Zirkularitätsverbot und zirkuläre Definitionen

Vortrag zur gleichnamigen Magisterarbeit

von

Özgür L. Özcep

Überblick

1 - Definitionen

Definitionskategorien

1. Wesensbestimmung/Sacherklärung (z.B. Aristoteles)
2. Begriffsbestimmung (Begriffskonstruktion- , Begriffsanalyse)
3. Definition als Feststellung der Bedeutung, die ein Zeichen besitzt, bzw. der Verwendung, die es findet
4. Definition als Festsetzung (stipulativ) der Bedeutung bzw. Verwendung eines (neu einzuführenden) Zeichens/Ausdrucks

Darüber hinaus gibt es Definitionssorten, bei denen es mehr bedarf als einer sprachlichen Äußerung bzw. schriftlichen Fixierung: z.B. ostensive Definitionen

Im folgenden Schwerpunkt auf 3. und 4.

Bei den verschiedenen Definitionskategorien 1-4 ist involviert ein Satz der Form:

Sprachlicher Ausdruck Definitionsoperator Sprachlicher Ausdruck

Definiendum: das zu Definierende (Ausdruck/Begriff „links“ vom Definitionsoperator)

Definiens: das, mit dem definiert wird (Ausdruck/Begriff „rechts“ vom Definitionsoperator)

Zirkularitätsverbot

Zirkularitätsverbot (grobe Fassung):

Definitionen dürfen nicht zirkulär sein i.S.v.

das Definiendum darf nicht im Definiens vorkommen

Anmerkung:

Je nach dem, wie die Ausdrücke ‚Definiendum' und ‚Definiens' ausgelegt werden, bekommt das Zirkularitätsverbot eine andere Bedeutung

Intuitive Begründungsversuche:

- *Mereologisches Argument:* Das Ganze (Definiendum) kann sich nicht als Teil im Definiens enthalten
- *Epistemisches Argument:* Man kann die Anwendungsbedingungen des Definiendum nicht lernen, indem man die Anwendungsbedingungen für das Definiens lernt, wenn in diesem das Definiendum vorkommt, indem man das Gleichsetzungsverfahren anwendet.
- *Vererbungsargument:* Das Definiendum kann nicht die Eigenschaften des Definiens erben, wenn in ihm das Definiendum vorkommt.

Historisches: Aristoteles und Port Royal

- Zirkularitätsverbot geht vermutlich zurück auf Aristoteles. Verstreute Bemerkungen in den „Analytica Posteriora“ und der „Topik“.
- Erstmalig explizit aufgeführt in einem Regelkanon der Schule von Port Royal:
  - Eine Definition muss das Wesen dessen wiedergeben, das definiert wird (Kategorie 1.)
  - Eine Definition darf nicht zirkulär sein
  - Eine Definition darf nicht negativ formuliert sein, wenn es positiv formuliert werden kann
  - Eine Definition darf nicht in figurativer oder obskurer Sprache ausgedrückt werden

Anmerkung:

Obiges Regelwerk ist nicht ausreichend. (Führt zu Widersprüchen)

Zirkularitätsverbot im Rahmen der klassischen Logik

Das Verbot zirkulärer Definitionen im Rahmen der klassischen Logik wird auf zwei Prinzipien zurückgeführt, die neu definierte Zeichen erfüllen sollen:

Nicht-Kreativitäskriterium (grobe Fassung)**:**

Die Einführung neuer Symbole soll nicht die Ableitung solcher Sätze gestatten, die vorher nicht ableitbar waren und die das neue Zeichen nicht enthalten. (Konservativität)

Eliminierbarkeitskriterium (grobe Fassung): Jedes Vorkommnis eines neu definierten Ausdrucks soll in allen Kontexten (eines bestimmten Typs) durch bereits zur Verfügung stehende Ausdrücke ersetzt werden können, so dass der Ausdruck mit dem neuen Zeichen und der Ausdruck ohne das neue Zeichen „gleichwertig“ sind.

Ausbuchstabierung der Kriterien (1)

Voraussetzung an die Sprache

- Zu Grunde liegende Sprache zu verstehen als angewandte prädikatenlogische Sprache erster Stufe: Sowohl Grammatik als auch beweistechnische Begriffe wie in der klassischen Prädikatenlogik erster Stufe
- Betrachtung lediglich von Definitionen von Prädikat-, Funktions- und Namensymbolen
- Betrachtung von Definitionen, die satzförmig sind und der Sprache angehören

Gegeben

- *Theorie*: Menge von Sätzen der zu Grunde gelegten Sprache erster Stufe
- *Vorliegende Definitionen*: Menge von Sätzen
- *Definition*: Satz aus der zu Grunde gelegten Sprache
- *Zeichen*: das neue Zeichen/Symbol, das definiert werden soll.

Ausbuchstabierung der Kriterien (2)

Eliminierbarkeitskriterium (EK)

Seien *Theorie, vorliegende Definitionen, Definition* und *Zeichen* wie oben erklärt.

Definition als Definition von *Zeichen* erfüllt das Eliminierbarkeitskriterium bzgl. *Theorie* und *vorliegende Definitionen* :gdw

Für alle (möglicherweise offenen) Sätze B in der Sprache von *Theorie, vorliegende Definitionen* und *Zeichen*, gibt es einen Satz B‘, so dass folgendes erfüllt ist:

- B‘ ist in der Sprache von *Theorie* und *vorliegende Definitionen* formuliert und
- *Zeichen* erscheint nicht in B‘ und
- B und B‘ sind äquivalent bzgl. *Theorie, vorliegende Definitionen* und *Definition*

Nicht-Kreativitätskriterium (NK)

Seien *Theorie, vorliegende Definitionen, Definition* und *Zeichen* wie oben erklärt.

Definition als Definition von *Zeichen* erfüllt das Nicht-Kreativitäskriterium bzgl *Theorie* und *vorliegende Definitionen* :gdw Für alle Sätze in der Sprache von *Theorie* und *vorliegende Definition* gilt: Wenn aus *Definition, Theorie* und vorliegende *Definitionen* zusammen B folgt, so folgt B schon aus *Theorie* und *vorliegende Definitionen*

Regeln für Definitionen

Regel für Definition von Prädikatbuchstaben

Definition ist eine Standarddefinition eines Prädikatbuchstabens R relativ zu *Theorie* und *vorliegende Definitionen* :gdw

es gibt natürliche Zahl n (Stelligkeit von R), Variablen $v_1,...,v_n$ und einen Satz A (das Definiens), so dass folgende Bedingungen erfüllt sind:

- *Theorie* und *vorliegende Definitionen* sind Mengen von Sätzen, R ist n-stelliger Prädikatbuchstabe und *Definition* ist ein Satz.
- *Definition* ist ein n-mal universell quantifiziertes Bikonditional der Form $(\forall v_1)...(\forall v_n)$ $(Rv_1,...,v_n \leftrightarrow A)$
- $v_1,...,v_n$ sind verschieden
- Das Definiens hat keine anderen freien Variablen als $v_1,...,v_n$
- Jedes nicht-logische Symbol in A gehört der Sprache von *Theorie* und *vorliegende Definitionen* an
- R kommt nicht in *Theorie* und *vorliegende Definitionen* vor

Regeln als Zweck zum Mittel

- Es lässt sich beweisen, dass mit den so formulierten Regeln (ähnlich für Funktionssymbole) das Nicht-Kreativitäts- und Eliminierbarkeitskriterium erfüllen lässt.
- Die Regeln sind Mittel zum Zweck. (vgl. Verhältnis Kalkül-semantisches System)
- Im Eliminierbarkeitskriterium steckt ein globales Zirkularitätsverbot, in den Regeln ein lokales.

Rekursive Definition

- **Problem mit (EK)**: Eliminierbarkeitskriterium gestattet keine rekursiven Definitionen.

  Bsp.: $x + y = z$:gdw $(y = 0 \;\&\; z = x) \vee (\exists v)(\exists w)(y = v' \;\&\; z = w' \;\&\; x + v = w)$

  Additionszeichen in $(\forall x)(\forall y)(x + y = y + x)$ nicht eliminierbar

- Es gibt neben den rekursiven Definitionen weitere Definitionsformen, die auch negative Vorkommnisse des Definiendums im Definiens gestatten

  Das Verfahren zur Bestimmung der Anwendungsbedingungen des Definiendums weicht bei solchen Definitionen sowohl vom Gleichsetzungsverfahren, als auch von dem Verfahren bei rekursiven Definitionen (kleinsten Fixpunkt zu wählen) ab. Am Ende erhält man aber ständig eine eindeutige Extension für das Definiendum.

- Hiervon weicht die Revisionstheorie der Definition ab.

Revisionstheorien der Definition und Wahrheit

- **Lit.:** Nuel Belnap/Anil Gupta: „The Revision Theory of Truth“ (1993)
- Vorarbeiten in den 70‘er (Herzberger)
- Formale Wahrheitstheorie RTT (Revision Theory of Truth) im Stile der Fixpunkttheorie der Wahrheit gemäß Kripke („An Outline of a Theory of Truth“)
- Benutzt Revisionstheorie der Definition (RTD)

Thesen und Ansprüche von RTD/RTT

- Einige Begriffe (natürlicher Sprachen) sind zirkulär. (Nicht: Begriffe, die rekursiv definiert werden)
- Standarderklärung von Definitionen sagt nichts Nützliches über zirkuläre Begriffe
- Man erhält „mächtige Theorie“ zirkulärer Begriffe durch Theorie zirkulärer Definitionen
- Wahrheit (im Englischen, Deutschen usw.) ist ein zirkulärer Begriff
- Mit der Revisionstheorie lassen sich bestimmte linguistische Phänomene (z.B. Paradoxie vom Lügner) im Zusammenhang mit dem Wahrheitsbegriff erklären

2 - Motivation: ein Vergleich

Für folgende Darstellung akzeptiere Regeln für Prädikat „ist wahr“

(WB) $\dfrac{\text{‚A‘ ist wahr}}{\text{A}}$ (WE) $\dfrac{\text{A}}{\text{‚A‘ ist wahr}}$

Fasse Regeln als Prozeduren zur Berechnung des semantischen Status auf

‚A‘ ist wahr $\downarrow$ A

‚A‘ ist nicht wahr $\downarrow$ Nicht-A

Motivation: Verhalten des Lügnersatzes

Bestimme das “semantische Verhalten“ der folgenden Sätze

(1) Der Satz (1) ist nicht wahr. („Lügnersatz“)

(2) Der Satz (2) ist wahr. („Wahrsagersatz“)

Ad (1)	Ad (2)
‚A‘ ist wahr	‚A‘ ist wahr
↓	↓
‚A‘ ist nicht wahr	‚A‘ ist wahr
↓	↓
‚A‘ ist wahr	‚A‘ ist wahr
↓	↓
...	...

Motivation: Verhalten einer zirkulären Definition

Ähnliche Regeln für Definition

$Gx =_{def} A(x)$

(G: einstelliges Prädikatsymbol, A(x): in x offene Formel der Prädikatenlogik)

(DB) $\begin{array}{c} Gt \\ \downarrow \\ A(t) \end{array}$ (DE) $\begin{array}{c} A(t) \\ \downarrow \\ Gt \end{array}$

Betrachte das „semantische Verhalten“ für folgende Definitionen

$\begin{array}{c} Gt \\ \downarrow \\ A(t) \end{array}$ $\begin{array}{c} \neg Gt \\ \downarrow \\ \neg A(t) \end{array}$

(1*) $Gx =_{def} Fx \vee (Hx \,\&\, \neg Gx)$

(2*) $Gx =_{def} Fx \vee (Hx \,\&\, Gx)$

Sei hierzu M = (W,I) ein beliebiges Modell mit $I(a) \in I(H)$ und $I(a) \notin I(F)$,
G unter I nicht interpretiert.

Motivation: Verhalten von zirkulären Definitionen

Ad (1*)	Vgl. Ad (1)	Ad (2*)	Vgl. Ad (2)
Ga	‚A' ist wahr	Ga	‚A' ist wahr
↓	↓	↓	↓
Fa v (Ha & ¬Ga)	‚A' ist nicht- wahr	Ga	‚A'wahr
↓	↓	↓	↓
¬Ga	‚A' ist wahr	Ga	‚A' ist wahr
↓	↓	↓	↓
¬ (Fa v (Ha & ¬Ga))	‚A' ist nicht-wahr	Ga	‚A' ist wahr
↓		↓	↓
Ga		…	…
↓			
¬Ga			
↓			
…			

Fazit: Der Wahrheitsbegriff und manche zirkuläre Definition sind „verhaltensgleich"

3 - Revisionstheorie der Definition (RTD)

Zur Präzisierung und Ausarbeitung der motivierenden Idee entwickle Theorie zirkulärer Definitionen, die für Definitionen angibt:

- die Art, wie die Bedeutung der Definienda festgelegt werden
- Regeln für den Umgang mit Definitionen

Die (RTD) von Belnap/Gupta hat folgende Eigenschaften/umfasst:

- Eliminierbarkeitskriterium in ursprünglicher Fassung verworfen
- Nicht-Kreativitätsforderung erfüllt
- Definitionen werden als metasprachliche Schemata aufgefasst, folglich: Definitionen nicht als materiale Bikonditionale aufgefasst
- Verschiedene semantische Systeme zirkulärer Definitionen und zugehörige Kalküle: S_0, S_n ($n \in N$),S^*,$S^\#$, C_0,Cn.

(RTD): S_0 und C_0

- Wie in der klassischen Definitionstheorie: Definitionen legen die Bedeutung des Definiendums fest.
- Bedeutung verstanden als: Extension unter allen möglichen Umständen (Intension)
- In dem „Wie“ der Bedeutungsfestlegung weicht RTD von der klassischen Theorie ab:
  - Nicht: absolute Anwendungsbedingungen für das Definiendum über das Definiens festgelegt
  - Sondern: hypothetische Anwendungsbedingungen

- Im folgenden: Darstellung des semantischen Systems S_0 und des zugehörigen Kalküls C_0. Hinweis auf S_n,C_n,S^*,$S^\#$

S_0: ein Beispiel

Voraussetzungen für das Beispiel:

- Prädikatenlogische Sprache mit:
  - einstelligen Prädikatbuchstaben: F, G, H
  - Namenbuchstaben: a,b,c,d
  - Variablen: x,y,z,…
- Modell M = (W,I) mit Wertebereich W und Interpretation I mit:
  - W = {a‘,b‘,c‘,d‘}
  - I(F) = {a‘,b‘}, I(H) = {a‘,c‘}
  - I(a) = a‘,I(b) = b‘, I(c) = c‘, I(d) = d‘
- Beachte: unter M erhält G keine Interpretation, M ist ein partielles Modell.

Die „Bedeutung“ des einstelligen Prädikatbuchstabens ‚G‘ soll durch das folgende metasprachliche Definitionsschema festgelegt sein:

(D) $Gx =_{def} (Fx \,\&\, Hx) \vee (Fx \,\&\, \neg Hx \,\&\, Gx) \vee (\neg Fx \,\&\, Hx \,\&\, \neg Gx)$

S_0: Beispiel (Forts.)

Idee zur Bestimmung der Bedeutung von G:

Unterscheide das Vorkommnis von G im Definiendum und die Vorkommnisse von G im Definiens A(G,x)

(D) $\quad G\,x =_{def} (Fx \,\&\, Hx) \vee (Fx \,\&\, \neg Hx \,\&\, Gx) \vee (\neg Fx \,\&\, Hx \,\&\, \neg\, Gx)$

Nimm für G im Definiens hypothetisch eine Extension an und werte unter der so von I auf I‘ erweiterten Interpretation A(G,x) aus.

Das Ergebnis der Auswertung wird als neue Extension für G verstanden.

Wiederhole das Verfahren mit der neuen Extension.

Ergebnis:

Für jede Anfangshypothese bzgl. der Extension von G erhält man eine Folge von revidierten Extensionen.

Diese Folgen sind bestimmt durch den so genannten Revisionsoperator; er liefert für eine hypothetische Extension die revidierte Extension:

$\delta_{M,D}$: Pot(W) $\rightarrow$ Pot(W)

Revisionsfolgen festgelegt durch: $\delta^0_{M,D}(X) = X$, $\delta^{n+1}_{M,D}(X) = \delta_{M,D}(\delta^n_{M,D}(X))$

S_0: Beispiel Fortsetzung

1. Anfangshypothese für G: $\emptyset$; gesucht: neue Extension $\delta_{M,D}(\emptyset)$ für G.
2. Auswertung von A(G,x):
   - Auswertung von Fx & Hx: {a‘}
   - Auswertung von (Fx & ¬Hx & Gx): $\emptyset$
   - Auswertung von (¬Fx & Hx & ¬ Gx): {c‘}
3. Ergebnis: $\delta_{M,D}(\emptyset)$ = {a‘} $\cup$ $\emptyset$ $\cup$ {c‘} = {a‘,c‘}

(D) G x $=_{def}$ (Fx & Hx) $\vee$ (Fx & ¬Hx & Gx) $\vee$ (¬Fx & Hx & ¬ Gx)

Modell M = (W,I) mit Wertebereich W und Interpretation I mit:

- W = {a‘,b‘,c‘,d‘}
- I(F) = {a‘,b‘}, I(H) = {a‘,c‘}
- I(a) = a‘,I(b) = b‘, I(c) = c‘, I(d) = d‘

S_0: Beispiel Fortsetzung

1. Revidierte Hypothese für G: {a‘,c‘} ; gesucht: $\delta_{M,D}$ ({a‘,c‘}).
2. Auswertung von A(G,x):
   - Auswertung von Fx & Hx: {a‘}
   - Auswertung von (Fx & ¬Hx & Gx): ∅
   - Auswertung von (¬Fx & Hx & ¬ Gx): ∅
3. Ergebnis: $\delta_{M,D}$ ({a‘,c‘}) = $\delta_{M,D}(\delta_{M,D}(\emptyset))$ = {a‘} ∪ ∅ ∪ ∅ = {a‘}

(D) G x = $_{def}$ (Fx & Hx) ∨ (Fx & ¬Hx & Gx) ∨ (¬Fx & Hx & ¬ Gx)

Modell M = (W,I) mit Wertebereich W und Interpretation I mit:

- W = {a‘,b‘,c‘,d‘}
- I(F) = {a‘,b‘}, I(H) = {a‘,c‘}
- I(a) = a‘,I(b) = b‘, I(c) = c‘, I(d) = d‘

S_0: Beispiel (Forts.)

Ergebnis

Für die Anfangshypothese ergibt sich die Folge revidierter Hypothesen bzgl. der Extension von G:

$\delta^0_{M,D}(\emptyset) = \emptyset$

$\delta^1_{M,D}(\emptyset) = \{a', c'\}$

$\delta^2_{M,D}(\emptyset) = \{a'\}$

$\delta^3_{M,D}(\emptyset) = \{a', c'\}$

$\delta^4_{M,D}(\emptyset) = \{a'\}$

…

Auf ähnliche Weise lassen sich für alle Teilmengen X der Domäne W Revisionsfolgen mit X als Anfangshypothese bilden.

Problem

Wie kommt man von hypothetischen Aussagen zu „absoluten“ / kategorischen Aussagen?

Antwort von Belnap/Gupta: Semantische Kategorien in den Systemen $S_0, S_n, S^*, S^\#$

Das semantische System S_0: Kategorien (1)

Die folgenden semantischen Kategorien werden unter Rückgriff auf den Revisionsoperator (hier für den Spezialfall einer Definition von G) getroffen.

(Def. „wahr in M+X")

Ein Satz A ist wahr in M+X :gdw A ist wahr in M relativ zur Hypothese, dass X die Extension von G ist.

(Def. „gültig in M")

Ein Satz A ist gültig in M relativ zur Definition D in S_0 :gdw Es gibt natürliche Zahl p: Für alle q >= p und alle Teilmengen X der Domäne von M gilt: A ist wahr in M + $\delta^q_{M,D}$ (X)

(Def. „kategorisch")

Ein Satz A ist kategorisch in M relativ zu D in S_0 :gdw Entweder A oder $\neg$A ist gültig in M relativ zu D in S_0

Das semantische System S_0: Kategorien (2)

(Def. „pathologisch“)

Ein Satz A ist pathologisch in M relativ zu D in S_0 :gdw A ist nicht kategorisch in M relativ zu D in S_0.

Pathologische Sätze, die bzgl. jeder Anfangshypothese in ihrem Wahrheitswert oszillieren, werden **paradox** genannt.

(Def. „gültig“)

Ein Satz ist gültig relativ zu D in S_0 :gdw A ist für alle klassischen Modelle M gültig in M relativ zu D in S_0

Das semantische System S_0: Beispiel

Modell M = (W,I) mit Wertebereich W und Interpretation I mit:

- W = {a‘,b‘,c‘,d‘}
- I(F) = {a‘,b‘}, I(H) = {a‘,c‘}
- I(a) = a‘,I(b) = b‘, I(c) = c‘, I(d) = d‘

(D) Gx = def
(Fx & Hx) v
(Fx & ¬Hx & Gx) v
(¬Fx & Hx & ¬Gx)

Bsp. 1:

Der Satz G(a) ist falsch in M+ $\emptyset$.

G(a) ist wahr in M + $\delta^n_{M,D}$ ($\emptyset$) für alle n>0.

G(a) ist wahr in M + $\delta^n_{M,D}$ (X) für alle Teilmengen X der Domäne von M für alle n>0.

Folglich: G(a) ist gültig in M relativ zu D (in S_0).

Man kann G(a) „ohne wenn und aber “ bejahen.

Bsp. 2:

Der Satz A:= H(a) & G(c) ist wahr in M + $\delta^n_{M,D}$ ($\emptyset$) , falls n ungerade, falsch sonst.

Folglich ist A nicht gültig. A ist sogar paradox.

Anmerkungen zu S_0

Obige Definitionen für S_0 lassen sich in mehrfacher Hinsicht verallgemeinern:

- Gestattet ist statt *einer* Definition auch eine (sogar unendliche) Menge von Definitionen. Der Revisionsoperator operiert dann auf einem kartesischen Produkt von Extensionen.
- Statt Prädikatbuchstaben lassen sich auch Namenbuchstaben und Funktionsbuchstaben definieren
- Die zugrundeliegende Sprache muss nicht zweiwertig sein: Gestattet sind u.a. dreiwertige Logiken mit dem starken oder schwachen Kleene-Schema oder einem vierwertigen Bewertungsschema.

In ähnlicher Weise wie S_0 werden die semantischen Systeme S_n erklärt. Bei S^* und $S^\#$ wird die Revision ins Transfinite fortgeführt.

Der Kalkül C_0

- Für den Kalkül C_0 werden die Eliminierbarkeits- und Einführungsregel der klassischen Definitionstheorie modifiziert. (Ansonsten droht ein Widerspruch).
- Idee: Mit einem Index wird Protokoll über die momentane Revisionsstufe geführt.

  Alle Formeln werden dafür mit einem ganzzahligen Index i versehen.
- Für Definitionen $Gx =_{def} A(x,G)$ erkläre

$$\textbf{(DE}_{\textbf{r}}\textbf{)} \quad \frac{A(t,G)^{i}}{G(t)^{i+1}} \qquad\qquad \textbf{(DB}_{\textbf{r}}\textbf{)} \quad \frac{G(t)^{i}}{A(t,G)^{i-1}}$$

Zu Grunde gelegt wird ein Kalkül des natürlichen Schließens.
Einschränkung für die Regeln: Alle Prämissen einer Regel haben denselben Index. Die Konklusion der Regel (verschieden von IS) hat denselben Index wie die Prämisse(n).

Kalkül C_0 (Forts.)

Regel IS (Index-Shift): In einer Ableitung darf bei Formeln, die G nicht enthalten, der Index auf eine beliebige ganze Zahl geändert werden.

Regeln von C_0:

(DE_r)+ (DB_r)+ IS + Ableitungskalkül der Prädikatenlogik (mit obiger Einschränkung)

(Def. Ableitung):

Eine Formel ist ableitbar auf der Basis einer Definition D :gdw es gibt eine Ableitung von A^0 in C_0 auf der Basis von D.

C_0: Beispiel

Bsp.:

$Gx =_{def} Fx \lor (Hx \,\&\, \neg Gx)$

Ableitung von $(\forall x)\,(Fx \rightarrow Gx)$

1*	(1)	$(Fa)^0$	Annahme
1*	(2)	$(Fa)^{-1}$	1, IS
1*	(3)	$(Fa \lor (Hx \,\&\, \neg Gx))^{-1}$	2, Theorem der klass. Logik
1*	(4)	$(Ga)^0$	3, DE_r
	(5)	$(Fa \rightarrow Ga)^0$	1,4 $\rightarrow$E
	(6)	$(\forall x)\,(Fx \rightarrow Gx)^0$	5, $\forall$E

C_0: Theoreme

- C_0 ist bzgl. S_0 korrekt und vollständig
- Bei nicht-zirkulären Definitionen kann man auf die Indizes verzichten

4 - Wahrheit als zirkulärer Begriff: RTT

Die Revisionstheorie der Wahrheit (RTT: Revision Theory of Truth) verwendet die RTD zur Plausibilisierung der folgenden Thesen:

- Der natürlichsprachliche Wahrheitsbegriff ist zirkulär
- Die Signifikation des Wahrheitsprädikats ist eine Revisionsregel und wird vollständig durch die Tarskischen Bikonditionale festgelegt

  (Signifikation: Verallgemeinerung des Extensionbegriffes)
- Ein Großteil des Verhaltens des Wahrheitsbegriffes, sowohl das gewöhnliche wie auch pathologische, kann in RTT erklärt werden.

Eingrenzung des Wahrheitsbegriffes

Mit *dem* Wahrheitsbegriff natürlicher Sprachen ist gemeint der

1. logische (im Gegensatz zu: nicht-logisch)
2. schwache (im Gegensatz zu: stark)
3. absolute (im Gegensatz zu: Modellrelativ)

Wahrheitsbegriff

Ad (1): ‚A' ist wahr in einer möglichen Welt w :gdw ‚A' ist korrekt in w in der Bedeutung, die A in unserer Welt. (U.a. ist gefordert: Bedeutung der Junktoren konstant über Welten w hinweg.)

Ad (2): Der semantische Status von Sätzen der Form „ „P" ist wahr" ist identisch mit dem semantischen Status von P.

Ad (3): nicht „wahr in M" (für Variable M die über Modelle läuft), sondern „wahr in der aktuellen Welt"

RTT: Voraussetzungen an die Sprache (Syntax)

Betrachtet werden Sprachen, die folgende universelle Züge aufweisen:

- Jede dieser Sprachen enthält ein Wahrheitsprädikat und
- Anführungsnahmen für alle in ihr enthaltenen Sätze

Syntax

- *Technische Zeichen*: Klammern und abzählbar viele Variablen (zur Quantifikation)
- *Logische Zeichen*: Alle bekannten Junktorenzeichen, Quantorensysmbole für Existenz- und Alloperator; Prädikatbuchstabe „T“ oder Ausdruck „ist wahr“ für das Wahrheitsprädikat
- *Nichtlogische Zeichen*: Möglicherweise leere Menge an Prädikat- und Namenbuchstaben. Zusätzlich spezielle Anführungsnamen, die links mit einem Komma beginnen, gefolgt von einem Satz (= geschlossene Formel) dieser Sprache, und mit einem rechten Hochkomma abgeschlossen werden.

„Term“, “wohlgeformte Formel“, “Satz“ wie üblich definiert.

RTT: Voraussetzungen an die Sprache (Semantik)

Begriff des Basismodells (für Sprachen L mit oben genannter Syntax)

M ist ein Basismodell einer Sprache L :gdw M = (W,I), wobei W eine Domäne, I eine partielle Interpretation von L ist und M folgende Bedingungen erfüllt:

- W enthält sämtliche Sätze aus L.
- Jeder Namenbuchstabe m von L wird durch ein Objekt I(m) aus W interpretiert.
- Für Anführungsnamen ‚s' (mit einem Satz s aus L) gilt: I(‚s') = s.
- Für jeder n-stelligen Prädikatbuchstaben ‚R' verschieden vom Wahrheitsprädikat ‚T' gilt:

  $I(R) \subseteq W^n$
- I wird wie üblich auf Formeln erweitert, so dass T-freie Sätze einen klassischen Wahrheitswert erhalten.

T-Bikonditionale

- In Basismodellen ist der Prädikatbuchstabe ‚T' uninterpretiert. Deren Bedeutung soll durch die so genannten T-Bikonditionale festgelegt werden:

  s ist wahr ↔ A   bzw. T(s) ↔ A

(s: Name eines Satzes der Sprache L; A: Übersetzung von s in dieselbe Sprache L)

- **Bsp.:**

  ‚ ‚Schnee ist weiß' ist wahr' ist wahr ↔ ‚Schnee ist weiß' ist wahr

- In der RTT werden die T-Bikonditionale als metasprachliche Definitionen aufgefasst, die nicht Teil der Sprache L sind:

  s ist wahr $=_{def}$ A

- Unterscheide daher im folgenden: Sätze aus L des Typs „s ist wahr ↔ A" und metasprachliche Ausdrücke: s ist wahr $=_{def}$ A für L

Der Tarski-Jump

Für L liegen unendlich viele partielle Definitionen für jeden Satz s_i aus L vor:

	s_1 ist wahr $=_{def} A_1$		$T(s_1) =_{def} A_1$
(D)	s_2 ist wahr $=_{def} A_2$	bzw. genauer:	$T(s_2) =_{def} A_2$
	s_3 ist wahr $=_{def} A_3$		$T(s_3) =_{def} A_3$
	...		...

Gemäß RTD wird durch diese Definitionen D in jedem Modell M ein Revisionsoperator $\delta_{D,M}$ bestimmt. Dieser wird als Tarski-Jump bezeichnet und mit τ_M abgekürzt wird. Er lässt sich folgendermaßen reformulieren:

Sei M Basismodell (W,I) und U Teilmenge von W. Dann ist für einen Satz s_i :

$s_i \in \tau_M(U)$:gdw A_i ist wahr in M (i.e. erhält den Wahrheitswert **t** unter I) unter der Hypothese, dass die Extension von ‚T' U ist.

Beispiel für RTT

Bsp. 1:

L mit Vokabular:

Namen: Schnee, Jones, Anführungsnamen für alle Sätze der Sprache L

Einstellige Prädikate: ist weiß

Zweistellige Prädikate: äußert

Wahrheitsprädikat: ist wahr

Modell M = (W,I) mit:

W = {Schnee, Jones} vereinigt mit Menge aller Sätze aus L

I für Namen: I(Schnee) = Schnee;I(Jones) = Jones; für Anführungsnamen:

I(‚S') = S (wobei S Satz aus L)

I für 1-stellige Prädikate: I(ist weiß) = {Schnee}

I für 2-stellige Prädikate: I{äußert) = {(Jones,Schnee ist weiß),(Jones,'Schnee ist weiß' ist wahr),(Jones, Etwas, was Jones äußert, ist wahr)}

Bsp. 1 (Forts.)

$\tau_M(\emptyset)$ = {Schnee ist weiß, ‚Schnee ist weiß' ist nicht wahr, , ‚Schnee ist weiß' ist wahr' ist nicht wahr,...,Alles ist nicht wahr, Alles, was Jones äußert, ist nicht wahr, ...}

Beobachtung:

- Betrachtung der Revisionsfolgen und aller Hypothesen U ergibt die Beobachtung: Es gibt zu jedem U ein n> 1, so dass der Satz ‚Schnee ist weiß' ist wahr in der Menge $\tau^n_M(U)$ enthalten ist. In der Terminologie von S_0 ist dieser Satz gültig in M.
- Für alle Sätze gilt, dass sich ihr semantischer Wert nach einer endlichen Anzahl von Revisionsschritten auf einen Wert stabilisiert; sogar für jede Anfangshypothese auf denselben Wert => Alle Sätze sind entweder kategorisch wahr oder kategorisch falsch.

Bsp. 2,3 (Forts.)

Bsp.2:

Betrachte Modell M‘, das wie M ist bis auf die Interpretation von äußert:

I‘(äußert) = {(Jones , Etwas von dem, was Jones äußert, ist wahr)}

Beobachtung:

Für jeden Anfangshypothese stabilisiert sich die Revisionsfolge auf eine bestimmte Menge. Jedoch ist die verschieden, je nachdem ob in der Anfangshypothese der Satz Etwas von dem, was Jones äußert, ist wahr vorkommt oder nicht. Dieser Satz ist pathologisch (aber nicht paradox).

Bsp. 3:

Betrachte Modell M‘‘, das wie M ist bis auf die Interpretation von äußert:

I‘‘(äußert) = {(Jones , Etwas von dem, was Jones äußert, ist nicht wahr)}

Beobachtung:

Oszillation in der Revisionsfolge. Der Satz Etwas von dem, was Jones äußert, ist nicht wahr ist paradox in S_0

Das Argument für die Zirkularität der Wahrheit

1. Für alle Sprachen L mit den „extensionalen“ Eigenschaften der Prädikatenlogik erster Stufe und Anführungsnamen für alle Sätze der Sprache L, für alle Prädikate G und H aus L gilt: Wenn H und G intensional äquivalent sind und H in L über eine wesentlich zirkuläre stipulative Definition definiert wird, dann drückt G einen zirkulären Begriff aus.
2. Für jede Sprache L mit den extensionalen Eigenschaften gilt: Man kann in einer wesentlich zirkulären, stipulativen Definition ein einstelliges Prädikat T in L definieren, für welches gilt: T in L und das Prädikat ‚ist wahr in L‘, welches den logischen, schwachen, absoluten Begriff der Wahrheit für L ausdrückt, sind intensional äquivalent.
3. Also: Für alle Sprachen mit obigen Eigenschaften gilt: Der logische, schwache, absolute Begriff der Wahrheit ist zirkulär.
4. „These von der konservativen Erweiterung der natürlichen Sprachen“
5. Also: Der logische, schwache, absolute Wahrheitsbegriff der deutschen (englischen etc.) Sprache ist zirkulär.

5 - Erläuterung der Lügnerparadoxie

Belnaps/Guptas Erläuterung der Lügnerparadoxie lässt sich knapp folgendermaßen formulieren:

Die Lügnerparadoxie ist darin begründet, dass der (logische,absolute,schwache) Wahrheitsbegriff natürlicher Sprachen zirkulär ist.

Anmerkung:

- Diese Erläuterung versucht nicht das normative Problem der Lügnerparadoxie zu lösen, sondern das deskriptive: sie kann eine konsistente Erklärung für die unterschiedliche Einteilung von unproblematischen Sätzen wie „Schnee ist weiß" und problematischen wie „Dieser Satz ist wahr" und „Dieser Satz ist nicht wahr" (ersteres pathologisch,nicht-paradox, letzteres paradox) geben.
- Sie hält an der Intuition fest, dass natürliche Sprachen ihr Wahrheitsbegriff enthalten
- Die unterliegende Logik für die Sprache L, in die das Wahrheitsprädikat eingeführt wird, wird nicht geändert.

6 - Kritik an RTD/RTT

Im folgenden eine kurze Aufzählung möglicher Kritikpunkte an RTT:

- Die Problematik einer stärkeren Version der Lügnerparadoxie: „Dieser Satz ist nicht kategorisch oder nicht wahr."
  - Existenz eine universellen Sprache?
- Wahrheitsprädikat zu komplex für den „naiven Wahrheitsbegriff" natürlicher Sprachen
- Klassische Gesetze wie das „Tertium non datur" stellen sich als gültig in RTD heraus.
- Spezielle Behandlung an Limespunkten
- Omega-Inkonsistenz
- ...

7 - Zusätzliche Anwendungen von RTD

- Theorie der Eigenschaften (s. Orilia)
- Theorie der Berechenbarkeit (s. Antonelli)
- Mengentheorie: Theorie nicht-fundierter Mengen

  (Vgl. Barwise/Moss: „Vicious Circles“)