MATHEMATICAL RESEARCH

MATHEMATISCHE FORSCHUNG

# Frege Conference 1984

edited by
G. Wechsung

Band 20

Akademie-Verlag · Berlin

## FREGES PERMUTATIONSARGUMENT. ZU § 10 DER "GRUNDGESETZE DER ARITHMETIK"

Peter Schroeder-Heister [1]

Wertverläufe und Wahrheitswerte bilden den Gegenstandsbereich der logisch-mathematischen Sprache von Freges "Grundgesetzen der Arithmetik" [2]. Wertverläufe sind dabei über das Abstraktionsprinzip

$$(\grave{\varepsilon}\,\Phi(\varepsilon) = \grave{\alpha}\,\Psi(\alpha)) \quad = \quad (-\!\!\overset{\mathfrak{a}}{\smile}\!\!-\,\Phi(\mathfrak{a}) \quad = \quad \Psi(\mathfrak{a}))$$

eingeführt (§ 3), das im formalen Aufbau des Fregeschen Systems als Grundgesetz V erscheint. Es bestimmt als notwendige und hinreichende Bedingung für die Gleichheit von Wertverläufen die extensionale Äquivalenz derjenigen Funktionen, aus denen die Wertverläufe abstrahiert sind. Für die Einführung von Wahrheitswerten als Bedeutungen von Aussagesätzen ist der Gedanke leitend, auch Begriffe als Funktionen auffassen zu können (§ 2).

Diese beiden Bereiche sind nun nicht notwendigerweise disjunkt. Vielmehr versucht Frege in der zweiten Hälfte des § 10 der "Grundgesetze" zu zeigen, daß es "immer möglich [ist] zu bestimmen, daß ein beliebiger Wertverlauf das Wahre und ein beliebiger anderer das Falsche sein solle" (S. 17). Eine solche Behauptung ist natürlich nur dann sinnvoll, wenn nicht schon "ontologisch" festgelegt ist, welche Gegenstände Wertverläufe und welche Wahrheitswerte sind, wenn also eine Manipulation innerhalb gewisser Grenzen möglich ist. Diese Grenzen werden durch das Abstraktionsprinzip gesetzt; die Identifizierung von Wahrheitswerten mit Wertverläufen soll geschehen, "ohne ... mit der Gleichsetzung von '$\grave{\varepsilon}\,\Phi(\varepsilon) = \grave{\varepsilon}\,\Psi(\varepsilon)$' mit '$-\!\!\overset{\mathfrak{a}}{\smile}\!\!-\,\Phi(\mathfrak{a}) = \Psi(\mathfrak{a})$' in Widerspruch zu geraten" (ebd.).

Daß die Bedeutung von Wertverlaufstermen durch das Abstraktionsprinzip nicht eindeutig festgelegt wird, hatte Frege zu Beginn von § 10 mit folgendem Argument gezeigt: Sei $X(\xi)$ eine von der Identität verschiedene bijektive Funktion. Dann ist für beliebige '$\Phi$' und '$\Psi$' '$X(\grave{\varepsilon}\,\Phi(\varepsilon)) = X(\grave{\alpha}\Psi(\alpha))$' gleichbedeutend mit '$\grave{\varepsilon}\,\Phi(\varepsilon) = \grave{\alpha}\,\Psi(\alpha)$' und damit mit '$-\!\!\overset{\mathfrak{a}}{\smile}\!\!-\,\Phi(\mathfrak{a}) = \Psi(\mathfrak{a})$'. D.h., die durch Namen der Form '$X(\grave{\varepsilon}\,\Phi(\varepsilon))$' bezeichneten Gegenstände haben dasselbe "Kennzeichen zur Wiedererkennung" wie die durch Namen der Form '$\grave{\varepsilon}\,\Phi(\varepsilon)$' bezeichneten Gegenstände, können also nicht von ihnen unterschieden werden.

---

1) Peter Schroeder-Heister, Universität Konstanz, Fachgruppe Philosophie, Postfach 5560, D-7750 Konstanz. - Ich danke Franz-Viktor Kuhlmann und Pirmin Stekeler-Weithofer für Anregungen. Eine ausführlichere Darstellung der hier präsentierten Thesen findet sich in meiner Arbeit [4].

2) Frege [3]. Seiten- und Paragraphenziffern im Text beziehen sich immer auf dieses Werk.

Dieses Argument bedarf der Rekonstruktion. Einerseits spricht Frege von "Gegenständen, deren Zeichen die Form '$\grave{\varepsilon}\,\Phi(\varepsilon)$' haben" (S. 16), setzt also voraus, daß Ausdrücke wie '$\grave{\varepsilon}\,\Phi(\varepsilon)$' für etwas Bestimmtes stehen, andererseits davon, daß ihre Bedeutung "keineswegs völlig bestimmt ist" (ebd.) durch die gegebenen Festlegungen. Ferner kann es ja wohl nicht die Absicht sein, neben Ausdrücken der Form '$\grave{\varepsilon}\,\Phi(\varepsilon)$' noch Ausdrücke der Form '$X(\grave{\varepsilon}\,\Phi(\varepsilon))$' einzuführen, d.h. eine neue Sprache zu konstruieren. Daß es eine zur Ausgangssprache isomorphe Sprache mit analogen Theoremen gibt, die jedoch über andere Gegenstände spricht, ist nicht das Problem, sondern daß die Ausgangssprache selbst in verschiedener Weise interpretiert werden kann.

Ich schlage zur Klärung eine modelltheoretische Deutung vor, nach der Zeichen eine Bedeutung relativ zu einer Interpretation I haben; die Nichteindeutigkeit der Bedeutung besteht dann darin, daß keine Interpretation I durch das Abstraktionsprinzip ausgezeichnet ist, sondern sich zu jeder Interpretation eine andere Interpretation I' angeben läßt, die dieselben Aussagen wahr bzw. falsch macht wie I; insbesondere ist dann I Modell des Abstraktionsprinzips genau dann, wenn I' Modell des Abstraktionsprinzips ist. Die von Frege benutzte Funktion X wäre dann eine Funktion zwischen den Gegenstandsbereichen von I und I', d.h. ein Objekt, über das man in der Metasprache redet. Ausdrücke der Form '$X(\grave{\varepsilon}\,\Phi(\varepsilon))$' wären als unzulässige Mischungen von meta- und objektsprachlichen Ausdrücken gar nicht mehr definiert. [1)]

Bei einer solchen modelltheoretischen Deutung ist zu beachten, daß in Freges System Wahrheitswerte (wenn auch ausgezeichnete) Gegenstände sind, die zum Variabilitätsbereich des Allquantors gehören. Ich gehe dabei von einer Sprache aus, wie sie Frege in den ersten 9 Paragraphen der "Grundgesetze" eingeführt hat, deren Terme also mit Hilfe von Waagerechtem ('——'), Negationszeichen ('—┬—'), Gleichheitszeichen ('='), Allquantor erster Stufe ('—◡—') und Wertverlaufsabstraktor ('ʾ') gebildet sind. Terme mit '——', '—┬—', '=' oder '—◡—' als Hauptzeichen sollen Formeln heißen, solche mit 'ʾ' als Hauptzeichen Wertverlaufsterme. Terme ohne freie Variablen (lateinische Buchstaben bei Frege) sollen geschlossen heißen, geschlossene Formeln auch Aussagen. Eine Interpretation I bezieht sich immer auf einen Gegenstandsbereich U, in dem zwei verschiedene Gegenstände w und f als die Wahrheitswerte (das Wahre und das Falsche) ausgezeichnet sind. Deshalb soll immer von einer Interpretation I über dem Tripel

---

1) Frege selbst scheint keine andere Möglichkeit gehabt zu haben, die Nichteindeutigkeit der Bedeutung von Wertverlaufstermen auszudrücken. Da er noch nicht über die Unterscheidung zwischen Objekt- und Metasprache verfügt und so (in moderner Ausdrucksweise) objektsprachliche Zeichen auch dazu verwendet, um in der Metasprache über Gegenstände zu reden, muß er zum Ausdruck der Mehrdeutigkeit zusätzliche objektsprachliche Zeichen (nämlich solche der Gestalt '$X(\grave{\varepsilon}\,\Phi(\varepsilon))$') anstelle von metasprachlichen Zeichen einführen.

(U,w,f) gesprochen werden (wobei $\{w,f\} \subseteq U$); ein solches I ist definiert als eine Abbildung von der Menge aller geschlossenen Wertverlaufsterme nach U, so daß höchstens w und f nicht zur Bildmenge von I gehören [1]. Wertverlaufsterme gelten also als kleinste semantische Einheiten. '$\top$' und '$\perp$' dienen als Abkürzungen für die ausgezeichneten Aussagen '$-\overset{\mathfrak{a}}{\smile}-\ \mathfrak{a} = \mathfrak{a}$' bzw. '$\top\!-\overset{\mathfrak{a}}{\smile}-\ \mathfrak{a} = \mathfrak{a}$', die als Standardnamen für die beiden Wahrheitswerte fungieren.

I wird wie folgt induktiv erweitert zu einer Abbildung I* von der Menge aller geschlossenen Terme nach U (mit '$\equiv$' als metasprachlichem Identitätssymbol):

$I^*(A) \equiv I(A)$ wenn A geschlossener Wertverlaufsterm

$I^*(\top) \equiv w$ $\qquad I^*(\perp) \equiv f$

$$I^*('\!-\!\!-\!\!-\,A') \equiv \begin{cases} w & \text{wenn } I^*(A) \equiv w \\ f & \text{sonst} \end{cases}$$

$$I^*('\!-\!\!\top\!\!-A') \equiv \begin{cases} f & \text{wenn } I^*(A) \equiv w \\ w & \text{sonst} \end{cases}$$

$$I^*('A = B') \equiv \begin{cases} w & \text{wenn } I^*(A) \equiv I^*(B) \\ f & \text{sonst} \end{cases}$$

$$I^*('\!-\overset{\mathfrak{a}}{\smile}-\Phi(\mathfrak{a})') \equiv \begin{cases} w & \text{wenn } I^*(\Phi(\top)) \equiv I^*(\Phi(\perp)) \equiv I^*(\Phi(B)) \equiv w \\ & \text{für jeden geschlossenen Wertverlaufsterm B } [2] \\ f & \text{sonst} \end{cases}$$

Ein geschlossener Term A <u>gilt</u> bei I über (U,w,f) ("I ist <u>Modell</u> von A über (U,w,f)"), falls $I^*(A) \equiv w$. Ein offener Term gilt bei I über (U,w,f), falls der durch Allquantifikation über alle in ihm vorkommenden freien Variablen entstehende geschlossene Term bei I über (U,w,f) gilt. Das Abstraktionsprinzip gilt bei I über (U,w,f), wenn alle seine Instanzen bei I über (U,w,f) gelten.

Freges oben erwähntes Argument über die nichteindeutige Festlegung der Bedeutung von Wertverlaufstermen kann man nun wie folgt reformulieren:

---

1) D.h., Wahrheitswerte und Wertverläufe (Bilder von Wertverlaufstermen unter I) erschöpfen den Gegenstandsbereich von I, wobei zugelassen ist, daß einer oder beide Wahrheitswerte zum Bereich der Wertverläufe gehören.

2) Der Variabilitätsbereich des Allquantors '$-\!\smile\!-$' besteht also aus den Namen <u>aller</u> Gegenstände einschließlich der durch $\top$ und $\perp$ bezeichneten Wahrheitswerte.

Argument 1 Sei I eine Interpretation über (U,w,f). Sei U' eine Menge, die w und f enthält, und sei X eine von der Identität verschiedene bijektive Abbildung von U nach U', für die gilt: X(w) = w, X(f) = f. Sei I' diejenige Interpretation über (U',w,f), die jedem geschlossenen Wertverlaufsterm A den Gegenstand X(I(A)) zuordnet. Dann läßt sich induktiv entsprechend der Definition von I* beweisen: Für jeden Term A gilt: I ist Modell von A über (U,w,f) genau dann, wenn I' Modell von A über (U',w,f) ist. Insbesondere ist I Modell des Abstraktionsprinzips über (U,w,f) genau dann, wenn I' Modell des Abstraktionsprinzips über (U',w,f) ist; d.h., das Abstraktionsprinzip zeichnet den Gegenstandsbereich U nicht vor U' aus.

In der zweiten Hälfte des § 10 kommt Frege dann zu der zitierten These, jedes Paar (r,s) verschiedener Wertverläufe könne als Paar (w,f) der beiden Wahrheitswerte festgesetzt werden. Seine Argumentation unterscheidet sich von der vorherigen darin, daß als X eine bestimmte Permutation von U (d.h. bijektive Abbildung von U nach U) gewählt wird, die (r,s) mit (w,f) vertauscht, ansonsten aber die Identität ist. Ich schlage folgende Rekonstruktion vor:

Argument 2 (Permutationsargument [1]) Sei I eine Interpretation über (U,w,f). Seien r und s untereinander und von w und f verschiedene Elemente von U. Sei X folgende Permutation von U: X(r) = w, X(w) = r, X(s) = f, X(f) = s, X(u) = u für alle sonstigen u aus U. Sei I' diejenige Interpretation über (U,r,s), die jedem geschlossenen Wertverlaufsterm A den Gegenstand X(I(A)) zuordnet. Dann läßt sich durch Induktion entsprechend der Definition von I* beweisen: Für jeden Term A gilt: I ist Modell von A über (U,w,f), wenn I' Modell von A über (U,r,s) ist. Insbesondere ist I Modell des Abstraktionsprinzips über (U,w,f), wenn I' Modell des Abstraktionsprinzips über (U,r,s) ist. [2]

eges Permutationsargument in dieser Reformulierung ist korrekt, wie der Detail durchgeführte induktive Beweis zeigt. [3] Jedoch ist es - dies t meine These - kein Argument für das, was es begründen soll, nämlich für, daß beliebige verschiedene Wertverläufe als das Wahre und das Falhe festgesetzt werden können. Auf den ersten Blick scheint dies zwar

---

Von "Permutationsargument" spreche ich im Anschluß an Dummett [2], S. 408.

Frege greift in seiner eigenen Formulierung des Permutationsarguments auf neue Wertverlaufsterme der Gestalt '$\grave{\eta}\,\Phi(\eta)$' zurück. Ich interpretiere das wieder als Folge der fehlenden Unterscheidung von Objekt- und Metasprache.

Vgl. Schroeder-Heister [4].

der Fall zu sein: Ist A ein geschlossener Wertverlaufsterm, für den $I(A) \equiv r$ gilt, so gilt $I'(A) \equiv w$, und entsprechend $I'(B) \equiv f$, falls $I(B) \equiv s$. D.h., die Wertverlaufsterme A und B bezeichnen unter I' die Wahrheitswerte w und f. Diese Ausdrucksweise ist jedoch irreführend. Wahrheitswert zu sein ist etwas Relatives und hängt von der gewählten Interpretation ab. Da I eine Interpretation über (U,w,f) ist, sind w und f die I zugehörigen Wahrheitswerte, die sich von den Wahrheitswerten r und s der Interpretation I' über (U,r,s) unterscheiden. Man kann also höchstens sagen: Die Wertverlaufsterme A und B bezeichnen unter I' die Wahrheitswerte w und f von I (und nicht zugleich die Wahrheitswerte r und s von I'). Damit ist allerdings in Hinblick auf Freges Ziel nichts gewonnen; man möchte ja Wahrheitswerte und Wertverläufe in einem Modell identifizieren.

Dies kann man sich auch so klar machen: Wenn ein unter I durch A bezeichneter Wertverlauf als das Wahre und ein durch B bezeichneter Wertverlauf als das Falsche festgesetzt werden soll, so müssen bei I' die Behauptungen '$A = \top$' und '$B = \bot$' gelten. Dies ist jedoch nicht der Fall. Wenn A und $\top$ bzw. B und $\bot$ durch I* verschieden interpretiert werden (d.h. wenn $I(A) \neq w$ und $I(B) \neq f$) - wovon man ja ausging - , dann werden sie auch durch I'* verschieden interpretiert (d.h. $I'(A) \neq r$ und $I'(B) \neq s$), da X als Permutation injektiv ist. Dieser Gedankengang, der allen mir bekannten Interpretationen widerspricht, auch der sehr sorgfältigen und detaillierten Darstellung bei Thiel [5] [1], läßt sich dahingehend verallgemeinern, daß sich Freges Ziel grundsätzlich nicht durch Erzeugung einer Interpretation I' aus einer Interpretation I durch Transformation des Gegenstandsbereichs erreichen läßt (vgl. [4]).

Das spricht natürlich nicht gegen Freges These über die Identifizierbarkeit von Wahrheitswerten mit Wertverläufen, sondern nur gegen Freges Begründung für diese These, d.h. gegen die Ansicht, das (korrekte) Permutationsargument stütze die Identifizierbarkeitsthese. Folgendes Beispiel zeigt allerdings, daß auch die These selbst in der von Frege beanspruchten Allgemeinheit nicht gilt. Sei I Modell des Abstraktionsprinzips über (U,w,f). Die Wertverlaufsterme A und B seien wie folgt definiert:

---

1) Auch Dummett, dessen kurze Skizze des Permutationsarguments (Dummett [2], S. 403f.) der vorgeschlagenen modelltheoretischen Interpretation sehr nahe kommt, scheint es als korrekte Begründung für die Identifizierbarkeit von Wertverläufen mit Wahrheitswerten anzusehen - ganz abgesehen von solchen Interpreten, denen die Gültigkeit von Freges Argumentationskette offensichtlich als so unproblematisch erscheint, daß sie Freges Behauptungen ohne weiteren Kommentar als erwiesen ansehen (z.B. Currie [1], S. 69).

$$A \equiv \begin{cases} '\acute{\varepsilon}(\bot = \varepsilon)' & \text{falls } I('\acute{\varepsilon}(\top = \varepsilon)') \equiv w \\ '\acute{\varepsilon}(\top = \varepsilon)' & \text{sonst} \end{cases}$$

$$B \equiv \begin{cases} '\acute{\alpha}(\neg\,\acute{\varepsilon}(\bot = \varepsilon) = \alpha)' & \text{falls } I('\acute{\varepsilon}(\top = \varepsilon)') \equiv w \\ '\acute{\alpha}(\acute{\varepsilon}(\top = \varepsilon) = \alpha)' & \text{sonst.} \end{cases}$$

Dann ist $I(A) \not\equiv I(B)$. Denn falls $I('\acute{\varepsilon}(\top = \varepsilon)') \not\equiv w$, ist $I^*('-\!\overset{\mathfrak{a}}{\smile}\!-(\top = \mathfrak{a}) = (\acute{\varepsilon}(\top = \varepsilon) = \mathfrak{a})') \equiv f$ (man instanziiere $\mathfrak{a}$ durch $\top$), also nach dem Abstraktionsprinzip $I^*('\acute{\varepsilon}(\top = \varepsilon) = \acute{\alpha}(\acute{\varepsilon}(\top = \varepsilon) = \alpha)') \equiv f$, d.h. $I('\acute{\varepsilon}(\top = \varepsilon)') \not\equiv I('\acute{\alpha}(\acute{\varepsilon}(\top = \varepsilon) = \alpha')$. Falls $I('\acute{\varepsilon}(\top = \varepsilon)') \equiv w$, ist $I('\acute{\varepsilon}(\bot = \varepsilon)') \not\equiv w$, da sich mit Hilfe des Abstraktionsprinzips $I^*('\acute{\varepsilon}(\top = \varepsilon) = \acute{\varepsilon}(\bot = \varepsilon)') \equiv f$ ergibt; damit ist $I^*('\neg\,\acute{\varepsilon}(\bot = \varepsilon)') \equiv w$, und somit $I^*('-\!\overset{\mathfrak{a}}{\smile}\!-(\bot = \mathfrak{a}) = (\neg\,\acute{\varepsilon}(\bot = \varepsilon) = \mathfrak{a})') \equiv f$ (man instanziiere '$\mathfrak{a}$' durch $\bot$), also nach dem Abstraktionsprinzip $I^*('\acute{\varepsilon}(\bot = \varepsilon) = \acute{\alpha}(\neg\,\acute{\varepsilon}(\bot = \varepsilon) = \alpha)') \equiv f$, d.h. $I('\acute{\varepsilon}(\bot = \varepsilon)') \not\equiv I('\acute{\alpha}(\neg\,\acute{\varepsilon}(\bot = \varepsilon) = \alpha)')$. Doch obwohl A und B unter I Verschiedenes bezeichnen, sind sie als Bezeichnungen für die beiden Wahrheitswerte nicht geeignet: Wäre nämlich ein Modell I' des Abstraktionsprinzips über einem beliebigen Gegenstandsbereich auch Modell von '$A = \top$' dann wäre es auch Modell von '$A = B$' und damit nicht von '$B = \bot$'. Falls nämlich $A \equiv '\acute{\varepsilon}(\top = \varepsilon)'$ und daher $B \equiv '\acute{\alpha}(A = \alpha)'$ ist, würde $I'(B) \equiv I'('\acute{\alpha}(\top = \alpha)') \equiv I'(A)$ folgen; falls $A \equiv '\acute{\varepsilon}(\bot = \varepsilon)'$ und daher $B \equiv '\acute{\alpha}(\neg A = \alpha)'$ ist, würde entsprechend $I'(B) \equiv I'('\acute{\alpha}(\neg\top = \alpha)')$ $\equiv I'('\acute{\alpha}(\bot = \alpha)') \equiv I'(A)$ folgen.

Damit ergibt sich als Aufgabe weiterführender Überlegungen, notwendige und/oder hinreichende Bedingungen für die Identifizierung von Wahrheitswerten mit Wertverläufen anzugeben. Im Hinblick darauf könnte dann auch Freges eigene Wahl von '$\acute{\varepsilon}(-\varepsilon)$' als Namen für das Wahre und von '$\acute{\varepsilon}(\varepsilon = (\neg -\!\overset{\mathfrak{a}}{\smile}\!- \mathfrak{a} = \mathfrak{a}))$' als Namen für das Falsche beurteilt werden. Dies würde möglicherweise den Anschein des (ansonsten Frege recht ferne liegenden) Konventionalismus von seiner Festsetzung der Wahrheitswerte bzw. Wahrheitswertnamen in § 10 der "Grundgesetze" nehmen.