ . Dies ist eine Instanz des ausgeschlossenen Dritten, da  $\phi \supset \phi$  äquivalent ist zu  $\neg\phi \vee \phi$ . Da  $\lambda$  in Kripkes Theorie keinen Wahrheitswert hat, können daher die entsprechenden Sätze, die aus diesem Konditional geformt werden, nicht *im Allgemeinen* gültig sein. Mit dieser Diagnose haben wir gleich mehrere der vorher besprochenen Probleme auf den Punkt gebracht (vgl. [24], S.73, [23], S.139, [37], S. 144ff, [35], S. 235).

- Da das gewöhnliche Konditional  $\supset$  das Gesetz des ausgeschlossenen Dritten beinhaltet, veranschaulicht dies das Problem der Asymmetrie von innerer und äußerer Logik
- Da  $\supset$  in Kripkes Theorie nicht allgemein gilt, kann auch das T-Bikonditional  $T(\ulcorner\phi\urcorner) \leftrightarrow \phi$  nicht allgemein gelten, da  $\leftrightarrow$  aus  $\supset$  und  $\wedge$  definiert ist.
- Da diese Probleme dadurch verursacht werden, dass wir ‘unbestimmte’ Sätze wie  $\neg T(\ulcorner\lambda\urcorner)$  äußern, veranschaulicht dies die Ausdrucksschwäche von Kripkes Theorie; kurz: wir können nicht alles äußern was wir äußern wollen.

Ausgehend von dieser Kritik führt Field sein eigenes Konditional ein, welches die oben besprochenen Kritikpunkte überwinden soll. Er beginnt damit, dass er der Sprache ein neues Konditional  $\rightarrow$  hinzufügt, woraus die Sprache  $\mathcal{L}_{\mathcal{T}\rightarrow}$  resultiert. Für diese Sprache konstruiert er partielle Modelle - deren Konstruktion wir später kurz besprechen werden -, welche für alle  $\phi \in \mathcal{L}_{\mathcal{T}\rightarrow}$  das Field-Bikonditional

$$\phi \rightleftharpoons T(\ulcorner\phi\urcorner)$$

erfüllt (vgl. [37], S. 145). Obwohl dieses Konditional eine ganze Liste klassischer Gesetze erfüllt (siehe [24] S. 266 ff), erfüllt es zwei wichtige Gesetze nicht, bzw nur beschränkt (vgl. [24], S. 269f). Diese sind der konditionale Beweis der Form

$$\frac{\Gamma, \phi \models \psi}{\Gamma \models \phi \rightarrow \psi}$$

und der Modus Ponens:

$$\models (\phi \wedge (\phi \rightarrow \psi)) \rightarrow \psi$$

Diese sind nur eingeschränkt gültig, d.i. in nicht-defekten Kontexten, in denen das Gesetz des ausgeschlossenen Dritten angenommen werden kann. Der gültige eingeschränkte konditionale Beweis hat die Form:

$$\frac{\Gamma, \phi \models \psi}{\Gamma, \phi \vee \neg\phi \models \phi \rightarrow \psi}$$

---

wurde, sollte es nicht schwierig sein die Details dahinter zu erkennen.

Analog dazu, hat der gültige eingeschränkte Modus Ponens die Form:

$$\phi \vee \neg\phi, \psi \vee \neg\psi \models (\phi \wedge (\phi \rightarrow \psi)) \rightarrow \psi$$

Die betreffenden Regeln sind daher nur *schwach* gültig. Obwohl der Eindruck besteht, dass die hier notwendigen Einschränkungen ähnlich natürlich sind wie die Anwendung einer dreiwertigen Logik, in welcher defekte Sätze den Wahrheitswert ‘-‘ (undefiniert) erhalten,<sup>28</sup> bezeichnet es Field als eine Schwäche seiner Theorie.

The fact that quite a few standard ‘laws’ (...) turn out invalid in their unrestricted form is, I think, an unpleasant feature of the current theory. In the case of many of these ‘laws’, they *must* fail to be unrestrictedly valid if *anything like* the current theory is correct. ([24], S. 270)

Ähnlich, jedoch weitaus kritischer, bemerkt auch Leon Horsten:

Formalized modus ponens is not only satisfied by the familiar material implication but also by all of the implication relations that have been proposed in the recent literature as candidates for capturing central aspects of the semantics of indicative conditionals in English. Thus it is questionable whether the Field-biconditionals are biconditionals in *our* sense of the word. ([37], S. 145)

Ausgehend von diesem Zitat scheint eine zentrale Kritik von Horsten zu sein, dass das Bikonditional in der Forderung (TD) sowohl klassisch (und daher aus  $\supset$  und  $\wedge$  definiert ist), als auch natürlich ist und daher Fields Theorie, aufgrund seines nichtklassischen Konditionals  $\rightarrow$ , unnatürlich ist und das selbst auferlegte Ziel, (TD) zu erfüllen, verfehlt. Obwohl es jedoch offensichtlich stimmt, dass Fields Konditional andere formale Eigenschaften hat als Tarskis, werde ich kurz dafür argumentieren, dass diese Kritik nicht definitiv ist.

Zuerst sei zur Wiederholung bemerkt, dass Fields Konditional den Modus Ponens erfüllt, solange für die betreffenden Sätze das Gesetz des ausgeschlossenen Dritten gilt. Der klassische Modus Ponens wird hier insbesondere in Kontexten, in welchen durch das Hinzufügen eines Wahrheitsprädikats zu unserer Sprache Paradoxien entstehen, eingeschränkt.<sup>29</sup> Dementsprechend ist auch Horstens Vergleich mit den indikativen Konditionalen verfehlt: Solange wir unserer Sprache kein Wahrheitsprädikat hinzugefügt haben, erfüllt Fields Konditional den klassische Modus Ponens. Meines besten Wissens nach ist die Logik indikativer Konditionale nicht mit dem Verhalten des Wahrheitsprädikats beschäftigt (siehe [6]) und somit erfüllt Fields Konditional die entsprechende

<sup>28</sup> Gegeben, dass dies natürlich ist

<sup>29</sup> Es ist sehr wichtig zu sehen, dass es sich hier um spezifisch Wahrheitstheoretische Probleme handelt und, dass es nicht a priori klar ist, dass diese analog zu anderen Problemen behandelt werden sollten.

Anforderung in allen relevanten Zusammenhängen. Insofern ist daher mit Fields Konditional alles in Ordnung, da es eher eine Verallgemeinerung als eine Einschränkung des klassischen Konditionals darstellt (siehe dazu auch [24], S.30f).

Obwohl man daher gute Gründe hat anzunehmen, dass diese Kritik Fields Theorie nicht gerecht wird, besteht ganz allgemein ein ähnlicher Einwand. Der Vorwurf wäre aber nicht, dass Fields Konditional unnatürlich ist, sondern vielmehr, dass die erwähnte Verallgemeinerung des klassischen Konditionals unmotiviert ist. So würde ein Vertreter der klassischen Logik behaupten, dass es keinen Grund gibt, angesichts der Paradoxien (welche durch das Hinzufügen eines Wahrheitsprädikats entstehen) ein neues Konditional einzufügen um diese zu blockieren. Die klassische Logik, so der Einwand, gilt *allgemein* für alle Sätze unserer Sprache und somit sind die entstandenen Paradoxien schlicht und einfach hinzunehmen. In diesem Sinn wäre Fields Theorie auch keine Verallgemeinerung einer klassischen Theorie (formuliert in klassischer Logik), da diese selbst im Allgemeinen gilt.

Dennoch ist dieser Einwand ebenfalls nicht definitiv und bedarf daher weiterer Diskussion. Die Standard - Antwort ist natürlich in erster Linie, dass die Gültigkeit der klassischen Logik im Allgemeinen nicht ausgemacht ist. Insbesondere ist zu bemerken, dass es sich hier um eine Sprache mit spezifisch wahrheitstheoretischen Inhalt handelt und es a priori nicht klar ist, wie diese zu behandeln sind. So ist es zum Beispiel nicht ausgemacht, dass Sprachen mit Wahrheitsprädikat auf dieselbe Art und Weise an die klassische Logik gebunden sind wie z.B Peano Arithmetik oder Mengenlehre erster Stufe (wo es sehr plausibel ist, dass sie an die klassische Logik erster Stufe gebunden sind). Da es die bezüglich meines besten Wissens nach keinen Konsens in der modernen Forschungslandschaft gibt, sei dies hier jedoch nur dahingestellt.

Dies angemerkt, fahren wir nun mit der Darstellung von Fields Theorie fort.

Fields Konditional  $\rightarrow$  erlaubt es nun einen Begriff der Determiniertheit einzuführen, durch welchen es möglich wird zu sagen, dass der Lügner Satz nicht determiniert wahr ist.

**Definition 5.1** Sei  $D$  ein Determiniertheitsoperator.  $D(\phi)$  ist definiert als

$$\phi \wedge (\top \rightarrow \phi)$$

So definiert, gilt in Field's Modellen  $\neg D(\ulcorner \lambda \urcorner)$ . Daraus ergibt sich jedoch, dass man einen verstärkten Lügner  $\lambda^*$  konstruieren kann, welcher von sich selbst sagt, dass er nicht determiniert wahr ist. Natürlich kann man ausgehend davon erneut einen superverstärkten Lügner  $\lambda^{**}$  konstruieren und erhält somit (wie zuvor) einen neuen Sinn, in dem dieser nicht determiniert wahr ist. Field begegnet diesem Problem, indem er eine Hierarchie der Determiniertheit einführt. Dieser Prozess wird ad infinitum fortgesetzt; wobei es nicht möglich ist auszudrücken, dass alle Sätze der Form  $\lambda, \lambda^*, \lambda^{**}, \dots$  nicht

determiniert wahr sind. (vgl. [37], S. 146)

Sehr interessant sind auch die Modelle von Fields Konditional  $\rightarrow$ ; und zwar deshalb, weil sie auf einer iterierten Kripke-Fixpunktkonstruktion basieren, die für die jeweilige Stufe, in welcher wir uns in der Determiniertheits-Hierarchie befinden, ein Modell liefert. Steven Yablo beschreibt diesen Prozess folgendermaßen:

Field uses a three-valued scheme in which sentences are assigned either 1 (determinate truth, or something like it), 0 (determinate falsity), or  $1/2$  (indeterminate). Values are generated by a transfinite series  $P^\alpha$  of Kripkean fixed points, each the Kripke-closure (...) of a ‘seed’ valuation  $S^\alpha$  that assigns values only to conditionals, defined here as statements whose main connective is  $\rightarrow$ . Each  $P^\alpha$  gives clues to the proper interpretation of  $\rightarrow$ , clues that guide the construction of  $S^{\beta+1}$ , which forms the basis of  $P^{\beta+1}$ . This results in a back-and-forth process (...). ([69], S. 314)

Zu bemerken ist hierbei, dass der Prozess, in dem wir in der Hierarchie der Determiniertheit, für das neue Konditional  $\rightarrow$ , aufsteigen, anders als in Kripkes Theorie, nicht monoton ist (vgl. [24], S. 249).

Diese hierarchisch aufgebaute Theorie ist jedoch nicht unproblematisch. Insbesondere deshalb, weil es ein zentrales Anliegen der Tradition nach Tarski war, genau diesen hierarchischen Aufbau zu vermeiden. Leon Horsten schreibt daher:

(...) this is unsatisfactory. The whole point of moving to type-free theories of truth was that it seemed implausible that what seems to be one notion (truth) on closer inspection turns out to be an irrevocably fragmented concept. But then we should be equally skeptical of the view that a closely related notion (determinate truth) that appears to be every bit as much unified on closer inspection proves to be irrevocably fragmented. ([37], S. 146)

Diese Kritik ist ziemlich treffend und bedarf weiterer Diskussion. Da die genauen formalen Eigenschaften von Fields Theorie jedoch noch nicht vollständig ausgearbeitet sind (siehe [67]), wird die Zukunft entscheiden, wie zutreffend die Kritik wirklich ist.

### 5.3 PKF und inferentieller Deflationismus

Eine weitere Theorie, welche, um das Revenge - Problem zu vermeiden, durchgehend in nichtklassischer Logik formuliert ist, ist PKF (= Partial Kripke Feferman). Diese geht zurück auf Volker Halbach und Leon Horsten ([34]) und ist eine Variante von KF. Die Idee dahinter ist, Kripkes Theorie vollständig in partieller starker Kleene Prädikaten-Logik zu formulieren; d.h. auch die Axiome der Theorie werden hier in nichtklassischer Logik formuliert. Daher werden auch sehr elementare Axiomenschemata der klassischen

Logik, wie z.B.  $\phi \rightarrow \phi$  nicht mehr allgemein gültig sein. Dies ist in diesem Fall genau so, wie es sein soll. Denn ähnlich wie zuvor bei Field, bemerkt Leon Horsten, dass wir (rückblickend auf die Probleme die sich bei Kripkes Theorie ergeben haben) keine Theoreme der Form  $\lambda \rightarrow \lambda$  haben wollen. In der Theorie PKF muss insbesondere der Konditionale Beweis der Form

$$\frac{\Gamma, \phi \models \psi}{\Gamma \models \phi \rightarrow \psi}$$

auf Wahrheitsdeterminierte Sätze eingeschränkt werden. Die restriktierte Form des konditionalen Beweises nimmt daher folgende Form an:

$$\frac{\Gamma, T(\phi) \vee T(\neg\phi), \phi \models \psi}{\Gamma \models \phi \rightarrow \psi}$$

Leon Horsten beschreibt PKF als ein System, das entsteht, wenn man die Axiome von KF durch Regeln ersetzt. Die Regeln von PKF sind:

|       |                                                      |                                                      |
|-------|------------------------------------------------------|------------------------------------------------------|
| PKF 1 | $\frac{val^+(t_1 = t_2)}{T(t_1 = t_2)}$              | $\frac{T(t_1 = t_2)}{val^+(t_1 = t_2)}$              |
| PKF 2 | $\frac{T(\phi) \wedge T(\psi)}{T(\phi \wedge \psi)}$ | $\frac{T(\phi \wedge \psi)}{T(\phi) \wedge T(\psi)}$ |
| PKF 3 | $\frac{T(\phi) \vee T(\psi)}{T(\phi \vee \psi)}$     | $\frac{T(\phi \vee \psi)}{T(\phi) \vee T(\psi)}$     |
| PKF 4 | $\frac{\forall x T(\phi(x))}{T(\forall x \phi(x))}$  | $\frac{T(\forall x \phi(x))}{\forall x T(\phi(x))}$  |
| PKF 5 | $\frac{\exists x T(\phi(x))}{T(\exists x \phi(x))}$  | $\frac{T(\exists x \phi(x))}{\exists x T(\phi(x))}$  |
| PKF 6 | $\frac{T(\phi)}{T(T(\phi))}$                         | $\frac{T(T(\phi))}{T(\phi)}$                         |
| PKF 7 | $\frac{\neg T(\phi)}{T(\neg\phi)}$                   | $\frac{T(\neg\phi)}{\neg T(\phi)}$                   |

Zuerst sei bemerkt, dass in diesen Regeln  $t_1, t_2, \phi$ , und  $\psi$  als *Variablen*, die über Terme und Formeln laufen, fungieren. In den Regeln wird implizit über eben diese Terme und Formeln ‘all’-quantifiziert und somit gleicht PKF in dieser Hinsicht eher CT (in welchem dies explizit getan wird) als DT.

Anders als die Theorien, die wir bisher kennengelernt haben, macht PKF keine Aussagen über den Wahrheitswert des Lügnersatzes; und beweist somit weder  $\lambda$  noch  $\neg\lambda$ , noch  $T(\lambda)$  oder  $T(\neg\lambda)$ . Somit ist PKF ebenfalls nicht auf den verstärkten Lügner, bzw. das Revenge-Problem anfällig. (vgl. [37], S. 144)

Die folgenden zwei Lemmata zeigen, dass PKF formal korrekt ist und minimale Adäquatheitsbedingungen erfüllt.

**Lemma 5.2** (*Korrektheit*) Wenn  $PKF \vdash \phi$ , dann gilt  $(\mathbb{N}, S^+, S^-) \models_{SK} \phi$  für alle Fixpunktmodelle.

*Beweis:* siehe [35], S. 239  $\square$

**Lemma 5.3** (*Materiale Adäquatheit*) Für alle Formeln  $\phi(x_1, \dots, x_n)$  von  $\mathcal{L}$  gilt:

$$PKF \vdash \forall t_1 \dots \forall t_n (T\phi(t_1, \dots, t_n) \leftrightarrow \phi(t_1, \dots, t_n))$$

*Beweis:* siehe [35], S. 241  $\square$

Lemma 5.3 zeigt, dass PKF die restriktierten Tarski-Bikonditionale beweist (vgl. [35], S. 241, [37], S.135). Für Sätze  $\phi$  aus  $\mathcal{L}_{\mathcal{T}}$  gilt dies jedoch nicht und PKF muss daher folgendermaßen eingeschränkt werden:

**Lemma 5.4** Für alle Sätze  $\phi$  aus  $\mathcal{L}$  gilt:

$$PKF \vdash \phi \text{ gdw } PKF \vdash T(\ulcorner \phi \urcorner).$$

*Beweis:* siehe [35], S. 241  $\square$

Wobei Sätze der Form  $T(\ulcorner \phi \urcorner)$  nur durch Anwendung der Regeln von PKF gebildet werden können. Weiters ist PKF symmetrisch:

**Lemma 5.5** ([37], S. 138) *NEC und CONEC sind zulässige Regeln von PKF.*

Außerdem gilt:

**Lemma 5.6** *PKF hat ein Standardmodell.*

*Beweis:* Da PKF ein Kripke-Fixpunktmodell hat, hat es auch ein Standardmodell.  $\square$

Leon Horsten behauptet nun, dass PKF sowohl die kompositionale Natur des Wahrheitsprädikats vollständig ausdrückt (vgl. [37], S. 135), als auch philosophisch korrekt ist ([37], S. 137). Sein Grund für die erste Behauptung ist, dass die Einführungs- und Eliminationsregeln von PKF symmetrisch sind. Der Grund für letztere Behauptung ergibt sich aus den formalen Eigenschaften von PKF, die wir oben aufgezeigt haben. Im Folgenden werde ich argumentieren, dass Horstens Diagnose zumindest teilweise verfehlt ist. Dafür muss jedoch noch etwas Vorarbeit geleistet werden. Zuerst ist zu bemerken, dass wir in PKF die Disquotationssätze aus Lemma 5.3 nicht auf Sätze, die das Wahrheitsprädikat beinhalten, ausweiten können. D.i.:

**Lemma 5.7**  $PKF \not\models \forall t_1 \dots \forall t_n (T\phi(t_1, \dots, t_n) \leftrightarrow \phi(t_1, \dots, t_n))$  für alle  $\phi(x_1, \dots, x_n) \in \mathcal{L}_{\mathcal{T}}$ .

Ganz allgemein gilt außerdem, dass keine einzige Regel von PKF durch ein ihr korrespondierendes Axiom ersetzt werden kann (vgl.[37], S. 144). Daraus erhalten wir insbesondere:

**Lemma 5.8**  $PKF \not\models \forall x \forall y ((Sent_T(x \dot{\wedge} y) \rightarrow (T(x \dot{\wedge} y) \leftrightarrow Tx \wedge Ty))$

Daraus folgt, dass Kompositionalität (ähnlich wie in Lemma 5.3 die Disquotationssätze) auf Sätze von  $\mathcal{L}$  eingeschränkt werden muss und Wahrheitsiterationen nur durch die dazu korrespondierenden Regeln (hier PKF 2) gebildet werden können. Bemerke weiters, dass KF die in Frage stehende quantifizierte Version des Kompositionalitätsprinzips als Axiom (KF 4) enthält.

Dies vorausgesetzt, beginne ich mit dem Einwand. Horsten scheint seine oben gegebenen Behauptung bezüglich der philosophischen Korrektheit von PKF darauf zu stützen, dass das System Tarskis materiale Adäquatheitsbedingung erfüllt und symmetrisch ist (und daher auch Kompositionalität voll ausdrückt). Um dies angemessen zu veranschaulichen, hier ein Zitat:

Like all the truth theories that we consider, PKF meets minimal adequacy conditions. For instance, PKF proves the restricted Tarski-biconditionals. ([37], S. 135)

(...) PKF does appear to explicate the compositional nature of the concept of truth in a natural and elegant way. Its introduction and elimination rules display a pleasing symmetry. ([37], S. 135)

Obwohl es natürlich stimmt, dass PKF sowohl minimal adäquat in Tarski's Sinn, als auch symmetrisch ist, behaupte ich, dass dies nicht ausreichend ist um in diesem Zusammenhang PKF als philosophisch korrekte Wahrheitstheorie zu qualifizieren.

Der erste Teil des Arguments ist, dass Tarskis minimale Adäquatheitsbedingung in diesem Kontext zu schwach ist. So war doch ein wesentlicher Teil der Motivation, welcher uns veranlasst hat, von Tarskis getypter Wahrheitstheorie zu einer ungetypten überzugehen, gerade derjenige, dass man (auf Stufe 0) die Restriktion der Anwendung des Wahrheitsprädikats auf Sätze ohne Wahrheitsprädikat (d.i. auf Sätze  $\phi \in \mathcal{L}$ ) überwinden wollte.<sup>30</sup> Damit ändern sich natürlich auch die Standards für eine adäquate Wahrheitstheorie und wir sollten nun nicht mehr nur die Ableitung von  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  für alle  $\phi \in \mathcal{L}$  fordern, sondern vielmehr  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  für alle  $\phi \in \mathcal{L}_{\mathcal{T}}$ . Daher ist es verfehlt, eine ungetypte Wahrheitstheorie mit einer getypten in dieser Hinsicht zu

<sup>30</sup>So ist es klarer weise auch ein Charakteristikum getypter Wahrheitstheorien, dass sie nichts über die Wahrheit von Sätzen aussagen, die das Wahrheitsprädikat beinhalten. ‘Sloppy speaking, these (typed, LS) theories do not say anything about the truth of sentences containing the truth predicate.’ ([35], S. 144)

vergleichen und die Standards letzterer auf erstere anzulegen. Daraus ergibt sich, dass PKF - obwohl adäquat in Tarskis Sinn - nicht *Kripke-adäquat* ist.

Der zweite Teil meines Arguments betrifft die Kompositionanlitätsanforderung und verläuft ähnlich wie das vorhergehende. Zuerst sei bemerkt, dass wir den Einwand mangelnder Generalisierung und Kompositionalität schon einmal im Kontext von Tarski's Theorie TB angetroffen haben. Diese Probleme wurden von der Theorie CT überwunden, in welcher Kompositionalität schon in den Axiomen vollständig ausgedrückt wird. Wie zuvor sollten wir jedoch die philosophische Korrektheit von PKF nicht dadurch rechtfertigen, dass sie wie CT Kompositionalität für alle Sätze  $\phi \in \mathcal{L}$  beweist. Vielmehr sollte aus den oben gegebenen Gründen PKF mit ihrem axiomatischen Pendant KF verglichen werden. Wie wir gesehen haben, beweist PKF die Verallgemeinerung der Kompositionalität, welche in KF jedoch als Axiom enthalten ist, nicht und wir sind daher berechtigt zu behaupten, dass PKF (ähnlich wie TB) Kompositionalität nicht vollständig ausdrückt.

Daraus ergibt sich, dass PKF weder *Kripke-adäquat*, noch *Kripke-kompositional* ist und wir können mit Recht beanstanden, dass Horstens Behauptung der philosophischen Korrektheit von PKF auf schwachen philosophischen Annahmen beruht.

Obwohl PKF mit diesen Argumenten eigentlich schon disqualifiziert ist, sein noch kurz die philosophischen Thesen skizziert, die Leon Horsten mit der Theorie PKF verbindet. Wie wir schon gesehen haben, besteht PKF aus den zu den Axiomen von KF korrespondierenden Regeln. Gegen die gerade vorgebrachten Einwände behauptet Leon Horsten jedoch, dass dies eine vorteilhafte Eigenschaft von PKF ist und Anlass zu einer neuen Theorie des Deflationismus gibt. Diese nennt er Inferentiellen Deflationismus. Nach Horsten passt PKF nämlich sehr gut zur Leitidee des Deflationismus (siehe These 1, S. 15), da es in ihr keine allgemeinen Wahrheitsprinzipien gibt, welche die Natur oder Essenz von Wahrheit ausdrücken könnten.

We have seen that, according to PKF, there are no unrestricted general principles of truth. This can be explained by the fact that there is no nature or essence of truth to be described by general principles. (Horsten, S. 143)

Da PKF außerdem aus Schlussregeln und nicht aus Axiomen besteht, zieht Horsten eine Analogie zum Inferentialismus. Dieser besagt in seiner schwachen Form, dass die logischen Konnektive durch unser inferentielles Verhalten bestimmt sind und in seiner starken Form, dass unsere Begriffe durch unser inferentielles Verhalten bestimmt sind (siehe z.B. [15], [8]). Der inferentielle Deflationismus besagt nun, dass die Bedeutung des Wahrheitsprädikats durch Schlussregeln bestimmt wird.

According to inferential deflationist, the meaning of the concept of truth is explained in terms of rules of inference. ([37], S.147)

We should not aim at describing the nature of truth because there is no such thing. Rather we should aim at describing the inferential behavior of truth. ([37], S.146)

Obwohl nun die Idee, dass die Bedeutung des Wahrheitsprädikats durch Schlussregeln bestimmt ist, plausibel erscheint, werde ich im Folgenden kurz dafür argumentieren, dass diese Interpretation in Zusammenhang mit PKF überstrapaziert ist.

Erstens haben wir es eigentlich als Schwäche interpretiert, dass PKF nur aus Schlussregeln besteht. Zweitens ist es keinesfalls ein spezielles Charakteristikum der Theorie PKF, dass sie aus Regeln besteht. Da die Regeln schwächer sind als die dazu korrespondierenden Axiome und deshalb konsistent sind, wenn die betreffenden Axiome konsistent sind, kann jede konsistente axiomatische Wahrheitstheorie in ein System von Regeln verwandelt werden.<sup>31</sup> Dementsprechend ist es mehr oder weniger willkürlich, PKF mit dem Namen ‘inferentieller Deflationismus’ zu versehen. Dies würde nur dann Sinn ergeben, wenn wir PKF im Vorhinein als gute Wahrheitstheorie auszeichnen könnten, was nicht der Fall ist. Weiters wird es aufgrund des bisher gesagten aus äußerst fraglich, ob es überhaupt Sinn macht, Axiomensysteme in Systeme von Schlussregeln umzuwandeln; und somit wird es ebenfalls fraglich, ob der ‘inferentielle Deflationismus’ eine attraktive philosophische Position sein kann.<sup>32</sup>

## 5.4 Zusammenfassung

Im Zuge dieser Arbeit wurden Tarskis und Kripkes formale Wahrheitstheorien gemeinsam mit den philosophischen Thesen und Weiterentwicklungen, die mit ihnen in Verbindung stehen, präsentiert. Zum Abschluss fasse ich die wichtigsten Punkte noch einmal zusammen.

Tarskis Theorem hat gezeigt, dass es nicht möglich ist, ein unrestrictiertes Wahrheitsprädikat für die Sprache der Arithmetik zu definieren. Seine Lösung war eine Typisierung des Wahrheitsprädikats, die es erlaubte, eine konsistente Wahrheitstheorie für die Basissprache PA zu formulieren. Die schwächste seiner Theorien, welche einzig aus den Tarski-Bikonditionalen bestand (TB), ist konservativ und kann keine infiniten Generalisierungen logischer Gesetze ausdrücken. Deshalb musste sie zu ausdruckstärkeren Theorien erweitert werden.

Dies führte uns zu Theorien, welche die geforderten Generalisierungen schon in ihren Axiomen enthalten. Dabei wurde zwischen der Theorie  $CT\upharpoonright$  (in welcher das Wahrheitsprädikat nicht im Induktionsaxiom vorkommen darf) und der Theorie CT (die

---

<sup>31</sup>Vielleicht wäre es sogar interessant dies wirklich zu tun und die formalen Eigenschaften dieser Theorien zu untersuchen. Höchstwahrscheinlich würde sich dabei jedoch zeigen, dass jedes derart formulierte System an ähnlichen Problemen wie PKF leidet, da die Regeln a priori zu schwach sind um z.B. Forderungen wie Kompositionalität voll auszudrücken.

<sup>32</sup>Ich bedanke mich an dieser Stelle bei Leon Horsten für eine Diskussion dieser Punkte.

das Wahrheitsprädikat im Induktionsaxiom enthält) unterschieden. Ausgehend davon wurde unter Verweis auf Volker Halbachs Theorem gezeigt, dass  $CT\uparrow$  konservativ ist über PA. Danach wurde bewiesen, dass CT nicht konservativ ist und sowohl das Globale Reflexionsprinzip als auch die Konsistenz von PA beweist. Dies hat veranschaulicht, dass die Ursache der Nichtkonservativität von CT durch das Vorkommen des Wahrheitsrädikats im Induktionsaxiom verursacht wird (und nicht durch die in den Axiomen von  $CT\uparrow$  und CT vorkommenden infiniten Generalisierungen).

In Kapitel 3 wurden die mit Tarskis Theorien in Verbindung stehenden Thesen des frühen Deflationismus ausgearbeitet. Wie wir gesehen haben, war die These des Deflationismus in diesem Kontext, dass die Theorie TB alles ausdrückt, was über Wahrheit zu sagen ist. Die zentrale Kritik an dieser Position war das Konservativitätsargument von Ketland und Shapiro. Dieses besagt in seiner Kurzfassung, dass eine konservative Wahrheitstheorie nicht adäquat ist, weil sie das Globale Reflexionsprinzip und daher auch die Konsistenz von PA nicht beweist; und, dass eine nichtkonservative Wahrheitstheorie gewichtige Konsequenzen (wie eben das globale Reflexionsprinzip und die Konsistenz von PA) mit sich bringt und deshalb für den Deflationisten nicht akzeptabel sein kann. Da es jedoch allgemein akzeptiert ist, dass eine konservative Wahrheitstheorie zu schwach ist, wurde der Deflationismus von dieser Anforderung wieder befreit. Das Argument dafür war, dass die Konservativitätsanforderung zu stark ist, weil mit ihr keine gute Wahrheitstheorie deflationistisch sein könnte. Im Kontext von Tarskis Theorien blieb deshalb auch offen, auf welche formalen Anforderungen der Deflationismus festgelegt werden sollte.

Danach wurde Kripkes Kritik an Tarski präsentiert. Diese war in ihren Grundzügen, dass ein getyptes Wahrheitsprädikat unnatürlich ist. Kripkes Projekt war dementsprechend, eine Wahrheitstheorie mit einem ungetypten Wahrheitsprädikat zu formulieren. Widersprüche durch Paradoxien werden dabei durch die Anwendung einer dreiwertigen Logik mit Wahrheitswertlücken vermieden. Ausgehend von dieser allgemeinen Darstellung wurde besonderer Wert auf eine genaue Ausarbeitung seiner Fixpunktsemantik gelegt, deren formale Eigenschaften eine Konsequenz der Anwendung der dreiwertigen Logik sind. Zuerst wurden allgemeine Tatsachen über positive induktive Definitionen und monotone Operatoren aus der Beiweistheorie präsentiert. Dabei wurde gezeigt, dass positive induktive Definitionen Anlass zu einem monotonen Operator geben. Ausgehend davon wurde bewiesen, dass monotone Operatoren einen Fixpunkt haben. Diese Beweise wurden dann auf Kripkes Theorie mit ihrem dreiwertigen Starken-Kleene Evaluierungsschema übertragen und es wurde bewiesen, dass es in Kripkes Semantik einen Fixpunkt gibt. Informell besagt dieses Lemma, dass der Prozess, in dem das Wahrheitsprädikat immer neue Extensionen erhält, ein Ende finden muss. Daraus resultiert eine Semantik, welche für jeden Satz  $\phi$  der Sprache mit Wahrheitsprädikat,  $T(\ulcorner \phi \urcorner)$  erfüllt genau dann, wenn sie  $\phi$  erfüllt.

Darauf aufbauend wurde die Axiomatierung KF von Kripkes Theorie präsentiert und deren wichtigsten formalen Eigenschaften dargestellt. Zuletzt wurde das zentrale Problem von Kripkes Theorie besprochen. Dieses ist, dass Kripkes Theorie nicht Revenue-immun ist. D.i.: obwohl durch die Anwendung einer dreiwertigen Logik Paradoxien ausgeschaltet werden, entstehen auf einer Metaebene erneut Widersprüche. Diese werden durch eine Asymmetrie von innerer und äußerer Logik verursacht. Denn obwohl die innere Logik von Kripkes Theorie nichtklassisch ist, ist die Metatheorie (bzw. das Axiomensystem KF) durchgängig in klassischer Logik formuliert.

In Kapitel 5 wurde ausgehend von Kripkes Arbeit veranschaulicht, auf welche These der zeitgenössische Deflationismus festgelegt ist. Diesbezüglich wurde auf Hartry Field verweisen und behauptet, dass eine deflationistische Theorie das Revenge Problem überwinden soll und alle Bikonditionale der Form  $T(\ulcorner \phi \urcorner) \leftrightarrow \phi$  im Allgemeinen ableiten können soll. Um zu zeigen, wie diese Anforderung formal umgesetzt werden kann, wurden Fields Theorie und die Theorie PKF von Halbach und Horsten präsentiert und kritisch begutachtet. Fields Theorie wurde dabei zwar gegen Horstens Einwände verteidigt, auf eine definitive Bewertung wurde jedoch verzichtet, da der Forschungsstand diesbezüglich noch weitgehend offen ist. Zuletzt wurde die Theorie PKF und die damit in Verbindung stehende philosophische Position des inferentiellen Deflationismus präsentiert; wobei dafür argumentiert wurde, dass sowohl PKF als auch der inferentielle Deflationismus von einem formalen Standpunkt aus betrachtet inadäquat sind.

## 6 Anhang

### 6.1 Literaturverzeichnis

- [1] Aczel, Peter: An Introduction to Inductive Definitions. *Handbook of Mathematical Logic*. Ed.: J. Barwise. Amsterdam, North-Holland Pub. 1977
- [2] Beall, J.C (Hrsg.): *Liars and Heaps: New Essays on Paradox*. Oxford University Press, Oxford, New York 2003
- [3] Beall, J.C. (Hrsg.): *Revenge of the Liar*. Oxford University Press, Oxford, New York 2007
- [4] Beall, J.C und Glanzberg, Michael: Where the Paths Meet: Remarks on Truth and Paradox. *Midwest Studies in Philosophy*. Vol. 32, 2008
- [5] Beklemishev, Lev, D.: Reflection principles and provability algebras in formal arithmetic. *Uspekhi Mat. Nauk*, Vol. 60, Nr. 2, 2005
- [6] Bennett, Jonathan: *A Philosophical Guide to Conditionals*. Oxford University Press, Oxford, New York 2003
- [7] Boolos, George, Burgess, John and Jeffrey, Richard: *Computability and Logic*. Fifth Ed; Cambridge University Press, Cambridge, New York 2001
- [8] Brandom, Robert: *Making It Explicit*. Harvard Universtiy Press, Cambridge Mass; London 1998
- [9] Burgess, John: The Truth is Never Simple. *The Journal of Symbolic Logic*, Vol. 51, No. 3, 1986
- [10] Burgess, John: Kripke on Truth. *forthcoming*

- [11] Burgess, John and Burgess, Alexis: *Truth*. Princeton University Press, Princeton, New Jersey 2011
- [12] Cieslinski, Cezary: Truth, Conservativeness, and Provability. *Mind*, Vol. 119, No. 474, 2010
- [13] Davidson, Donald: *Inquiries into Truth and Interpretation*. Second. Ed; Clarendon Press, Oxford 2001
- [14] Davidson, Donald: *Truth and Predication*. The Belknap Press of Harvard University, Cambridge Mass; London 2005
- [15] Dummett, Michael: *The Logical Basis of Metaphysics*. Harvard University Press, Cambridge Mass. 1993
- [16] Ebbinghaus, Heinz-Dieter: *Einführung in die Mengenlehre*. Spectrum Akademischer Verlag, Heidelberg, Berlin 2003
- [17] Ebbinghaus, H.D; Flum, J; Thomas, W.: *Einführung in die mathematische Logik*. 5. Aufl; Spectrum Akademischer Verlag, Heidelberg, Berlin 2007
- [18] Enderton, Herbert, B.: *A Mathematical Introduction to Logic*. Second Ed. Harcord Academic Press, San Diego, New York 2001
- [19] Feferman, Solomon: Reflecting on Incompleteness. *The Journal of Symbolic Logic*, Vol. 56, No.1, 1991
- [20] Field, Hartry: Tarski's Theory of Truth. *The Journal of Philosophy*, Vol. 69, No. 13, 1972
- [21] Field, Hartry: Deflationist Views of Meaning and Content. *Mind*, Vol. 103, No. 411, 1994
- [22] Field, Hartry: Deflating the Conservativeness Argument. *The Journal of Philosophy*, Vol. 96, No. 11 1999
- [23] Field, Hartry: A Revenge Immune Solution to the Semantic Paradoxes. *Journal of Philosophical Logic*, Vol. 32, 2003
- [24] Field, Hartry: *Saving Truth from Paradox*. Oxford University Press, Oxford, New York 2008
- [25] Field, Hartry: Precise of *saving truth from paradox*. *Philosophical Studies*, Vol. 147, 2010
- [26] Fischer, Martin: Minimal Truth and Interpretability. *The Review of Symbolic Logic*, Vol. 2, No. 4, 2009
- [27] Fitting, Melvin: Notes on the Mathematical Aspects of Kripke's Theory of Truth. *Notre Dame Journal of Formal Logic*. Vol. 27, No. 1, 1986
- [28] Glanzberg, Michael: A Contextual-Hierarchical Approach to Truth and the Liar Paradox. *Journal of Philosophical Logic*. Vol. 33, 2004
- [29] Glanzberg, Michael: Truth, Reflection, and Hierarchies. *Synthese* Vol. 142, 2004
- [30] Glanzberg, Michael: Truth. *The Stanford Encyclopedia of Philosophy* (Spring 2009 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2009/entries/truth/>.
- [31] Halbach, Volker: Tarskian and Kripkean Truth. *Journal of Philosophical Logic*, Vol. 26, No. 1, 1997
- [32] Halbach, Volker: Truth and Reduction. *Erkenntnis*, Vol. 53, 2000
- [33] Halbach, Volker: How Innocent is Deflationism? *Synthese*, Vol. 126, 2001
- [34] Halbach, Volker, Leon Horsten: Axiomatizing Kripke's Theory of Truth. *Journal of Symbolic Logic*, Vol. 71, No. 1, 2006
- [35] Halbach, Volker: *Axiomatic Theories of Truth*. Cambridge University Press, Cambridge 2011
- [36] Horsten, Leon: Levity. *Mind*, Vol. 118, No. 471, 2009
- [37] Horsten, Leon: *The Tarskian Turn. Deflationism and Axiomatic Truth*. MIT Press, Cambridge Mas; London 2011
- [38] Horwich, Paul: A Defence of Minimalism. *Synthese*, Vol. 126, 2001
- [39] Horwich, Paul: *Truth-Meaning-Reality*. Oxford University Press, Oxford, New York 2010
- [40] Ketland, Jeffrey: Deflationism and Tarski's Paradise. *Mind*, Vol. 108, No. 429, 1999

- [41] Ketland, Jeffrey: Deflationism and the Gödel Phenomena: Reply to Tennant. *Mind*, Vol. 114, No. 453, 2005
- [42] Ketland, Jeffrey: Truth, Conservativeness, and Provability: Reply to Cieslinski. *Mind*, Vol. 119, No. 474, 2010
- [43] Kripke, Saul: Outline of a Theory of Truth. *The Journal of Philosophy*. Vol. 72, No. 19, 1975
- [44] Leitgeb, Hannes: Theories of Truth which have no Standard Models. *Studia Logica*, Vol. 68, 2001
- [45] Leitgeb, Hannes: What Truth Depends On. *Journal of Philosophical Logic*, Vol. 34, 2005
- [46] Leitgeb, Hannes: On the Metatheory of Field's 'Solving the Paradoxes, Escaping Revenge'. In: Beall, J.C. (Hrsg.): *Revenge of the Liar*. Oxford University Press, Oxford, New York 2007
- [47] Leitgeb, Hannes: What Theories of Truth Should be Like (but Cannot be). *Philosophy Compass*, Vol.2, No. 2, 2007
- [48] Leitgeb, Hannes und Welch, Philip: A Theory of Propositional Functions and Truth. *forthcoming*
- [49] Moschovakis, Yiannis: *Elementary Induction on Abstract Structures*, North Holland Pub. and American Elsevier, Amsterdam, New York 1974
- [50] Pohlers, Wolfram: *Proof Theory. The First Step into Impredicativity*. Springer, Berlin 2009
- [51] Priest, Graham: *An Introduction to Non-Classical Logic*. Second Ed; Cambridge University Press, Cambridge, New York 2009
- [52] Quine, W.V.O.: *Word and Object*. The MIT Press, Cambridge Mass. 1960
- [53] Quine, W.V.O.: *Philosophy of Logic*. Second Ed; Harvard University Press, Cambridge Mass, London 1986
- [54] Ramsey, F.P.: *The Foundations of Mathematics*. Routledge and Keagan Paul, London 1950
- [55] Rathjen, Michael: The art of ordinal analysis. *Proceedings of the International Congress of Mathematicians*, Madrid 2006
- [56] Shapiro, Stewart: Proof and Truth: Through Thick and Thin. *The Journal of Philosophy*. Vol. 95, No. 10, 1998
- [57] Shapiro, Stewart: The Guru, the Logician, and The Deflationist: Truth and Logical Consequence. *Noûs*, Vol. 37, No. 1, 2003
- [58] Shapiro, Stewart: So truth is safe from paradox. now what? *Philosophical Studies*, Vol. 147, 2010
- [59] Smith, Peter: *An Introduction to Gödel's Theorems*. Cambridge University Press, Cambridge 2007
- [60] Soames, Scott: *Understanding Truth*. Oxford University Press, Oxford, New York 1999
- [61] Stoljar, Daniel und Damjanovic, Nic: The Deflationary Theory of Truth. *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), Edward N. Zalta (ed.), forthcoming URL = <http://plato.stanford.edu/archives/sum2012/entries/truth-deflationary/>.
- [62] Tarski, Alfred: Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, Vol.1, 1935
- [63] Tarski, Alfred: Die semantische Konzeption der Wahrheit und die Grundlagen der Semantik. In.: Gunnar Skirbekk, (Hrsg.): *Wahrheitstheorien*. Suhrkamp, Frankfurt am Main 1977
- [64] Tarski, Alfred: A Lattice-Theoretical Fixpoint Theorem and its Applications. *Pacific Journal of Mathematics*, Vol. 5, 1955
- [65] Tennant, Neil: Deflationism and the Gödel Phenomena. *Mind*, Vol. 111, No. 443, 2002
- [66] Tennant, Neil: Deflationism and the Gödel Phenomena: Reply to Ketland. *Mind*, Vol. 114, No. 453, 2005
- [67] Welch, Philip: Truth, Logical Validity, and Determinateness: a commentary on Field's *Saving Truth from Paradox*. *Review of Symbolic Logic*, Vol. 4, No.3, 2011
- [68] Yablo, Steven: Grounding, Dependence, and Paradox. *Journal of Philosophical Logic*, Vol. 11, 1982
- [69] Yablo, Steven: New Grounds for Naive Truth Theory. In: Beall, J.C (Hrsg.): *Liars and Heaps: New Essays on Paradox*. Oxford University Press, Oxford, New York 2003

## 6.2 Abstract

Die Arbeit stellt eine kritische Auseinandersetzung mit den formalen Wahrheitstheorien von Tarski und Kripke dar. Dabei wird sowohl der formal-mathematische Apparat, als auch die jeweils daran anschließende philosophische Diskussion präsentiert. Darüber hinaus wird ein besonderer Schwerpunkt auf die formale Ausarbeitung von Kripkes Fixpunktsemantik gelegt. Dies bildet eine solide Grundlage für ein Verständnis der Forschungsergebnisse nach Kripke, welche ebenfalls präsentiert werden.

Im ersten Teil der Arbeit werden Tarskis Theorien in ihrer modernen Formulierung diskutiert. Es wird gezeigt, dass die schwächste seiner Theorien (TB), welche einzig aus den Tarski-Bikonditionalen besteht, keine infiniten Generalisierungen beweisen kann und konservativ über Peano-Arithmetik ist; was als Schwäche dieser Theorie interpretiert wird. Danach wird zu der stärkeren Theorie ‘Compositional Truth’ übergegangen, welche die geforderten Generalisierungen schon in den Axiomen enthält. Anhand der Unterscheidung zwischen der Theorie  $CT\upharpoonright$  (ohne dem Wahrheitsprädikat im Induktionsaxiom) und der Theorie CT (mit dem Wahrheitsprädikat im Induktionsaxiom) wird gezeigt, dass  $CT\upharpoonright$  konservativ ist, CT jedoch nicht. Dies veranschaulicht, dass der Grund für die Nichtkonservativität von CT im unrestriktierten Induktionsaxiom liegt. Danach wird, ausgehend von Tarskis Theorien, die frühe Deflationismusdebatte präsentiert und herausgearbeitet, dass der frühe Deflationismus auf Tarskis Theorie (TB) festgelegt ist. Der zentrale Einwand dagegen wird sein, dass eine konservative Wahrheitstheorie wie TB formal nicht adäquat ist, eine nichtkonservative Wahrheitstheorie - da sie gewichtige Konsequenzen mit sich bringt - für einen Deflationisten jedoch nicht akzeptabel ist. Anschließend daran wird dafür argumentiert, dass der Deflationismus von dieser Anforderung wieder befreit werden sollte, da diese zu stark ist.

Um Kripkes Theorie zu motivieren wird seine Kritik an Tarski präsentiert. Diese ist in ihren Grundzügen, dass ein getyptes Wahrheitsprädikat nicht natürlich ist. Ausgehend davon wird Kripkes formale Theorie präsentiert, wobei ein besonderer Schwerpunkt auf positive induktive Definitionen, monotone Operatoren und den minimalen Fixpunkt gelegt wird. So wird die Feinstruktur von Kripkes Semantik detailliert ausgearbeitet. Daran anschließend wird das Axiomensystem KF (Kripke Feferman) und dessen formale Eigenschaften präsentiert. Zuletzt wird das zentrale Problem von Kripkes Theorie herausgearbeitet. Dieses ist, dass seine Theorie nicht revenge-immun, bzw. nicht symmetrisch ist; was sich aus der Asymmetrie von innerer und äußerer Logik ergibt (d.i.: obwohl die innere Logik seiner Theorie nichtklassisch ist, ist die Metatheorie, bzw. das Axiomensystem KF, durchgehend in klassischer Logik formuliert).

Im letzten Teil der Arbeit wird, ausgehend von Kripkes Theorie, eine zeitgenössische Form des Deflationismus in formaler und philosophischer Hinsicht ausgearbeitet. Es wird gezeigt, dass der Deflationismus auf die Ableitung aller T-Bikonditionale im All-

gemeinen festlegt ist; was eine revenge-immune Theorie, durchgehend in nichtklassischer Logik formuliert, verlangt. Als Beispiele der Umsetzung dieser Forderung werden Fields Theorie und die Theorie PKF (Partial Kripke Feferman) von Halbach und Horsten herangezogen und kritisch begutachtet.

### **6.3 Curriculum Vitae**

#### **Persönliche Daten**

- Name: Leo Stadlmüller
- Geburtsdatum: 5. 4. 1986
- Nationalität: Österreich

#### **Ausbildung**

- 2010 - 2012: Masterstudium Philosophie an der Fakultät für Philosophie und Bildungswissenschaft der Universität Wien
- 2011: Erasmus Aufenthalt an der Ludwig-Maximilians-Universität München
- 2007 - 2010: Bachelorstudium Philosophie an der Fakultät für Philosophie und Bildungswissenschaft der Universität Wien
- 2000 - 2006: Handelsakademie Wiener Neustadt