R. M. Sainsbury
Paradoxien

Erweiterte Ausgabe

Aus dem Englischen übersetzt
von Vincent C. Müller

Philipp Reclam jun. Stuttgart

Titel der englischen Originalausgabe:
R. M. Sainsbury: Paradoxes. Second Edition.
Cambridge / New York / Melbourne:
Cambridge University Press, 1995.

Umschlagabbildung:
Paul Klee, Revolution des Viaduktes, 1937, 153 (R 13);
Öl auf Baumwolle; 60 × 50 cm; Hamburger Kunsthalle
© VG Bild-Kunst, Bonn 2000

RECLAMS UNIVERSAL-BIBLIOTHEK Nr. 18135

Die Übersetzung erscheint mit Genehmigung von
The Syndicate of the Press of the University of Cambridge,
Cambridge, England

Satz: Utesch GmbH, Hamburg
Druck und Bindung: Reclam, Ditzingen. Printed in Germany 2005
RECLAM, UNIVERSAL-BIBLIOTHEK und
RECLAMS UNIVERSAL-BIBLIOTHEK sind eingetragene
Marken der Philipp Reclam jun. GmbH & Co., Stuttgart
ISBN 3-15-018135-6

www.reclam.de

Inhalt

5 Klassen und Wahrheit

Die in diesem Kapitel zu behandelnden Paradoxien sind wahrscheinlich die schwierigsten von allen, aber auch die fruchtbarsten. Russells Paradoxie über Klassen, die er 1901 entdeckte, hat eine gewaltige Anzahl von Arbeiten über die Grundlagen der Mathematik hervorgerufen. Russell meinte, seine Paradoxie sei von derselben Art wie die Lügnerparadoxie, welche in ihrer einfachsten Form in der Behauptung besteht »Ich lüge jetzt (hiermit!)«. Die Lügnerparadoxie ist für Theorien der Wahrheit von größter Wichtigkeit gewesen. Alles, was mit diesen Paradoxien zu tun hat, ist höchst umstritten, einschließlich der Frage, ob Russell damit Recht hatte zu meinen, dass seine Paradoxie über Klassen und die Lügnerparadoxie derselben Quelle entspringen (siehe Abschnitt 5.9).

5.1 Russells Paradoxie

Wenn Sokrates ein Mensch ist, dann ist er ein Element der Klasse der Menschen. Wenn er Element der Klasse der Menschen ist, dann ist er ein Mensch. Können *Klassen* Elemente von Klassen sein? Die Antwort scheint Ja zu sein. Die Klasse der Menschen hat mehr als 100 Elemente, also ist die Klasse der Menschen Element der Klasse der Klassen mit mehr als 100 Elementen. Im Gegensatz dazu gehört die Klasse der Musen nicht zu der Klasse der Klassen mit mehr als 100 Elementen, da die Klasse der Musen unserer Tradition zufolge nur neun Elemente hat.

Die meisten Klassen sind nicht Elemente ihrer selbst. Die Klasse der Menschen ist eine Klasse und kein Mensch, also ist sie kein Element der Klasse der Menschen, d. h. kein Element ihrer selbst. Einige Klassen jedoch sind Elemente ihrer selbst: Die Klasse aller Klassen ist es wohl und ebenso ist es die Klasse aller Klassen mit mehr als 100 Elementen. Und desgleichen ist es die Klasse aller Nicht-Menschen, die Klasse aller und

nur der Dinge, die keine Menschen sind. Keine Klasse ist ein Mensch, also ist die Klasse der Nicht-Menschen kein Mensch und ist daher ein Element der Klasse der Nicht-Menschen.

Betrachten wir die Klasse aller Klassen, die nicht Elemente ihrer selbst sind. Nennen wir diese Klasse *R*. Die notwendige und hinreichende Bedingung für etwas, zu *R* zu gehören, ist, eine Klasse und nicht Element ihrer selbst zu sein. Frage: Ist *R* ein Element ihrer selbst?

Angenommen, sie ist es. Dann muss *R* die (notwendige) Bedingung dafür erfüllen, zu *R* zu gehören: nicht Element ihrer selbst zu sein. Wenn sie also Element ihrer selbst ist, dann ist sie nicht Element ihrer selbst.

Angenommen, sie ist es nicht. Als eine sich selbst nicht angehörende Klasse erfüllt sie nun die (hinreichende) Bedingung dafür, zu *R* zu gehören: nicht Element ihrer selbst zu sein. Wenn sie also nicht Element ihrer selbst ist, dann gehört sie zu *R* und ist also Element ihrer selbst.

Zusammenfassend: *R* ist genau dann Element ihrer selbst, wenn sie nicht Element ihrer selbst ist. Das ist ein Widerspruch.[1]

Ein Widerspruch bedeutet nicht notwendigerweise eine Paradoxie. Man erinnere sich an die Paradoxie des Barbiers aus der Einleitung. Wer rasiert den Barbier? Durch einen Gedankengang, der dem in der Ableitung von Russells Paradoxie ähnlich ist, finden wir, dass der Barbier sich genau dann selbst rasiert, wenn er es nicht tut.

1 Wie Russell gesehen hat, ist die Klassenparadoxie einer Paradoxie sehr ähnlich, die Eigenschaften betrifft. Die meisten Eigenschaften können nicht auf sich selbst angewandt werden. Die Eigenschaft, ein Mensch zu sein, ist eine Eigenschaft und kein Mensch, also trifft sie nicht auf die Eigenschaft zu, ein Mensch zu sein; d. h., sie ist nicht auf sich selbst anwendbar. Einige Eigenschaften jedoch sind auf sich selbst anwendbar: So die Eigenschaft, eine Eigenschaft zu sein, und ebenso die Eigenschaft, eine Eigenschaft zu sein, die auf mehr als 100 Dinge zutrifft, etc. Wie würden Sie den Widerspruch ausformulieren?

Wir antworten auf die Paradoxie des Barbiers, indem wir einfach sagen, dass es keinen solchen Barbier gibt. Warum sollten wir auf Russells Paradoxie nicht einfach antworten, es gibt keine solche Klasse? Der Unterschied ist folgender: Nichts spricht für die Annahme, es gäbe einen solchen Barbier, aber wir scheinen durch unsere Auffassung von Klassen zu der Existenz von *R* verpflichtet zu sein. Natürlich zwingt uns die Paradoxie zu akzeptieren, dass es keine solche Klasse geben kann. Das ist paradox, weil es zeigt, dass einige sehr zwingende Ansichten darüber, was Existenz für eine Klasse bedeutet, aufgegeben werden müssen.

Der erste Absatz dieses Abschnittes sollte die natürliche oder intuitive Auffassung von Klassen einführen, welche ich jetzt weiter explizit machen muss. Ich hatte gesagt, wenn Sokrates *ein Mensch ist*, dann ist er ein Element der Klasse der Menschen. Verwenden wir das Wort »Bedingung« für das, was zum Beispiel von dem soeben verwendeten kursiven Ausdruck gesagt wird. Ein Mensch zu sein, ist die Bedingung, und eine, die Sokrates erfüllt, während sie der Mont Blanc nicht erfüllt. Die natürliche Auffassung von Klassen beinhaltet folgendes Prinzip der Existenz einer Klasse:

KE Zu jeder verständlichen Bedingung gibt es eine Klasse. Alle und nur die Dinge, welche die Bedingung erfüllen (falls es welche gibt), sind ihre Elemente.

Der Bedingung entsprechend, ein Mensch zu sein, gibt es die Klasse der Menschen. Selbst wenn eine Bedingung widersprüchlich ist – etwa die Bedingung, sowohl quadratisch als auch nicht quadratisch zu sein –, so entspricht ihr doch eine Klasse; auch wenn es eine Klasse ohne Elemente ist (die leere oder »Nullklasse«), da nichts die Bedingung erfüllt.

KE scheint zu Russells Paradoxie zu führen. Es beinhaltet die Existenz einer Klasse wie *R*, wenn es die verständliche Bedingung gibt: eine Klasse zu sein, die nicht Element ihrer selbst ist. Die Bedingung erscheint verständlich; wir haben jedoch schon gesehen, dass es so eine Klasse wie *R* nicht geben kann.

Wir könnten diesen Punkt wie folgt auf formalere und klarere Weise ausdrücken. Verwenden wir »$\in$« als Abkürzung für »ist Element von« (und »gehört zu«). Wir können KE nun wie folgt reformulieren:

KE Für jede verständliche Bedingung F gibt es eine Klasse x, so dass gilt: Für jedes y, $y \in x$, genau dann, wenn y F erfüllt.

Für »x erfüllt F« können wir einfach schreiben »x ist F«, und für »genau dann, wenn« können wir »gdw« schreiben. Indem wir »R« für Russells Klasse setzen, »$\neg$« für »nicht« und »$\neg$ ein Element seiner selbst« für »F«, erzeugt KE:

Für jedes y, $y \in R$ gdw $\neg$ (y ist ein Element seiner selbst).

Ein Element seiner selbst zu sein, ist für ein y wohl, zu y zu gehören, wir können das Obige also wie folgt schreiben:

Für jedes y, $y \in R$ gdw $\neg\,(y \in y)$.

Was für alles gilt, muss auch für R gelten, also erhalten wir das explizit widersprüchliche:

RP $R \in R$ gdw $\neg\,(R \in R)$.

Es ist ein nahe liegender Gedanke, dass die Bedingung »nicht Element seiner selbst zu sein« nicht wirklich verständlich ist – und das ist es, was die meisten Erwiderungen auf Russells Paradoxie im Endeffekt nahe legen. Folgen wir diesem Vorschlag, dann kann KE erhalten werden, solange wir unsere Auffassung darüber, was eine Bedingung ausmacht, eng genug gestalten. Es ist jedoch alles andere als offensichtlich, was diese engere Auffassung sein soll, *falls* wir auch einige wohl bekannte Resultate aus der Mathematik erhalten wollen. Insbesondere kommen Gedankengänge und Annahmen, die denen in der Ableitung von Russells Paradoxie sehr ähnlich sind, auch in einem berühmten Beweis von Cantor vor, den ich nun darlegen werde. Das Studium von Cantors Beweis war es, was Russell zur Entdeckung der Paradoxie führte. Es ist sehr schwer zu sehen, wie man die Paradoxie blockieren könnte, während man den Beweis gestattet.

Zu beweisen ist, dass die Potenzmenge jeder Klasse mehr Elemente hat als die Klasse selbst. (Für eine Definition von Potenzmenge siehe die Literaturhinweise.) Cantors Beweis könnte nicht-formal wie folgt skizziert werden:

(1) Eine Klasse muss *mindestens* so viele Teilklassen wie Elemente haben, da bei jedem Element die Klasse, zu der es alleine gehört, eine Teilklasse ist.
(2) Es gibt also entweder so viele Teilklassen wie Elemente oder mehr.
(3) Angenommen, es sind gleich viele. Das bedeutet, es gibt eine eins-zu-eins Funktion f, die Elemente der Klasse den Teilklassen der Klasse zuordnet.
(4) Nun bilden wir die folgende Teilklasse T: $x \in T$ gdw $\neg\, [x \in f(x)]$. Der Annahme in (3) zufolge gilt für ein α: $T = f(\alpha)$. Indem wir die Definition von T anwenden, erhalten wir:
$\alpha \in T$ gdw $\neg\, [\alpha \in f(\alpha)]$.

Also, da $\mathrm{T} = f(\alpha)$:

(*) $\alpha \in T$ gdw $\neg\, (\alpha \in T)$.
(5) Dieser Widerspruch zeigt, dass wir die Annahme in (3) verwerfen müssen. Also müssen wir die einzige in (2) zur Verfügung stehende Alternative annehmen: Eine Klasse hat mehr Teilklassen als Elemente.

In (4) haben wir einen Widerspruch, der dem in Russells Paradoxie ähnelt (man vergleiche RP und die Zeile mit dem Sternchen), hier jedoch zu ernsthaften und informativen Zwecken im Beweis verwendet. Der Beweis geht von der Annahme aus, *wenn* es eine Funktion f gibt, dann gibt es eine Teilklasse T. Wenn wir gleichzeitig Cantors Beweis erhalten und den gegenwärtigen Vorschlag verwenden wollen, dass eine allzu liberale Auffassung von KE an der Paradoxie schuld ist, dann müssen wir eine Einschränkung des Begriffes der »Bedingung« finden, welche mittels KE die hypothetische Existenz von T gestattet, die Existenz von R jedoch verbietet.

Damit die Angelegenheit außerdem noch philosophisch befriedigend ist, muss es eine philosophische Rechtfertigung für die Einschränkung geben, welche es uns ermöglicht, den Ursprung der Paradoxie zu verstehen und zu meinen, wir hätten etwas Besseres als ein *ad hoc* Blockademanöver. Es ist Russell zugute zu halten, dass er versucht hat, genau dies mit seinem Prinzip Teufelskreis (PT) zur Verfügung zu stellen. Seine Idee ist, dass die Bedingung in der Spezifikation von R zirkulär und deshalb nicht verständlich ist. Er meinte auch, eine Erklärung mittels des PT würde eine Reihe von anderen Paradoxien beseitigen, einschließlich (und vor allem) der Lügnerparadoxie. Ich werde das PT (in Abschnitt 5.8 und 5.9) nach der Beschreibung der Lügnerparadoxie (beginnend mit Abschnitt 5.2) behandeln.

Für Russell bot das PT die philosophische Motivation für seine Typentheorie, derzufolge Klassen derart in einer Hierarchie angeordnet sind, dass jede Klasse auf einer höheren Stufe steht als ihre Elemente. Die Theorie stellt sicher, dass keine Klasse sich selbst angehört: Kein Ausdruck der Form $x \in x$ zählt als sinnvoll. Vereinfachte Versionen von Russells Typentheorie waren die dominante Tradition in der mathematischen Arbeit zu Klassen. Man muss jedoch unterscheiden zwischen der Typentheorie auf der einen Seite – ein Mittel, das geschaffen wurde, um sicherzustellen, dass Paradoxien mathematisches Arbeiten nicht behindern – und der Rechtfertigung für eine solche Theorie auf der anderen Seite, wie jene, die, Russell zufolge, von PT geboten wird. Die Rechtfertigung sollte uns dabei helfen, zu verstehen, was an Klassen eine Typentheorie erfordert. Die Suche nach einem Verständnis dieser Art ist ein entschieden philosophisches Projekt, von dem sich der praktische Mathematiker mit Recht abwenden darf.

5.2 Der Lügner: Semantische Mängel

Der Stoff der folgenden vier Abschnitte ist gefährlich. (Erinnern Sie sich an das in der Einleitung erwähnte Schicksal von Philetas?)

Eine relativ späte Version der Lügnerparadoxie erscheint im Brief des Apostel Paulus an Titus (Tit. 1,12f.).[2] Jene Version beinhaltet die Insel Kreta, den Begriff der Lüge, und Lügen beinhaltet eine Absicht zu täuschen. Diese Aspekte sind für die Paradoxie irrelevant. Wenn wir solche irrelevanten Dinge beseitigen, erhalten wir so etwas wie:

Was ich jetzt sage, ist falsch.

Die einfachste Version von allen, welche den Ausgangspunkt unserer Diskussion bilden wird, ist:

L_1 L_1 ist falsch.

Hier haben wir einen L_1 genannten Satz, der von sich selbst sagen soll, dass er falsch ist. Man kann folgendermaßen etwas scheinbar Paradoxes ableiten. Angenommen, der Satz ist wahr, dann ist er so, wie er sagt, dass er ist: falsch. Also ist er falsch. Nehmen wir jedoch an, er sei falsch. Nun, *falsch* ist genau das, was er sagt, dass er ist, und ein Satz, der sagt, wie es ist, ist wahr. Also ist er wahr. Zusammengefasst: Wenn L_1 wahr ist, ist er falsch, und wenn er falsch ist, ist er wahr.

2 Es ist nicht klar, dass der Apostel zusätzlich zu den moralischen Problemen noch irgendwelche logischen sieht. Der betreffende Text lautet wie folgt:
12. Es hat einer von ihnen gesagt, ihr eigener Prophet: Die Kreter sind immer Lügner, böse Tiere und faule Bäuche.
13. Dieses Zeugnis ist wahr. Aus diesem Grund weise sie scharf zurecht, damit sie gesund werden im Glauben.
Die Version des Paulus hängt von der Annahme ab, dass alle anderen Kreter Lügner sind. Konstruieren Sie ein explizites Argument für den Widerspruch (eventuell dem nachgebildet, was unten für L_1 gegeben ist), welches diese Abhängigkeit klarmacht.

Ist das paradox? Vielleicht hört es sich so an, aber sehen wir genauer hin. Wir haben zwei konditionale Aussagen:

(a) Wenn L_1 wahr ist, dann ist er falsch.
(b) Wenn L_1 falsch ist, dann ist er wahr.

Wir setzen voraus, dass alles, was falsch ist, nicht wahr ist und alles, was wahr ist, nicht falsch; also erzeugen (a) und (b):

(a') Wenn L_1 wahr ist, dann ist er nicht wahr.
(b') Wenn L_1 falsch ist, dann ist er nicht falsch.

Wenn etwas seine eigene Negation impliziert, dann können wir auf diese Negation schließen. (Dieses Prinzip heißt *consequentia mirabilis*. Es besagt die Gültigkeit der Ableitung: $\mathrm{A} \rightarrow \neg\, \mathrm{A} \models \neg\, \mathrm{A}$.) Sowohl (a') als auch (b') bieten Zugang zu dem Prinzip. Ersteres versichert, »L_1 ist wahr« impliziere seine eigene Negation, also sagt uns das Prinzip, wir können schließen, dass L_1 nicht wahr ist. Das zweite ermöglicht uns, in genau paralleler Weise zu schließen, dass L_1 nicht falsch ist. Fassen wir das wie folgt zusammen:

G L_1 ist weder wahr noch falsch.

Ist *das* paradox? Nicht, solange wir nicht einen unabhängigen Grund für die Annahme haben, dass L_1 entweder wahr oder falsch ist. Wir könnten zum Beispiel in der Lage sein, ein *Bivalenzprinzip* zu rechtfertigen, etwa in dem Sinne, dass *jeder* Satz, also insbesondere auch L_1, entweder wahr oder falsch ist. Außerdem könnten wir auch einfach G *akzeptieren* und sagen, dass L_1 in einer *Lücke* zwischen Wahrheit und Falschheit liegt (daher »G«). Damit wäre an sich noch keine vollständige Auffassung der Paradoxie geboten, denn es wären noch allgemeine Prinzipien zu entdecken, die erklären, warum L_1 weder wahr noch falsch sein soll. Eine Akzeptanz von G würde aber immerhin die allgemeine Herangehensweise retten.

Wir könnten G dann nicht akzeptieren, wenn es einen unwiderstehlichen Grund für die Annahme gäbe, dass L_1 entweder wahr oder falsch sein muss. Könnte es einen unwiderstehlichen Grund für die Annahme irgendeines Bivalenzprinzips geben? Die im vorigen Absatz gegebene Version ist zweifellos nicht wahr. Fragen werden in Sätzen ausgedrückt, aber keine Frage ist entweder wahr oder falsch. Angenommen also, wir schränken das Prinzip auf deklarative Sätze im Indikativ ein. Dennoch gibt es angebliche Gegenbeispiele, etwa:

> Sie haben aufgehört, Ihre Frau zu schlagen.

Wenn Sie Ihre Frau nie geschlagen haben, dann ist der Satz sicherlich nicht wahr; aber zu sagen, er sei falsch, oder zu sagen, Sie hätten nicht aufgehört, Ihre Frau zu schlagen, gibt wohl zu verstehen, dass Sie sie nach wie vor schlagen. Man betrachte ferner den Fall, in dem jemand sagt:

> Der Elefant greift gleich an

während kein Elefant in der Nähe ist. Wir können den Satz sicherlich nicht als wahr betrachten; aber können wir ihn als falsch ansehen? Wenn wir das täten, sollte der folgende Satz dann wahr sein?

> Der Elefant greift *nicht* gleich an.

Wenn es jedoch keinen Elefanten gibt, dann ist das ein ebenso schwacher Kandidat für Wahrheit wie der vorige.

Trotz der scheinbaren Gegenbeispiele ist es schwer, die Attraktion des Gedankens nicht zu spüren, dass irgendein Prinzip der Bivalenz, zweifellos angemessen raffiniert, korrekt sein müsste. Die zugrunde liegende Idee könnte folgendermaßen ausgedrückt werden: Jede nicht-mangelhafte Darstellung dessen, wie sich die Dinge in der Welt verhalten, muss entweder richtig oder unrichtig sein, wahr oder falsch. Einige Sätze, wie Fragen und Befehle, sind nicht dazu geschaffen, die Welt darzustellen, es stellt sich also nicht die Frage, ob sie diese richtig oder unrichtig darstellen. Anderen Sätzen, auch wenn sie dazu geschaffen sind, die Welt darzustellen, gelingt es aufgrund eines semantischen Mangels nicht, überhaupt als

Darstellungen zu gelten, richtig oder unrichtig. Der Fall des fehlenden Elefanten ist ein angebliches Beispiel. Damit dieser Satz die Welt überhaupt abbilden kann, muss er sich auf einen Elefanten beziehen (das ist jedenfalls plausibel anzunehmen). Da ihm dies nicht gelingt, zählt er als semantisch mangelhaft und daher als weder wahr noch falsch.

Zusammengefasst ist die natürlichste und unmittelbarste Reaktion auf die Lügnerparadoxie, den Gedankengang zu akzeptieren, der zu dem Schluss kommt, der paradoxe Satz L_1 sei weder wahr noch falsch.[3] Da es schwer fällt anzunehmen, ein semantisch nicht-mangelhafter Satz könnte nicht entweder wahr oder falsch sein, bringt dieser Ansatz die Verpflichtung mit sich zu erklären, worin der Mangel von L_1 besteht. Es würde überhaupt nichts erklären, wenn man sagen würde, der Mangel bestehe im Potential des Satzes für Paradoxien, denn es ist eben jenes Potential, das wir verstehen müssen.

Die meisten Auffassungen der Lügnerparadoxie trachten danach, plausible allgemeine Prinzipien aufzustellen, denen zufolge die Lügner-Sätze mangelhaft sind. Wir werden einige von diesen in den folgenden Abschnitten in Betracht ziehen.

5.3 Fundierung und Wahrheit

Ein Ansatz zur Identifikation eines semantischen Defekts in L_1 beginnt mit der Vorstellung, die Wahrheit eines Satzes müsse in irgendetwas außerhalb des Satzes selbst fundiert sein. Wir können den Gedanken lebendiger gestalten, wenn wir uns vorstellen, wie man jemanden in den Begriff der Wahrheit einführen könnte.

Wir könnten annehmen, der Lernende versteht gut

3* Zeigen Sie, wie die Prinzipien dieses Abschnittes anscheinend dazu verwendet werden können, abzuleiten, dass L_1 sowohl wahr als auch falsch ist. Diese Ableitung zeigt, dass man L_1 nicht als eine Basis für einen direkten Beweis von G ansehen kann.

Deutsch, aber nicht das Wort »wahr«. Wir könnten versuchen, ihm den Begriff der Wahrheit durch folgende Anweisung zu erklären:

> Sie sollten einen Satz wahr nennen, gdw Sie gewillt wären, ihn auszusagen.

(»gdw« ist die Abkürzung für »genau dann, wenn«.) Der Lernende könnte diese Erklärung dazu gebrauchen, um zum Beispiel auf »Schnee ist weiß« zu erwidern »Wahr!« und auf »Gras ist rot« »Nicht wahr!«. Er könnte jedoch die Erläuterung zunächst nicht dazu verwenden, herauszufinden, wie er auf einen Satz wie diesen reagieren soll:

> (1) »Schnee ist weiß« ist wahr.

Solange er »wahr« nicht verstanden hat, kann er nicht wissen, was es bedeuten würde, gewillt zu sein, diesen Satz für eine Aussage zu verwenden. In einem späteren Stadium, wenn ihm bewusst wird, dass er auf »Schnee ist weiß« mit »Wahr« antworten soll, wird er verstehen, dass er (1) zustimmen sollte, und also verstehen, dass (1) selbst etwas ist, worauf das Wort »wahr« zutrifft. Das Bild gleicht jemandem, der eine Leiter hochsteigt: An der Basis sind Sätze, die das Wort »wahr« nicht enthalten. Er kann lernen, das Wort auf Sie anzuwenden. Während er das tut, kann er verstehen lernen, wie man das Wort auf Sätze wie (1) von der nächsten Sprosse darüber anwendet, auf Sätze, die »wahr« von Sätzen der Basis aussagen. Er kann sich die Leiter unendlich weit hinaufarbeiten. Wenn *S* ein Satz ist, der »wahr« nicht enthält, dann kann er diesen Prozess anwenden, um jeden Satz der Form

> ... »*S* ist wahr« ... ist wahr

zu verstehen, wo die zweite Auslassung für jede Anzahl weiterer Vorkommnisse von »ist wahr« steht.[4]

Zu lernen, wie man den Begriff der Wahrheit anwendet,

4 Wofür steht die erste Auslassung?

erfordert, dass es Sätze gibt, die nicht selbst diesen Begriff enthalten: die *Basis*sätze. Diese Lernsituation spiegelt die mutmaßliche metaphysische Tatsache wider, dass *Wahrheit von etwas außerhalb von sich selbst abhängt*. Die Behauptung, L_1 sei weder wahr noch falsch, ließe sich mit der Begründung verteidigen, dass L_1 jene Tatsache nicht respektiere. Ich werde versuchen, das zu erklären.

Ob »Schnee ist weiß« wahr ist oder nicht, hängt davon ab, ob Schnee weiß ist oder nicht. Ob etwas wahr ist oder nicht, hängt in diesem einfachen Fall von einer Tatsache ab, die ohne Einbeziehung des Begriffes der Wahrheit ausgedrückt werden kann: ob Schnee weiß ist oder nicht. Dies ist ein Beispiel dafür, was ich damit meine, dass Wahrheit von etwas außerhalb von sich selbst abhängt. In komplexeren Fällen ist die Abhängigkeit weniger direkt. Fassen wir zum Beispiel (1) wieder ins Auge (»Schnee ist weiß« ist wahr). Ob (1) wahr ist, hängt davon ab, ob »Schnee ist weiß« wahr ist oder nicht. Dies wiederum hängt davon ab, ob Schnee weiß ist. Ob (1) wahr ist oder nicht, hängt also mit einem Zwischenschritt davon ab, ob Schnee weiß ist. Am Ende kommen wir zu einer Frage zurück, die Wahrheit nicht mit enthält. In diesem Gedankengang gehen wir die Leiter abwärts zur Basis; in der Betrachtung des Lernens gingen wir aufwärts von der Basis aus – die gleiche Leiter, verschiedene Richtung.

Um diesen Vorschlag zu bekräftigen, betrachte man folgende Reihe von Sätzen:

> (S2) (S1) ist wahr.
> (S3) (S2) ist wahr.
> (S4) (S3) ist wahr.

Macht diese Reihe wirklich irgendeinen Sinn für uns? Alles hängt davon ab, was (S1) ist. Ist es zum Beispiel »Schnee ist weiß«, dann gibt es kein Problem: Wir erreichen die Basis. Wir würden diese jedoch nie erreichen, wenn (S1) zum Beispiel wäre:

(S1) (S4) ist wahr.

Hier geht die Wahrheit im Kreise, ohne je den Boden zu berühren. In diesem Fall müssen wir sagen, dass keiner der Sätze wahr ist und auch keiner falsch – aus demselben Grund (Kripke, 1975).

Diese Denkweise liefert einen allgemeinen Grund dafür, G zu akzeptieren. (L_1 ist weder wahr noch falsch). Er ist weder wahr noch falsch, weil es kein Erreichen der Basis gibt. Man kommt nicht zu einer Tatsache, die Wahrheit nicht enthält und von der die Wahrheit oder Falschheit von L_1 abhängen könnte. Wir kommen immer zu L_1 zurück, was kein Basissatz ist. Zusammenfassend: Das Problem an L_1 ist, er ist *unfundiert*.

Dieselbe Auffassung passt auch gut auf

W_1 W_1 ist wahr.

Hier haben wir einen Satz, der von sich selbst sagt, dass er wahr ist. Er ist nicht paradox: Die Annahme, er sei wahr, führt nicht zu dem Schluss, dass er falsch ist. Die Annahme, er sei falsch, führt nicht zu dem Schluss, dass er wahr ist. Dennoch scheint intuitiv etwas faul an W_1, und L_1 teilt diesen Mangel. Die soeben gegebene Auffassung will das Problem benennen: W_1 ist unfundiert. Ebenso wie L_1 hat er keinen Kontakt mit einer nicht Wahrheit enthaltenden Basis, daher sind beide Sätze weder wahr noch falsch.

So kann G verteidigt werden. Wir können Gründe dafür angeben zu meinen, L_1 sei weder wahr noch falsch, die von der drohenden Paradoxie unabhängig sind: Es ist unfundiert. Aber selbst wenn all dies akzeptiert wird, die Paradoxie bleibt.

5.4 Der Verstärkte Lügner

G sagt, dass L_1 weder wahr noch falsch ist, und akzeptiert also den Gedankengang, den wir zu Beginn des Abschnitts 5.2 in Erwägung gezogen hatten. G scheint jedoch selbst eine Paradoxie zu gestatten.

G impliziert, dass L_1 nicht falsch ist. Das aber ist die Negation von L_1 selbst. Also impliziert G:

nicht-L_1 L_1 ist nicht falsch.

Also ist nicht-L_1 wahr (unter Verwendung des Prinzips, dass alles, was einen Satz impliziert, die Wahrheit dieses Satzes impliziert). Dies wiederum impliziert, dass L_1 falsch ist (unter Verwendung des Prinzips, dass jeder Satz, dessen Negation wahr ist, falsch ist). Also scheint G einen Widerspruch zu implizieren: dass L_1 nicht falsch und dass L_1 falsch ist.[5] G kann also keine Lösung der Paradoxie darstellen.

Eine damit zusammenhängende Schwierigkeit ist, dass G nicht in der Lage ist, einem ähnlichen paradoxen Satz gerecht zu werden:

L_G L_G ist entweder falsch oder es ist weder wahr noch falsch.

Wir können wie folgt argumentieren: Angenommen, L_G ist weder wahr noch falsch; dann ist er wahr (da eine Oder-Aussage wahr ist, wenn eine ihrer Alternativen wahr ist), und also ist er entweder wahr oder falsch. Wir können dann so argumentieren, wie wir es bei L_1 getan hatten, um zu zeigen, dass er weder wahr noch falsch ist. Wenn wir die Ergebnisse kombinieren, zeigen wir, dass er weder wahr noch falsch ist und also entweder wahr oder falsch.

Die einfachste Möglichkeit zu verstehen, was in Argumenten dieser Sorte passiert, ist, einen weiteren paradoxen Satz zu betrachten:

5* Zeigen Sie, wie man (scheinbar) *ableiten* kann, dass L_1 nicht falsch ist, ohne an G zu appellieren.

L_2 L_2 ist nicht wahr.

Angenommen, L_2 ist wahr. Dann ist er so, wie er sagt, d. h. nicht wahr; also ist er nicht wahr. Nun, *nicht wahr* ist eben, was er sagt, dass er ist, und ein Satz, der sagt, wie es ist, ist wahr; also ist er wahr. Zusammenfassend: Wenn L_2 wahr ist, dann ist er nicht wahr, wenn er nicht wahr ist, dann ist er wahr.

Das scheint ein echter Widerspruch zu sein, und zwar einer, der von G nicht bereinigt werden kann. Wenn L_2, wie G behauptet, weder wahr noch falsch ist, dann ist er insbesondere nicht wahr. Der soeben vorgestellte Gedankengang hingegen soll zeigen, dass man diese Behauptung widerlegen kann: Wenn L_2 nicht wahr ist, dann ist er wahr (denn *nicht wahr* zu sein, ist eben das, was er von sich selbst sagt).

Ebenso wenig könnten wir G zu einer Ansicht modifizieren wie:

G' L_2 ist weder wahr noch nicht wahr.

Zum Ersten scheint das ein Widerspruch zu sein. Der übliche Gedankengang würde es uns gestatten, aus G' zu schließen, dass L_2 wahr *und* nicht wahr ist.[6] Zum Zweiten impliziert G' ebenso wie G, direkt, dass L_2 nicht wahr ist: Es impliziert den paradoxen Satz selbst.

L_2 und der dazugehörige Gedankengang werden gelegentlich als der »Verstärkte Lügner« bezeichnet. Eine Standardauffassung ist, dass er die Unangemessenheit von Theorien zur Lösung von Lügnerparadoxien zeigt, wie etwa der auf dem Begriff der Fundierung basierenden. Allgemeiner könnte man meinen, er zeige, dass jedes Herangehen, das die Paradoxie dadurch zu lösen versucht, dass es einen semantischen

6 Der Gedankengang hängt von der Äquivalenz von
weder *P* noch *Q*
mit
sowohl nicht-*P* als auch nicht-*Q*
ab. Von welchem anderen Prinzip hängt der Gedankengang ab?

Mangel in L_2 ausmacht, zum Scheitern verurteilt ist, denn was semantisch mangelhaft ist, ist nicht wahr.

Lassen wir diese Behauptung zunächst so stehen. Ich wende mich nun einem auf Tarski zurückgehenden Ansatz zur Lösung dieser Paradoxien zu, für den der Verstärkte Lügner kein besonderes Problem ist, kein Problem, das nicht bereits durch den gewöhnlichen Lügner gestellt würde. Dieser Ansatz findet auch an Sätzen wie L_1 (»L_1 ist falsch«) und L_2 (»L_2 ist nicht wahr«) etwas semantisch mangelhaft, aber von einer ganz anderen Art.

5.5 Ebenen

Um oberflächlich unannehmbare Schlüsse aus L_1 und L_2 abzuleiten, verließen wir uns auf zwei Prinzipien:

Wenn ein Satz wahr ist, dann sind die Dinge so, wie er sagt, dass sie sind;
wenn die Dinge so sind, wie ein Satz sagt, dass sie sind, dann ist der Satz wahr.

Tarski betonte den Zug der Wahrheit, den diese Prinzipien erfassen. Er drückte es auf eine etwas formalere Weise aus. Wir verwenden σ für den Namen eines Satzes und p für einen Satz. Tarski behauptete, wir müssten dann für jede akzeptable Sprache jede Instanz von

W σ ist wahr, gdw p

akzeptieren, vorausgesetzt, der von σ benannte Satz bedeutet das Gleiche wie jener Satz, der p ersetzt. Im Grenzfall kann dies derselbe Satz sein; also ist eine Instanz von W (»›Schnee ist weiß‹« für σ gesetzt und »Schnee weiß ist« für p).

»Schnee ist weiß« ist wahr, gdw Schnee weiß ist.

W mag äußerst platt erscheinen, aber der Verstärkte Lügner zeigt, ganz im Gegenteil, dass es widersprüchliche Instanzen hat. Indem wir »L_1« für σ setzen und »L_2 ist nicht wahr« für p, bekommen wir:

(*) L_2 ist wahr, gdw L_2 nicht wahr ist.

Da L_2 lautet, »L_2 ist nicht wahr«, erfüllt (*) wohl die Anforderung, dass der von L_2 genannte Satz (nämlich »L_2 ist nicht wahr«) dasselbe bedeutet wie der Satz, der p ersetzt (nämlich »L_2 ist nicht wahr«).

Ein Aspekt des durch den Lügner gestellten Problems ist, dass die scheinbare Plattitüde W durch einen scheinbar korrekten Gedankengang zu dem widersprüchlichen (*) führt. Tarskis Erwiderung lautet, dass der gewöhnliche Wahrheitsbegriff, jener, den wir jeden Tag verwenden, inkohärent ist und verworfen werden muss. Tarski zufolge muss er durch eine Reihe von Wahrheitsbegriffen in hierarchischer Ordnung ersetzt werden, die jeweils in einer nicht-natürlichen Sprache ausgedrückt werden (d. h. in einer Sprache, die nicht natürlich entstanden ist).

Angenommen, eine gewisse Sprache Σ_0 enthält ein Prädikat W_1, das auf alle und nur die wahren Sätze von Σ_0 zutrifft. Angenommen, Σ_0 enthält auch einen Satz σ, der von sich selbst sagt, dass er nicht W_1 ist. Dann haben wir eine Version des Lügners, wenn wir W zugestehen: Wenn W_1 auf σ zutrifft, dann sagt σ wahrerweise, dass W nicht auf es zutrifft – in welchem Fall W_1 in Anbetracht dessen, was er sagt, nicht auf ihn zutrifft. Wenn aber W_1 nicht auf ihn zutrifft, dann ist er wahr, da es das ist, was er sagt, und also trifft W_1 auf ihn zu. Tarski nahm den Widerspruch als Widerlegung der Annahme, dass σ zu Σ_0 gehört. Die natürliche Erklärung dafür ist, dass W_1 kein Ausdruck von Σ_0 ist. Also gehört kein Satz zu Σ_0, wenn er W_1 enthält. Das blockiert die Paradoxie im folgenden Sinne: Da die vorgeschlagene Sprache kein Prädikat enthält, das genau auf ihre wahren Sätze zutrifft, ist sie eine Sprache, in der der paradoxe Satz nicht formuliert werden kann. Man kann die Worte hinschreiben, aber von ihnen wird behauptet, sie hätten keinen Wert: Sie sind semantisch gänzlich mangelhaft.

Wir können eine Sprache erweitern, indem wir neue Ausdrücke hinzufügen. Insbesondere könnten wir Σ_0 – von dem wir annehmen, sie enthält kein Vorkommnis von W_1 – erweitern, indem wir W_1 hinzufügen. Wir könnten die neu gebildete Sprache Σ_1 nennen: Sie enthält alle Sätze von Σ_0 sowie alle Sätze, die aus diesen mittels W_1 gebildet werden können; sie enthält also σ. Eine Paradoxie ist nach wie vor vermieden: σ gehört nicht zu Σ_0, und da W_1 nur für Sätze aus Σ_0 definiert ist, steht nicht zur Debatte, ob W_1 auf σ zutrifft. Der Ausdruck σ (= »σ ist nicht W_1«) gehört nicht zu Σ_0 und ist also keiner, dem W_1 sinnvoll zugesprochen oder abgesprochen werden könnte.

Das heißt nicht, dass es kein Prädikat gäbe, was auf eben die Sätze von Σ_1 zutrifft. Es gibt eines: Nennen wir es W_2.[7] Aus inzwischen bekannten Gründen jedoch kann es nicht zu Σ_1 gehören.[8] Allgemein kann ein Prädikat W_n nicht zu einer Sprache Σ_{n-1} gehören, sondern nur zu einer Sprache von mindestens Ebene n.

In keiner der Sprachen in Tarskis Hierarchie kann ein paradoxer Lügner-Satz formuliert werden. Wie soll das eine »Lösung« für die Paradoxie bieten? Die Paradoxie tritt in unserer Sprache auf, also muss eine angemessene Auflösung etwas über unsere Sprache sagen und nicht bloß einen Ersatz anbieten.

Was Tarski über unsere Sprache sagt, ist, dass die Lügnerparadoxie ihre Inkohärenz zeige. Wir müssen unseren gegenwärtigen, aber inkohärenten Wahrheitsbegriff durch eine Familie von Begriffen ersetzen, die jeweils in der soeben beschriebenen Weise an eine Ebene in der Hierarchie gebunden sind. Viele haben nach etwas weniger Radikalem gesucht, nach einer Lösung, die mehr von unserem gewöhnlichen Sprechen und Denken erhält.

Eine solche weniger radikale Erwiderung verwendet einen

7 Ist σ W_2?

8 Wie führt die Annahme, dass W_2 zu Σ_1 gehört, zu einer Paradoxie?

Begriff von Hierarchie im Sinne Tarskis, behauptet aber, dieser sei bereits in unserem tatsächlichen Gebrauch von »wahr« enthalten. Anders als Tarskis Auffassung, welche meinte, die gewöhnliche Sprache sei unheilbar mangelhaft, behauptet diese Alternative, die Mängel seien bloßer Schein: Die darunter liegende Wirklichkeit sei, dass wir bereits eine Hierarchie von Wahrheitsbegriffen à la Tarski verwenden.

Eine wesentliche Schwierigkeit bei diesem Vorschlag besteht darin, dass nichts in unserem Sprachgebrauch die angemessene Empfindlichkeit für tarskische im Voraus festgelegte Ebenen wiederzugeben scheint. Angenommen, ich sage zum Beispiel:

Was Sie soeben gesagt haben, ist nicht wahr.

Auf den ersten Blick könnte jeder, mich eingeschlossen, recht gut wissen, was ich gesagt habe, ohne zu wissen, was Sie gesagt haben. (Stellen Sie sich ein Spiel analog zu Stein-Schere-Licht vor: Ich muss raten, ob das, was Sie gerade aufgeschrieben haben, wahr oder falsch ist.) Einer hierarchischen Ansicht zufolge, nach der Ebenen im Voraus festgelegt sind, bestimmt irgendetwas in meinem Gebrauch des Satzes eine Verknüpfung von »wahr« mit einer Ebene. Die normale (Grund-)Ebene wäre wohl 1. Wenn Sie »Schnee ist weiß« gesagt haben, gibt es kein Problem; angenommen aber, Sie haben gesagt: »Was M. S. sagen wird, ist wahr.« Nach der gegenwärtigen Theorie verlangt die Verständlichkeit meiner Äußerung, dass sich mein »wahr« auf einer höheren Ebene befindet als Ihres. Wenn aber meine Äußerung verständlich ist, ohne zu wissen, was Sie gesagt haben, dann muss Ihre Wahrheitsebene unabhängig vom Inhalt dessen festgelegt werden, was Sie gesagt haben. Das deutet darauf hin, dass der Versuch unplausibel ist, diese Art der Hierarchie-Antwort auf natürliche Sprache anzuwenden. (Vgl. jedoch Burge, 1979.)

Bisher haben wir zwei zentrale Möglichkeiten erörtert, die Behauptung einzulösen, Sätze der Lügnerparadoxie seien semantisch mangelhaft. Die eine verwendete den Begriff der

Fundierung, dank dessen es Hoffnung für eine Verteidigung der Auffassung zu geben schien, L_1 sei weder wahr noch falsch; auch wenn die Hoffnung, dies würde zu einer Lösung aller Versionen des Lügners führen, anscheinend durch den Verstärkten Lügner (L_2: L_2 ist nicht wahr) zunichte gemacht wurde. Die andere war Tarskis Behauptung, dass jeder nicht-hierarchische Begriff von Wahrheit inkohärent ist. Der Verstärkte Lügner stellt für diese Auffassung kein besonderes Problem dar.[9] Sie hat jedoch ihre Schwierigkeiten. Es scheint zu radikal, unseren gewöhnlichen Wahrheitsbegriff zu verwerfen; andererseits erscheint es nicht korrekt anzunehmen, unser Begriff enthalte implizit bereits die geforderte Trennung in Ebenen. Wo sonst sollte man nach einer Auffassung der semantischen Mängel von Lügner-Sätzen suchen?

5.6 Selbstbezüglichkeit

Es ist nahe liegend zu meinen, irgendetwas am selbstbezüglichen Charakter der paradoxen Lügner-Sätze sei die Hauptquelle ihrer paradoxen Natur. An diesem Gedanken mag et-

9 Warum nicht? Sie könnten etwa antworten wollen, indem Sie eine oder beide Versionen des folgenden Gedankenganges kritisieren:
Version 1:
Selbst wenn Ebenen der Wahrheit explizit gemacht werden, wie Tarski es fordert, können wir einen Lügner-Satz formulieren, z. B.:
L_N: L_N ist nicht wahr$_n$.
Wenn das mangelhaft ist, weil es die Forderung nach Ebenen nicht erfüllt, dann ist es nicht wahr$_n$, aber da es das ist, was es sagt, muss es doch wahr$_n$ sein.
Version 2:
Ein Satz, der Ebenen durchbricht, ist semantisch mangelhaft und also nicht wahr; man kann also immer einen Verstärkten-Lügner-Satz konstruieren, um einen Ansatz über Ebenen zu widerlegen. (Vergleiche das gegen Ende von Abschnitt 5.4 erwähnte Argument.)

was dran sein, aber so ausgedrückt ist er zunächst sowohl unrichtig als auch unangemessen.

Er ist unrichtig, weil man Lügnerparadoxien konstruieren kann, ohne einen Satz zu verwenden, der auf sich selbst Bezug nimmt. Ein Beispiel für dieses Phänomen beinhaltet Lügner-Zirkel wie den folgenden:

(A) (Von α am Montag gesagt:) Alles, was β am Dienstag sagen wird, ist wahr.

(B) (Von β am Dienstag gesagt:) Nichts von dem, was α am Montag gesagt hat, ist wahr.

Wenn α und β nichts anderes als (A) bzw. (B) am Montag bzw. Dienstag gesagt haben, dann haben wir eine Paradoxie im Wesentlichen von der Art des Lügners. Angenommen, (B) ist wahr; dann ist (A) nicht wahr und β wird am Dienstag nichts Wahres sagen. Da β nur (B) sagt, ist (B) nicht wahr. Wenn (B) also wahr ist, dann ist es nicht wahr. Angenommen, (B) ist nicht wahr; dann hat α am Montag etwas Wahres gesagt. Da α nur (A) gesagt hat, ist (A) wahr, d. h., alles, was β am Dienstag sagen wird, ist wahr. Das schließt (B) ein, folglich ist (B) wahr. Wenn (B) also wahr ist, dann ist es nicht wahr.

Keiner der Sätze in der Geschichte bezieht sich buchstäblich auf sich selbst. Es gibt vielmehr eine Art Zirkel, vielleicht sollten wir also eher von »zirkulärem Bezug« anstatt von Selbstbezüglichkeit sprechen. Da die Zirkularität genau genommen gar keinen Bezug beinhaltet, sondern vielmehr Quantifikation, wird es nach wie vor sicherer sein, nur von Zirkularität zu sprechen.

Wir könnten die Geschichte von (A) und (B) erweitern, indem wir uns eine dritte Äußerung vorstellen:

(C) (Von γ am Dienstag gesagt:) Nichts von dem, was α am Montag gesagt hat, ist wahr.

Die Tatsache, dass β und γ genau denselben Satz verwenden, aber nur einer von ihnen auf die relevante Weise zirkulär ist, zeigt, dass Zirkularität keine Eigenschaft von Sätzen an sich

ist. Bedeutung zu haben oder keine Bedeutung zu haben, ist eine Eigenschaft von Sätzen. Da an (C) nichts paradox ist, gibt es keinen Grund zu sagen, er habe keine Bedeutung, und da (B) derselbe Satz ist, folgt, dass die durch Zirkularität verhinderte Eigenschaft nicht ist, Bedeutung zu haben. Wir benötigen einen ausgefeilteren Begriff, der für den Gebrauch empfindlich ist, der von einem Satz zu einer bestimmten Gelegenheit gemacht wird. Ein solcher Begriff ergibt sich ganz natürlich aus einer Betrachtung der Indexikalität – und die verbleibenden beiden Erwiderungen auf den Lügner, die ich in Erwägung ziehen will, behaupten beide, ein Element der Indexikalität in Gedankengängen zu entdecken, die in Verbindung mit dem Verstärkten Lügner stehen. Die eine lokalisiert Indexikalität in der spezifischen Selbstbezüglichkeit, die beim Lügner auftritt, die andere (kurz erwähnt im letzten Absatz von Abschnitt 5.8) lokalisiert sie im Prädikat »wahr«.

Fassen wir zusammen: Wenn wir Selbstbezüglichkeit als den Schurken des Stückes überführen wollen, dann muss die einschlägige Art von Selbstbezüglichkeit ein indexikalisches Element beinhalten. Wenn Indexikalität jedoch erlaubt ist, dann müssen wir den Weg für eine Hierarchie der Ebenen im Tarski-Stil öffnen, die durch indexikalische Eigenschaften von »wahr« hervorgerufen wird.

5.7 Indexikalität

Aus von Lügnerparadoxien unabhängigen Gründen ist es notwendig, Sätze – als Dinge betrachtet, die von verschiedenen Personen und zu verschiedenen Gelegenheiten geäußert werden können – von den Dingen zu unterscheiden, die Personen sagen oder ausdrücken können, indem sie Sätze gebrauchen. Der Grund ist die »Indexikalität« der Sprache: Die Tatsache, dass dieselben Wörter – ohne eine Mehrdeutigkeit auszunutzen – zu verschiedenen Gelegenheiten dazu gebraucht werden können, verschiedene Dinge zu sagen. In-

dexikalität von Pronomen bietet ein gängiges Beispiel: Wenn Sie den Satz »Ich bin hungrig« verwenden, dann sagen Sie etwas, und wenn ich ihn verwende, dann sage ich etwas anderes. Was gesagt wird, ist verschieden, weil es möglich ist, dass das, was Sie sagen, wahr ist, während das, was ich sage, falsch ist.

Eine *Aussage* nenne ich das, was mit einem Satz zu einer bestimmten Gelegenheit gesagt oder ausgedrückt wird. Indexikalität zeigt, dass es nur Aussagen sind und nicht Sätze, die man eigentlich wahr oder falsch nennen kann. Ich werde annehmen, dass für Aussagen Bivalenz gilt, so dass jede Aussage entweder wahr oder falsch ist. Wir haben effektiv bereits gesehen, dass ein Satz Bedeutung haben und dennoch zu einer bestimmten Gelegenheit so verwendet werden kann, dass er keine Aussage macht (»Der Elefant greift gleich an«). Auch wenn Sätze selbstbezüglich sein können, oder allgemeiner, die Art von Zirkularität haben können, die mit Paradoxien einhergeht, ist es möglich, dass Aussagen das nicht können. Somit – um zu dem Beispiel in Abschnitt 5.6 oben zurückzukommen – könnte es uns gelingen, die Behauptung zu rechtfertigen, dass β und γ zwar denselben Satz verwenden, es aber nur γ gelingt, eine Aussage zu machen. Der Begriff einer Aussage scheint also die Eigenschaften zu haben, nach denen wir suchen: Er hängt nicht nur von der Bedeutung eines Satzes ab, sondern auch von dem Gebrauch, der von dem Satz zu einer bestimmten Gelegenheit gemacht wird.

Der Verstärkte Lügner muss an die Unterscheidung zwischen Satz und Aussage angepasst werden. Eine Möglichkeit dafür ist die folgende:

L_2 L_2* drückt keine wahre Aussage aus.

Eine erneute Betrachtung des Gedankenganges im Verstärkten Lügner unterstützt die Auffassung, irgendeine Art von Indexikalität sei hier am Werke. Wir untersuchen L_2* und betrachten ihn als mangelhaft. Wenn wir dies ausdrücken, dann könnten wir dazu Worte verwenden, die selbst L_2* sind oder

beinhalten, zum Beispiel: »L_2* ist semantisch mangelhaft, also drückt er (*a fortiori*) keine wahre Aussage aus; das könnte nur der Gebrauch eines nicht-mangelhaften Satzes erreichen.« Intuitiv scheint dies auf den ersten Blick vollkommen vernünftig (bis uns klar wird, dass wir selbst eben jene Worte wieder gebraucht haben, von denen wir sagen wollten, sie seien mangelhaft). Diese Intuition könnte gerettet werden, wenn wir zeigen könnten, dass dieselben Worte, selbst wenn sie sich auf dasselbe Ding beziehen und dasselbe Prädikat darauf anwenden, bei zwei Gelegenheiten des Gebrauchs nicht dasselbe sagen können. Wir wollen sagen, dass der erste Gebrauch von L_2* mangelhaft ist, es aber der zweite Gebrauch derselben Worte nicht ist, weil sie dort dazu verwendet werden, eine Wahrheit auszudrücken.

Erwägungen wie die folgende legen die allgemeine Machbarkeit eines solchen Ansatzes nahe. Angenommen, der folgende Satz ist der einzige Satz der in Raum 101 an der Tafel steht:

Der Satz, der in Raum 101 an der Tafel steht, drückt keine wahre Aussage aus.

Es scheint vollkommen konsistent, wenn ich auf dieser Seite schreibe, der Satz, der in Raum 101 an der Tafel steht, drückt keine wahre Aussage aus (aufgrund eines semantischen Mangels). Ich gebrauche die Worte, die, so wie sie in Raum 101 an der Tafel stehen, mangelhaft sind, unter Umständen, unter denen an ihrem Gebrauch nichts mangelhaft ist. Das legt nahe, dass dieselben Worte, selbst wenn sie sich auf dasselbe Ding beziehen und dasselbe Prädikat darauf anwenden, nicht notwendigerweise dieselbe Aussage machen. Der Satz an der Tafel in Raum 101 macht, bei seiner Verwendung in Raum 101, keine Aussage; während ich diese Worte auf dieser Seite dazu verwende, eine wahre Aussage zu machen. Ich hätte nicht dieselben Worte verwenden müssen. Unter passenden Umständen hätte ich einfach sagen können: »*Dieser* Satz drückt keine wahre Aussage aus.« Die Tatsache, dass es spezielle Umstände gibt, unter denen ich dieselben Worte verwenden kann, ist ein Zufall unseres Sprachgebrauches und hat keinen Einfluss auf die Wahrheit der Aussage, die ich machen will.

Wir sind von unserem Ziel noch recht weit entfernt, denn es bleiben drei Aufgaben: (1) das Anbieten einer detaillierteren Auffassung davon, worin die problematische Zirkularität besteht; (2) eine unabhängige Rechtfertigung dafür, zu sagen, eine Aussage könne diese nicht aufweisen; (3) eine Rückkehr zu den Problemen, die der Verstärkte Lügner aufwirft.

5.8 Indexikalische Zirkularität

Für eine Spezifikation und Rechtfertigung der einschlägigen Art von Zirkularität ist es wert, auf eine Idee von Bertrand Russell zurückzukommen, sein so genanntes Prinzip Teufelskreis [*Vicious Circle Principle*]. Er gibt mehr als eine Darstellung dessen, was das Prinzip sein soll, aber ein angemessener Ausdruck (kein Zitat) wäre wie folgt:

PT Keine Gesamtheit kann Elemente enthalten, die nur über diese Gesamtheit selbst vollständig zu spezifizieren sind.

Dies ist als allgemeines metaphysisches Prinzip gedacht, anwendbar auf Klassen wie auf alles andere, und also auch anwendbar auf Aussagen (oder, wie Russell manchmal sagte, Propositionen).

Das PT erfasst nicht Gesamtheiten von gewöhnlichen materiellen Dingen, denn keine von diesen enthält Elemente, die nur über diese Gesamtheit selbst spezifizierbar wären. Wir mögen Friedrich als den größten Mann im Regiment spezifizieren, ihn also über eine Gesamtheit spezifizieren, der er angehört, aber es kann nicht sein, dass dies die *einzige* Art und Weise ist, wie er spezifiziert werden kann. Das PT hat keine Tendenz dazu, Regimenter zu beseitigen.

Das PT scheint Sätze nicht zu erfassen, wenn man sie sich als Zeichen oder Formen vorstellt, denn diese können, wie Regimentsangehörige, auf alle möglichen unabhängigen Arten und Weisen spezifiziert werden, von denen nicht alle irgendeine Gesamtheit beinhalten. Im Falle von Aussagen jedoch scheint es, dass einige nur über eine Gesamtheit vollständig spezifiziert werden können. So kann meine Aussage, *Alles, was Sie in Ihrer Radioansprache gesagt haben, war Quatsch*, nur über die Gesamtheit Ihrer Aussagen vollständig spezifiziert werden. Es kann unvollständig oder indirekt auf andere Weise spezifiziert werden, etwa als die Aussage, die das Ende unserer Freundschaft bedeutete, oder die Aussage, deren verbaler Ausdruck auf Seite 187 von *Paradoxien* in kursiv auftaucht. Eine vollständige Spezifikation jedoch scheint die gemeinsame Erwähnung der Dinge zu beinhalten, die Sie gesagt haben. In diesem Falle liegt keine Verletzung von PT vor, da die Gesamtheit Ihrer Aussagen meine nicht beinhaltet, und die Gesamtheit Ihrer Aussagen zusammen mit meiner nicht die Gesamtheit ist, über die meine Aussage spezifiziert ist (sondern vielmehr eine größere).

Gibt es eine Gesamtheit von Aussagen, die durch

$L_1^\dagger$ (Die Aussage) $L_1^\dagger$ ist falsch

spezifiziert wird? Wenn ja, enthält sie wohl die Aussage $L_1^\dagger$ als ihr einziges Element. $L_1^\dagger$ kann aber nur über dieses Element vollkommen spezifiziert werden. PT urteilt also, dass es keine solche Gesamtheit gibt. Es gibt daher keine solche Aussage $L_1^\dagger$.

Es scheint eine gewisse Hoffnung zu geben, dass PT erweitert werden könnte, um Lügner-Zirkel zu handhaben, und dass wir derart unsere erste Aufgabe erledigen könnten, die Spezifikation des Wesens der problematischen Zirkularität. Die zweite Aufgabe besteht darin, eine unabhängige Motivation für die Ansicht anzubieten, dass Aussagen die relevante Art von Zirkularität nicht besitzen können. Zu diesem Punkt bietet uns Russell wenig Argumentation. Das PT soll unab-

hängig akzeptabel sein und soll angemessene Einschränkungen bieten. Es ist, wie ich meine, recht schwierig, intuitiv plausible Erwägungen zu finden, die direkt zeigen, dass Aussagen nicht in der problematischen Weise zirkulär sein können. Ein besonderer Grund zur Sorge ist, dass sich eine Art von Zirkularität, die von PT ausgeschlossen werden würde, systematischer mathematischer Behandlung zugänglich gezeigt hat, die Art von Zirkularität, die in nicht-fundationaler Mengenlehre verwendet wird. (Siehe die Literaturhinweise für eine kurze Beschreibung und Verweise.) Die Entwicklung der Idee, die Wurzel der Paradoxien liege bei der Selbstbezüglichkeit von Aussagen, oder allgemeiner bei Zirkularität von Aussagen, erfordert Arbeit an diesem Punkt.

Angenommen, dies kann erfolgreich erledigt werden, wird der Verstärkte Lügner nicht auf der Lauer liegen, um die Anstrengungen zunichte zu machen? Das Problem erscheint besonders drohend, wenn wir den Satz L_2^* noch einmal betrachten:

L_2^* L_2^* drückt keine wahre Aussage aus.

»L_2^*« bezeichnet einen Satz, und die hier diskutierte Herangehensweise an die Paradoxien will behaupten, dass er ein Satz ist, der keine Aussage ausdrückt: Er ist dank Zirkularität semantisch mangelhaft. Nichts, was semantisch mangelhaft wäre, drückt eine wahre Aussage aus und also gilt, *L_2^* drückt keine wahre Aussage aus*. Die Worte in kursiv sind eben L_2^* selbst und wir scheinen die Anfänge der bekannten Paradoxie zu haben.

Der Theoretiker muss behaupten, dass seine Verwendung der (kursiv geschriebenen) Worte eine Aussage ausdrückt, und zwar eine wahre, und sich also von der ursprünglichen Verwendung dieser Worte unterscheidet. Er hätte dies unter Verwendung der nicht-paradoxen Worte ausdrücken können: »Der oben ausgeschriebene Satz drückt keine wahre Aussage aus.« Es ist ein Zufall unserer Konventionen der Bezeichnung, dass eben dieselben zu verdammenden Worte dazu verwendet

werden können, diese zu verdammen. In seiner mangelhaften Verwendung verlangt L_2^* das Unmögliche, die Existenz einer selbstbezüglichen Aussage; in seiner nicht-mangelhaften Verwendung tut er das nicht.

Eine Lösung dieser Art sollte zu mindestens zwei Arten von Befürchtungen Anlass geben. Die eine ist spezifisch: Könnte ein gewitzter Gegner nicht Lügner-Sätze konstruieren, für die diese Erwiderung ausgeschlossen ist?[10] Die andere ist allgemeiner: Ist die Auffassung mit der Möglichkeit einer wirklich *formalen* Logik verträglich, in der die logischen Relationen durch bloß syntaktische abgebildet werden? Wenn es zwischen Sätzen und Aussagen eine Kluft gibt, die theoretisch nicht ausgeräumt werden kann, wird dieses Projekt zum Scheitern verurteilt sein.

Der Begriff der Indexikalität kann auf eine andere Art und Weise ausgenutzt werden. Wir haben die Möglichkeit untersucht, dass Indexikalität die Subjektterme in Sätzen wie L_2 und L_2^* betreffen könnte: Ob das Subjekt auf eine Art und Weise Bezug nimmt, welche die Existenz einer selbstbezüglichen Aussage verlangt, wurde als abhängig von den Umständen seiner Verwendung angesehen. Man sollte auch die Möglichkeit in Betracht ziehen, dass Indexikalität den Prädikatsterm betrifft: »wahr«. Tyler Burge (1979) hat einen Vorschlag in dieser Richtung entwickelt. Es gibt verschiedene Ebenen der Wahrheit, und welches die einschlägige Ebene ist, wird nicht von der Bedeutung des Satzes festgelegt, sondern von der Aussage, die mit ihm bei der betreffenden Gelegenheit gemacht wird. Diese Auffassung auf der Grundlage indexikalischer Ebenen vermeidet viele der mit Tarskis Hierarchie verknüpften Schwierigkeiten, welche sich auf Sätze statt auf Aussagen bezieht. Burges Konstruktion ist komplex, und ich biete hier keine Details, weil ich folgenden Verdacht hege: Es ist

10* Zeigen Sie, wie die Erwiderung unangemessen erscheint für: Keine Verwendung eben dieses Satzes drückt eine wahre Aussage aus.

schwierig bis unmöglich, die Behauptung zu rechtfertigen, dass »wahr« indexikalisch ist, unabhängig von der scheinbaren Notwendigkeit, indexikalisch zu sein, um den Paradoxien gerecht zu werden, während wir unabhängige Gründe dafür haben, an die Indexikalität in Verbindung mit Selbstbezüglichkeit zu glauben.

5.9 Vergleich: Wie ähnlich sind Russells Paradoxie und der Lügner?

Sind die zwei Paradoxien dieses Kapitels vollkommen verschieden oder im Wesentlichen gleich? Liegt die Wahrheit vielleicht irgendwo zwischen diesen Extremen?

Ramsey meinte, dass die Paradoxien von verschiedener Art sind, und seine Ansicht war die vorherrschende, zumindest bis vor kurzem. Seine Unterscheidung gründete sich auf ihren verschiedenen Gegenstandsbereich: Die logischen Paradoxien, unter denen er Russells Paradoxie einordnete, wurzeln in logischen Begriffen, wie dem der Klasse; die semantischen Paradoxien, unter denen er den Lügner einordnete, wurzeln in semantischen Begriffen, wie dem der Wahrheit.

Es gibt außerdem strukturelle Unähnlichkeiten, die auf die Unterschiede der an den verschiedenen Paradoxien beteiligten Begriffe zurückgehen. In Russells Paradoxie gibt es keine Entsprechung für den Verstärkten Lügner. Die Idee, *es gebe keine* von L_2 ausgedrückte Aussage, ist unmittelbar problematisch, denn es scheint zu folgen, dass L_2 nicht wahr ist. Die Behauptung, *es gebe keine* solche Klasse R hat keine solche Schwierigkeit zur Folge. Bei der Klassenparadoxie gibt es recht weitgehende Einigkeit darüber, was wir sagen müssen – dass es keine solche Klasse R gibt. Unklar ist, wie man das rechtfertigen kann. Beim Lügner ist nicht klar, was man sagen soll, ganz zu schweigen davon, wie es zu rechtfertigen ist.

Die Paradoxien sind sich aber auch in vielen Hinsichten ähnlich. Ich zähle fünf auf.

(1) Die Klassenparadoxie ähnelt einer Paradoxie über Eigenschaften und die Eigenschaftsparadoxie wiederum ähnelt dem Lügner. Die meisten Eigenschaften treffen nicht auf sich selbst zu. Die Eigenschaft, ein Mann zu sein, zum Beispiel, trifft nicht auf sich selbst zu, da dieser Eigenschaft die Eigenschaft fehlt, ein Mann zu sein. Aber die Eigenschaft, ein Nicht-Mann zu sein, trifft auf sich selbst zu, da die Eigenschaft, ein Nicht-Mann zu sein, die Eigenschaft hat, ein Nicht-Mann zu sein. Das in der Klassenparadoxie verwendete Argumentationsmuster würde zu dem Schluss führen:

> Die Eigenschaft, *auf sich selbst nicht zuzutreffen*, trifft auf sich selbst zu, genau dann, wenn sie nicht auf sich selbst zutrifft.

Es gibt eine oberflächliche Ähnlichkeit zwischen diesem Widerspruch und dem Widerspruch, dass L_2 Wahrheit von sich selbst wahr genau dann prädiziert, wenn er es nicht tut. Wo der Widerspruch über Eigenschaften den Begriff *nicht zutreffen* verwendet, eine Relation, die zwischen einer Eigenschaft und etwas anderem (eventuell ebenfalls eine Eigenschaft) bestehen kann, verwendet der Lügner-Widerspruch den Begriff *nicht wahr*, eine Eigenschaft, die ein Satz oder eine Aussage besitzen kann.

(2) Sowohl die Klassenparadoxie als auch der Lügner beinhalten Selbstbezüglichkeit oder etwas Ähnliches.

(3) Die Prinzipien, an die bei der Ableitung der beiden Paradoxien appelliert wird (KE: Für jede verständliche Bedingung F gibt es eine Klasse x, so dass gilt: Für jedes y, $y \in x$, genau dann, wenn y F erfüllt; und W: σ ist wahr gdw p), sind strukturell ähnlich und scheinen eine ähnlich konstitutive Rolle bezüglich der intuitiven Begriffe von *Klasse* und *Wahrheit* zu spielen.

Auf der Seite der Ableitung ist der Vergleich folgender. Das Schema

> Für jedes y, $y \in x$, gdw y F erfüllt

erzeugt einen Widerspruch, wenn x durch einen Namen ersetzt wird, etwa R für die Russell-Klasse, und F durch die mit Hilfe dieses Namens ausgedrückte Bedingung, die angeblich Zugehörigkeit zu dieser Klasse definiert, »$\neg \in R$«.

Auf ähnliche Weise erzeugt das Schema

> σ ist wahr gdw p

einen Widerspruch, wenn σ durch einen Namen ersetzt wird, etwa L_2 für den Lügner-Satz, und p durch die mit Hilfe dieses Namens ausgedrückte Bedingung, die angeblich Wahrheit für diesen Satz definiert, »L_2 ist nicht wahr«.

Was ihre Rollen betrifft, ist vergleichbar, dass W ebenso konstitutiv für unseren vortheoretischen Begriff von Wahrheit erscheint wie KE für unseren vortheoretischen Begriff von Klasse. KE bestimmt, was es für eine Klasse bedeutet zu existieren; W bestimmt, was es für eine Wahrheitsbedingung bedeutet, zu existieren.

(4) In Reaktion auf beide Arten von Paradoxien sind Hierarchien verwendet worden, beginnend mit den frühesten systematischen Behandlungen bei Russell (1908). Es ist natürlich, anzunehmen, dass wir uns Klassen als aus Nicht-Klassen konstruiert vorstellen, wobei jeder konstruktive Schritt nur auf Entitäten beruht, die bereits konstruiert worden sind. Ebenso ist es natürlich, anzunehmen, dass wir uns Wahrheit zuschreibende Aussagen als aus Aussagen konstruiert vorstellen, die von dem Wahrheitsbegriff frei sind, wobei jeder konstruktive Schritt nur auf Aussagen beruht, die bereits konstruiert worden sind.

(5) Russells Einordnung der Klassenparadoxie und des Lügners als von derselben Art beruht auf der Behauptung, dass sie beide gleichermaßen von einem Verstoß gegen das *Prinzip Teufelskreis* (PT) herrühren.

> PT Keine Gesamtheit kann Elemente enthalten, die nur über diese Gesamtheit selbst vollständig zu spezifizieren sind.

Wir haben bereits gesehen, wie dies dazu verwendet werden kann, gewisse Arten von Zirkularität bei Aussagen zu verhindern. Die Art und Weise, wie es Zirkularität bei Klassen verhindert, ist einfacher. Die Spezifikation von Klasse R, der sich-nicht-selbst-enthaltenden Klassen, war diese:

> Für jede Klasse x, $x \in R$ gdw $\neg\, x \in x$.

Die Spezifikation handelt von etwas, was Russell eine Gesamtheit nennen würde: die durch den Ausdruck »jede Klasse« eingeführte Gesamtheit aller Klassen. Russell meint, dass dies die einzig mögliche Spezifikation von R darstellt, und das PT soll uns sagen, dass R nicht zu der durch den Ausdruck »jede Klasse« eingeführten Gesamtheit gehören kann, weil diese in der Spezifikation von R vorkommt. Denn angenommen, R gehörte zu dieser Gesamtheit: Dann würde die Gesamtheit ein Element enthalten, welches nur mittels dieser Gesamtheit spezifizierbar wäre, was dem PT zufolge unmöglich ist. Wenn R jedoch nicht zu der durch »jede Klasse« eingeführten Gesamtheit gehört, dann können wir den üblichen Schritt zum Widerspruch nicht machen. Der übliche Schritt geht folgendermaßen: Wenn die Definition auf jede Klasse zutrifft, dann trifft sie insbesondere auf R zu, und also können wir schließen:

> $R \in R$ gdw $\neg\, R \in R$.

Es liegt im PT, dass die von *jeder Klasse* eingeführte Gesamtheit R ausschließt, also ist der Schluss fehlerhaft. Im Endeffekt ist das Ergebnis von PT, dass wir R nicht so spezifizieren können, wie wir es ursprünglich vorhatten – das heißt derart, dass die Frage, ob es seine definierende Eigenschaft hat oder nicht, überhaupt aufkommen kann.

Eine formale Verteidigung von Russells Behauptung, dass die Paradoxien zu einer wichtigen gemeinsamen Art gehören, hat Priest (1994) bereitgestellt.

Die Klassen- und die Lügnerparadoxie sind sich, wie die meisten Dinge, in einigen Hinsichten ähnlich, in anderen nicht. Na und? Klassifikation ist hier deshalb bedeutsam, weil sie Einschränkungen für angemessene Erwiderungen auf die Paradoxien bietet. (Viele weitere Paradoxien müssen in einer Klassifikation noch ihren Platz finden; siehe Priest, 1994.) Wenn zwei Paradoxien *wesentlich* ähnlich sind, ähnlich in dem, worauf es wirklich ankommt, dann ist es angemessen, auf sie im Wesentlichen ähnlich zu erwidern. Wenn zum Beispiel die Klassenparadoxie nach einer Hierarchie von Ebenen verlangt und der Lügner im Wesentlichen ähnlich ist, dann verlangt er ebenfalls nach einer Hierarchie von Ebenen. Wenn eine angemessene Auffassung von Klassen nicht wohl fundierte Klassen erlauben sollte, etwa die Klasse α, deren einziges Element α ist, und der Lügner im Wesentlichen ähnlich ist, dann sollte unsere Erwiderung eine analoge Zirkularität gestatten. Russells angeblich gemeinsame Lösung der beiden Paradoxien, seine Verzweigte Typentheorie, hat eine leicht vorgetäuschte Einheitlichkeit, da einige ihrer komplexeren Eigenschaften durch Dinge motiviert sind, die eher auf den Lügner zurückgehen als auf die Klassenparadoxie.

Um zu dem Schluss zu kommen, dass die Paradoxien eine einheitliche Lösung erfordern, müssen wir mehr zeigen, als dass es *eine* Art gibt, der die Paradoxien angehören. Wir müssen auch zeigen, dass diese Art ihre gemeinsame essentielle Natur aufdeckt. Ich bezweifle, dass dies ganz unabhängig von Auffassungen dazu geschehen kann, welche Erwiderung jeder einzelnen Paradoxie angemessen ist.

Literaturhinweise

Abschnitt 5.1

Russells früheste veröffentlichte Darstellung der Klassenparadoxie findet sich in (1903). Für einen historischen Überblick siehe van Heijenoort (1967). Die meisten Einführungstexte in die Logik beinhalten eine Diskussion. Für eine weniger technische Darstellung, die die Paradoxien stets in Blick behält, siehe Copi (1971). Eine gute und zugängliche Einführung in die Mengenlehre bietet Thomason (1970), Kap. 13.

Der gegenwärtige Gebrauch unterscheidet Mengen und Klassen: Alle Mengen sind Klassen, aber nicht alle Klassen sind Mengen. Intuitiv ist eine Menge eine sich ordentlich verhaltende Klasse. Die Unterscheidung setzt eine bestimmte Herangehensweise an die Lösung von Russells Paradoxie voraus und ist daher für die gegenwärtige Erörterung nicht angemessen.

Die *Potenzklasse* einer Klasse ist die Klasse, die aus jeder Teilklasse jener Klasse besteht. Man betrachte etwa die Klasse, welche nur aus den drei Elementen a, b und c besteht und die wir $\{a, b, c\}$ schreiben können. Eine Klasse α ist Teilklasse einer Klasse β genau dann, wenn jedes Element von α Element von β ist. Die Klasse $\{a, b, c\}$ hat die folgenden Teilklassen: (1) $\emptyset$ (die Nullklasse) – da $\emptyset$ keine Elemente hat, ist jedes Element von $\emptyset$ ein Element von $\{a, b, c\}$; (2) $\{a\}$; (3) $\{b\}$; (4) $\{c\}$; (5) $\{a, b\}$; (6) $\{a, c\}$; (7) $\{b, c\}$; (8) $\{a, b, c\}$ – da jedes Element dieser Klasse ein Element von $\{a, b, c\}$ ist.

Es gibt also acht Teilklassen einer Klasse mit drei Elementen. Die Elemente der Klasse aller Teilklassen sind daher zahlreicher (um fünf Elemente) als $\{a, b, c\}$ selbst. Cantors Theorem gilt offensichtlich für Klassen mit endlich vielen Elementen. Das Interessante an dem Theorem liegt darin, dass es für Klassen jeder Kardinalzahl gilt.

Eine eins zu eins Funktion zwischen zwei Klassen α und β ordnet jedem Element von α genau ein Element von β zu und jedem Element von β genau ein Element von α. Cantor nimmt an, dass zwei Klassen genau dann dieselbe Anzahl von Elementen (dieselbe Kardinalzahl) haben, wenn es eine eins zu eins Funktion zwischen ihnen gibt.

Eine gute Einführung zu den beiden in diesem Kapitel diskutierten Paradoxien ist Priest (1987), Kap. 2 (mengentheoretische Paradoxien) und Kap. 1 (semantische Paradoxien).

Abschnitt 5.2

Gute Ausgangspunkte für den Lügner: Mackie (1973); Prior (1961); Martin (1984), Einleitung des Herausgebers; Barwise/Etchemendy (1987), Kap. 1.

Für eine Diskussion der Bivalenz siehe Haack (1978) und Burge (1984).

Abschnitt 5.3

Das Wort »Fundierung« [*grounding*] und seine Verwandten leihe ich mir von Kripke (1975). Ich behaupte nicht, erfasst zu haben, was er damit meint, da sein Begriff von Fundierung in eine mathematische Theorie eingebettet ist, der ich nicht ansatzweise gerecht werden kann. Der erste Teil dieses Aufsatzes jedoch ist dem Nicht-Mathematiker durchaus zugänglich und sollte gelesen werden. Eine klassische Stelle für Fundierung ist Herzberger (1970).

Abschnitt 5.4

Der Ausdruck »Verstärkter Lügner« geht auf van Fraassen (1968) zurück, auch wenn das Problem selbst viel älter ist.

Abschnitt 5.5

Tarski (1969) enthält eine halbwegs allgemein verständliche Darstellung seiner Ansichten. Der klassische Text ist Tarski (1935). Trotz des formalen Hauptteils dieses Werkes sind die ersten beiden Teile, die Zusammenfassung und das Nachwort, nicht formal, zugänglich und sehr lesenswert. Er gibt eine *Konvention W* genannte Adäquatheitsbedingung für eine formale Wahrheitsdefinition (als »Wr« symbolisiert):

> Eine formal korrekte, [...] Definition des Symbols »Wr« werden wir eine *zutreffende Definition der Wahrheit* nennen, wenn sie folgende Folgerungen nach sich zieht:
>
> (α) alle Sätze, die man aus dem Ausdruck »$x \in$ Wr dann und nur dann, wenn p« gewinnt, indem man für das Symbol x einen strukturell-deskriptiven Namen einer beliebigen Aussage der betrachteten Sprache und für das Symbol »p« den Ausdruck, welcher die Übersetzung dieser Aussage in die Metasprache bildet, einsetzt;
>
> (β) [...] (S. 45 f.)

Tarski gestattet es, wie manchmal gesagt wird, uns auf unsere intuitive Auffassung von Bedeutung (Übersetzung) zu beziehen, während wir

die Bedingungen für eine korrekte Definition der Wahrheit spezifizieren.

Gupta (1982), Abschnitt II, bietet eine sorgfältige Formulierung von Tarskis Prämissen, gemeinsam mit einem Angriff auf die ganze Allgemeinheit des von Tarski gezogenen Schlusses.

Tarski akzeptiert für jedes Wahrheitsprädikat in der Hierarchie jede Instanz von W. Die Tatsache, dass W eine inkonsistente Instanz hat, wenn Instanzen durch jede grammatisch akzeptable Konstruktion des Deutschen gebildet werden können – insbesondere durch L_2 –, soll zeigen, dass es keine kohärente Sprache Deutsch gibt. Das Ersetzen des einzigen deutschen Prädikates »wahr« durch eine Hierarchie von Wahrheitsprädikaten beinhaltete in Tarskis Augen auch, sich des Deutschen, so wie es gewöhnlich verstanden wird, zu entledigen.

Abschnitt 5.6

Für einen Angriff auf Selbstbezüglichkeit siehe Jørgensen (1953). Für die Behauptung, dass einige Arten von Selbstbezüglichkeit unschuldig sind, siehe Barwise/Etchemendy (1987), bes. S. 15 f. Für einen Sammelband zum Thema siehe Bartlett (1992).

Abschnitt 5.7

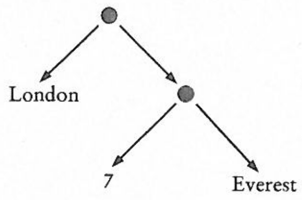
Die klassische Quelle für die Unterscheidung zwischen Satz und Aussage, aus vom Lügner ganz unabhängigen Gründen vorgetragen, ist Strawson (1950).

Abschnitt 5.8

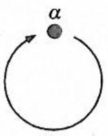
Die auf Seite 186 gegebene Formulierung des PT ist der in Russell (1908), S. 75, am nächsten. Die Rechtfertigung für das »nur« stammt aus der Formulierung auf Seite 63 jenes Werkes; vgl. Russell/Whitehead (1910–13), S. 37. Ich habe »vollständig spezifizierbar« anstelle von Russells »definiert« verwendet.

Hazen (1987) verfolgt den Ursprung der in diesem Abschnitt erörterten allgemeinen Art von Erwiderungen bis auf den Philosophen Jean Buridan aus dem 14. Jahrhundert. Hazen verteidigt eine Version der hier betrachteten Ansicht, und dieser Ansatz ist von anderen aufgenommen worden, zum Beispiel Goldstein (1992). Für Kritik siehe Hinkfuss (1991), Smiley (1993) (auch wenn er einen Ansatz vorschlägt, dessen zentraler Punkt ist, dass Lügner-Sätze »nicht funktionieren«) und Priest (1993).

Um nicht-fundationale Mengenlehre einzuführen, denken wir uns eine Menge wie ein Diagramm. Die Menge, welche nur aus London und der Menge, deren Elemente die Zahl 7 und der Mount Everest sind, besteht (konventionell geschrieben: {London, {7, Everest}}), kann also durch das Diagramm repräsentiert werden:



Jeder fette Punkt repräsentiert hier eine Menge und die Äste darunter repräsentieren ihre Elemente. Das folgende Diagramm würde dann eine Menge α repräsentieren, deren einziges Element α ist:



Die Theorie solcher Diagramme ist relativ zur klassischen Mengenlehre konsistent. Eine allgemeine Auffassung des Übels der Zirkularität, wie Russells PT, täte gut daran, die Verknüpfung von Diagrammen und Mengen zu brechen.

Nicht-fundationale Mengenlehre verdankt Aczel (1987) sehr viel. Eine gute Darstellung, den gegenwärtigen Fragen wohl angepasst, ist in Barwise/Etchemendy (1987), Kap. 3.

Abschnitt 5.9

Russell (1908) argumentiert für die gemeinsame Natur der Klassen-, Lügner- und anderer Paradoxien. Zur Unterscheidung zwischen logischen und semantischen Paradoxien siehe Ramsey (1925), S. 171 f. Die beste Entwicklung der Argumentation für eine einzige Familie von Paradoxien bietet Priest (1994).

6 Gibt es akzeptable Widersprüche?

In den bisherigen Kapiteln habe ich an zahlreichen Stellen gesagt, wenn etwas zu einem Widerspruch führt, dann muss es, bzw. der betreffende Gedankengang, verworfen werden. Die Annahme, dass man Widersprüche stets verwerfen müsse, ist in der Geschichte der Philosophie verschiedentlich angegriffen worden. Seit kurzer Zeit hat dieser Angriff eine subtile Form angenommen und beeindruckende technische Mittel in Anschlag gebracht.

Ich werde die folgenden Auffassungen erörtern:

(1) Einige Widersprüche sind wahr.
(2) Bei einigen Widersprüchen ist es rational, sie für wahr zu halten.

Die einzige Version von (1), die ich betrachten werde, behauptet auch, dass jeder Widerspruch falsch ist; das ist der »Dialethismus«. Die Auffassung, dass einige Widersprüche sowohl wahr als auch falsch sind, ist allerdings noch nicht die Auffassung, dass einige Widersprüche akzeptabel sind, denn man könnte weiterhin darauf bestehen, dass alles, was man für falsch hält, verworfen werden muss. Wenn dem so ist, dann bliebe die Annahme der bisherigen Kapitel dieses Buches unkritisiert: Wir sollten alles verwerfen, was zu einem Widerspruch führt. Ich werde in diesem Kapitel also die Verbindung der beiden Auffassungen (1) und (2) erörtern. Diese Kombination will ich »rationalen Dialethismus« nennen.

Die Behauptung des rationalen Dialethisten, dass einige Widersprüche wahr sind, ist, unter natürlichen Annahmen, mit der Behauptung äquivalent, dass einige Sätze sowohl wahr als auch falsch sind. Jeder solche Satz soll eine *Dialetheia* genannt werden. Ich werde in Abschnitt 6.4 vorschlagen, rationaler Dialethismus impliziere, unter natürlichen Annahmen, dass einige Sätze, die keine Widersprüche sind (d. h. nicht von der Form: A und nicht-A), sowohl wahr als auch falsch sind.