Mathematik - ein Reiseführer

Bearbeitet von
Ingrid Hilgert, Joachim Hilgert

1. Auflage 2012. Taschenbuch. x, 278 S. Paperback
ISBN 978 3 8274 2931 5
Format (B x L): 15,5 x 23,5 cm
Gewicht: 439 g

1 Vom Wesen der Mathematik

Übersicht

In diesem Kapitel beginnt die eigentliche Vorstellung mathematischer Denkweisen. Dabei werden keine mathematischen Kenntnisse jenseits des normalen Schulstoffes der Mittelstufe vorausgesetzt. Die Beispiele sind aber durchaus anspruchsvoll, und im Verlauf des Textes wird klar werden, dass die Einführung einer formalen Sprache das Mitdenken erleichtert.

1.1 Modelle, Gleichungen, Prognosen

Man kann die Mathematik zunächst einmal als eine Verfeinerung der Alltagssprache auffassen, eine Art „sprachliche Lupe". Sie dient dazu, beobachtbare Vorgänge so präzise zu beschreiben, dass es möglich wird, Gesetzmäßigkeiten zu erkennen und durch das Lösen von Gleichungen Prognosen zu erstellen. Diese Beschreibungen nennt man *Modelle*.

Die Flugbahn einer Kugel beim Kugelstoßen könnte man in der Alltagssprache etwa so beschreiben: „Erst fliegt sie relativ gerade nach oben, dann flacht sich die Bahn ab, erreicht die maximal Höhe und fällt dann immer schneller zu Boden." Erste mathematische Konzepte hat man schon verwendet, wenn man über horizontale und vertikale Geschwindigkeiten spricht, um die Beschreibung zu verfeinern. Stärker formalisiert kann man sagen, dass die Flugbahn (in sehr guter Näherung) eine *Parabel* ist, die durch eine Gleichung der Bauart $y = ax^2 + bx + c$ beschrieben werden kann. Dabei ist y die Höhe über dem Boden, x der Abstand

I. Hilgert, J. Hilgert, *Mathematik – ein Reiseführer*,
DOI 10.1007/978-3-8274-2932-2_1, © Spektrum Akademischer Verlag Heidelberg 2012

vom Kugelstoßer und a, b, c sind Zahlen, die von physikalischen Parametern wie dem Gewicht der Kugel, der Stoßrichtung und der Schnellkraft des Sportlers abhängen. Aus dieser Beschreibung lassen sich Vorhersagen gewinnen: Man kann die Flugbahn verlässlich rekonstruieren. Insbesondere kann man den Aufschlagpunkt am Boden berechnen, wenn man die physikalischen Parameter kennt. Mehr noch, wenn man genügend Messwerte hat, kann man daraus die physikalischen Parameter und die optimale Stoßrichtung bestimmen. Die Möglichkeit, auf eine Beschreibung (Modell) Rechentechniken (oft Gleichungen) anzuwenden und so zu Vorhersagen (Prognosen) zu kommen, hebt die Mathematik über eine reine Sprache hinaus.

Kompliziertere Beispiele sind die Bewegungen von Himmelskörpern wie Sonne, Mond, Sternen und Planeten. Eine rein verbale Beschreibung der beobachtbaren täglichen Bahnen würde zum Beispiel die fundamentalen Unterschiede zwischen Fixsternen und Planeten nicht offenlegen. Diese treten erst zutage, wenn man die unterschiedlichen Perioden beobachtet, in denen sich Phänomene wiederholen. Relativ früh hat man periodisches Verhalten mit Bewegungen auf Kreisen assoziiert, also ein mathematisches Konzept zur präziseren Beschreibung eines Phänomens verwendet. Der Weg von diesen ersten mathematischen Beschreibungen bis zu den Kepler'schen Gesetzen, nach denen sich die Planeten auf elliptischen Bahnen um die Sonne bewegen, führt über einige Zwischenstufen. Kepler konnte ihn erst gehen, als er Zugriff auf die Messergebnisse des Astronomen Tycho Brahe hatte. Diese sind aber Beschreibungen von Beobachtungen durch Zahlen. Daraus hat Kepler dann Beschreibungen von Beobachtungen durch Gleichungen gemacht. Im nächsten Schritt konnte Newton erkennen, dass die Flugbahnen von Kanonenkugeln und Planeten denselben universellen Gesetzen genügen. Mit jedem der beschriebenen Schritte wuchs die Komplexität der benutzten mathematischen Konzepte. Die zunehmend präziseren Beschreibungen oder, anders gesagt, zunehmend besseren Modelle der Himmelsmechanik erlauben auch zunehmend präzisere Prognosen: Die Vorhersage, dass die Sonne nach 24 Stunden wieder aufgehen wird ist noch nicht sehr beeindruckend, ganz im Gegensatz zur Vorhersage einer Sonnenfinsternis.

Mathematische Beschreibungen sollten nicht als Abbilder, sondern als Modelle realer Situationen betrachtet werden. Man kann den gleichen Vorgang auf ganz unterschiedliche Weisen modellieren. Je nach Zielsetzung kann man sich auf beschreibende Aspekte beschränken oder aber dynamische und kausale Aspekte berücksichtigen. So braucht man beispielsweise nur relativ wenig Mathematik, um ein vorliegendes Wahlergebnis statistisch so aufzubereiten, dass man es auf einer Zeitungsseite wiedergeben kann; präzise Hochrechnungen aus relativ wenigen ausgezählten Wahlkreisen zu erstellen, erfordert jedoch sehr viel mehr Modellierungs- und auch mathematischen Aufwand.

Der Modellcharakter mathematischer Beschreibungen kann an einem fundamentalen Beispiel aus der Physik näher erläutert werden, vgl. Hawking und Mlodinow (2010): Anfang des 20. Jahrhunderts hat man festgestellt, dass Newtons Beschreibung der Bewegungsgesetze mit den mathematischen Beschreibungen der Lichtausbreitung nicht verträglich ist. Albert Einstein entwickelte ein neuartiges Modell, die *Relativitätstheorie*. Er zeigte auch, dass seine neuen Bewegungsgleichungen für Bewegungen, die im Vergleich zur Lichtgeschwindigkeit relativ langsam ablaufen, sehr gut durch die Newton'schen Bewegungsgleichungen angenähert werden. Für höhere Geschwindigkeiten verliert das Newton'sche Modell seine Gültigkeit, das heißt seine Prognosen werden nachweislich falsch. Man ersetzt es durch das Einstein'sche Modell, das auch für hohe Geschwindigkeiten verifizierbare Prognosen liefert. Für kleine Geschwindigkeiten benutzt man auch heute noch die Newton'sche Mechanik, da Berechnungen im Einstein'schen Modell sehr aufwendig und die Fehler bei den Prognosen so klein sind, dass sie nicht ins Gewicht fallen. Ebenfalls Anfang des 20. Jahrhunderts hat man entdeckt, dass die Newtonsche Mechanik auch bei der Beschreibung von Vorgängen in Atomen falsche Vorhersagen liefert. In Verbindung mit der mathematischen Beschreibung des Elektromagnetismus sagt sie nämlich voraus, dass Atome nicht stabil sein können, sondern Elektronen in kürzester Zeit in die Atomkerne stürzen müssten. Diese falschen Vorhersagen konnte auch die Relativitätstheorie nicht korrigieren. Für kleine Skalen verliert also auch die Relativitätstheorie ihre Gültigkeit als mathematisches Modell für das Verhalten von Teilchen. In diesen Bereichen kann man ein völlig anderes mathematisches Modell benutzen, die *Quantenmechanik* nach Niels Bohr und Werner Heisenberg. Allerdings enthält die Quantenmechanik keine Modellierung von Gravitation. Das heißt, die Quantenmechanik ist keine Verallgemeinerung der Newton'schen oder Einstein'schen Theorie, die diese beiden Theorien ersetzen könnte. Es ist eines der großen offenen Probleme der Physik, ein mathematisches Modell zu finden, das sowohl die Quantenmechanik als auch die allgemeine Relativitätstheorie als Spezialfälle enthält.

Um zu testen, ob ein mathematisches Modell brauchbar ist, vergleicht man die aus ihm abgeleiteten Prognosen mit Versuchsergebnissen. Findet man hohe Übereinstimmung zwischen Vorhersage und Messung, so ist das Modell gut. Je geringer die Übereinstimmung, desto mehr Anlass gibt es, das Modell infrage zu stellen, zu verbessern oder gegebenenfalls ganz zu verwerfen. Ein prominentes Beispiel für ein mathematische Modell, das ganz verworfen werden musste, weil es falsche Prognosen lieferte, ist das Kepler'sche Planetenmodell aus seinem Jugendwerk *Mysterium Cosmographicum* von 1596 (Abbildung 1.1).
Es beschreibt das Sonnensystem mit den sechs damals bekannten Planeten Merkur, Venus, Erde, Mars, Jupiter und Saturn. Diese dachte man sich auf Kugelschalen angeheftet, die sich in wohlbestimmten Abständen um die Sonne drehen. Kepler versuchte hier rein spekulativ die Abstandsverhältnisse durch die Abmes-

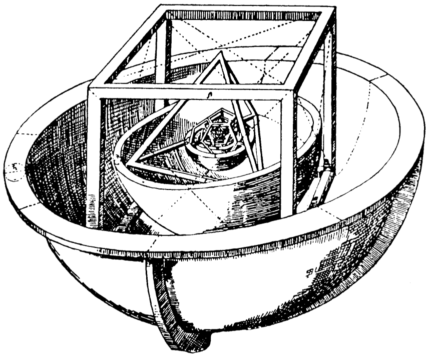


Abb. 1.1 Kepler'sches Planetenmodell

sungen der fünf platonischen Körper (siehe Abbildung 2.1) zu erklären: Es gebe nach göttlichem Plan genau sechs Planeten und genau fünf platonische Körper, die in die Zwischenräume passten. Spätestens mit der Entdeckung des siebten Planeten Uranus im Jahr 1781 war das Kepler'sche Modell obsolet. Es hatte den Realitätstest nicht bestanden und war falsifiziert worden.

Mathematische Modelle im Bankenwesen, welche die Finanzkrise von 2008 nicht prognostiziert hatten, haben sich ebenfalls als fehlerhaft erwiesen. Eine wissenschaftliche Vorgehensweise verlangt, die Abweichung zwischen Prognose und tatsächlicher Entwicklung genau zu studieren und verbesserte Modelle zu entwickeln. Die Falsifizierbarkeit von Modellen ist ein Vorteil der wissenschaftlichen Methode gegenüber anderen Vorgehensweisen.

Ein mathematisches Modell eines komplexen Systems beschreibt in der Regel gesetzmäßige Zusammenhänge zwischen verschiedenen Eigenschaften des Systems. In einer Modellierung des Stromkreises sind zum Beispiel Spannung und Stromstärke in einem Stromkreis über eine Materialeigenschaft, den Widerstand, gekoppelt. Das *Ohm'sche Gesetz* besagt nämlich, dass der Quotient

$$R = \frac{U}{I}, \quad \text{Spannung durch Stromstärke}$$

bei konstanter Temperatur konstant bleibt. Diesen Quotienten nennt man den Widerstand. Kennt man zwei der Größen in der Gleichung $U = RI$, kann man die dritte berechnen. Hat man eine Batterie mit gegebener Spannung U und kennt den Widerstand R eines Drahtes, so lässt sich durch Auflösung der Gleichung $U = RI$ nach I, das heißt durch $I = \frac{U}{R}$, vorhersagen, welcher Strom fließen wird, wenn man die beiden Pole der Batterie mit dem Draht verbindet. Die Verwendung von Buchstaben statt konkreter Zahlenwerte erlaubt es uns, alle Rechnungen dieser Form auf einen Schlag abzuhandeln. Es ist so auch einfacher zu erkennen, wenn verschiedene Modelle mit identischen Methoden behandelbar sind (siehe

Beispiele 1.2 und 1.3). Prognose bedeutet für den Mathematiker hier, wie in vielen anderen Fällen, das Herausrechnen einer unbekannten Größe aus einer Reihe von bekannten Größen. Da sich viele der Prognoseaufgaben in mathematischen Modellen so formulieren lassen, wird das Lösen von Gleichungen gern als die zentrale Problemstellung der Mathematik beschrieben.

Das Lösen von Gleichungen ist keineswegs eine automatisch zu bewerkstelligende Aufgabe. Die meisten Gleichungen lassen sich nicht explizit lösen, das heißt, sie führen nicht zu der gesuchten konkreten Information. Auch einfachen Gleichungen ist normalerweise nicht anzusehen, ob sie überhaupt eine Lösung haben. Man prüft zum Beispiel leicht nach, dass die Gleichung

$$x^2 + y^2 = z^2$$

in den Unbekannten x, y, z von $x = 3$, $y = 4$ und $z = 5$ gelöst wird. Eine solche ganzzahlige Lösung dieser Gleichung heißt ein *pythagoreisches Zahlentripel*. Mit etwas elementarer Geometrie kann man sogar ein Konstruktionsverfahren angeben, das alle pythagoreischen Zahlentripel liefert (Seite 81). Für die zunächst ganz ähnlich aussehende Gleichung

$$x^n + y^n = z^n \tag{1.1}$$

in den Unbekannten x, y, z mit einer vorgegebenen natürlichen Zahl $n > 2$ dagegen gibt es keine positiven ganzzahligen Lösungen. Das war von Pierre de Fermat (1601–1665) behauptet worden, wurde aber trotz vieler Anläufe erst 1995 von dem englischen Mathematiker Andrew Wiles (geb. 1953) vollständig bewiesen. Es bedurfte dazu einer Vielzahl von mathematischen Begriffsbildungen, die zu Zeiten von Fermat noch nicht zur Verfügung standen und von denen teilweise noch vor 30 Jahren niemand ahnte, dass sie eine wesentliche Rolle in der Lösung des Fermat'schen Problems spielen würden.

Zusammenfassung

Die genannten Beispiele illustrieren, dass die Rolle der Mathematik als sprachliche Lupe besonders in der Modellierung zum Tragen kommt. Mit der Fachsprache der Mathematik lassen sich Phänomene präzise beschreiben. Um aus Modellen Prognosen abzuleiten, bedarf es diverser Rechentechniken, insbesondere solcher, mit deren Hilfe Gleichungen gelöst werden können. Während die Modellierung eine Aufgabe ist, die nicht von der Mathematik oder dem Mathematiker allein geleistet werden kann, sondern auf Fachwissen in anderen Disziplinen zurückgreifen muss, ist die Entwicklung von Rechentechniken und Begriffsbildungen der zentrale Gegenstand der Wissenschaft Mathematik. Losgelöst von der Modellierung beobachteter Phänomene schafft sich die Mathematik eine eigene Begriffswelt und

Werkzeuge, mit denen man diese Begriffswelt untersuchen kann. Der Versuch, Gleichungen zu lösen, hat viele neue mathematische Entwicklungen in Gang gesetzt, aber den Begriffen und Techniken, die heute zum Repertoire des Mathematikers gehören, ist diese Abstammung oft nicht mehr anzusehen.

1.2 Begriffsbildung

Am Beispiel des Zahlbegriffs erläutert dieser Abschnitt, wie der Versuch, Gleichungen zu lösen, auf neue mathematische Begriffe führen kann. Als Leitprinzip soll dabei gelten: „Wenn eine Gleichung nicht lösbar ist, erweitere das Zahlensystem so, dass es eine Lösung gibt.“ Diese Darstellung entspricht nicht der historischen Entwicklung des Zahlbegriffs. Diese verlief nicht so zielgerichtet und kontinuierlich entlang eines einheitlichen Leitprinzips und war auch nicht frei von Sackgassen.

Als Ausgangspunkt betrachten wir die *natürlichen Zahlen* $1, 2, 3, \ldots$, ohne ihre Existenz weiter zu begründen.[1] Durch Abzählen von Objekten gelangt man intuitiv zum Konzept der *Addition* natürlicher Zahlen: Zu zwei natürlichen Zahlen a und b bildet man die Summe $a + b$. Stellt man jetzt die Gleichung

$$x + 1 = 1$$

auf, so findet man keine natürliche Zahl x, die sie löst. Jeder heutige Leser sieht sofort, dass $x = 0$ die Gleichung lösen würde, aber die Null wird ursprünglich nicht zu den natürlichen Zahlen gezählt. Welcher enorme Abstraktionsschritt die Erfindung der Null ist, lässt sich erahnen, wenn man bedenkt, dass man viele Jahrhunderte mit den natürlichen Zahlen gerechnet hat, ohne die Null zu kennen. Dann führte man sie zunächst als Leerstellen in der Zahlendarstellung durch Ziffern ein und nochmals Jahrhunderte später begann man mit der Null wirklich zu rechnen, vgl. Kaplan (2003). Wir kürzen den Prozess hier drastisch ab und führen einfach eine Zahl 0, genannt *Null*, ein, für die gelten soll

$$0 + a = a = a + 0$$

für jede natürliche Zahl a und auch für $a = 0$. Insbesondere können wir jetzt die Gleichung $x + 1 = 1$ lösen.

[1] Letztendlich muss man natürlich auch das hinterfragen. In Anhang A.2 (Seite 231) findet sich eine formale Beschreibung der natürlichen Zahlen durch ein Axiomensystem. Auf Seite 187 wird die Existenz der natürlichen Zahlen und die Sinnhaftigkeit dieser formalen Beschreibung im Licht der Gödel'schen Unvollständigkeitssätze diskutiert.

Was, wenn wir stattdessen jetzt die Gleichung

$$x + 2 = 1$$

betrachten? Dann stellen wir fest, dass keine natürliche Zahl x, und auch nicht $x = 0$, die Gleichung löst. Wieder ist dem modernen Leser klar, dass man hier $x = -1$ braucht. Allgemeiner, wenn man in der Lage sein will, alle Gleichungen der Bauart

$$x + a = b$$

mit zwei fest vorgegebenen natürlichen Zahlen zu lösen, dann braucht man die *ganzen Zahlen*

$$\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$$

An dieser Stellen müsste man jetzt Addition und Multiplikation auf den ganzen Zahlen einführen. Dies ist in Anhang A.3 sauber durchgeführt; hier sei an die Intuition oder die Vorbildung der Leser appelliert, um das eigentliche Argument weiterführen zu können. Dies ist eine gängige Vorgehensweise in der Mathematik: „Nehmen wir an, wir hätten dieses Problem schon gelöst, dann könnten wir folgendermaßen vorgehen.“ Auf diese Weise zerteilt man komplexe Probleme in leichter überschaubare Teilprobleme, die man lösen zu können glaubt oder hofft.

Die ganzen Zahlen reichen nicht aus, um die Gleichung

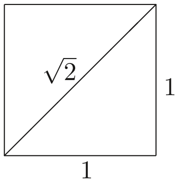
$$2x + 2 = 1$$

zu lösen. Diesmal bräuchte man $x = -\frac{1}{2}$, also muss man Bruchzahlen einführen. Ähnlich wie im Falle der Konstruktion der ganzen Zahlen aus den natürlichen Zahlen würde eine ausführliche Diskussion der Konstruktion der Bruch- oder *rationalen Zahlen* sowie der Addition solcher Zahlen den Gedankengang zu sehr unterbrechen. Dies wird in Anhang A.4 durchgeführt.

Schon die alten Griechen wussten, dass es keine rationale Zahl $x = \frac{m}{n}$ gibt, die die Gleichung

$$x^2 = 2$$

löst. Anders ausgedrückt, die Zahl $\sqrt{2}$ ist *irrational*. Dafür gibt es zahlentheoretische, aber auch geometrische Beweise. Der einfachste zahlentheoretische Beweis (Seite 53) basiert auf der Tatsache, dass man jede natürliche Zahl in eindeutiger Weise als Produkt von Primzahlen schreiben kann. Der Zusammenhang mit der Geometrie ergibt sich aus dem Satz von Pythagoras. Dieser liefert, dass das Quadrat der Länge der Diagonale eines Quadrats mit Seitenlänge 1 gleich $1^2 + 1^2 = \sqrt{2}^2 = 2$ sein muss.



Da die Diagonale eines Quadrats „offensichtlich“ eine Länge haben muss, die Griechen aber keine allgemeineren Zahlen als Bruchzahlen kannten, führte dieses Argument zu einer tiefen Grundlagenkrise der Mathematik. Die Griechen lösten diese Krise, indem sie Geometrie und Arithmetik als separate Gebiete betrachteten, die übereinander nichts zu sagen hatten.

Diese Trennung von Geometrie und Algebra wurde erst von Pierre de Fermat und René Descartes (1596–1650) aufgehoben. Heute sind beide Gebiete eng miteinander verwoben. Allerdings dauerte es bis ins 19. Jahrhundert, bis man die Zahlbereiche so erweitern konnte, dass auch $\sqrt{2}$ und andere aus der Geometrie bekannte Längenverhältnisse zu klar definierbaren Zahlen wurden, die sich auch addieren und multiplizieren ließen. Der resultierende Zahlbegriff ist die *reelle* Zahl. Man führt reelle Zahlen als *Grenzwerte* von rationalen Zahlen (z. B. als unendliche Dezimalbrüche) ein. Grenzwerte sind kein offensichtliches mathematisches Konzept, da sie eine Art Unendlichkeit beinhalten, die nicht leicht präzise zu fassen ist. Eine mathematisch saubere Konstruktion der reellen Zahlen wird im Anhang vorgestellt. Hier begnügen wir uns mit einer sehr brauchbaren Visualisierung der reellen Zahlen, dem *Zahlenstrahl* (Abbildung 1.2). Jeder Punkt auf dem Zahlenstrahl entspricht einer reellen Zahl. Die ganzen Zahlen sind äquidistant wie auf einer Perlenschnur aufgereiht. Die rationalen Zahlen sind auf dem Zahlenstrahl dicht verteilt, das heißt, zwischen zwei beliebigen Punkten auf dem Zahlenstrahl gibt es immer eine rationale Zahl. Andererseits findet sich zwischen zwei rationalen Zahlen immer auch eine, die nicht rational ist.

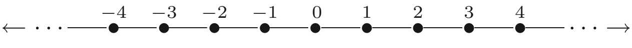


Abb. 1.2 Der Zahlenstrahl

Die reellen Zahlen sind noch nicht der Abschluss der Entwicklung. Die Gleichung

$$x^2 = -1$$

hat keine reelle Lösung x, weil Quadrate reeller Zahlen immer positiv sind. Dieses Mal kommt man auch mit Grenzwerten nicht weiter. Die entscheidende Idee ist, sich vom Zahlenstrahl zu lösen und zu einer Zahlenebene (Abbildung 1.3) überzugehen: Man führt die *komplexen Zahlen* als Paare (a, b) reeller Zahlen ein und stellt sich a und b als *Koordinaten* („Abszisse“ und „Ordinate“) vor. Der reelle Zahlenstrahl ist dann die horizontale Koordinatenachse.

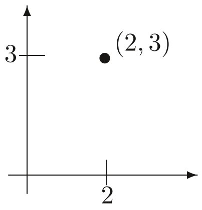


Abb. 1.3 Die Zahlenebene

Man führt auf dieser Zahlenebene wie folgt eine Addition und eine Multiplikation ein:

$$(a,b)+(a',b')=(a+a',b+b') \quad \text{und} \quad (a,b)\cdot(a',b')=(aa'-bb',ab'+ba').$$

Punkte als Symbol für Multiplikation schreibt man nur, wenn es die Notation übersichtlicher macht, meist lässt man sie weg. Damit Schreibweise und Rechenregeln den gewohnten Regeln gleichen, schreibt man $a+\mathrm{i}b$, $a+b\mathrm{i}$ oder auch $a+\mathrm{i}\cdot b$ für (a,b). In obiger Abbildung hätte man dann zum Beispiel $(2,3)=2+3\mathrm{i}$. Mit dieser Notation werden Addition und Multiplikation zu

$$\begin{aligned}(a+\mathrm{i}b) \;+\; (a'+\mathrm{i}b') &= (a+a')+\mathrm{i}(b+b'),\\ (a+\mathrm{i}b) \;\cdot\; (a'+\mathrm{i}b') &= (aa'-bb')+\mathrm{i}(ab'+ba').\end{aligned}$$

Durch die Einführung des Begriffs der komplexen Zahlen wird es möglich, eine Lösung für die Gleichung $x^2+1=0$ explizit anzugeben. Für $(a,b)=(0,1)$ erhält man $a+\mathrm{i}b=0+\mathrm{i}\cdot 1$, was man einfach als i schreibt. Das Quadrat von i ist

$$\mathrm{i}^2=\mathrm{i}\cdot\mathrm{i}=-1+\mathrm{i}\cdot 0=-1.$$

Mit den komplexen Zahlen ist das Leitprinzip dieses Abschnitts im Wesentlichen ausgereizt. Man findet keine aus Additionen und Multiplikationen zusammengesetzten Gleichungen mehr, die sich nicht lösen lassen. Der Grund dafür ist das folgende mathematische Gesetz, für das zuerst Carl Friedrich Gauß (1777–1855) einen strengen Nachweis geliefert hat.

Satz 1.1 (Fundamentalsatz der Algebra)
Jede Gleichung der Form

$$a_k x^k + a_{k-1}x^{k-1} + \ldots + a_1 x + a_0 = 0 \tag{1.2}$$

mit einer natürlichen Zahl k und komplexen Zahlen $a_1,\ldots,a_k$ hat eine komplexe Lösung x, wenn $a_k \neq 0$ gilt.

Verfolgt man die oben vorgestellte Entwicklung des Zahlenbegriffs, erkennt man Prinzipien, die in der Mathematik immer wieder zum Einsatz gelangen:

(1) Man erweitert den begrifflichen Rahmen für ein Problem. Im vorliegenden Fall startet man mit einer Gleichung für natürliche Zahlen ($x+1=1$) und gelangt durch ständige Erweiterung des Zahlbegriffs zu den komplexen Zahlen.
(2) Man erweitert den Lösungsbegriff. Lösungen werden aus den neu gefundenen, erweiterten Zahlenräumen gewonnen.
(3) Man untersucht, ob sich unter den verallgemeinerten Lösungen (im erweiterten Zahlenraum) eine Lösung des Problems im ursprünglichen Zahlenraum befindet. Im Falle des Fermat'schen Problems (1.1) von Seite 5 findet man alle komplexen Lösungen und hat damit Objekte in der Hand, die sich genauer studieren lassen. Letztendlich zeigt sich, dass diese komplexen Lösungen niemals natürliche Zahlen sein können.
In anderen Problemen ist es so, dass man tatsächlich Bedingungen beschreiben kann, unter denen eine verallgemeinerte Lösung schon eine Lösung des ursprünglichen Problems ist. Solche Phänomene findet man zum Beispiel auch für Differenzialgleichungen, wie sie in der Physik bei der Beschreibung von Wellen oder Wärmeleitung vorkommen. In diesem Kontext nennt man die verallgemeinerten Lösungen auch *schwache Lösungen* (Seite 194).

Die zusätzlichen Einsichten in die Natur der Zahlen, die sich aus der Untersuchung von Gleichungen ergeben, lieferten auch ganz unerwartete Lösungen klassischer geometrischer Probleme: Gleichungen der Form wie (1.2) heißen *Polynomgleichungen*. Man nennt die komplexen Zahlen, die Lösungen von Polynomgleichungen mit rationalen Koeffizienten $a_1, \ldots, a_k$ sind, auch *algebraische Zahlen*. Zahlen, die nicht algebraisch sind, heißen *transzendent*. Die Kreiszahl π, die das Verhältnis von Umfang und Durchmesser eines Kreises angibt, ist eine solche transzendente Zahl. Dies wurde zuerst von Carl Louis Ferdinand von Lindemann (1852–1939) gezeigt. Mit der Transzendenz von π lässt sich relativ leicht die Unmöglichkeit der Kreisquadrierung mit Zirkel und Lineal beweisen (Seite 206).

1.3 Die Rolle der Abstraktion in der Mathematik

Der grundlegende Ansatz der modernen Mathematik ist es, Dinge über ihre Eigenschaften zu beschreiben und sich dabei möglichst auf diejenigen Eigenschaften zu beschränken, die für die zu behandelnde Frage wirklich relevant sind. Wenn man die Umlaufbahn eines Raumgleiters beschreiben will, dann betrachtet man ihn abstrahierend als einen bewegten Punkt (den Schwerpunkt). Will man den Gleiter von einer Umlaufbahn in eine andere steuern, muss man ihn als dreidimensionales

Objekt auffassen mit ausgezeichneten Richtungen, in die die Steuerraketen Schub ausüben. Beim Andocken an eine Raumstation spielt die genaue Form des Gleiters eine Rolle und beim Eintauchen in die Atmosphäre auch noch die Hitzebeständigkeit des Materials. Dabei ist jeweils zu testen, ob man das Modell fein genug gewählt hat. So spielen bei Satellitenbahnen Luftreibungseffekte durchaus eine Rolle und führen auf die Dauer zu Abstürzen, wenn man nicht gegensteuert.

Die Abstraktion vom Raumgleiter auf ein Objekt mit einigen klar festgelegten Eigenschaften erleichtert es, Ähnlichkeiten mit in anderen Zusammenhängen gefundenen Beschreibungen und Lösungen zu erkennen und zu benutzen. Auch Wurfgeschosse oder Planeten kann man durch bewegte Massepunkte modellieren, in der Aerodynamik von Tragflächen verfügt man über eine lange Erfahrung, und die Hitzebeständigkeit von Kacheln betrachtet man auch nicht erst seit dem Eintritt ins Raumfahrtzeitalter.

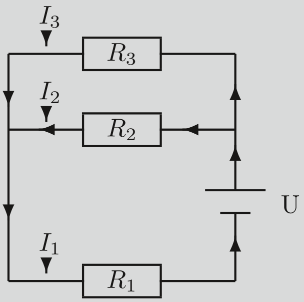
Der hohe Abstraktionsgrad ist für viele die höchste Hürde bei der Annäherung an die Mathematik. Das gilt auch für Studierende der Mathematik, die oft davon überrascht sind, dass an der Universität nur selten gerechnet, dafür aber sehr viel definiert und bewiesen wird. Der folgende Abschnitt gibt den Lesern die Gelegenheit, anhand einer Reihe von Beispielen die Vorteile der Abstraktion in der Mathematik zu erkennen.

Elemente der Abstraktion: Vereinfachung und Übertragbarkeit

Die zwei folgenden Beispiele zeigen, dass Modellierungen von Phänomenen aus völlig unterschiedlichen Bereichen auf identische mathematische Fragestellungen führen können. In Beispiel 1.2 wird beschrieben, wie in einem Stromkreis Spannung, Widerstände und Stromstärken zusammenhängen. In Beispiel 1.3 geht es dagegen um ein einfaches Produktionsmodell aus der Volkswirtschaft, in dem drei Firmen jeweils ein Produkt anbieten. Man möchte wissen, wie viel jede Firma produzieren muss, damit sowohl die externe Nachfrage als auch der Bedarf der beiden anderen Firmen an dem Produkt befriedigt werden können, aber nichts auf Halde produziert wird. In beiden Modellen setzt man gewisse Größen als bekannt voraus und möchte die anderen daraus berechnen. Im Stromkreis nehmen wir die Widerstände und die Spannungsquelle als bekannt an und wollen die Stromstärken bestimmen. Im Produktionsmodell soll bekannt sein, wie hoch die externe Nachfrage ist und welche Mengen der Produkte jede Firma einsetzen muss, um eine Einheit ihres eigenen Produkts herzustellen. Beide Systeme lassen sich durch Gleichungen modellieren, die die bekannten und die unbekannten Größen miteinander verknüpfen. Für den Stromkreis resultieren sie aus physikalischen Gesetzen, im Falle des Produktionsmodells aus mehr oder weniger plausiblen Modellannahmen.

Wir werden sehen, dass diese Gleichungen eine sehr ähnliche Struktur haben und mit denselben mathematischen Methoden gelöst werden können.

Beispiel 1.2 (Stromkreise)
Gegeben sei ein Stromkreis mit einer Spannungsquelle U und drei Widerständen R_1, R_2, R_3:



Der Zusammenhang zwischen der Spannung U, den Widerständen R_1, R_2, R_3 und den resultierenden Stromstärken I_1, I_2, I_3 ist durch die *Kirchhoff'schen Gesetze* gegeben, die hier auf folgende Gleichungen führen:

$$\begin{aligned} I_1 &= I_2 + I_3, \\ U &= R_1 I_1 + R_2 I_2, \\ 0 &= -R_2 I_2 + R_3 I_3. \end{aligned} \tag{1.3}$$

Dabei sind Spannung und Widerstände bekannt und die Stromstärken zu berechnen. Man kann das machen, indem man eine Gleichung nach einer der gesuchten Größen auflöst, das Ergebnis in die anderen Gleichungen einsetzt und so die Zahl der Gleichungen und der Unbekannten um eins reduziert. Dieses Verfahren wiederholt man und löst die letzte Gleichung nach der einzig verbliebenen Unbekannten auf. Dann setzt man das Ergebnis wieder in die anderen Gleichungen ein und findet so sukzessive auch die anderen Unbekannten. Im physikalisch gesehen realistischen Fall $R_1, R_2, R_3 > 0$ findet man so

$$\begin{aligned} I_2 &= \frac{R_3}{R_2} I_3, \\ I_1 &= I_2 + I_3 = \left(\frac{R_3}{R_2} + 1\right) I_3 = \left(1 + \frac{R_2}{R_3}\right) I_2, \\ U &= \frac{R_1 R_2 + R_2 R_3 + R_3 R_1}{R_2} I_3 \end{aligned}$$

und durch Umformen schließlich

$$
\begin{aligned}
I_3 &= \frac{R_2 U}{R_1 R_2 + R_2 R_3 + R_3 R_1}, \\
I_2 &= \frac{R_3 U}{R_1 R_2 + R_2 R_3 + R_3 R_1}, \\
I_1 &= \frac{(R_2 + R_3) U}{R_1 R_2 + R_2 R_3 + R_3 R_1}.
\end{aligned}
$$

Die folgenden drei Punkte sind charakteristisch für Beispiel 1.2:

(1) Es werden Gleichungen aufgestellt und gelöst.
(2) Die Lösung ist nicht wirklich algorithmisch, das heißt automatisierbar, weil wir nicht vorgeschrieben haben, welche Gleichung nach welcher Unbekannten aufgelöst werden soll.
(3) Für jeden Schaltkreis wird neu entschieden, wie die Gleichungen gelöst werden.

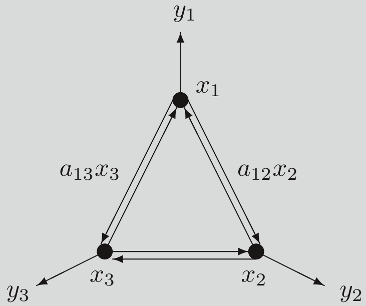
Es ist offensichtlich, dass dieses Vorgehen in (2) und (3) Schwachstellen aufweist. Nach der Betrachtung des Produktionsmodells wird hierfür eine Lösung erkennbar werden.

Beispiel 1.3 (Produktionsmodell nach Leontief)
Betrachten wir ein Produktionssystem bestehend aus drei Produzenten mit je einem Produkt, das nach außen (an den Markt) und untereinander (zur Ermöglichung der Produktion) geliefert wird. Mit x_1, x_2, x_3 werden die produzierten Mengen (einheitlich gemessen z. B. in Geldwert) bezeichnet. Weiter sei y_i die Nachfrage des Marktes nach dem Produkt von i. Wir machen die folgende Annahme: Wenn der j-te Produzent die Menge x_j produziert, dann kauft er vom i-ten Produzenten die Menge $a_{ij}x_j$.

Mit dieser Annahme modelliert man folgenden Umstand: j braucht das Produkt von i als Rohstoff für seine eigene Produktion, und zwar proportional zur Produktionsmenge.

Ziel ist dann die Herstellung eines Gleichgewichts von Produktion und Nachfrage. Das heißt, jede Firma produziert genau so viel, wie Markt und Partnerfirmen ihr abnehmen.

In der folgenden Grafik ist dieses Gleichgewicht dargestellt. Um sie übersichtlich zu halten, sind nur die von Produzent 1 an die anderen Produzenten gelieferten Mengen eingetragen.



Es ergeben sich die Gleichungen

$$
\begin{aligned}
y_1 &= x_1 - a_{12}x_2 - a_{13}x_3,\\
y_2 &= x_2 - a_{21}x_1 - a_{23}x_3,\\
y_3 &= x_3 - a_{31}x_1 - a_{32}x_2,
\end{aligned}
\tag{1.4}
$$

wobei die zum (technischen) Produktionsprozess gehörige Kenngrößen a_{ij} und die Nachfragen y_i als bekannt vorausgesetzt werden. Gesucht sind dann die Produktionsmengen x_i.

Die mathematischen Modellierungen in den Beispielen 1.2 und 1.3 liefern die folgenden Gleichungen, die nun näher verglichen werden. Eine Umformung macht die Analogien deutlich sichtbar:

$$
\begin{aligned}
I_1 &= I_2 + I_3 & \qquad y_1 &= x_1 - a_{12}x_2 - a_{13}x_3\\
U &= R_1I_1 + R_2I_2 & \qquad y_2 &= x_2 - a_{21}x_1 - a_{23}x_3\\
R_2I_2 &= R_3I_3 & \qquad y_3 &= x_3 - a_{31}x_1 - a_{32}x_2
\end{aligned}
$$

$\downarrow$ umstellen $\downarrow$

$$
\begin{aligned}
1\cdot I_1 + (-1)I_2 + (-1)I_3 &= 0 & \qquad 1\cdot x_1 + (-a_{12})x_2 + (-a_{13})x_3 &= y_1\\
R_1I_1 + R_2I_2 + 0I_3 &= U & \qquad (-a_{21})x_1 + x_2 + (-a_{23})x_3 &= y_2\\
0\cdot I_1 + R_2I_2 + (-1)R_3I_3 &= 0 & \qquad (-a_{31})x_1 + (-a_{32})x_2 + x_3 &= y_3
\end{aligned}
$$

$\downarrow$ die gesuchten Größen weglassen $\downarrow$

$$\begin{array}{ccc|c} 1 & -1 & -1 & 0 \\ R_1 & R_2 & 0 & U \\ 0 & R_2 & -R_3 & 0 \end{array} \qquad \begin{array}{ccc|c} 1 & -a_{12} & -a_{13} & y_1 \\ -a_{21} & 1 & -a_{23} & y_2 \\ -a_{31} & -a_{32} & 1 & y_3 \end{array}$$

Solche rechteckigen Zahlenschemata nennt man *Matrizen*. Sie enthalten die vollständige Information über die Gleichungssysteme, aus denen sie gewonnen wurden: Wir können die Gleichungssysteme aus den Matrizen rekonstruieren, weil die Unbekannten alle nur linear in die Gleichung eingehen, das heißt keine Produkte oder noch kompliziertere Funktionen von Variablen vorkommen. Man nennt solche Gleichungssysteme daher *linear*. Für diese Systeme gibt es als Berechnungsmethode eine systematisierte Variante des sukzessiven Variableneliminierens aus Beispiel 1.2, den sogenannten *Gauß-Algorithmus*, mithilfe dessen man durch Manipulation dieser Matrizen die Unbekannten bestimmen kann. Der Gauß-Algorithmus funktioniert für beliebige Matrizen mit beliebig vielen Zeilen und Spalten. Das ist durchaus relevant, denn in praktischen Anwendungen wie der Simulation von Crashtests kommen Gleichungssysteme mit mehr als einer Million Variablen vor. Ehrlicherweise muss man zugeben, dass der Gauß-Algorithmus in solchen Anwendung nicht eingesetzt wird, weil die beschränkte Genauigkeit von Zahlendarstellungen im Computer zwangsläufig zu Rundungsfehlern führt und der Gauß-Algorithmus bei großen Systemen zuviele solcher Rundungsfehler produziert (siehe Beispiel 2.38). Für die oben gewählten Beispiele führen sie jedoch zu den gesuchten Lösungen.

Zusammenfassend kann man sagen, dass die Beispiele 1.2 und 1.3 auf dieselbe mathematische Problemstellung führen: *lineare Gleichungssysteme* und *Matrizenrechnung*. Der Übergang von den Beispielen zu diesen abstrakten Strukturen trägt sowohl den Aspekt der Vereinfachung als auch den Aspekt der Übertragbarkeit in sich:

(1) Mit dem Übergang von den Gleichungen zu den Matrizen hat man nur redundante Information aus den Gleichungen entfernt und gleichzeitig die Übersichtlichkeit erhöht.
(2) Die Anzahl der Gleichungen und der Unbekannten ist für das Vorgehen hier völlig unerheblich, es lassen sich die Lösungsstrategien also sofort auf beliebig große andere Beispiele übertragen.

Die angeführten Beispiele hätten möglicherweise auch durch ganz andere mathematische Beschreibungen modelliert werden können. Dass man die hier betrachteten Gleichungen eingesetzt hat, liegt an einer strategischen Abwägung: Wir haben in Abschnitt 1.2 schon gesehen, dass man Gleichungen keineswegs immer lösen kann, nicht einmal im Prinzip. Daher ist ein Modell, das auf Gleichungen führt, die mit dem heutigen Wissen tatsächlich lösbar sind, einem Modell vorzuziehen, für dessen Gleichungen man keine Lösungsmethoden hat. Es sind durchaus Situationen denkbar, in denen man ein grobes Modell, dessen Prognosen leicht zu bekommen sind, einem sehr akkuraten Modell vorzieht, aus dem man aber keine Prognosen ableiten kann. Selbst wenn sich die Prognosen des besseren Modells näherungsweise bestimmen lassen, aber vielleicht nur unter hohem Kostenaufwand, wird man das gröbere Modell oft vorziehen. Wie viel Auswahl unter verschiedenen Modellen man hat, hängt natürlich stark vom Gebiet ab. In der Physik sind da weit engere Grenzen gesetzt als in der Ökonomie.

Ergebnis von Abstraktion: Neue Strukturen

Konkrete Problemstellungen führen oft in natürlicher Weise auf neue, abstrakte (algebraische) Strukturen. Dieser Sachverhalt lässt sich gut mit der Problemstellung „Teilbarkeitsregeln" illustrieren.
Schüler lernen, dass eine Zahl genau dann durch 2 bzw. 5 teilbar ist, wenn die letzte Ziffer der Dezimaldarstellung durch 2 bzw. 5 teilbar ist. Um festzustellen, ob eine Zahl durch 4 teilbar ist, muss man die letzten beiden Ziffern betrachten. Allgemein bekannt ist auch die *Quersummenregel*, mit der man feststellen kann, ob eine Zahl durch 3 bzw. 9 teilbar ist: nämlich genau dann, wenn die Quersumme durch 3 bzw. 9 teilbar ist. In der Regel wird in der Schule aber nichts darüber gesagt, wie man einer Zahl ansieht, ob sie durch 7 teilbar ist. Es stellt sich die Frage: Gibt es eine Teilbarkeitsregel für 7?

Die Quersummenregel für die Teilbarkeit durch 3 legt nahe, dass Teilbarkeitsregeln mit der Darstellung der Zahlen im Zehnersystem zusammenhängen, denn schließlich addiert man bei einer Quersumme die Einer, die Zehner, die Hunderter etc. Der Schlüssel zu den Teilbarkeitsregeln ist dann die Operation des *Teilens mit Rest*, die auch oft *Division mit Rest* genannt wird. Wenn man 10 durch 3 teilt, bleibt ein Rest von 1, denn $10 = 3 \cdot 3 + 1$. Teilt man 100 durch 3, bleibt wegen $100 = 33 \cdot 3 + 1$ wieder ein Rest von 1. Wir werden sehen, dass für alle Zehnerpotenzen beim Teilen durch 3 ein Rest von 1 bleibt und sich daraus die Quersummenregel ergibt.

Um eine Teilbarkeitsregel für 7 zu finden, teilen wir jede Zehnerpotenz durch 7 und betrachten den Rest. Als Reste kommen nur $0, 1, 2, 3, 4, 5, 6$ infrage. Wenn man 10 durch 7 teilt, bleibt ein Rest von 3, denn $10 = 1 \cdot 7 + 3$. Teilt man 100 durch 7, bleibt wegen $100 = 14 \cdot 7 + 2$ ein Rest von 2. Für $1\,000 = 142 \cdot 7 + 6$ erhalten wir 6 als Rest. Die Restefolge ist also komplizierter als bei der Division durch 3. Es ergibt sich aber trotzdem ein Schema: Für $10\,000$ findet man den Rest 4, für $100\,000$ den Rest 5. In Beispiel 1.4 wird ein Argument dafür angegeben, dass sich ab da die Reste wiederholen. Das heißt, die Restefolge für $1, 10, 100, 1\,000, 10\,000, 100\,000, 1\,000\,000$ etc. ist $1, 3, 2, 6, 4, 5, 1$ etc.

$$\begin{aligned}
1 &= 0+1 = 0\cdot 7+1\\
10 &= 7+3 = 1\cdot 7+3\\
100 &= 98+2 = 14\cdot 7+2\\
1\,000 &= 994+6 = 142\cdot 7+6\\
10\,000 &= 9\,996+4 = 1\,428\cdot 7+4\\
100\,000 &= 99\,995+5 = 14\,285\cdot 7+5\\
1\,000\,000 &= 999\,999+1 = 142\,857\cdot 7+1
\end{aligned}$$

Aus dieser Information lässt sich ein Analogon der Quersummenregel ableiten. Die Details der Herleitung finden sich in den Beispielen 1.4 und 1.5. Der wesentliche Punkt dabei ist die Einteilung der ganzen Zahlen in sieben Klassen, die sogenannten *Restklassen modulo* 7. Zwei Zahlen gehören dabei zur selben Restklasse, wenn sie bei Division durch 7 denselben Rest ergeben oder, was dasselbe ist, wenn die Differenz der Zahlen durch 7 teilbar ist (Abbildung 1.4).

Gewichtete Quersummenregel für 7: Man multipliziere die

Einer mit 1,
Zehner mit 3,
Hunderter mit 2,
Tausender mit 6,
Zehntausender mit 4,
Hunderttausender mit 5
und dann von vorn etc.

Durch Aufaddieren der Ergebnisse erhält man eine gewichtete Quersumme. Die Zahl ist durch 7 teilbar genau dann, wenn die gewichtete Quersumme durch 7 teilbar ist.

In Bezug auf Teilbarkeitsfragen sind Zahlen in derselben Restklasse gleichwertig (äquivalent), das heißt, für solche Fragen muss man statt der unendlich vielen ganzen Zahlen nur noch sieben Klassen betrachten, was eine dramatische Vereinfachung darstellt.
Es stellt sich heraus, dass die Menge der Restklassen zusätzliche Struktur hat. Man kann nämlich darauf eine Addition und eine Multiplikation einführen (Abbildung 1.5). Diese neue Struktur ermöglicht es, die angesprochene periodische Struktur der Reste der Zehnerpotenzen zu finden.

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
-14	-13	-12	-11	-10	-9	-8
-7	-6	-5	-4	-3	-2	-1
0	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Abb. 1.4 Aufteilung der ganzen Zahlen in Restklassen modulo 7

Beispiel 1.4 (Teilbarkeitsregeln I)
Sei n eine ganze Zahl und m eine natürliche Zahl. Dann kann man immer zwei ganze Zahlen k und r mit $0 \leq r < m$ finden, für die gilt

$$n = k \cdot m + r. \tag{1.5}$$

Das bedeutet, wenn man n durch m teilt, bleibt der Rest r. Wir werden diese Tatsache in Abschnitt 1.6 beweisen (siehe Satz 1.14), aber hier benutzen wir sie einfach. Man schreibt

$$n \equiv r \mod m \quad \text{oder} \quad n \equiv_m r,$$

wenn (1.5) gilt (auch wenn r nicht zwischen 0 und $m-1$ liegt), und liest „n ist äquivalent zu r *modulo* m“. Eine Zahl n ist durch m *teilbar*, wenn sie bei Teilung durch m den Rest 0 liefert, das heißt wenn $n \equiv 0 \mod m$ gilt.

Die entscheidende Beobachtung für die Herleitung von Teilbarkeitsregeln ist, dass die Operation Teilen mit Rest mit Addition und Multiplikation verträglich ist: Aus $n = k \cdot m + r$ und $n' = k' \cdot m + r'$ folgt

$$n + n' = (k + k') \cdot m + (r + r') \quad \text{und} \quad nn' = (kk'm + kr' + k'r) \cdot m + rr'.$$

Wenn man also $n + n'$ durch m teilt, bleibt derselbe Rest, wie wenn man $r + r'$ durch m teilt. Analog gilt für die Multiplikation: Wenn man nn' durch m teilt, bleibt derselbe Rest, wie wenn man rr' durch m teilt. In Kurzform:

$$n + n' \equiv r + r \mod m \quad \text{und} \quad nn' \equiv rr' \mod m. \tag{1.6}$$

Sei $[k]$ die *Restklasse* $\{n \mid n \equiv k \mod m\}$ aller Zahlen, die bei Division durch m denselben Rest haben wie k. Diese m Restklassen modulo m lassen sich als m Punkte auffassen, für die wegen der beiden Gleichungen in (1.6) eine Addition und eine Multiplikation wie folgt definiert werden können (Abbildung 1.5):

$$[k] + [k'] = [k + k'] \quad \text{und} \quad [k] \cdot [k'] = [kk']. \tag{1.7}$$

Das heißt, man kann für festes m bei Division durch m mit Resten rechnen.

Insbesondere gilt für jede natürliche Zahl n mit $n \equiv r \mod 7$, dass $10 \cdot n \equiv 3 \cdot r \mod 7$. Das erklärt die periodische Struktur der Reste von Zehnerpotenzen bei Division mit Rest, weil man beim Übergang von einer Zehnerpotenz zur nächsten immer nur den Rest mit 3 multiplizieren und dann den Rest bei Division durch 7 bestimmen muss:

$$\begin{array}{rcrlcll}
1 & = & 1 & \equiv_7 & & & \mathbf{1} \\
10 & = & 10 \cdot 1 & \equiv_7 & 3 \cdot \mathbf{1} & \equiv_7 & \mathbf{3} \\
100 & = & 10 \cdot 10 & \equiv_7 & 3 \cdot \mathbf{3} & \equiv_7 & \mathbf{2} \\
1\,000 & = & 10 \cdot 100 & \equiv_7 & 3 \cdot \mathbf{2} & \equiv_7 & \mathbf{6} \\
10\,000 & = & 10 \cdot 1\,000 & \equiv_7 & 3 \cdot \mathbf{6} & \equiv_7 & \mathbf{4} \\
100\,000 & = & 10 \cdot 10\,000 & \equiv_7 & 3 \cdot \mathbf{4} & \equiv_7 & \mathbf{5} \\
1\,000\,000 & = & 10 \cdot 100\,000 & \equiv_7 & 3 \cdot \mathbf{5} & \equiv_7 & \mathbf{1}
\end{array}$$

Ab hier wiederholen sich die Reste der Zehnerpotenzen modulo 7:

$$\begin{aligned}
10\,000\,000 &= 10 \cdot 1\,000\,000 \equiv 3 \cdot \mathbf{1} \mod 7 \equiv \mathbf{3} \mod 7 \\
100\,000\,000 &= 10 \cdot 10\,000\,000 \equiv 3 \cdot \mathbf{3} \mod 7 \equiv \mathbf{2} \mod 7 \\
&\text{etc.}
\end{aligned}$$

Man schreibt jetzt eine beliebige Zahl im Zehnersystem, das heißt als gewichtete Summe von Zehnerpotenzen, zum Beispiel

$$94\,325 = 9 \cdot 10\,000 + 4 \cdot 1\,000 + 3 \cdot 100 + 2 \cdot 10 + 5 \cdot 1,$$

und rechnet die Reste modulo 7 aus:

$$94\,325 \mod 7 \equiv 9 \cdot 4 + 4 \cdot 6 + 3 \cdot 2 + 2 \cdot 3 + 5 \cdot 1 \mod 7.$$

Die Zahl 94 325 ist also durch 7 teilbar, weil

$$9 \cdot 4 + 4 \cdot 6 + 3 \cdot 2 + 2 \cdot 3 + 5 \cdot 1 = 77$$

durch 7 teilbar ist.

Anders ausgedrückt, multipliziert man die Einer, Zehner, Hunderter etc. mit den Gewichten $1, 3, 2$ etc., dann hat die gewichtete Quersumme denselben Rest modulo 7 wie die ursprüngliche Zahl. Damit ist man bei der oben beschriebenen gewichteten Quersummenregel angekommen.

+	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

·	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

Abb. 1.5 Addition und Multiplikation der Restklassen modulo 7

Die beschriebene Vorgehensweise zur Bestimmung einer gewichteten Quersummenregel lässt sich sofort auf beliebige andere Zahlen übertragen. Für die Zahl 11 ergeben sich als Reste der Zehnerpotenzen zum Beispiel abwechselnd immer 1 und 10. Die resultierende gewichtete Quersummenregel ist also:

> *Gewichtete Quersummenregel* für 11: Man multipliziere die
>
> Einer, Hunderter, Zehntausender etc. mit 1,
> Zehner, Tausender, Hunderttausender etc. mit 10.
>
> Durch Aufaddieren der Ergebnisse erhält man eine gewichtete Quersumme. Die Zahl ist durch 11 teilbar genau dann, wenn die gewichtete Quersumme durch 11 teilbar ist.

Die 11er-Quersummenregel lässt sich noch vereinfachen, wenn man sich klar macht, dass 10 und -1 bei Division durch 11 den selben Rest ergeben. Damit bekommt man die alternierende Quersummenregel:

> *Alternierende Quersummenregel* für 11: Man multipliziere die
>
> Einer, Hunderter, Zehntausender etc. mit 1,
> Zehner, Tausender, Hunderttausender etc. mit -1.
>
> Durch Aufaddieren der Ergebnisse erhält man die *alternierende* Quersumme. Die Zahl ist durch 11 teilbar genau dann, wenn die alternierende Quersumme durch 11 teilbar ist.

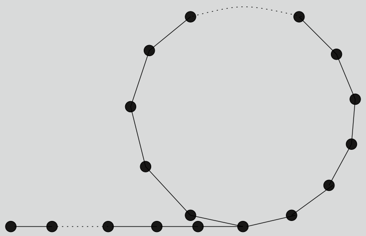
Auch die Teilbarkeitsregel für die Zahl 2 ergibt sich aus der gewichteten Quersummenregel, die die oben vorgestellte Methode für 2 liefert. Da nämlich bis auf

$1 = 10^0$ alle Zehnerpotenzen durch 2 teilbar sind, sind deren Gewichte alle gleich 0. Das heißt, die gewichtete Quersumme ist genau der Einer.

Für die Zahl 4 weiß man, dass sich die Teilbarkeit einer Zahl an den letzten beiden Ziffern ablesen lässt. Unsere Methode liefert ab dem Hunderter das Gewicht 0, das heißt, in der gewichteten Quersumme kommen nur Einer und Zehner vor. Die Gewichte sind dabei 1 und 2. Die Regel aus der Schule würde man erhalten, wenn man als Gewicht für den Zehner 10 statt 2 nähme, was ja bei Division mit 4 denselben Rest liefert.

Beispiel 1.5 (Teilbarkeitsregeln II)
Für eine beliebige natürliche Zahl n findet man nach Beispiel 1.4 die passende gewichtete Quersummenregel wie folgt:

1. Schritt: Bestimme die Reste der 10er-Potenzen bei Division mit m. Da es nur endlich viele Reste gibt, ergibt sich nach einem endlichen „Anlauf“ eine periodische Struktur.



Dass es für 7 keinen Anlauf in der periodischen Struktur gibt, liegt daran, dass 7 eine Primzahl ist und kein Teiler von 10.

2. Schritt: Schreibe eine Zahl n in Dezimaldarstellung, das heißt in der Form

$$n = a_j 10^j + a_{j-1} 10^{j-1} + \ldots + a_1 10 + a_0,$$

und berechne den Rest von n bei Division durch m über die Reste der $a_0, a_1, \ldots, a_j$, gewichtet mit den Resten der $1, 10, \ldots, 10^j$.

Wendet man das in Beispiel 1.5 beschriebene Verfahren auf die Zahlen $5, 8$ und 9 an, findet man ohne Probleme auch die üblichen Teilbarkeitsregeln für

$$\begin{aligned} &5 \ \ \text{(letzte Ziffer)},\\ &8 \ \ \text{(letzten drei Ziffern)},\\ &9 \ \ \text{(Quersumme)}. \end{aligned}$$

Auf die gleiche Weise erhält man auch solche Teilbarkeitsregeln für jede andere Zahl.

Die Diskussion von Teilbarkeitsregeln hat auf neue Strukturen geführt, nämlich die Restklassen modulo m zusammen mit ihren Additionen und Multiplikationen. Die Eigenschaften dieser Strukturen erlaubten es, das Ausgangsproblem und weitreichende Verallgemeinerungen davon zu lösen. Außerdem helfen sie, die unterschiedlichen Teilbarkeitsregeln als verschiedene Ausprägungen ein und desselben Prinzips zu verstehen. Man nennt diese Art von Struktur, die aus einer *Menge*, das heißt einer Ansammlung von verschiedenen Objekten, und *Verknüpfungen* wie $\cdot$ oder $+$ darauf besteht, eine *algebraische Struktur*. Je nach Anzahl der Verknüpfungen und ihren spezifischen Eigenschaften bekommen die algebraischen Strukturen unterschiedliche Namen wie beispielsweise *Gruppen*, *Ringe* und *Körper*. Man sollte nicht versuchen, aus diesen Namen Eigenschaften der Strukturen abzulesen, die Namensgebung entspringt oft einer spontanen Laune der Namensgeber und spiegelt manchmal auch nur deren Humor wider. Die Restklassen sind Beispiele für Ringe, ebenso wie die ganzen Zahlen $\mathbb{Z}$ mit ihrer Addition und ihrer Multiplikation (siehe Beispiel 2.9).

Algebraische Strukturen spielen eine ganz zentrale Rolle in der modernen Mathematik, sind aber eine relativ neue Erfindung. Erste Beispiele – neben den Zahlen – tauchten im 19. Jahrhundert auf; seit dem frühen 20. Jahrhundert erlebt die Mathematik eine Algebraisierung aller Teilbereiche. Im Folgenden zeigen wir wie man den geometrischen Begriff der Symmetrie algebraisieren kann, wodurch er auch für andere Bereiche nutzbar gemacht wird.

Jeder erkennt Symmetrien in der Natur, der Architektur, der Malerei oder auch in der Musik. Dabei ist es gar nicht so einfach zu definieren, was Symmetrie ist, ohne auf mathematische Begriffsbildungen zurückzugreifen. Ein guter Einstieg in diese Frage ist die elementare Geometrie. Wir betrachten regelmäßige Vielecke:
Wie kann man die offensichtlichen Symmetrieeigenschaften dieser Figuren präzise beschreiben? Hier ist eine erste Antwort: Eine *Symmetrieoperation* ist eine Bewegung, die die Figur mit sich selbst zur Deckung bringt. Das können zum Beispiel Spiegelungen an den eingezeichneten Achsen sein oder Drehungen um die Achsenschnittpunkte. Solche Symmetrieoperationen kann man hintereinander ausführen

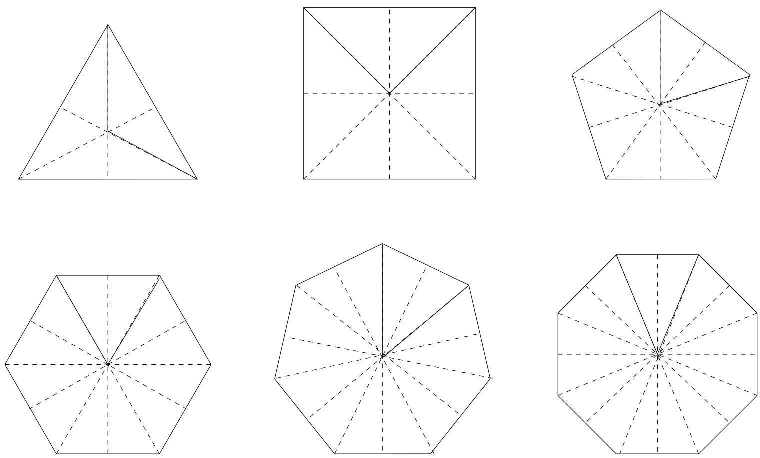


Abb. 1.6 Symmetrien von Vielecken

und findet dann wieder eine Symmetrieoperation. Verknüpft man in dieser Weise zwei Spiegelungen eines Vierecks wie in Abbildung 1.7, erhält man eine Rotation, und zwar um 90 Grad, das Doppelte des Winkels, in dem die beiden Achsen aufeinander stehen.

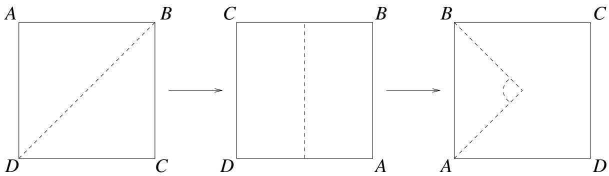


Abb. 1.7 Verknüpfung von Symmetrien

Die Gesamtheit aller Symmetrieoperationen beschreibt dann die *Symmetrie* des Objekts. Wieder ergibt sich eine algebraische Struktur, diesmal mit nur einer Verknüpfung, der Hintereinanderausführung der Symmetriebewegungen. Es handelt sich bei der Menge der Symmetrieoperationen um eine *Gruppe*, weswegen man auch von der *Symmetriegruppe* des Objekts spricht. Es ist allerdings bisher nicht wirklich klar, was genau man unter einer Bewegung zu verstehen hat, und so bleibt der Begriff „Symmetrie" zunächst immer noch vage. Abhilfe schafft hier der Begriff der *Abbildung*, ein Begriff, der in der Mathematik absolut zentral ist.

Wenn A und B zwei Mengen sind, dann ist eine Abbildung $f\colon A \to B$ eine Vorschrift, die jedem Element a von A ein Element $f(a) = b$ von B zuordnet, das man dann das *Bild* von a unter f nennt. Dies ist eine Abstraktion der Vorstellungen „Vergleichen", „Messen" und „Klassifizieren". So könnten die Elemente der Mengen zum Beispiel Punkte auf einem Vieleck oder auch reelle Zahlen sein. Wenn A und B Mengen ähnlicher Natur sind, entspricht das eher der Vorstellung vom

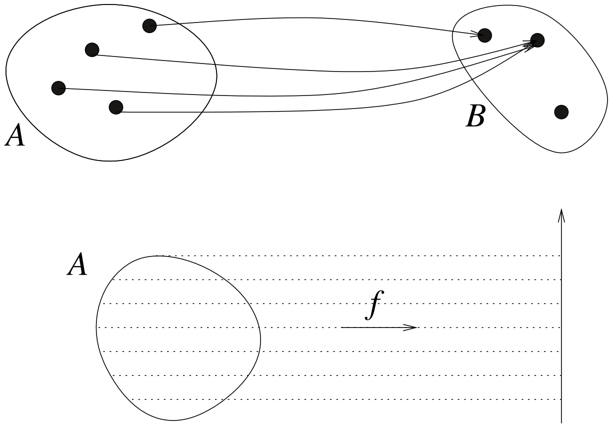


Abb. 1.8 Abbildungen

Vergleichen. Man kann sich beispielsweise vorstellen, dass A und B Flächen sind, die übereinandergelegt werden. Wenn B eine Menge von Zahlen ist, dann ist die Abbildung eher so etwas wie das Anlegen eines Maßstabs. Hier kann man sich A zum Beispiel als Menge von Personen vorstellen, deren Körpergröße durch die Abbildung angegeben wird. Wenn schließlich A eine Menge von Personen ist und B eine Menge von Merkmalen wie etwa Haarfarben, dann teilt die Abbildung, die jeder Person ihre Haarfarbe zuordnet, die Personen in Klassen ein.

Natürlich sind auch alle Funktionen wie sin, cos, exp, x^k etc. Abbildungen. In diesen Beispielen sind A und B Mengen von Zahlen. Weitere Beispiele findet man in Abschnitt 1.4, wo die elementaren Begriffe der Mengenlehre etwas genauer diskutiert werden und dann auch mehr über Abbildungen gesagt wird.

Mithilfe des Begriffs der Abbildung lässt sich eine präzise Definition von Symmetrie formulieren: Eine Symmetrie von A ist eine Abbildung $f\colon A \to A$, für die jedes Element von A als Bild von genau einem Element von A auftaucht. In einer Visualisierung durch Punkte und Pfeile wie in Abbildung 1.8 wird dann jedes Element von genau einem Pfeil getroffen. An dieser Stelle kann man dann noch Zusatzforderungen stellen. Man kann zum Beispiel verlangen, dass Geraden auf Geraden abgebildet werden oder Winkel und Abstände erhalten bleiben.

Die Menge aller dieser möglichen Abbildungen stellt eine algebraische Struktur dar, die mit den Restklassen aus dem vorhergehenden Beispiel vergleichbar ist. Auch auf der Menge der Symmetrieabbildungen kann man eine Verknüpfung definieren, nämlich die Hintereinanderausführung (Verknüpfung) der Abbildung wie in Abbildung 1.7.

Bisher haben wir die Entwicklung von abstrakten algebraischen Strukturen aus zwei konkreten Phänomenen (Teilbarkeit und Symmetrie) betrachtet. Nun wird ersichtlich, wie die nächste Stufe der Abstraktion funktioniert. Wir *vergleichen* die beiden algebraischen Strukturen, die wir in den so unterschiedlichen Kontexten gefunden haben.

Als Beispiel nehmen wir die Rotationssymmetrien des n-Ecks und die Addition der Restklassen modulo n. Die Rotation kann man sich vorstellen als die Darstellung der Uhrzeit auf einem Ziffernblatt. In diesem Kontext ist man es gewohnt, dass man nur zwölf verschiedene Stunden hat. 2 Uhr und 14 Uhr sind auf dem Ziffernblatt nicht zu unterscheiden. Man rechnet modulo 12. Geometrisch kann man sich ein Ziffernblatt, auf dem nur die Stunden markiert sind, als ein regelmäßiges Zwölfeck vorstellen. Ließe man den Stundenzeiger fest und rotierte dafür das Ziffernblatt, so entspräche jeder Stundenwechsel einer Rotation von 30 Grad. Nach zwölf solcher Rotationen ist man wieder bei der Ausgangsposition.

Allgemein liefert n-malige Anwendung der Rotation um $\frac{360}{n}$ Grad auf ein regelmäßiges n-Eck die Identität, das heißt die Abbildung, die jedes Element auf sich selbst abbildet. Das entspricht dann einer Rotation um 0 Grad, was an die n-fache Addition von 1 in den Restklassen modulo n aus Beispiel 1.4 erinnert. Die liefert ja gerade n, was modulo n auch wieder 0 ist.

Die Ähnlichkeiten gehen aber noch weiter. Als algebraische Strukturen sind die Rotationssymmetrien des regelmäßiges n-Ecks und die Restklassen modulo n *isomorph* (gleichgestaltig). Beispiel 1.6 liefert eine mathematisch präzise Formulierung dieser Behauptung. Um es nachvollziehen zu können, sollte man sich die Definition von Restklassen und ihrer Addition aus (1.7) in Beispiel 1.4 in Erinnerung rufen.

Beispiel 1.6 (Isomorphie)
Wir vergleichen zwei Strukturen:

(i) die Rotationen R_k des n-Ecks um $k \cdot \frac{360}{n}$ Grad, zusammen mit der Hintereinanderausführung von Rotationen,
(ii) die Restklassen $[k] = \{m \mid m \equiv k \mod n\}$ modulo n, zusammen mit der Addition.

Jeder Rotation R_k kann man die Restklasse $[k]$ zuordnen. Dann gehört jede Restklasse zu genau einer Rotation. Die durch $\alpha(R_k) = [k]$ definierte Abbildung α

führt die algebraische Struktur der Rotationen in die Addition auf den Restklassen über: Es gilt nämlich wegen $R_k \circ R_{k'} = R_{k+k'}$ und $[k + k'] = [k] + [k']$, wobei $\circ$ die Hintereinanderausführung von Abbildungen bezeichnet, dass

$$\alpha(R_k \circ R_{k'}) = \alpha(R_{k+k'}) = [k + k'] = [k] + [k'] = \alpha(R_k) + \alpha(R_{k'}).$$

Man sagt, die beiden Strukturen sind *isomorph* (hier als Gruppen), und stellt fest, dass Aussagen, die die Verknüpfungen betreffen, mit dieser Isomorphie von einer zur anderen Struktur einfach transferiert werden können.

Es mag beim ersten Betrachten gekünstelt und als irrelevante Spielerei erscheinen, hier die Ähnlichkeiten zwischen Rotationen um n-te Teile von 360 Grad und der Teilung durch n mit Rest zu thematisieren. Das Konzept der Abbildung und insbesondere der Isomorphie ist aber der Schlüssel dafür, Einsichten, die man in einem Kontext gewonnen hat, in einem anderen Kontext anwenden zu können.

Zusammenfassung: Abstraktion in der Mathematik

Ein wesentlicher Punkt in jeder mathematischen Betrachtung ist die Konzentration auf die für die gegebene Problemstellung wesentlichen Eigenschaften. Durch Ausblendung aller anderen Aspekte tritt die Ähnlichkeit zwischen Problemstellungen zutage, was die Übertragung von bekannten Lösungsstrategien auf neue Problemstellungen ermöglicht. Mathematische Erkenntnisse veralten nicht. Einmal erreichte Einsichten, auch die der alten Griechen, werden immer wieder neu eingesetzt. Es ergeben sich aber immer auch dieselben Schwierigkeiten:

(1) Es ist in der Regel nicht von vornherein klar, was die wesentlichen Eigenschaften einer Problemstellung sind.
(2) Die Reduktion auf wenige Eigenschaften, das heißt der hohe Abstraktionsgrad, macht die Beschreibungen unanschaulich.
(3) Die Ähnlichkeiten bestehen oft nicht zwischen den Objekten, die in den primären Problemstellungen auftauchen, sondern in den Relationen zwischen diesen Objekten oder noch komplizierteren abgeleiteten Strukturen. Beim Vergleich der Division mit Rest mit der Geometrie von Vielecken sind zum Beispiel die Struktur der Addition der Restklassen und die Struktur der Symmetrien ähnlich.

Der hohe Abstraktionsgrad ist einerseits das Erfolgsgeheimnis der Mathematik, andererseits aber auch der Grund für die hohen Hürden, die der Neuling oder der interessierte Laie zu überwinden hat, wenn er verstehen will, wieso die Mathematik so erfolgreich ist. Dass selbst Physikern dieser Erfolg oft unheimlich ist, kann man in Wigner (1960) nachlesen.

1.4 Sprechen über mathematische Objekte: Die Sprache der Mengenlehre

Die Diskussion von Teilbarkeit und Symmetrie in Abschnitt 1.3 zeigt, dass die Untersuchung und der Vergleich auch relativ einfacher Problemstellungen sehr schnell zu Strukturen führen, die komplex und wenig anschaulich sind. Allein die Symmetrien eines Vierecks sind schon Ansammlungen von Abbildungen, die Punkte und Geraden auf andere Punkte und Geraden abbilden. Bei vielen Fragen, die man zu den Symmetrien haben kann, ist es aber nicht relevant, wie genau diese Symmetrien zustande kommen. Weil man sich dafür interessiert, wie sie sich untereinander verhalten, fasst man sie zu einem neuen mathematischen Objekt zusammen und beschränkt sich auf bestimmte Eigenschaften, zum Beispiel darauf, dass man zwei Symmetrien verknüpfen kann und damit wieder eine Symmetrie erhält. Die Mengenlehre ist eine Sprache, mit der man sehr effizient Aussagen über Ansammlungen von Objekten formulieren kann. Man verwendet sie als ein vereinfachendes Hilfsmittel. Dass die Mengenlehre auch Teil der philosophischen Grundlegung der Mathematik ist und grundsätzliche Probleme aufwirft, erläutern wir hier nur kurz. Wir gehen aber in Abschnitt 4.2 näher auf die Problematik ein.

Die ersten Objekte mathematischer Überlegungen waren Zahlen und geometrische Figuren. Was diese Objekte eigentlich sind, war praktisch von Anfang an Gegenstand intensiver Überlegungen und kontroverser Philosophien. Im Laufe der Zeit kamen weitere mathematische Objekte ganz unterschiedlicher Natur hinzu, wie zum Beispiel Variablen, Gleichungen und Funktionen. Mengen betrachtet man erst seit gut hundert Jahren, aber die moderne Mathematik bedient sich ihrer bei der Beschreibung aller mathematischen Objekte.

Zahlen, Figuren, Funktionen sind Mengen mit gewissen Eigenschaften. Dabei ist keineswegs klar, was eine Menge eigentlich ist. Die „Definition“, die auf Seite 22 verwendet wird, ist im Wesentlichen die des Begründers der Mengenlehre, Georg Cantor (1845–1918): „Unter einer Menge verstehen wir jede Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen.“ Dies ist keine wirkliche Erklärung, weil der unbekannte Begriff *Menge* auf den ebenfalls unbekannten Begriff *Zusammenfassung zu einem Ganzen* zurückgeführt wird. Widersprüchliche Bildungen solcher Zusammenfassungen führen zu logischen Problemen wie dem vom Dorfbarbier, der alle Männer des Dorfes rasiert, die sich nicht selbst rasieren: Wenn sich der Barbier selbst rasiert, darf er sich als Barbier nicht rasieren. Wenn er sich aber nicht selbst rasiert, so muss er sich als Barbier rasieren (vgl. Seite 183). Solche Antinomien führten Anfang des 20. Jahrhunderts in eine Diskussion über die Grundlagen der Mathematik, die bis heute nicht abgeschlossen ist. Für die gegenwärtige Praxis der Mathematik ist vor allem bedeutsam, dass sich die axiomatische Methode, eine Theorie aus nicht hinterfragten Grundtatsachen (Axiomen) unter Benützung festgelegter logischer Regeln aufzubauen, universell durchgesetzt hat. Im Anhang findet man ein exemplarische Darstellung nach diesem Muster: ein Axiomensystem für die natürlichen Zahlen, die Konstruktion der anderen in Abschnitt 1.1 erwähnten Zahlbereiche und die Herleitung deren fundamentaler Eigenschaften.

Auch für die Mengenlehre und die Logik gibt es solche axiomatischen Zugänge, allerdings sind sie ohne mathematische Ausbildung praktisch nicht nachvollziehbar. Daher stützt man sich bei Einführungen in die Mathematik in der Regel auf (möglichst wenige) intuitive Konzepte, aus denen man dann das mathematische Gebäude aufbaut. Diese intuitiven Konzepte werden im Verlauf des Mathematikstudiums in Vorlesungen wie *Axiomatische Mengentheorie* oder *Mathematische Logik* hinterfragt und durch eigene axiomatische Gebäude ersetzt. Der Besuch solcher Vorlesungen ist allerdings selten verpflichtend, und man muss davon ausgehen, dass nur ein Bruchteil der Universitätsabsolventen in den mathematischen Studiengängen je eine solche Veranstaltung besucht hat.

Naive Mengenlehre

Ausgangspunkt für den „naiven" Zugang zur Mengenlehre ist, dass eine *Menge* durch ihre Elemente festgelegt wird: Eine Menge ist gebildet, wenn feststeht, welche Objekte dazugehören. Eine Menge ist also eine Art „Sack", der dadurch bestimmt wird, was er enthält. Die Objekte, die zu einer Menge gehören, heißen *Elemente* der Menge. Wenn M eine Menge ist und a ein Element von M, dann schreibt man $a \in M$. Eine Menge kann man beschreiben, indem man alle ihre Elemente aufzählt oder aber indem man ihre Elemente durch eine Eigenschaft charakterisiert:

$$\{a, b, c, d, e\}$$

ist die Menge der ersten fünf (kleinen) Buchstaben des Alphabets, und

$$\left\{x \;\middle|\; x \in \mathbb{Z} \text{ und } \frac{x}{2} \in \mathbb{Z}\right\}$$

ist die Menge der durch 2 teilbaren ganzen Zahlen (wenn wir akzeptieren, dass $\mathbb{Z}$ die Menge der ganzen Zahlen ist). Die Klammern $\{\quad\}$, die in dieser Schreibweise vorkommen, nennt man *Mengenklammern*. Will man klarstellen, dass eine Menge aus Elementen einer vorgegebenen Menge X besteht, schreibt man auch

$$\{x \in X \mid \text{Eigenschaften von } x\} \quad \text{oder} \quad \{x \in X : \text{Eigenschaften von } x\},$$

zum Beispiel

$$\left\{x \in \mathbb{Z} \;\middle|\; \frac{x}{2} \in \mathbb{Z}\right\},$$

gelesen als „die Menge der Elemente von $\mathbb{Z}$ mit der Eigenschaft, dass $\frac{x}{2} \in \mathbb{Z}$", für die geraden Zahlen. Wenn a kein Element von M ist, schreibt man $a \notin M$. Entsprechend unserem Ausgangspunkt nennen wir zwei Mengen *gleich*, wenn sie die gleichen Elemente enthalten. Also sind die Mengen

$$\{a, b, c, d, e\} \quad \text{und} \quad \{e, d, c, b, a\}$$

gleich, nicht aber die Mengen

$$\{a, b, c, d, e\} \quad \text{und} \quad \{e, d, b, a\}.$$

Wenn A und B Mengen sind, dann heißt A eine *Teilmenge* von B, wenn jedes Element von A auch Element von B ist. Man schreibt dann $A \subseteq B$. Es gilt also

$$\{e, d, b, a\} \subseteq \{a, b, c, d, e\}.$$

Manchmal schreibt man $\subset$ statt $\subseteq$. Es gibt auch Autoren, die nur $A \subset B$ schreiben, wenn $A \subseteq B$ und $A \neq B$ gilt, das heißt wenn A echt kleiner als B ist. Möchte man die „kleinere" Menge rechts stehen haben, schreibt man auch $B \supseteq A$ statt $A \subseteq B$ und $B \supset A$ statt $A \subset B$.

Für jede Teilmenge $A \subseteq B$ kann man ihr *Komplement*

$$\{b \in B \mid b \notin A\}$$

betrachten. Es wird mit $B \setminus A$ – gelesen als „B ohne A" – oder (wenn B aus dem Kontext klar ist) $\complement A$ bezeichnet. Wenn A keine Teilmenge von B ist, schreibt man $A \nsubseteq B$.

Eine sehr wichtige neue Menge, die man aus einer gegebenen Menge konstruieren kann, ist die *Potenzmenge*

$$\mathfrak{P}(M) := \{N \mid N \subseteq M\}$$

von M, deren Elemente alle Teilmengen von M sind. Die Schreibweise $N := B$ verwendet man häufig, um eine neue Bezeichnung N für ein bekanntes Objekt B einzuführen.

Hat man zwei Mengen A und B, so gibt es verschiedene Möglichkeiten, daraus neue Mengen zu konstruieren:

$$\{x \mid x \in A \text{ oder } x \in B\}$$

heißt die *Vereinigung* von A und B und wird mit $A \cup B$ bezeichnet. Eine Doppelung von Elementen ist in der Vereinigungsmenge nicht vorgesehen. Wenn c ein Element sowohl von A als auch von B ist, dann ist c ein Element von $A \cup B$, kommt aber darin nicht doppelt vor.

Die Menge

$$\{x \mid x \in A \text{ und } x \in B\}$$

heißt der *Schnitt* von A und B und wird mit $A \cap B$ bezeichnet. Um sicherzustellen, dass der Schnitt zweier Mengen immer gebildet werden kann, muss man eine besondere Menge zulassen: die *leere Menge*, die überhaupt kein Element enthält und mit $\emptyset$ oder $\{\ \}$ bezeichnet wird. Letztere Bezeichnung unterstützt die Intuition, dass $\{\ \}$ ein leerer Sack ist.

Die Definition von Schnitt und Vereinigung von zwei Mengen lässt sich problemlos auf mehr als zwei Mengen verallgemeinern. Nehmen wir an, wir haben eine ganze Ansammlung von Mengen, von denen jede ein „Namensschild“ bekommt. Die Namen bilden zusammen wieder eine Menge, die hier mit I für *Indexmenge* bezeichnet sei. Jetzt kann man die Ansammlung von Mengen beschreiben: Es sind die Mengen A_i mit $i \in I$. Die *Vereinigung* der A_i ist dann die Menge, die als Elemente alles enthält, was in *irgendeiner* der einzelnen Mengen A_i Element ist:

$$\bigcup_{i \in I} A_i := \{x \mid \text{es gibt ein } i \in I \text{ mit } x \in A_i\}.$$

Man packt quasi die Elemente der A_i alle in einen großen Sack. Dagegen ist der *Schnitt* der A_i die Menge, die als Elemente nur das enthält, was in *jeder* der einzelnen Mengen A_i Element ist:

$$\bigcap_{i \in I} A_i := \{x \mid \text{für alle } i \in I \text{ gilt } x \in A_i\}.$$

Die Vereinigung von *disjunkten* Mengen, das heißt Mengen, deren Schnitt leer ist, bezeichnet man auch mit $A \mathbin{\dot\cup} B$ bzw. $\dot\bigcup_{i \in I} A_i$.

Manchmal möchte man zwei Mengen A und B, die nicht näher bekannt sind, als disjunkt betrachten, das heißt, jedes Element von A bekommt eine unsichtbare Kennzeichnung „zu A gehörig,“ und entsprechend bekommen die Elemente von B die B-Kennzeichnung. Die so präparierten Mengen sind disjunkt, und man kann ihre disjunkte Vereinigung betrachten, die man dann mit $A \amalg B$ bezeichnet und auch die *mengentheoretische Summe* nennt.

Eine weitere Menge, die man aus A und B bauen kann, ist das *kartesische Produkt*, das nach René Descartes benannt ist. Es besteht aus allen geordneten Paaren (a, b) mit $a \in A$ und $b \in B$ und wird mit $A \times B$ bezeichnet:

$$A \times B := \{(a, b) \mid a \in A, b \in B\}.$$

Das kartesische Produkt ist nützlich, wenn es darum geht, Beziehungen zwischen Elementen einer oder mehrerer Mengen zu modellieren. Betrachtet man als Beispiel die Menge H aller Hörer der Vorlesung *Analysis 1* und S, die Menge aller an der Universität Paderborn angebotenen Studienrichtungen, so lässt sich aus der Teilmenge

$$\{(x, F) \in H \times S \mid \; x \,\text{studiert}\, F\}$$

von $H \times S$ ablesen, welcher Hörer für welche Studienrichtung eingeschrieben ist. Allgemein bezeichnet man jede Teilmenge R eines kartesischen Produkts $A \times B$ als eine *Relation zwischen A und B*. Die Interpretation von $(a, b) \in R$ ist: „a steht in Relation R zu b." Man schreibt auch oft aRb statt $(a, b) \in R$. Wenn A gleich B ist, das heißt die Relation aus Elementen von $A \times A$ besteht, spricht man auch von einer Relation *auf* A.

Die Sprache der Mengenlehre hat es ermöglicht, die Situation im Hörsaal mathematisch präzise zu beschreiben. Darüber hinaus gelingt es in der Notation der Mengenlehre, mathematisch präzise Beschreibungen auch von Beziehungen zwischen Objekten zu geben.

Beispiele 1.7 und 1.8 illustrieren solche Beziehungen nicht zwischen Studierenden und ihren Studienfächern, sondern zwischen Zahlen.

Beispiel 1.7
Sei $A = \{1, 3, 5\}$ und $B = \{2, 3, 4\}$. Dann ist

$$R := \{(1, 2), (1, 3), (1, 4), (3, 4)\}$$

eine Relation. Bei genauerem Hinsehen stellt man fest, dass $(a, b) \in A \times B$ genau dann Element der Relation ist, wenn a kleiner als b ist. Wenn man jetzt die Relation $<$ statt R nennt, wird die Schreibweise aRb zu $a < b$. Auf diese Weise erhält man eine saubere mengentheoretische Beschreibung der Relation, ohne auf die gewohnte intuitive Schreibweise verzichten zu müssen.

Beispiel 1.8 (Teilbarkeit)
Sei $\mathbb{N}$ die Menge der natürlichen Zahlen und $a, b \in \mathbb{N}$. Wir sagen a *teilt* b, geschrieben $a \mid b$, falls b ein ganzzahliges Vielfaches von a ist, das heißt falls es ein $c \in \mathbb{N}$ mit $b = c \cdot a$ gibt. Damit definiert $R := \{(a, b) \in \mathbb{N} \times \mathbb{N} : a \mid b\}$ eine Relation auf $\mathbb{N}$.

(i) Die Relation R ist *reflexiv*, das heißt, es gilt $a \mid a$.
(ii) Die Relation R ist *transitiv*, das heißt, aus $a \mid b$ und $b \mid c$ folgt $a \mid c$.
(iii) Aus $a \mid b$ und $b \mid a$ folgt $a = b$.
(iv) Aus $a \mid b$ folgt $a \leq b$. Insbesondere hat jede natürliche Zahl nur endlich viele Teiler.

An dieser Stelle fügen wir einen kleinen Exkurs über *Äquivalenzrelationen* und *Äquivalenzklassen* ein, denn die mengentheoretische Beschreibung von Relationen erlaubt es auch, das im Kontext der Teilbarkeitsregeln diskutierte Prinzip der Vereinfachung durch Beschränkung auf bestimmte Eigenschaften mathematisch präzise zu formulieren. Äquivalenzrelationen sind eine Abstraktion der „Äquivalenz modulo m“ aus Beispiel 1.4, und die dort beschriebenen Restklassen sind Beispiele für Äquivalenzklassen.

Beispiel 1.9 (Restklassen)
Sei $m \in \mathbb{N}$ und $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid a \equiv b \mod m\}$. Man schreibt $m\mathbb{Z}$ für die Menge aller Zahlen der Form $a = mk$ mit $k \in \mathbb{Z}$, das heißt, $m\mathbb{Z}$ ist die Menge der durch m teilbaren ganzen Zahlen. Dann gilt

$$R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid (a - b) \in m\mathbb{Z}\},$$

weil zwei Zahlen bei Division durch m genau dann den gleichen Rest haben, wenn ihre Differenz den Rest 0 hat. Die Relation R ist reflexiv, transitiv und *symmetrisch*, das heißt, aus aRb folgt bRa. Die Restklasse $[k]$ von k modulo m lässt sich auch folgendermaßen beschreiben:

$$[k] = \{a \in \mathbb{Z} \mid aRk\} = k + m\mathbb{Z} = \{k + \ell \mid \ell \in m\mathbb{Z}\}.$$

In der abstrakten Version der *Äquivalenzrelation* werden die Restklassen dann zu *Äquivalenzklassen* (Beispiel 1.10).

Beispiel 1.10 (Äquivalenzrelationen)
Sei M eine Menge. Eine reflexive und transitive Relation R auf M heißt *Äquivalenzrelation*, wenn sie außerdem *symmetrisch* ist, das heißt, aus xRy folgt yRx. Wenn R eine Äquivalenzrelation auf M ist, dann nennt man die Menge

$$\{y \in M \mid xRy\}$$

aller Elemente, die *äquivalent* zu x sind, das heißt bezüglich R in Relation zu x stehen, die *Äquivalenzklasse* von x bezüglich R. Man bezeichnet sie oft mit $[x]_R$ oder einfach nur mit $[x]$, wie bei den Restklassen aus Abschnitt 1.3, wenn aus dem Kontext klar ist, über welche Relation man spricht.

Die Menge aller Äquivalenzklassen wird mit M/R bezeichnet. Die Elemente einer Äquivalenzklasse heißen auch *Repräsentanten* der Äquivalenzklasse. Insbesondere sind zwei Elemente genau dann äquivalent, wenn ihre Äquivalenzklassen gleich sind. Daher ist M die disjunkte Vereinigung der Äquivalenzklassen (siehe Abbildung 1.4).

Wenn man für eine Menge von Objekten einen Katalog von Eigenschaften definiert und Objekte mit gleichen Eigenschaften äquivalent nennt, erhält man eine Äquivalenzrelation, und das Prinzip der Vereinfachung durch Beschränkung auf bestimmte Eigenschaften lässt sich jetzt so formulieren, dass man die Menge der Äquivalenzklassen dazu untersucht.

Abbildungen lassen sich in der Sprache der Mengenlehre als zwischen den Elementen zweier Mengen bestehende Relationen beschreiben. Wenn jedes Element von A zu genau einem Element von B in Relation steht, nennt man so eine Relation eine *Abbildung* oder *Funktion* von A nach B. Die Idee hinter dieser Setzung ist, dass man jedem Element a von A genau ein Element b von B zuordnen will, nämlich dasjenige mit $(a, b) \in R$. Man schreibt dann $R\colon A \to B, a \mapsto R(a)$ und $R(a) = b$ für $(a, b) \in R$. Auch hier dient die veränderte Schreibweise dazu, mengentheoretische Definitionen (in denen nicht von Variablen etc. die Rede ist) und die traditionelle Notation für Funktionen unter einen Hut zu bringen.

Die oben betrachtete Relation zwischen Hörern der Analysis 1 und Studienrichtungen (Bachelor Mathematik, Lehramt an Gymnasien für Mathematik und Physik, Master Germanistik etc.) der Universität Paderborn ist also genau dann eine Funktion, wenn jeder Hörer für eine Studienrichtung eingeschrieben ist, aber nicht für mehrere (jedem Hörer lässt sich in eindeutiger Weise eine Studienrichtung zuordnen). Nicht ausgeschlossen ist durch die Definition der Funktion, dass mehrere Hörer für die gleiche Studienrichtung eingeschrieben sind. Würde man hier von Studienfächern und nicht von Studienrichtungen ausgehen, würde man keine Funktion erhalten, denn alle Lehramtskandidaten studieren mehrere Fächer. Wenn es in der Vorlesung Teilnehmer gibt, die ein echtes Doppelstudium machen, ist die Relation auch keine Funktion. Ebenso verhält es sich, wenn es in der Vorlesung „Schülerstudenten“ gibt, das heißt Schüler, die vollwertige Teilnehmer des Kurses sind, aber mangels Abitur noch nicht als Studierende an der Universität eingeschrieben werden dürfen. In diesem Fall kann nämlich nicht jedem Hörer eine Studienrichtung zugeordnet werden.

Die formale Definition einer Abbildung durch Relationen erlaubt es, die Idee eines Größenvergleichs, die in Abbildung 1.8 anklang, präziser zu fassen: Sei $f\colon A \to B$ eine Abbildung. Man nennt sie *injektiv*, wenn aus $f(a_1) = f(a_2)$ folgt $a_1 = a_2$. In dem Bild mit den Pfeilen bedeutet das, dass es keinen Punkt gibt, an dem zwei Pfeile enden. Anders ausgedrückt, man kann sich A in B eingebettet vorstellen, indem man jeden Punkt a mit seinem Bild $f(a)$ unter f identifiziert (Abbildung 1.9). Dies ist dann eine Möglichkeit, „A kleiner oder gleich B“ zu sagen. Diese Betrachtungsweise ermöglicht es auch, den Größenunterschied noch weiter zu quantifizieren, zum Beispiel indem man zählt, wie viele Elemente von B nicht im Bild von A sind.

Wenn auf der anderen Seite zu jedem $b \in B$ ein $a \in A$ mit $f(a) = b$ existiert, nennt man f *surjektiv*. In dem Bild mit den Pfeilen bedeutet das, in jedem Punkt von B endet mindestens ein Pfeil. Hier kann man sich vorstellen, dass B von A via f vollkommen überdeckt wird, und das ist eine Möglichkeit, „A größer oder gleich B“ zu sagen. Auch hier kann man den Größenunterschied noch weiter quantifizieren, indem man beispielsweise für jedes $b \in B$ zählt, wie viele Elemente von A auf b abgebildet werden.

Wenn eine Abbildung f sowohl injektiv als auch surjektiv ist, dann heißt sie *bijektiv*. Für $A = B$ findet man damit genau den Begriff der Symmetrie aus Beispiel 1.6. In dem Bild mit den Pfeilen bedeutet die Bijektivität, dass von jedem Punkt von A genau ein Pfeil ausgeht und in jedem Punkt von B genau ein Pfeil eingeht. Durch Umdrehen der Pfeile lässt sich also eine Abbildung $B \to A$ definieren, die man die *Umkehrabbildung* von f nennt und mit f^{-1} bezeichnet. Im Ergebnis entspricht dann jedem Punkt von A genau ein Punkt von B, das heißt, dies ist eine Möglichkeit, „A und B sind gleich groß“ zu sagen.

Die Formalisierung des Funktionsbegriffs als Teilmenge eines kartesischen Produkts mit gewissen Eigenschaften ist ein erstes Beispiel dafür, dass die Mengenlehre in der Lage ist, einen einheitlichen Rahmen für zunächst als ganz verschieden

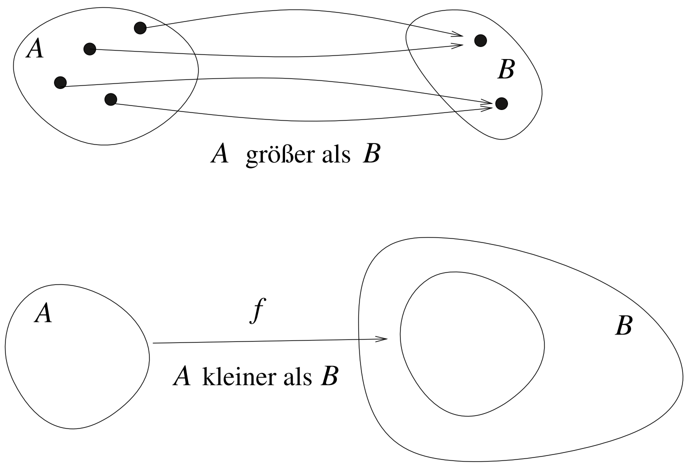


Abb. 1.9 Abbildungen als Vergleich von Mengen

betrachtete mathematische Begriffe zu schaffen. Sie ist ein wesentlicher Schritt im stufenweisen Aufbau eines immer komplexer werdenden mathematischen Universums, in dem nach immer den gleichen Prinzipien aus Mengen Funktionen zwischen Mengen, dann Mengen von Funktionen zwischen Mengen, dann Funktionen zwischen Mengen von Funktionen zwischen Mengen werden, und so weiter.

Zusammenfassung

Die naive Mengenlehre erlaubt es, in einer einheitlichen Sprache mathematische Objekte und Beziehungen zwischen diesen zu beschreiben. Durch iterativen Gebrauch lassen sich darüber hinaus auch komplexere Beziehungsgeflechte wie „Relationen auf Mengen von Abbildungen zwischen Mengen von Funktionen“ darstellen. So leistet die Mengenlehre nicht nur einen Beitrag zur Präzisierung der Sprache in der Mathematik, sondern trägt auch ganz entscheidend zur Vereinfachung der Darstellung bei.

1.5 Kommunikation über gesicherte Erkenntnisse

Auch wenn Objekte und ihre Beziehungen präzise beschrieben werden, ist die Kommunikation zwischen Mathematikern über wissenschaftliche Erkenntnisse und Einsichten kompliziert. Problematisch ist bereits die Frage, wann eine Einsicht als gesichert betrachtet werden kann und zur Veröffentlichung geeignet ist. Diese Frage wird in unterschiedlichen Disziplinen unterschiedlich beantwortet, vgl. Engelen et al. (2010). In der Mathematik legt man sehr strenge Kriterien an. Mathematiker werden im Studium dazu angehalten, nichts als offensichtlich anzusehen, was nicht durch abgesicherte logische Schlussfolgerungen aus den zugrunde liegenden Axiomen abgeleitet werden kann. Dennoch bestehen die mathematischen Arbeiten, durch die Mathematiker ihre Ergebnisse kommunizieren, nicht aus einer reinen Aneinanderreihung solcher logischen Schlussketten. Vielmehr finden sich in den Veröffentlichungen eine Reihe von Argumenten dafür, dass ein bestimmter Sachverhalt zutrifft. Nichtmathematiker haben in der Regel keine Chance, eine solche Argumentationskette nachzuvollziehen, schon weil sie die verwendeten Begriffe nicht kennen. Ähnlich geht es sogar professionellen Mathematikern, wenn es sich nicht gerade um ihr Fachgebiet handelt. Sie sind nicht vertraut mit den Konstruktionen und Schlüssen, die in spezifischen Kontexten verwendet und daher nicht separat eingeführt oder begründet werden. Selbst für Spezialisten ist die Lektüre von Fachliteratur oft mühsam.

Worin liegt die Schwierigkeit bei der Kommunikation über mathematische Sachverhalte? Angenommen, jemand macht sich die Mühe, alle Definitionen und beliebig viele Details einzufügen, könnte er dann einen Text erstellen, der mit einem Lexikon der Axiome und einer Aneinanderreihung von logischen Wahrheitstafeln für jedermann nachvollziehbar wäre? In den allermeisten Fällen nicht. Der Text wäre so lang und unübersichtlich, dass er im besten Falle nur noch lokal verständlich wäre. Ein Leser könnte letztendlich vielleicht noch entscheiden, ob der beschriebene Sachverhalt zutrifft oder nicht. Die Ideen jedoch, die zu seiner Auffindung geführt haben, wäre auf diese Weise kaum mitteilbar. Es stellt sich also die Frage, wie über Mathematik so kommuniziert werden kann, dass Einsichten übermittelt werden.

Um eine Antwort auf diese Frage zu finden, muss man sich Gedanken darüber machen, wie ein Individuum mathematische Sachverhalte erfasst. Der Physiker Richard Feynman (1918–1988) schreibt in seiner (sehr lesenswerten) Autobiografie Feynman (1997) dazu:

„I had a scheme, which I still use today when somebody is explaining something that I'm trying to understand: I keep making up examples. For instance, the mathematicians would come in with a terrific theorem, and they're all excited. As they're telling me the condition of the theorem, I construct something which fits all the conditions. You know, you have a set (one ball) – disjoint (two balls). Then the balls turn colors, grow hairs, or whatever, in my head as they put more conditions on. Finally they state the theorem, which is some dumb thing about the ball which isn't true for my hairy green ball thing, so I say, ‚False!'. “

Eine mathematische Arbeit sollte ebenso wie jedes Lehrbuch eine Anleitung zu einem individuellen Erkenntnisprozess sein. Ein Mangel an Details macht es dem

Leser unmöglich, brauchbare interne Modelle zu entwickeln. Andererseits lässt übergroße Detailfülle den Leser manchmal den Wald vor lauter Bäumen nicht erkennen.

Vor diesem Hintergrund ist es geradezu erstaunlich, dass die Spezialisten auf einem Gebiet sich so oft einig sind, wenn es darum geht zu beurteilen, ob ein Beweis vollständig ist, was so viel heißen soll wie: „Der Beweis versetzt einen Leser mit hinreichenden Kenntnissen in die Lage, interne Modelle zu entwickeln, aus denen er einen Beweis des angegebenen Sachverhalts produzieren kann, wenn er willens ist, sich ein bisschen anzustrengen." Wenn wir dies als „Definition" eines formalen Beweises akzeptieren, dann geht der Streit darum, ob anzustreben sei, dass eine mathematische Arbeit für die in ihr aufgestellten Thesen formale Beweise anführt. Insbesondere Spitzenmathematiker vertreten bisweilen den Standpunkt, dass der strenge Beweiszwang den freien Geist einschränke und den Fortschritt der Mathematik behindere. Bei dem Ausdruck *rigorous proof* fühlen sich manche an *rigor mortis* – Totenstarre – erinnert. Dementsprechend gibt es viele unbewiesene Behauptungen in der Literatur.

Schon immer haben Mathematiker unbewiesene Vermutungen aufgestellt, die dann die Fantasie der Nachwelt beflügelt haben. Das Fermatsche Problem (Seite 5) ist nur ein Beispiel von vielen. Einige der Spekulationen und skizzierten Programme sind zu mathematischem Weltruhm gelangt und ihre Schöpfer mit der Fields-Medaille ausgezeichnet worden, dem angesehensten Preis, den es in der Mathematik gibt. Das mag berechtigt sein, weil einige dieser Vermutungen große Wirkung zeigen und Einfluss auf die Entwicklung der Mathematik ausüben. Ein Beispiel dafür ist das *Geometrisierungsprogramm* von William Thurston (geb. 1946, Fields-Medaille 1982), das letztendlich den Beweis der im Jahr 1904 aufgestellten Poincaré-Vermutung durch Grigori Perelman (geb. 1966, hat die Fields-Medaille 2006 abgelehnt) ermöglicht hat.

Bei vielen Mathematikern löst jedoch die Vorstellung, dass Beweis und Spekulation untrennbar verwoben werden, Unbehagen aus. Es gibt warnende Beispiele: Im 19. Jahrhundert gab es eine Schule von hervorragenden italienischen Geometern, deren von brillanter Intuition geprägte Arbeiten zuletzt eben nicht mehr nachvollziehbar waren, was dem Gebiet der *Algebraischen Geometrie* einen längeren Dornröschenschlaf bescherte. Erst Mitte des 20. Jahrhunderts wurde es von André Weil (1906–1988) und anderen auf ein Fundament mit präzisen mathematischen Begriffsbildungen gestellt. Diesem Unbehagen haben amerikanische Mathematiker Ende des 20. Jahrhunderts Ausdruck verliehen, als sie eine „theoretische Mathematik" fordern, die analog der Rolle der theoretischen Physik Spekulationen bereitstellt, die dann von der strengen Mathematik, in Analogie zur experimentellen Physik, überprüft werden. Auch wenn man den Namen „theoretische Mathematik“ nicht in diesem Sinne akzeptiert, ist es unabdingbar, dass man Spekulation und Beweis trennt und jedes als das bezeichnet, was es ist. Tut man das nicht, so setzt man eine wichtige Errungenschaft des mathematischen Wissenschaftsbetriebs aufs Spiel: die Verlässlichkeit der veröffentlichten Ergebnisse und damit die Bereitschaft der Mathematiker, sich gegenseitig zuzuhören und so ständige Wiederholungen zu vermeiden.

Bisher bot das Peer Review System, das heißt eine Begutachtung durch Fachkollegen vor der Veröffentlichung, eine gewisse Garantie für die Verlässlichkeit mathematischer Veröffentlichungen. Der nicht zuletzt durch eine überzogene Preispolitik verursachte schleichende Niedergang der von kommerziellen Verlagen herausgegebenen Fachzeitschriften könnte diese Kontrolle nun schwächen. Die elektronischen Preprintserver verlassen sich auf eine reine Qualitätsselbstkontrolle.

In diesem Zusammenhang stellt sich die Frage nach einem Bewertungssystem für wissenschaftliche Leistungen. Insbesondere ist zu fragen, welcher Wert Konzepten und Vermutungen gegenüber Beweisen zugemessen wird. Gibt es keinen Druck mehr, Beweise zu veröffentlichen, weil man schon für die Formulierung eines Satzes oder für heuristische Argumente die entsprechende Anerkennung erfährt, so besteht die Gefahr, dass die allgemeine Verfügbarkeit von verlässlicher Information noch weiter abnimmt. In noch stärkerem Maß als bisher werden dann die Zentren der mathematischen Forschung wie Paris, Boston, Moskau, Zürich oder Bonn zu geschlossenen Schulen, weil man nur dort erfahren kann, was wirklich bekannt ist und was nicht. Das hat verheerende Auswirkungen auf die Themenwahl in der „Provinz“. Wer fasst schon ein heißes Thema an, wenn er immer gewärtig sein muss, dass er nach getaner Arbeit zu hören kriegt: „Wissen wir schon lange.“

Umgekehrt, wenn einer der Großen ein Gebiet systematisch und mit formalen Beweisen beackert, dann traut sich keiner mehr dorthin, weil er mit dem Meister ohnehin nicht konkurrieren kann. Fields-Medaillist Thurston gibt dieses als einen entscheidenden Grund für seine „Nichtveröffentlichungspolitik“ an.

Darüber hinaus darf man kreative Geister auch nicht mit Forderungen nach mathematischer Strenge ersticken, denen sie nicht nachkommen können. Benoît Mandelbrot (1924–2010), der für seine Arbeiten zur fraktalen Geometrie berühmt wurde, wies die Forderung folgendermaßen zurück:

„Philip Anderson describes mathematical rigor as irrelevant and impossible. I would soften the blow by calling it besides the point and usually distracting, even where possible.“

Anzustreben ist ein Wertesystem, das mathematisches Teamwork als eine vernünftige Arbeitsteilung zulässt: Eine neue mathematische Einsicht entsteht mit Sicherheit nicht nach dem bekannten Definition-Satz-Beweis-Schema. Spekulation und Intuition spielen eine wichtige Rolle. Aber es geht nicht nur um die Erweiterung der Mathematik, es geht auch darum, das angehäufte Wissen unter allen Mathematikern verfügbar zu machen. Die Erkenntnisse der wirklich innovativen Köpfe, von denen es in jeder Generation immer nur ein paar gibt, müssen so dokumentiert werden, dass sie erhalten bleiben, wenn die Wissenschaftler, an deren internen Modellen diese Erkenntnisse „abgelesen“ wurden, nicht mehr da sind. Es muss auch derjenige Anerkennung finden, der „nur“ beweist, was andere schon „gewusst“ haben. Selbst wer Bekanntes vereinfacht oder vereinheitlicht und damit besser greifbar macht, hat angesichts der Informationsfülle einen wichtigen Beitrag geleistet.

Wenn man die Masse der Mathematiker, die eben nicht an den Zentren der Forschung arbeiten und dementsprechend keinen Zugang zu den informellen Kanälen der Informationsverbreitung haben, in die Gemeinschaftsaufgabe „Weiterentwicklung der Mathematik“ einbeziehen will, braucht man verlässliche, allgemein zugängliche Medien, welche die verfügbare Information verbreiten. Man muss wissen, welche Aussagen schon überprüfte Teile des Mathematikgebäudes sind, geeignet, um darauf aufzubauen, und welche nur Markierungen darstellen, und darauf hinweisen, wo sich die Anstrengungen konzentrieren sollten. Andernfalls laufen wichtige Entwicklungen Gefahr, mit dem Ausscheiden ihrer Hauptproponenten einfach abzusterben, statt in ein Gesamtgebäude eingebaut zu werden.

So vollzieht sich die Kommunikation über mathematische Einsichten in unterschiedlichen Spannungsfeldern: Sie muss nachvollziehbar und überprüfbar sein, soll aber auch Intuition und Spekulation nicht aussparen. Sie muss Erkenntnisse so festhalten, dass sie immer wieder nachvollzogen werden können, aber es bleibt unvermeidbar, dass die Lektüre dem Leser ein hohes Maß an individuellem Engagement abfordert.

1.6 Beweise

Die eingehende Beschäftigung mit Beweisen unterscheidet die Ausbildung zum Mathematiker von der Mathematikausbildung für Ingenieure, Naturwissenschaftler und andere Mathematikanwender. Die Allgegenwärtigkeit von Beweisen in den Grundvorlesungen macht auch vielen Studienanfängern Schwierigkeiten, weil sie es aus der Schule gewohnt sind, dass im Mathematikunterricht dem Rechnen die meiste Beachtung geschenkt wird. Beweise markieren die Grenze zwischen Mathematik als Kulturtechnik und Mathematik als Wissenschaft. In diesem Abschnitt werden beispielhaft verschiedene Beweise vorgestellt und erläutert. Es geht darum, die abstrakte Diskussion des Beweisbegriffs in Abschnitt 1.5 mit konkreten Beispielen zu unterlegen und einen Einblick in die Vielfalt von Beweisführungen zu geben.

Von den hier vorgestellten fundamentalen Beweistypen lassen sich einige über die Zielsetzung beschreiben, wie *konstruktive Beweise* und *Existenzbeweise*. Andere charakterisiert man eher über ihre logische Struktur, wie *Widerspruchsbeweise* oder *Induktionsbeweise*. Man erhält so eine sehr grobe Einteilung von Beweistypen. Normalerweise sieht man einer mathematischen Aussage wie „Es gibt unendlich viele Primzahlen“ nicht von vornherein an, mit welchem Beweistyp man sie verifizieren oder widerlegen kann. Aber selbst wenn die gestellte Aufgabe lautet, man solle diese Aussage mit einem Widerspruchsbeweis verifizieren, enthält das noch keinen Hinweis auf die konkrete Beweisidee. In der Regel gibt es auch mehr als eine Methode, eine Aussage zu beweisen. Auch wenn erfahrene Mathematiker einem Problem oft Ähnlichkeiten mit einem schon vorher behandelten Problem ansehen, gibt es keine Automatismen. Nicht umsonst haben sich die Hoffnungen der KI-Forscher (KI = Künstliche Intelligenz) auf die Automatisierung mathematischer Beweisführungen bisher nur in sehr bescheidenem Maße erfüllt. Mathematiker empfinden das auch nicht als deprimierend, sondern im Gegenteil als sehr erfreulich. Es bestätigt ihre Auffassung, mathematische Forschung sei eine sehr kreative Tätigkeit.

Im Lichte der Unvollständigkeitssätze von Kurt Gödel (1906–1978), die auf Seite 187 beschrieben sind, kann man auch philosophisch argumentieren und zu dem Schluss kommen, dass die Mathematik sich nicht auf ein formales System reduzieren lässt und daher für einen Computer nicht vollständig erfassbar ist. Diese Frage wird in (Goldstein, 2006, 199 ff.) eingehender diskutiert. Gödels Resultate zeigen insbesondere, dass man in einem mathematischen Formalismus Aussagen formulieren kann, die sich innerhalb dieses Formalismus weder beweisen noch widerlegen lassen. Dieser Aspekt ist nicht Thema dieses Abschnitts. Er wird in Abschnitt 4.2 näher beleuchtet.

Da es keine systematische „Theorie der Beweisführung“ und kein *proof engineering* gibt, kann man zwar allgemeine Prinzipien mathematischer Beweisführung beschreiben, aber nur anhand gut gewählter Beispiele kann man lernen, wie man sie konkret einsetzt. Um solche Beispiele beschreiben zu können, muss man ein Mindestmaß an mathematischen Objekten zur Verfügung haben. Es ist daher nicht zu vermeiden, dass in diesem Abschnitt mehr Zahlen, Formeln und Mengen vorkommen als in den übrigen Teilen dieses Buches. Dementsprechend ist die Lektüre dieses Abschnitts fordernder als alle bisherigen Teile. Die technischen Aspekte der Beispiele sind wieder grau hinterlegt und können beim ersten Lesen übersprungen werden. Selbstverständlich kann man auch den ganzen Abschnitt bei der ersten Lektüre weglassen, aber ein gewisses Grundverständnis dafür, wie Mathematik funktioniert, ist ohne ein wenig Mühe ebenso wenig zu haben wie Hausmusik ohne Übung.

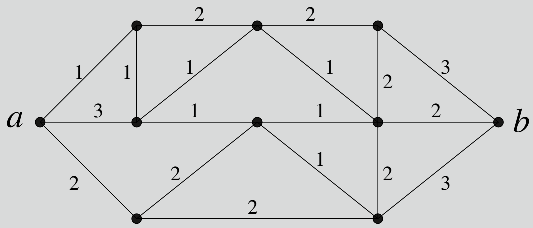
Mathematische Texte sind üblicherweise strikt gegliedert. Man führt neue Begriffe unter der Überschrift *Definition* ein und beschreibt mathematische Gesetzmäßigkeiten als *Sätze* wie im Falle des Fundamentalsatzes der Algebra (Satz 1.1). Sätze können unterschiedliches Gewicht und unterschiedliche Funktionen haben. Einen Satz, den man (gemessen am Kontext) relativ leicht beweisen kann, bezeichnet man oft als *Proposition*. Sätze, die vorbereitender Natur sind, werden gerne *Lemma* (Plural: Lemmata) genannt. Dagegen heißen Sätze, die mehr oder weniger unmittelbare Konsequenzen eines vorausgehenden Satzes sind, *Korollar*. Mathematische Gesetzmäßigkeiten, die man ohne Beweis voraussetzen will, nennt man *Axiome*. Die Axiome stehen am Anfang einer mathematischen Theorie. Aus ihnen werden mithilfe von logischen Schlussfolgerungen die Sätze abgeleitet. Man muss sich also immer im Klaren darüber sein, dass die Gültigkeit der Sätze nur unter der Bedingung nachgewiesen ist, dass alle Axiome gültig sind. Wenn die Axiome schlecht gewählt sind, widersprechen sie einander und es ist nicht möglich, dass sie alle gültig sind. Dann bricht die Theorie zusammen. Beweise für mathematische Sätze werden normalerweise auch klar gekennzeichnet. Meistens beginnt ein Beweis mit „Beweis“ und endet mit einem abschließenden Symbol wie $\square$ oder *q.e.d.*, das für *quod erat demonstrandum* (lateinisch „was zu beweisen war“) steht.

Konstruktive Beweise

Die einfachste Variante eines konstruktiven Beweises besteht darin, alle denkbaren Konstellationen einer vorgegebenen Situation aufzulisten und dann nachzuweisen, dass es eine Konstellation mit einer bestimmten Eigenschaft gibt. Dabei muss man jede einzelne Konstellation auf diese Eigenschaft testen. Man markiert die Konstellationen mit der gesuchten Eigenschaft und kann zum Schluss sagen, ob es solche Konstellationen gibt, und sie gegebenenfalls alle auflisten.

Als Beispiel betrachten wir die folgende typische mathematische Aussage, die die Fragestellung nach der Lösung von Gleichungen verallgemeinert: In der Menge M gibt es ein Element x mit der Eigenschaft E. In günstigen Fällen kann man solche Elemente durch eine Konstruktionsvorschrift tatsächlich bestimmen. Manchmal kann man sogar zeigen, dass es nur ein solches Element geben kann. In der Regel erhöht so eine Eindeutigkeit die Chance, das Element auch wirklich zu finden.

Beispiel 1.11 (Diskrete Optimierung)
Zwischen zwei Städten a und b soll eine Straße gebaut werden. Die möglichen Bauabschnitte seien durch den folgenden Graphen gegeben. Die Knoten könnten dabei zum Beispiel Dörfer sein, die potenziell an das Straßennetz angebunden werden.



Dabei stellen die Kantenbewertungen, das heißt die Zahlen an den Verbindungslinien, die Kosten für den jeweiligen Bauabschnitt dar. Beispiele für solche Kosten wären Baukosten, ökologische Kosten, Kosten für den Erwerb von Grundstücken oder Kombinationen solcher Kosten, jeweils bewertet in Geld oder einer anderen Berechnungseinheit. Die Frage ist jetzt: Wie muss die Straße verlaufen, damit die gesamten Baukosten minimal werden? Man sucht also diejenigen Wege von a nach b, für die die Summe der Bewertungen der durchlaufenen Kanten minimal ist.

Behauptung: Die minimalen Baukosten sind 6 Berechnungseinheiten, und es gibt drei verschiedene Streckenführungen, für die diese Baukosten realisiert werden.

Beweisstrategie: Man muss nur Wege betrachten, die keinen Punkt mehr als einmal ansteuern. Von dieser Bauart gibt es nur endlich viele Wege, die von a nach b führen. Also kann man sie alle auflisten und nachsehen, für welchen die Summe der Kantenbewertungen minimal ist. Damit ist Folgendes klar: Es gibt einen Weg der gesuchten Art und die beschriebene Methode findet alle solchen Wege.
Ebenfalls klar ist, dass schon bei diesem kleinen Beispiel eine ganz erhebliche Menge von verschiedenen Wegen von a nach b existieren. Das heißt, man muss eine lange Liste aufstellen und überprüfen, worauf hier aus Platzgründen verzichtet werden soll.

Die beschriebene Methode stellt einen konstruktiven Beweis dar, denn man hat am Ende nicht nur herausgefunden, dass es eine Konstellation der gesuchten Art gibt, sondern man kann auch eine solche Konstellation angeben. Allerdings stößt diese *brute force*-Methode bei der Suche nach optimalen Lösungen sehr schnell an Kapazitätsgrenzen und verwandelt sich praktisch in einen reinen Existenzbeweis (Seite 58), bei dem es nicht mehr darum geht, eine Lösung zu präsentieren, sondern nur darum, die Existenz einer Lösung nachzuweisen. Es gibt sehr viel elegantere Methoden, das oben gestellte Problem anzugehen, diese sind aber deutlich komplizierter.

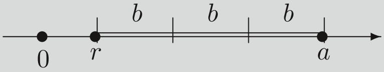
In unserem nächsten Beispiel geht es um den *größten gemeinsamen Teiler* $\operatorname{ggT}(a, b)$ zweier natürlicher Zahlen a und b, das heißt die größte Zahl, durch die man sowohl a als auch b teilen kann. Dieses Beispiel ist zweistufig. Zunächst wird gezeigt, dass es eine Zahl der gesuchten Art gibt (reiner Existenzbeweis). Erst danach wird eine Methode angegeben, dieses Element konkret zu bestimmen, indem man zeigt, dass das Resultat eines Algorithmus die beschriebene Eigenschaft hat. Beide Schritte zusammen ergeben einen konstruktiven Existenzbeweis, der die Existenz durch das Angeben des gesuchten Objekts nachweist.

Der Schlüssel zur Bestimmung des ggT liegt in der Division mit Rest, die wir schon mehrfach verwendet haben. Wir formulieren sie noch einmal ganz explizit:

Lemma 1.12 (Division mit Rest)
Seien $a, b \in \mathbb{N}$. Dann existieren q, $r \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ mit $a = q \cdot b + r$ und $0 \leq r < b$.

Verbalisierung: Man zieht von einer Zahl a die Zahl b so oft ab, wie man als Ergebnis eine natürliche Zahl oder 0 erhält. Das letzte solche Ergebnis ist r, und q ist die Anzahl der durchgeführten Subtraktionen.

Visualisierung:



Für den Beweis dieses Lemmas braucht man das Axiomensystem für die natürlichen Zahlen, das in Anhang A.2 besprochen wird. Wir führen den Beweis auf Seite 242.

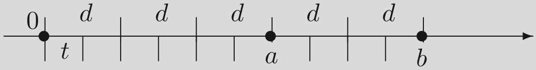
Satz 1.13 (Existenz des größten gemeinsamen Teilers)
Seien $a, b \in \mathbb{N}$. Dann gibt es ein eindeutiges $d \in \mathbb{N}$ derart, dass

(i) *$d \mid a$ und $d \mid b$ (d teilt a und d teilt b).*
(ii) *Für jeden gemeinsamen Teiler t von a und b gilt $t \mid d$.*

Wegen Eigenschaft (iv) in Beispiel 1.8 zeigt Satz 1.13(ii), dass d die größte Zahl $d \in \mathbb{N}$ ist, die a und b teilt. Daher nennt man sie den *größten gemeinsamen Teiler* von a und b. Sie wird mit $\mathrm{ggT}(a, b)$ bezeichnet.

Verbalisierung: Es gibt eine Zahl d, die a und b teilt und von jeder anderen Zahl, die a und b teilt, geteilt wird.

Visualisierung:



Beweis. Als Erstes zeigen wir die Eindeutigkeit des ggT. Dazu nehmen wir an, d und d' seien zwei Zahlen, die die Eigenschaften (i) und (ii) erfüllen. Ziel ist es dann zu zeigen, dass $d = d'$ gelten muss. Aus (i) für d' und (ii) für d folgt, dass $d' \mid d$. Genauso folgt aus (i) für d und (ii) für d', dass $d \mid d'$. Nach der Eigenschaft (iii) in Beispiel 1.8 liefert dies $d = d'$.

Wir beginnen den Nachweis der Existenz des ggT mit einer Vorüberlegung: Wenn eine Zahl d die Zahlen a und b teilt, dann teilt sie auch jede Zahl der Form $ax + by \in \mathbb{N}$, wobei x, y ganze Zahlen sind. Wenn nämlich $a = dm$ und $b = dn$, dann folgt $ax + by = dmx + dny = d(nx + my)$. Insbesondere kann d nicht größer sein als die kleinste solche Zahl. Da jede Menge von natürlichen Zahlen, die nicht leer ist, ein kleinstes Element hat (siehe Anhang A.2), existiert also das Minimum

$$d := \min\{ax + by \in \mathbb{N} \mid x, y \in \mathbb{Z}\}.$$

Behauptung: d erfüllt (i) und (ii).

Um diese Behauptung zu beweisen, wählen wir $x_0, y_0 \in \mathbb{Z}$ mit $d = ax_0 + by_0$. Um (ii) zu zeigen, nehmen wir an, dass t ein Teiler von a und b ist, das heißt $t \mid a$ und $t \mid b$. Dann gilt $t \mid ax_0$ und $t \mid by_0$, also auch $t \mid (ax_0 + by_0) = d$. Damit ist (ii) gezeigt.

Für den Nachweis von (i) brauchen wir das Lemma 1.12: Danach existieren $q, r \in \mathbb{Z}$ mit $0 \leq r < d$ und $a = qd + r$. Insbesondere gilt $r \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Es folgt

$$r = a - qd = a - q(ax_0 + by_0) = ax' + by',$$

wobei wir $x' = 1 - qx_0$ und $y' = -qy_0$ setzen. Da $d = ax_0 + by_0$ die kleinste natürliche Zahl der Form $ax + by$ mit $x, y \in \mathbb{Z}$ ist, muss $r = 0$ gelten. Aber dann gilt $a = qd$, das heißt $d \mid a$. Analog zeigt man $d \mid b$, indem man d mit Rest durch b teilt und den Rest betrachtet. Damit ist dann auch (i) gezeigt. $\square$

Satz 1.13 sagt, dass es den $\mathrm{ggT}(a, b)$ wirklich gibt. Der Beweis von Satz 1.13 sagt uns im Prinzip auch, wie man ihn findet: Er ist das kleinste Element der Menge $M = \{ax + by \in \mathbb{N} \mid x, y \in \mathbb{Z}\}$. Er sagt uns aber nicht, wie man einer Zahl ansieht, ob sie zu M gehört oder nicht. Das heißt, bis jetzt ist unser Beweis nicht wirklich konstruktiv. Aber schon Euklid wusste, wie man den ggT berechnet:

Satz 1.14 (Euklidischer Algorithmus)
Gegeben seien $a, b \in \mathbb{N}$, gesucht ist $\operatorname{ggT}(a,b)$. Division mit Rest (Lemma 1.12) liefert $q_1, r_1 \in \mathbb{Z}$ mit $a = q_1 b + r_1$ und $0 \leq r_1 < b$.

Ist $r_1 = 0$, dann gilt $b = \operatorname{ggT}(a,b)$.

Ist $r_1 \neq 0$, so existieren $q_2, r_2 \in \mathbb{Z}$ mit $b = q_2 r_1 + r_2$ und $0 \leq r_2 < r_1$.

Ist $r_2 = 0$, dann gilt $r_1 = \operatorname{ggT}(a,b)$.

Ist $r_2 \neq 0$, so existieren $q_3, r_3 \in \mathbb{Z}$ mit $r_1 = q_3 r_2 + r_3$ und $0 \leq r_3 < r_2$.

$\vdots$

Der Algorithmus bricht ab, sobald $r_j = 0$ auftritt. Dies geschieht spätestens nach b Schritten, denn $b > r_1 > r_2 > \cdots > r_j \geq 0$.
Mit den obigen Bezeichnungen gilt: Ist $k+1$ der kleinste Index mit $r_{k+1} = 0$, so haben wir $r_k = \operatorname{ggT}(a,b)$.

Beweis. Sei d der ggT von a und b. Wegen $a = q_1 b + r_1$ ist dann $r_1 = a - q_1 b$ durch d teilbar. Im nächsten Schritt liefert $b = q_2 r_1 + r_2$, dass $r_2 = b - q_2 r_1$ durch d teilbar ist. Sukzessive liefern die vom Algorithmus produzierten Gleichungen

$$\begin{array}{rcl@{\qquad}c@{\qquad}l}
a & = & q_1 b + r_1 & \Rightarrow & d \mid r_1 \\
b & = & q_2 r_1 + r_2 & \Rightarrow & d \mid r_2 \\
r_1 & = & q_3 r_2 + r_3 & \Rightarrow & d \mid r_3 \\
\vdots & & & & \vdots \\
r_{k-2} & = & q_k r_{k-1} + r_k & \Rightarrow & d \mid r_k.
\end{array}$$

Wenn wir umgekehrt auch $r_k \mid d$ zeigen können, dann gilt $d = r_k$ und wir sind fertig. Wir gehen so vor, dass wir $r_k \mid a$ und $r_k \mid b$ zeigen, was nach Satz 1.13 die Relation $r_k \mid d$ liefert. Wieder arbeiten wir die Gleichungen aus dem Algorithmus ab, allerdings in umgekehrter Reihenfolge, beginnend mit der letzten Gleichung $r_{k-1} = q_{k+1} r_k$:

$$\begin{array}{rcl@{\qquad}c@{\qquad}l}
r_{k-1} & = & q_{k+1} r_k & \Rightarrow & r_k \mid r_{k-1} \\
r_{k-2} & = & q_k r_{k-1} + r_k & \Rightarrow & r_k \mid r_{k-2} \\
\vdots & & & & \vdots \\
r_1 & = & q_3 r_2 + r_3 & \Rightarrow & r_k \mid r_1 \\
b & = & q_2 r_1 + r_2 & \Rightarrow & r_k \mid b \\
a & = & q_1 b + r_1 & \Rightarrow & r_k \mid a.
\end{array}$$

Wir haben also $d = r_k$ gezeigt und damit den Satz bewiesen. $\square$

Zur Illustration berechnen wir $\text{ggT}(18, 48)$ mithilfe des euklidischen Algorithmus:

$$\begin{aligned} 48 &= 2 \cdot 18 + 12, \\ 18 &= 1 \cdot 12 + 6, \\ 12 &= 2 \cdot 6 + 0 \end{aligned}$$

liefert $\text{ggT}(18, 48) = 6$.

Um zu zeigen, dass der euklidische Algorithmus wirklich den ggT liefert, haben wir die Charakterisierung des ggT aus Satz 1.13 benutzt, nicht die ursprüngliche Definition. Das heißt, nur zusammen liefern Satz 1.13 und 1.14 einen konstruktiven Beweis für die Existenz des größten gemeinsamen Teilers.

Weil wir in den Beweisen von Satz 1.19 und Satz 2.12 darauf zurückgreifen müssen, halten wir hier gleich noch eine Folgerung (Korollar) aus dem Beweis von Satz 1.13 fest:

Korollar 1.15
Wenn $a, b \in \mathbb{N}$ teilerfremd *sind, das heißt* $\text{ggT}(a, b) = 1$ *gilt, dann gibt es ganze Zahlen* x, y *mit* $ax + by = 1$.

Beweis. Im Beweis von Satz 1.13 haben wir gesehen, dass

$$\text{ggT}(a, b) = \min\{ax + by \in \mathbb{N} \mid x, y \in \mathbb{Z}\}.$$

Wenn also $\text{ggT}(a, b) = 1$ gilt, muss es $x, y \in \mathbb{Z}$ mit $ax + by = 1$ geben. □

Beweis durch Widerspruch

Um das Prinzip des *Beweises durch Widerspruch* zu erklären, muss man sich der elementaren *Aussagenlogik* bedienen. In der Aussagenlogik weist man beliebigen Aussagen die Werte *wahr* oder *falsch* zu und untersucht dann die Wahrheitswerte von Aussagen, die man durch Negation oder Kombination von Aussagen mit bekannten Wahrheitswerten erhält. Ist zum Beispiel eine Aussage A wahr, dann ist ihre Negation $\neg A$ falsch. Ist dagegen A falsch, dann ist $\neg A$ wahr. Solche Zusammenhänge stellt man oft in „Wahrheitswertetabellen“ dar:

A	w	f
$\neg A$	f	w

Wahrheitswertetabelle für $\neg A$

Die Wahrheitswertetabellen sehen komplizierter aus, wenn man mehrere Aussagen miteinander kombiniert, zum Beispiel durch das logische „und“, das mit $\wedge$ bezeichnet wird, bzw. das logische „oder“, das man mit $\vee$ bezeichnet.

$A \backslash B$	w	f
w	w	f
f	f	f

Wahrheitswertetabelle für $A \wedge B$

$A \backslash B$	w	f
w	w	w
f	w	f

Wahrheitswertetabelle für $A \vee B$

Erfahrungsgemäß fällt es einem Großteil der Studierenden in den ersten Semestern nicht leicht, kombinierte Aussagen zu negieren. Für Aussagen, wie „$A \vee B$“ prüft man durch Einsetzen von w und f in die Wahrheitswertetabellen sofort nach, dass $\neg(A \vee B)$ und $(\neg A) \wedge (\neg B)$ die gleichen Wahrheitswertetabellen haben, also logisch äquivalent sind. Man schreibt dann einfach $\neg(A \vee B) = (\neg A) \wedge (\neg B)$. Schwieriger wird es, wenn man logische *Quantoren* wie $\forall$, „für alle“, oder $\exists$, „es existiert ein“, ins Spiel bringt. Zum Beispiel ist die Negation von „Kein Mathematiker macht jemals etwas Nützliches“ „Es gibt einen Mathematiker, der irgendwann etwas Nützliches macht“.

Zur Beschreibung von Schlussfolgerungen benutzt man die logische „Implikation“, die mit $\Rightarrow$ bezeichnet wird. Ihre Wahrheitswertetabelle ist wie folgt gegeben:

$A \backslash B$	w	f
w	w	f
f	w	w

Wahrheitswertetabelle für $A \Rightarrow B$

Die logische Implikation ist insoweit ein wenig kontraintuitiv, als die Implikation $A \Rightarrow B$ wahr sein kann, obwohl B falsch ist. Dies passiert, wenn die *Prämisse* A falsch ist. Insbesondere ist $A \Rightarrow B$ wahr, wenn sowohl A als auch B falsch sind. Die Setzung ist sinnvoll, weil man aus falschen Voraussetzungen in logisch korrekter Art und Weise beliebige Aussagen ableiten kann. Zum Beispiel lässt sich aus der falschen Aussage $1 = 2$ problemlos in korrekter Weise ableiten, dass der Papst ein Eskimo ist: Man bildet eine Menge, die höchstens zwei Elemente hat und den Papst sowie einen Eskimo als Elemente enthält. Wenn aber $2 = 1$ gilt, dann hat diese Menge nur ein Element, das heißt, Eskimo und Papst müssen dieselbe Person sein. So albern dieses Beispiel ist, es illustriert doch ein Phänomen, das sich immer wieder beobachten lässt: Eine lange plausible Argumentationskette führt zum gewünschten Ergebnis, weil an einer Stelle eine falsche Annahme verwendet wird.

Der *Umkehrschluss* zu einer Implikation $A \Rightarrow B$ (dem *Schluss* „Aus A schließe ich B") ist die Implikation $\neg B \Rightarrow \neg A$ („Aus $\neg B$ schließe ich $\neg A$"). Man überprüft sofort, dass $\neg B \Rightarrow \neg A$ die gleiche Wahrheitstabelle hat wie $A \Rightarrow B$, das heißt, Schluss und Umkehrschluss sind logisch gleichwertig. Die Redewendung „Das bedeutet im Umkehrschluss ...," die gebraucht wird, um die Bedeutung logischer Konsequenzen zu erläutern, ist daher völlig gerechtfertigt. Man darf den Umkehrschluss $\neg B \Rightarrow \neg A$ von $A \Rightarrow B$ aber nicht mit der *umgekehrten Implikation* $B \Rightarrow A$ verwechseln. Diese hat eine andere Wahrheitswertetabelle. Der Umkehrschluss von „Politiker sind Plaudertaschen" ist also „Wer keine Plaudertasche ist, ist auch kein Politiker" und keineswegs „Plaudertaschen sind Politiker". Und „Wer ein Herz hat, ist ein Mensch" ist auch keine Konsequenz von „Wer ein Mensch ist, hat ein Herz".

Ein Vergleich der Wahrheitswertetabellen zeigt, dass $A \Rightarrow B$ nicht nur zu $\neg B \Rightarrow \neg A$, sondern auch zu $(\neg A) \vee B$ logisch äquivalent ist. Es gibt also diverse Möglichkeiten, eine Implikation darzustellen.

Wenn man eine Aussage A durch einen Widerspruchsbeweis nachweisen will, dann leitet man aus $\neg A$ logisch korrekt eine Aussage B ab, von der man weiß, dass sie falsch ist, das heißt im *Widerspruch* zu schon bekannten Tatsachen steht. Man benutzt also eine Implikation $\neg A \Rightarrow B$, deren Wahrheitswert w ist, obwohl B den Wahrheitswert f hat. Das geht das nur, wenn A wahr ist, wie man an der Wahrheitswertetabelle für $\neg A \Rightarrow B$ ablesen kann:

A \ B	w	f
w	w	w
f	w	f

Wahrheitswertetabelle für $\neg A \Rightarrow B$

Diese Vorgehensweise wirkt auf den Laien und den Anfänger zunächst befremdlich und umständlich. Man setzt sie normalerweise dann ein, wenn die Annahme, dass $\neg A$ gilt, zusätzliche Argumentationsmöglichkeiten liefert. Wenn man zum Beispiel zeigen möchte, dass es unendlich viele Objekte einer bestimmten Art gibt, dann eröffnet die Annahme, es gäbe nur endlich viele solche Objekte, die Möglichkeit, Abzählargumente einzusetzen. Euklids Beweis für die Aussage

$$A = \text{„Es gibt unendlich viele Primzahlen“}$$

aus Beispiel 1.16 funktioniert nach diesem Muster.

Beispiel 1.16 (Primzahlen)
Jede natürliche Zahl n hat die Teiler 1 und n. Üblicherweise definiert man eine *Primzahl* als eine natürliche Zahl $n > 1$, die keine weiteren Teiler hat. Das bedeutet, eine natürliche Zahl n ist genau dann eine Primzahl, wenn sie genau zwei verschiedene Teiler hat. Euklid hat als Erster einen Beweis dafür aufgeschrieben, dass die Menge aller Primzahlen unendlich sein muss. Sein Beweis ist ein Widerspruchsbeweis: Wir nehmen an, dass die Menge aller Primzahlen endlich ist, und leiten daraus eine falsche Aussage ab. Sei also $\{p_1, \ldots, p_k\}$ die Menge aller Primzahlen.

1. Schritt: Der kleinste Teiler $d > 1$ einer natürlichen Zahl $n > 1$ ist eine Primzahl. Andernfalls hätte man $d = d_1 d_2$ mit $1 < d_1 < d$ und $1 < d_2 < d$ und d wäre nicht minimal.

2. Schritt: Nach dem 1. Schritt hat die Zahl $n := p_1 \cdot \ldots \cdot p_k + 1 > 1$ einen Primteiler p. Dann muss p eine der Zahlen $p_1, \ldots, p_k$ sein. Also teilt p sowohl n als auch $p_1 \cdot \ldots \cdot p_k$, das heißt, es gibt Zahlen $a, b \in \mathbb{Z}$ mit

$$ap = n \quad \text{und} \quad bp = p_1 \cdot \ldots \cdot p_k.$$

Aber dann gilt

$$(a - b)p = ap - bp = n - p_1 \cdot \ldots \cdot p_k = 1.$$

Daraus folgt aber nach Eigenschaft (iv) in Beispiel 1.8, dass $p \leq 1$ gilt. Diese Aussage ist falsch, da jede Primzahl $p > 1$ erfüllt.

Wir haben also aus der Endlichkeit der Menge der Primzahlen die falsche Aussage abgeleitet, dass es eine Primzahl gibt, die kleiner als 2 ist, und damit die Negierung der Annahme, das heißt die Unendlichkeit der Menge der Primzahlen, gezeigt.

Vollständige Induktion

Die nächste Beweistechnik, die hier vorgestellt werden soll, ist der *Beweis durch Induktion*. Unter Induktion (vom Lateinischen *inducere* für „hineinführen") versteht man generell das Schließen vom Speziellen auf das Allgemeine. In der Mathematik hat das Wort eine sehr viel spezifischere Bedeutung: Es handelt sich um eine Methode, die in Situationen angewendet wird, in denen man eine Familie von Aussagen beweisen möchte, die durch natürliche Zahlen durchnummeriert werden. Sie beruht auf folgendem Prinzip: *Man findet alle natürlichen Zahlen, wenn man bei* 1 *anfängt und immer weiter zählt.* Wenn man also für jedes $n \in \mathbb{N}$ eine Aussage $A(n)$ hat und weiß, dass $A(1)$ wahr ist, dann genügt es, $A(n) \Rightarrow A(n+1)$ zu zeigen, um zu garantieren, dass *alle* $A(n)$ wahr sind. Man spricht in diesem Kontext vom *Induktionsanfang* $A(1)$, von der *Induktionsannahme* $A(n)$ und vom *Induktionsschluss* oder *Induktionsschritt* $A(n) \Rightarrow A(n+1)$.

Satz 1.17 (Vollständige Induktion)
Sei $\mathbb{N}$ die Menge der natürlichen Zahlen und $X \subseteq \mathbb{N}$ eine Teilmenge mit folgenden Eigenschaften:

(i) $1 \in X$.
(ii) *Wenn $x \in X$, dann gilt auch $x + 1 \in X$.*

Dann ist $X = \mathbb{N}$.

Wie im Falle der Division mit Rest (Lemma 1.12), findet man die Herleitung von Satz 1.17 aus den Axiomen in Anhang A.2.

Um vollständige Induktion als Beweismethode einsetzen zu können, muss man die Induktionsannahme präzise formulieren können, das heißt, man muss sehr genau wissen, was man beweisen möchte. Typischerweise probiert man zunächst die ersten paar Aussagen aus. Wenn man aber die allgemeine Induktionsannahme formuliert hat, ist die Verifikation von $A(n)$ für $n > 1$ nicht mehr nötig. Wir beginnen mit einem Beispiel, in dem die Induktionsannahme von Anfang an gegeben ist.

Beispiel 1.18
Für jedes $n \in \mathbb{N}$ ist die Zahl $5^n - 1$ durch 4 teilbar. Die Induktionsannahme $A(n)$ ist hier einfach: $4 \mid (5^n - 1)$. Wegen $5^1 - 1 = 4$ ist der Induktionsanfang $A(1)$ offensichtlich. Man könnte jetzt noch $5^2 - 1 = 25 - 1 = 24$ und $5^3 - 1 = 125 - 1 = 124$ auf Teilbarkeit durch 4 testen, aber es ist nicht nötig, wenn man den Induktionsschritt $A(n) \Rightarrow A(n+1)$ beweist. Wir nehmen also an, dass $5^n - 1$ durch 4 teilbar ist, das heißt von der Form $4k$ mit $k \in \mathbb{N}$. Dann rechnen wir

$$5^{n+1} - 1 = 5^{n+1} - 5^n + 5^n - 1 = 5^n(5 - 1) + 4k = 4(5^n + k),$$

das heißt, auch $5^{n+1} - 1$ ist durch 4 teilbar. Mit vollständiger Induktion (Satz 1.17) ist die Aussage also für alle $n \in \mathbb{N}$ bewiesen.

Induktionsbeweise kommen in der Mathematik auf jedem Niveau vor, in der Regel, wenn man durchnummerierte Aussagen hat, bei denen es einen inhaltlichen Zusammenhang zwischen der Nummer und der Aussage gibt. Die Nummern können Anzahlen von Elementen sein, aber auch Dimensionen (der Begriff der Dimension wird auf Seite 82 besprochen) oder die Anzahl von Teilmengen mit einer gewissen Eigenschaft. Es ist dabei nicht nötig, die Induktion bei 1 zu beginnen. Fängt man mit 2 an, so hat man eben die Aussage nur für alle $n \geq 2$ gezeigt, so wie im nachfolgenden *Fundamentalsatz der Zahlentheorie*, der besagt, dass jede Zahl eine eindeutige Zerlegung in Primfaktoren hat. Im Beweis dieses Satzes kommen sogar mehrere Induktionen vor, zum Teil mit recht ausgeklügelten Induktionsannahmen.

Satz 1.19 (Fundamentalsatz der Zahlentheorie)
Jede natürliche Zahl $n > 1$ besitzt eine bis auf die Reihenfolge der Faktoren eindeutige Darstellung als Produkt von Primzahlen.

Beweis.

1. Schritt: Existenz der Zerlegung

Wir beweisen zunächst mit Induktion, dass sich jedes $n > 1$ als Produkt von endlich vielen Primzahlen schreiben lässt.

Induktionsanfang: Der Fall $n = 2$ ist klar, denn 2 ist eine Primzahl und somit das Produkt von Primzahlen – auch wenn das Produkt nur einen Faktor hat.

Induktionsschritt: Wir nehmen an, dass jede natürliche Zahl m mit $1 < m < n$ sich als Produkt von endlich vielen Primzahlen schreiben lässt. Wegen $n > 1$ hat n einen Primfaktor p_1 (siehe 1. Schritt in Beispiel 1.16), für den dann insbesondere $\frac{n}{p_1} \in \mathbb{N}$ gilt. Ist $n = p_1$, so hat man die gewünschte Aussage. Andernfalls gilt $1 < \frac{n}{p_1} < n$, das heißt, mit der Induktionsannahme ergibt sich $\frac{n}{p_1} = p_2 \cdot \ldots \cdot p_k$, also $n = p_1 \cdot p_2 \cdot \ldots \cdot p_k$. Damit ist die Existenz einer *Primzahlfaktorisierung* der gewünschten Art gezeigt.

2. Schritt: Eindeutigkeit der Zerlegung

Die Eindeutigkeit der Primzahlfaktorisierung bis auf die Reihenfolge lässt sich ebenfalls mit Induktion zeigen, allerdings ist die Induktionsannahme $A(n)$ etwas komplizierter:

$A(n)$**:** Für $n \leq k$ seien $p_1, \ldots, p_n$ und $q_1, \ldots, q_k$ Primzahlen mit $p_1 \cdot \ldots \cdot p_n = q_1 \cdot \ldots \cdot q_k$. Dann gilt $n = k$ und es gibt eine umkehrbare Abbildung $\varphi\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ mit $p_j = q_{\varphi(j)}$ für $j = 1, \ldots, n$.

Um den Beweis führen zu können, brauchen wir noch eine weitere Aussage.

Behauptung: *Ist p eine Primzahl und gilt $p \mid mn$ für $m, n \in \mathbb{N}$, so folgt $p \mid m$ oder $p \mid n$.*

Zum Beweis der Behauptung nehmen wir an, dass p *kein* Teiler von m ist. Nach Satz 1.13 haben p und m einen größten gemeinsamen Teiler $\mathrm{ggT}(p, m)$. Der muss dann also der einzige andere Teiler von p sein, nämlich 1. Nach Korollar 1.15 existieren dann $x, y \in \mathbb{Z}$ mit $px + my = 1$, und es gilt $pnx + mny = n$. Wegen $p \mid mn$ finden wir ein $z \in \mathbb{N}$ mit $pz = mn$, und es folgt $p(nx + zy) = pnx + mny = n$, das heißt $p \mid n$. Damit ist die Behauptung bewiesen.

Induktionsanfang: Wenn $n = 1$, dann gibt es nach der Behauptung maximal zwei Möglichkeiten: $p_1 \mid q_1$ oder, falls $k \geq 2$ ist, $p_1 \mid q_2 \cdot \ldots \cdot q_k$. Wenden wir die Behauptung wiederholt an, finden wir ein $i \in \{1, \ldots, k\}$ mit $p_1 \mid q_i$ und weil auch q_i prim ist, gilt $p_1 = q_i$. Aber dann haben wir $1 = \frac{1}{q_i} q_1 \cdot \ldots \cdot q_k$, das heißt $k = 1 = i$. Damit ist $A(1)$ bewiesen.

Induktionsschritt: Wir nehmen an, $A(n)$ sei bewiesen, und wollen zeigen, dass auch $A(n+1)$ gilt. Wenn $p_1 \cdot \ldots \cdot p_n \cdot p_{n+1} = q_1 \cdot \ldots \cdot q_k$, dann gilt $p_{n+1} \mid q_1 \cdot \ldots \cdot q_k$, und das Argument aus dem Induktionsanfang liefert ein $i \in \{1, \ldots, k\}$ mit $p_{n+1} \mid q_i$. Wir setzen $\varphi(n+1) = i$ und stellen fest, dass

$$p_1 \cdot \ldots \cdot p_n = \frac{1}{q_i} q_1 \cdot \ldots \cdot q_k,$$

wobei die rechte Seite ein Produkt von nur $k-1$ Primzahlen ist, weil sich q_i herauskürzt. Jetzt wenden wir $A(n)$ an und finden $k-1=n$ sowie eine umkehrbare Abbildung $\varphi\colon \{1,\dots,n\} \to \{1,\dots,n+1\} \setminus \{i\}$ mit $p_j = q_{\varphi(j)}$. Zusammen liefern diese Aussagen $A(n+1)$. $\square$

Dieser Beweis ist ein Beispiel für eine Kombination aus mehreren Beweistechniken. Er enthält mehrere Induktionen, greift aber auch auf vorher schon bewiesene Resultate zurück, die nicht mit Induktion bewiesen wurden. Das ist eine ganz typische Situation, die auch gut illustriert, warum man in der Mathematik verloren ist, wenn man früher Gelerntes nach bestandener Prüfung wieder vergisst. In dieser Hinsicht hat Mathematik Ähnlichkeit mit einer Fremdsprache.

Als Anwendung des Fundamentalsatzes der Zahlentheorie kann man zeigen, dass $\sqrt{2}$ keine rationale Zahl ist (Seite 7). Es handelt sich bei dem Beweis von Satz 1.20 um einen Widerspruchsbeweis.

Satz 1.20
$\sqrt{2}$ ist keine rationale Zahl, das heißt, man kann keine Zahlen $n, m \in \mathbb{N}$ mit $\frac{n^2}{m^2} = 2$ finden.

Beweis. Seien $m = p_1 \cdot \ldots \cdot p_r$ und $n = q_1 \cdot \ldots \cdot q_s$ jeweils Darstellungen als Produkte von Primzahlen. Wir nehmen an, dass gilt

$$2m^2 = 2p_1^2 \cdot \ldots \cdot p_r^2 = q_1^2 \cdot \ldots \cdot q_s^2 = n^2.$$

Jetzt zählt man ab, wie oft links und rechts der Primfaktor 2 vorkommt. Da jedes p_i und jedes q_j quadriert wird, kommen in m^2 und n^2 jeweils eine gerade Anzahl von Zweien als Primfaktoren vor. In $2m^2$ ist es dann eine ungerade Anzahl von Zweien. Dies steht im Widerspruch zur Eindeutigkeitsaussage in Satz 1.19. $\square$

Nicht alles, was wie ein Induktionsbeweis aussieht, ist auch einer. Man muss die Aussagen sauber formulieren, sonst drohen Fehlschlüsse wie in Beispiel 1.21.

Beispiel 1.21
Man betrachte die Aussagen: „Alle auf der Erde lebenden Pferde haben dieselbe Farbe.“
„Beweis durch Induktion“: Es sei

$$X := \{n \in \mathbb{N} \mid \text{Je } n \text{ Pferde haben dieselbe Farbe}\}.$$

Da jedes Pferd dieselbe Farbe hat wie es selbst, gilt $1 \in X$. Nun sei $n \in X$, und wir müssen zeigen, dass auch $n+1 \in X$ ist. Man nehme eines der $n+1$ Pferde heraus; die restlichen n Pferde haben dieselbe Farbe (da $n \in X$). Nun füge man das herausgenommene Pferd hinzu und nehme ein anderes heraus. Dann ist der Rest wieder einfarbig. Also haben alle $n+1$ Pferde dieselbe Farbe.
Man erkennt den Fehler leicht, wenn man das Argument auf eine Menge mit nur zwei Pferden, einem Rappen und einem Schimmel, anwendet.

Näherungsweise konstruktive Beweise

Der euklidische Algorithmus liefert den ggT zweier Zahlen in endlich vielen Schritten. Es gibt auch Beweise, die die Existenz von Elementen in unendlich vielen Schritten zeigen, was dann automatisch bedeutet, dass man das Ergebnis nur näherungsweise angeben kann. Damit ist schon angedeutet, dass es neuer Begriffsbildungen bedarf, um solche Resultate zu erzielen. Zumindest wird man erklären müssen, was es heißt, eine Lösung „näherungsweise“ anzugeben. Soll das nur heißen, dass es approximative Lösungen gibt, die das gegebene Problem nur ungefähr lösen, aber mit immer höherer Präzision? Oder soll es heißen, dass es eine exakte Lösung des Problems gibt, an die man sich immer näher heranarbeitet?

Wir führen zwei Beispiele vor, die beide den Aspekt der Vollständigkeit der reellen Zahlen verdeutlichen, der schon auf Seite 8 angesprochen wurde. Dazu brauchen wir den Begriff einer *Folge* von Zahlen. Das ist eigentlich nichts anderes als eine Abbildung a von $\mathbb{N}$ in die zu betrachtende Menge von Zahlen (zum Beispiel ganze oder rationale Zahlen). Anstatt $a(n)$ schreibt man aber oft a_n, weil man sich bei einer Folge einfach eine durchnummerierte Menge von Zahlen vorstellt. Die ganze Folge wird als $(a_n)_{n\in\mathbb{N}}$ geschrieben. Die Folge $(a_n)_{n\in\mathbb{N}}$ mit $a_n = 2n-1$ liefert zum Beispiel eine Aufzählung aller ungeraden natürlichen Zahlen.

Betrachten wir zunächst wie die alten Griechen die Länge der Diagonale eines Quadrats der Seitenlänge 1. Dann kann man nach einer Folge von (rationalen) Zahlen a_n suchen, deren Quadrate immer näher an 2 liegen. So eine Folge kann man in der Tat finden. Will man aber eine Zahl finden, deren Quadrat genau 2 ist, muss man den Zahlenbegriff erst erweitern, weil, wie in Satz 1.20 gezeigt, $\sqrt{2}$ keine rationale Zahl sein kann. Danach lässt sich dann die Frage stellen, ob die a_ns selbst immer näher an die neue Zahl $\sqrt{2}$ herankommen. Sowohl für die Frage, ob die a_n^2 die Zahl 2 approximieren, als auch für die Frage, ob die a_n die Zahl $\sqrt{2}$ approximieren, braucht man aber eine Vorstellung von „Abstand“ und eine Vorstellung von „immer näher“. Für den Abstandsbegriff betrachtet man den *Absolutbetrag* $|x|$ einer reellen Zahl x, der durch $|x| = x$ für $x \geq 0$ und durch $|x| = -x$ für $x \leq 0$ definiert ist. Für die präzise Formulierung von „immer näher“ braucht man nicht nur Folgen, sondern auch den Begriff des *Grenzwerts*, der in Anhang A.3 ausführlich besprochen wird. Man sieht schon an diesen Ausführungen, dass näherungsweise konstruktive Beweise per se begrifflich zusätzliche Schwierigkeiten im Vergleich mit den bisher vorgeführten konstruktiven Beweisen aufweisen.

Beispiel 1.22 (Approximation von $\sqrt{2}$)
Wir betrachten die Folge

$$a_1 = 1,\ a_2 = \frac{a_1^2+2}{2a_1} = \frac{3}{2},\ a_3 = \frac{a_2^2+2}{2a_2} = \frac{17}{12},\ a_4 = \frac{a_3^2+2}{2a_3} = \frac{577}{508},\ \ldots .$$

Das heißt, $(a_n)_{n\in\mathbb{N}}$ ist durch $a_1 = 1$ und $a_{n+1} := \frac{a_n^2+2}{2a_n} = \frac{a_n}{2} + \frac{1}{a_n}$ definiert. Zunächst stellen wir für $n \geq 1$ fest, dass

$$a_{n+1}^2 - 2 = \frac{1}{4}\Big(a_n + \frac{2}{a_n}\Big)^2 - 2 = \frac{1}{4}\Big(a_n - \frac{2}{a_n}\Big)^2 \geq 0.$$

Aber dann gilt für $n \geq 2$

$$a_n - a_{n+1} = \frac{a_n}{2} - \frac{1}{a_n} = \frac{1}{2a_n}(a_n^2 - 2) \geq 0,$$

das heißt, die Folge ist ab $n \geq 2$ *monoton fallend*, besteht aber aus lauter Zahlen, die positiv sind. Aus der *Vollständigkeit* der reellen Zahlen kann man jetzt ableiten, dass so eine Folge einen Grenzwert $a := \lim_{n\to\infty} a_n$ hat. Intuitiv bedeutet das, dass der Abstand $|a_n - a|$ beliebig klein wird, wenn n nur groß genug ist. In diesem Beispiel liefert die Monotonieeigenschaft, dass der Grenzwert die größte reelle Zahl ist, die kleiner oder gleich jeder der Zahlen a_n für $n \geq 2$ ist. Man nennt diese Zahl die „größte untere Schranke“ oder das *Infimum* der Folge.

Während wir nachgerechnet haben, dass zum Beispiel 0 eine untere Schranke für die Folge ist, ist die Existenz einer *größten* unteren Schranke keineswegs selbstverständlich: Die Existenz solcher Infima ist eine der wesentlichen Eigenschaften reeller Zahlen. Für rationale Zahlen gilt sie nicht, denn der Grenzwert a erfüllt die Gleichung $a^2 = 2$, kann also keine rationale Zahl sein. Um zu sehen, dass a wirklich die Gleichung $a^2 = 2$ erfüllt, betrachten wir den Grenzwert der linken Seite der Formel

$$a_{n+1} = \frac{a_n}{2} + \frac{1}{a_n}. \qquad (*_n)$$

Er ist a. Man kann zeigen, dass der Grenzwert der rechten Seite von $(*_n)$ gleich $\frac{a}{2} + \frac{1}{a}$ ist. Die beiden Grenzwerte müssen gleich sein, also gilt $a = \frac{a}{2} + \frac{1}{a}$, das heißt $a^2 = 2$. Da wegen $a_n \geq 0$ für alle $n \in \mathbb{N}$ auch $a \geq 0$ gilt, haben wir $a = \sqrt{2}$. Insbesondere hat die Folge a_n^2 den Grenzwert $a^2 = 2$.

Die Vollständigkeit der reellen Zahlen garantiert die Existenz einer Lösung auch in Satz1.23, einem Spezialfall des Banach'schen Fixpunktsatzes, der nach dem polnischen Mathematiker Stefan Banach (1892–1945) benannt ist. Er ist prototypisch für näherungsweise konstruktive Existenzsätze, die man zum Gleichungslösen benutzt. Sie werden auch in viel komplizierteren Situationen eingesetzt, in denen es zum Beispiel um Differenzialgleichungen geht.

Satz 1.23 (Banach'scher Fixpunktsatz)
Sei $[-1,1] := \{x \in \mathbb{R} \mid -1 \leq x \leq 1\}$ *und* $f\colon [-1,1] \to [-1,1]$ *eine Abbildung mit der folgenden Kontraktionseigenschaft:*

$$\textit{Für alle } x, y \in M \textit{ gilt}: \quad |f(x) - f(y)| \leq \tfrac{1}{2}|x - y|.$$

Dann gibt es genau ein $x_\infty \in [-1,1]$ *mit* $f(x_\infty) = x_\infty$.

Beweis. Wir zeigen zunächst die Eindeutigkeit eines solchen *Fixpunkts.* Wenn $f(y) = y$ und $f(z) = z$ gilt, dann haben wir

$$|y - z| = |f(y) - f(z)| \leq \tfrac{1}{2}|y - z|,$$

was nur für $|y - z| = 0$, also für $y = z$, möglich ist.

Um die Existenz des Fixpunktes zu zeigen, wählt man ein $x_0 \in [-1,1]$ und setzt

$$x_1 := f(x_0), x_2 := f(x_1), x_3 := f(x_2), \ldots$$

Dann gilt

$$|x_{n+1} - x_n| \leq \tfrac{1}{2}|x_n - x_{n-1}| \leq \tfrac{1}{2^2}|x_{n-1} - x_{n-2}| \leq \ldots \leq \tfrac{1}{2^n}|x_1 - x_0|,$$

und mit der sogenannten *Dreiecksungleichung* $|a+b| \leq |a|+|b|$ bekommt man

$$\begin{aligned}
|x_{n+k}-x_n| &= |x_{n+k}-x_{n+k-1}+x_{n+k-1}-x_{n+k-2}+\ldots \\
&\quad \ldots + x_{n+2}-x_{n+1}+x_{n+1}-x_n| \\
&\leq |x_{n+k}-x_{n+k-1}|+|x_{n+k-1}-x_{n+k-2}|+\ldots \\
&\quad \ldots + |x_{n+2}-x_{n+1}|+|x_{n+1}-x_n| \\
&\leq \tfrac{1}{2^n}\Big(\tfrac{1}{2^{k-1}}+\tfrac{1}{2^{k-2}}+\ldots+\tfrac{1}{2}+1\Big)|x_1-x_0| \\
&\leq \tfrac{1}{2^{n-1}}|x_1-x_0|,
\end{aligned}$$

wobei man für die letzte Ungleichung die Abschätzung

$$\frac{1}{2^{k-1}}+\frac{1}{2^{k-2}}+\ldots+\frac{1}{2}+1 \leq 2$$

braucht, die man auch in der Form

$$2+\ldots+2^{k-1}+2^k \leq 2^{k+1}$$

schreiben kann. Um diese Ungleichung einzusehen, multipliziert man beide Seiten mit $\frac{1}{2}$, auf der linken Seite aber in der Form $-\frac{1}{2}+1$. Das liefert rechts 2^k und links

$$-\frac{2}{2}+2-\frac{2^2}{2}+2^2-\ldots-\frac{2^{k-1}}{2}+2^{k-1}-\frac{2^k}{2}+2^k = 2^k-1,$$

was die gewünschte Ungleichung beweist.

An dieser Stelle muss man die *Vollständigkeit* von $\mathbb{R}$ benutzen, um sicherzustellen, dass die Folge einen Grenzwert $x_\infty = \lim_{n\to\infty} x_n$ hat (siehe Anhang). Es bleibt zu zeigen, dass x_∞ der gesuchte Fixpunkt ist, das heißt $f(x_\infty)=x_\infty$ gilt. Da aber die Folge der $f(x_n)$ nichts anderes ist als die Folge der x_{n+1}, haben beide Folgen denselben Grenzwert, nämlich x_∞. Wieder mit der Dreiecksungleichung findet man

$$\begin{aligned}
|f(x_\infty)-x_\infty| &\leq |f(x_\infty)-f(x_n)|+|f(x_n)-x_\infty| \\
&\leq \tfrac{1}{2}|x_\infty-x_n|+|x_{n+1}-x_\infty|.
\end{aligned}$$

Da $|x_\infty-x_n|$ und $|x_{n+1}-x_\infty|$ nach Definition des Grenzwerts für große n beliebig klein werden, muss auch die linke Seite der Ungleichung beliebig klein werden. Da dieser Ausdrucks aber gar nicht von n abhängt, muss er gleich 0 sein, und es gilt $f(x_\infty)=x_\infty$. □

Existenz- und Nichtexistenzbeweise

Existenz- und Nichtexistenzbeweise weisen nach, dass Objekte einer bestimmten Art existieren bzw. nicht existieren. Wir haben schon mehrere Beispiele für Existenzbeweise (ggT, Primzahlfaktorisierung, Fixpunkte von Kontraktionen), aber auch schon einen Nichtexistenzbeweis (Lösung von $x^2 = 2$ in $\mathbb{Q}$, der Menge der rationalen Zahlen) gesehen.

Es gibt deutliche Abstufungen im Grad der Information, die ein Existenzbeweis über das Objekt liefert, dessen Existenz nachgewiesen wird. Tatsächlich berechnen konnte man unter den bisher vorgestellten Beispielen insbesondere den ggT. Prinzipiell berechenbar war auch die Primzahlfaktorisierung. Es sind aber alle Algorithmen, die man bis heute auf Computern implementieren kann, so langsam, dass man die Sicherheit im bargeldlosen Bankverkehr Verschlüsselungssystemen anvertraut, deren Funktionsweise darauf basiert, dass man zwar Primzahlen beliebiger Größe schnell miteinander multiplizieren kann, die Primfaktorzerlegung des Resultats aber für große Zahlen länger dauern würde, als man die Lebensdauer des Universums schätzt. (Das könnte sich übrigens ändern, wenn es gelänge, die von der theoretischen Physik postulierten „Quantencomputer" zu bauen. Deren Funktionsweise ließe andere mathematische Algorithmen zu, die so viel schneller wären, dass man die Verschlüsselungssysteme fundamental umstellen müsste.) Auch vom Prinzip her nur näherungsweise berechnen konnte man die Fixpunkte von Kontraktionen (siehe Beispiel 1.23).

Die reine Existenz von Lösungen für Polynomgleichungen liefert der Fundamentalsatz der Algebra (siehe Satz 1.1). Er sagt nichts darüber aus, wo man die Lösungen zu suchen hat. Andere Existenzsätze sind mehr oder weniger komplizierte Umformungen von Existenzaxiomen. Dazu gehören das *Zorn'sche Lemma* und der *Wohlordnungssatz*, die beide äquivalent zum sogenannten *Auswahlaxiom* sind, das in Abschnitt 4.2 näher diskutiert wird.

Es wurde schon angesprochen, dass man Nichtexistenzaussagen wie die Transzendenz von π, das heißt die Nichtexistenz einer Polynomgleichung mit rationalen Koeffizienten, deren Lösung π ist, in geometrische Nichtexistenzaussagen ummünzen kann: Es existiert keine Konstruktion mit Zirkel und Lineal, die zu einem Kreis ein flächengleiches Quadrat liefert. Das ist gemeint, wenn man von der Unmöglichkeit der Quadratur des Kreises spricht (siehe Seite 77). Es gibt viele unmögliche Konstruktionen, die der „Quadratur des Kreises" verwandt sind, etwa die *Winkeldreiteilung* und die *Würfelverdoppelung*.

Verwandt mit diesen Problemen ist die Frage des Gleichungslösens durch *Wurzelziehen*, wie man es von den quadratischen Gleichungen kennt, für die man

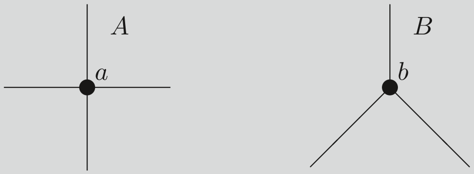
Lösungsformeln hat, in denen Quadratwurzeln vorkommen. Diese Formeln waren schon den Griechen bekannt, aber erst in der italienischen Renaissance hat man analoge Formeln für Gleichungen dritten und vierten Grades gefunden. An den Gleichungen fünften Grades versuchten sich Generationen von Mathematikern vergeblich. Erst im frühen 19. Jahrhundert konnte der norwegische Mathematiker Niels Hendrik Abel (1802–1829) nachweisen, dass ab Grad fünf solche Lösungen nicht mehr möglich sind, siehe Pesic (2005). Kurz darauf zeigte der Franzose Evariste Galois (1811–1832), wie eng diese Frage mit der Gruppentheorie zusammenhängt - eine absolut bahnbrechende Entdeckung, deren Relevanz die mathematische Fachwelt erst 30 Jahre nach dem gewaltsamen Tode Galois' (er wurde in einem Duell verwundet und erlag seinen Verletzungen) zu erkennen begann.

Auf einen speziellen Typus von Unmöglichkeitsbeweisen trifft man, wenn man untersucht, ob zwei verschieden aussehende Mengen (wie die beiden Gruppen aus Beispiel 1.6) durch eine strukturerhaltende bijektive Abbildung (Isomorphismen) ineinander übergeführt werden können. Eine Strategie zu zeigen, dass das *nicht* geht, besteht darin, solchen Mengen *Invarianten*, das heißt eine Art „Kennzahlen", zuzuordnen, die sich unter Isomorphismen nicht ändert; daher der Name Invariante. Wenn man jetzt von zwei Mengen zeigen kann, dass ihre Kennzahlen unterschiedlich sind, dann sind sie auch nicht durch Isomorphismen ineinander überführbar. Als Mathematiker sucht man nach Invarianten, die fein genug sind, die Objekte, die man unterscheiden will, auseinanderzuhalten, aber einfach genug, um bestimmbar zu sein.

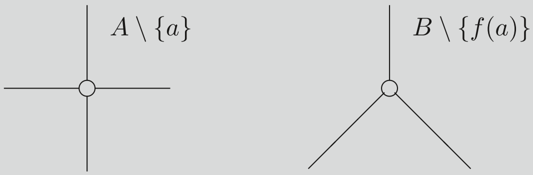
Dieses Prinzip sei an einem simplen Beispiel illustriert: Um als Zuschauer eines Fußballspiels die beiden Mannschaften auseinanderhalten zu können, braucht man nicht die Namen der Spieler; es genügt, wenn die Mannschaften unterschiedliche Trikots tragen. Der Schiedsrichter muss auch die einzelnen Spieler der Mannschaften unterscheiden können, daher tragen die Trikots Nummern.

Das folgende Beispiel ist näher am mathematischen Alltag. Um es verstehen zu können, braucht man eine Vorstellung von *stetigen* Abbildungen, ein Begriff, der bisher nicht vorkam. Anschaulich bedeutet Stetigkeit von Abbildungen, dass zusammenhängende Mengen durch die Abbildung nicht zerrissen, sondern höchstens gebogen, gestaucht und gedehnt werden können. Etwas präziser ausgedrückt: Man betrachtet Mengen, auf denen es eine Vorstellung von „benachbart" gibt und die man *topologische Räume* nennt. Stetige Abbildungen sind solche, die benachbarte Gegenden benachbart lassen, auch wenn sie die Abstände ändern. Damit sind stetige Abbildungen die „strukturerhaltenden" Abbildungen topologischer Räume. Ein topologischer Isomorphismus zwischen zwei solchen Räumen ist dann eine stetige Bijektion, deren Umkehrabbildung auch stetig ist. Man betrachtet in der Topologie zwei Räume als gleich, wenn man einen topologischen Isomorphismus zwischen ihnen finden kann. Das bedeutet, man kann das Objekt A so verbiegen, dass das Objekt B daraus wird, und zwar so, dass die Verbiegung auch wieder rückgängig gemacht werden kann. Die Idee hinter Beispiel 1.24 ist, dass man die Existenz solcher Verbiegungen ausschließen kann, wenn die Objekte aus unterschiedlich vielen zusammenhängenden Stücken bestehen.

Beispiel 1.24
Betrachte die beiden folgenden Figuren, die jeweils eine Menge von Punkten in der Ebene beschreiben:



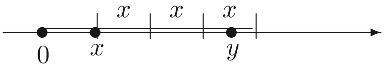
Die Frage ist, ob es eine stetige Bijektion zwischen den beiden Mengen gibt, deren Umkehrung auch stetig ist. Die Antwort ist Nein. Um das einzusehen, nehmen wir an, wir hätten so eine Abbildung $f\colon A \to B$, und betrachten das Bild $f(a)$ des Punktes a. Wir nehmen die beiden Punkte aus der jeweiligen Menge heraus (das Bild steht für den Spezialfall $f(a) = b$):



Dann bildet f die Menge $A \setminus \{a\}$ bijektiv auf die Menge $B \setminus \{f(a)\}$ ab und die Stetigkeit bleibt erhalten. Die Menge $A \setminus \{a\}$ besteht aus vier zusammenhängenden Stücken. Je nachdem, ob $f(a)$ gleich b ist oder nicht, besteht die Menge $B \setminus \{f(a)\}$ aber nur aus drei oder zwei Stücken: Wenn $f(a) = b$, dann sind das die drei Strahlen des „Sterns“, und wenn $f(a) \neq b$, dann ist $B \setminus \{f(a)\}$ ein „Stern“ mit einem verkürztem Strahl und das restliche Stück des Strahls. Da stetige Abbildungen aber zusammenhängende Stücke nicht zerreißen können, kann die Umkehrabbildung $B \setminus \{f(a)\} \to A \setminus \{a\}$ auch nicht stetig sein.

Geometrische Beweise

Geometrische Beweise sind „bildbasierte“ Beweise. Sie können unterschiedlicher Natur sein, und meist stellt sich heraus, dass die Bezeichnung „Beweis“ für die entsprechenden Argumente ein wenig zu hoch gegriffen ist. Betrachten wir zum Beispiel das *archimedische Axiom*. Es besagt, dass man zu zwei positiven reellen Zahlen $x, y > 0$ eine natürliche Zahl $n \in \mathbb{N}$ mit $nx > y$ finden kann. Stellt man sich x und y auf dem Zahlenstrahl vor, dann ergibt sich das aus der folgenden Skizze, die an die Division mit Rest erinnert, und zeigt, wie man irgendwann y „überholt“, wenn man Stücke der Länge x wiederholt aneinanderlegt:

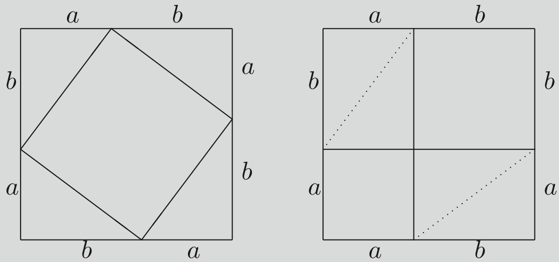


Dieser „Beweis“ ist problematisch, denn er appelliert an unsere Intuition von den reellen Zahlen. Genauer gesagt, benutzt er einfach unsere Visualisierung. Das ist gefährlich. Es stellt sich in der Tat heraus, dass es neben den reellen Zahlen, für die wir diese Eigenschaft wirklich haben (siehe Proposition A.8), auch andere Zahlbereiche gibt, für die das archimedische Axiom nicht gilt, beispielsweise für die sogenannten p-adischen Zahlen, die ebenfalls die rationalen Zahlen enthalten. Will man die reellen Zahlen von diesen anderen Zahlbereichen unterscheiden, muss man die Eigenschaft, die das archimedische Axiom beschreibt, voraussetzen. Das erklärt, warum man hier von einem *Axiom* spricht.

Deutlich überzeugender ist der folgende geometrische Beweis des Satzes von Pythagoras, dessen wesentlicher Gehalt einfach nur in zwei Bildern besteht.

Satz 1.25 (Pythagoras)
In einem rechtwinkligen Dreieck ist die Summe der Quadrate über den Katheten gleich dem Quadrat über der Hypothenuse.

Beweis. Man betrachtet die beiden folgenden Bilder.



Links hat man aus dem Quadrat der Seitenlänge $a + b$ das Quadrat über der Hypothenuse herausgeschnitten, rechts die beiden Kathetenquadrate. Übrig bleiben auf beiden Seiten vier rechtwinklige Dreiecke mit den Katheten a und b. □

So faszinierend einfach dieser geometrische Beweis des Satzes von Pythagoras auch ist. Man muss doch einschränkend sagen, dass man hier stillschweigend einige Tatsachen der euklidischen Geometrie benutzt hat, die zwar sehr einleuchtend sind, aber doch bewiesen werden müssen, zum Beispiel, dass man links wirklich ein Quadrat herausgenommen hat, die Winkel in dem inneren Viereck also wirklich alle 90 Grad sind oder dass alle rechtwinkligen Dreiecke mit den Katheten a und b die gleiche Fläche haben. Ersteres folgt, weil die Winkelsumme in einem Dreieck immer 180 Grad ist, letzteres weil Bewegungen, das heißt Abbildungen, die Längen und Winkel erhalten, automatisch auch flächenerhaltend sind. Dass es im Gegensatz zur intuitiven Vorstellung keine Selbstverständlichkeit ist, beim Zerschneiden, Bewegen und Neuzusammensetzen, die Flächen unverändert zu lassen, zeigt das *Banach-Tarski-Paradoxon*, siehe Wapner (2008). Es besagt, dass man für zwei vorgegebene Mengen A und B im dreidimensionalen Raum $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, falls sie in einer großen Kugel enthalten sind und beide mindestens eine kleine Kugel enthalten, eine Zerlegung von A in Teilmengen finden kann, die man so herumschieben kann, dass die geschobenen Teilmengen sich zu B vereinigen. Insbesondere sagt dieser Satz also, dass man eine Kugel vom Radius 1 zerschneiden und dann die Teile so wieder zusammensetzen kann, dass eine Kugel vom Radius 2 herauskommt. Das widerspricht jeder Intuition, und es ist auch so, dass die „Schnitte“, die man hier setzen muss, extrem irregulär sind. Das Banach-Tarski-Paradoxon gehört in den Zoo mengentheoretischer Pathologien, und wir wollen hier gar nicht versuchen zu erläutern, was die Idee hinter dem Beweis ist. Es geht uns eher darum, die Vorsicht zu motivieren, die man als Mathematiker walten lässt, wenn anschauliche Argumente ins Spiel kommen.

Ein weiteres Problem bei geometrischen Argumenten ist, dass man sich zuerst auf eine Geometrie festlegen muss: Euklids Geometrie galt bis ins 19. Jahrhundert als die Geometrie der realen Welt. Eines der Axiome dieser Geometrie, das sogenannte *Parallelenpostulat* (Seite 71 und 179), wirkte im Axiomensystem wie ein Fremdkörper, wurde aber für viele Beweise gebraucht. 2000 Jahre hat man vergeblich versucht, das Parallelenpostulat aus den anderen Axiomen herzuleiten. Schließlich stellte man fest, dass es auch eine Geometrie gibt, die alle Axiome bis auf dieses eine erfüllen, und dieses Axiom somit nicht aus den anderen Axiomen abgeleitet werden kann. In dieser *nichteuklidischen Geometrie* ist die Winkelsumme eines Dreiecks kleiner als 180 Grad und der Satz von Pythagoras falsch.

Wie schreibt man einen Beweis auf?

Von einer Bierparty an der University of Chicago Mitte des 20. Jahrhunderts ist folgendes Lied überliefert, in dem der Siegeszug eines gewissen „Bourbaki“ durchaus ironisch konstatiert wird:

Analysts, topologists, geometers agree:
When it comes to generality, there's none like Bourbaki!
One theorem by them will equal N by you and me,
Bourbaki goes marching on!

Glory, glory halleluja,
Their generality will fool ya–
They're axiomatically peculiah!
Bourbaki goes marching on!

Algebras and groups and rings suffice for you and me –
cogebras and bigebras abound in Bourbaki;
functors contravariant, defined by their degree,
Bourbaki goes marching on!

In einer Reihe weiterer Strophen wird noch über die Abstraktheit und Allgemeinheit in Bourbaki gelästert. Bourbaki ist das Pseudonym für eine Gruppe junger französischer Mathematiker (darunter der auf Seite 37 bereits erwähnte André Weil), die als Reaktion auf die verkrusteten Strukturen in Frankreichs akademischer Ausbildung ein neues, grundlegendes mathematisches Universallehrwerk zu schreiben begannen. Das war in den 1940er Jahren. Der letzte von etwa 30 Bänden erschien 1982. Das Vorhaben wurde dann aufgegeben, weil es sich als nicht realisierbar entpuppte. Die Schwierigkeit in der Konzeption solcher Werke liegt nicht zuletzt darin, dass ein Netzwerk von Zusammenhängen sich nicht vernünftig in einer baumartig angeordneten Buchreihe abbilden lässt.

Bourbaki und ihre Methode, einen Satz und seinen Beweis immer mit der minimalen Menge von Voraussetzungen, das heißt in der größtmöglichen Allgemeinheit zu formulieren, waren zwischenzeitlich sehr in Misskredit geraten. Dem Argument, dass man einen mathematischen Sachverhalt am besten versteht, wenn man sieht, welche Tatsachen wirklich verwendet werden, hielt man entgegen, dass mathematisches Verständnis, wie zuvor schon angedeutet, auf internen Modellen des Lesers oder Zuhörers beruht, deren Bildung durch übertriebene Allgemeinheit eher behindert denn gefördert wird. Umgekehrt produzieren allzu spezielle Varianten der zu behandelnden Resultate beim Lernenden zu einfache interne Modelle und die Querverbindungen zwischen Teilgebieten sind schwerer zu erkennen.

Es geht also wie immer darum, einen geeigneten Mittelweg zu finden. Dabei muss man auf die spezielle Zielgruppe Rücksicht nehmen. Bücher, in denen Redundanz nach Möglichkeit vermieden wird, haben relativ wenig Fließtext und sind voller Binnenzitate. Das macht sie gut geeignet für kundige Leser, die sich von der Richtigkeit einer bestimmten Aussage überzeugen wollen. Der Neuling auf dem behandelten Gebiet dagegen erschließt sich in solchen Texten die Einordnung der Ergebnisse nicht. Am anderen Ende der Skala stehen Bücher, die sich an reine Anwender richten und mehr Methodensammlungen sind, gut dafür geeignet bestimmte Aufgabentypen einzutrainieren. Auch hier kommt das Verständnis für den größeren Zusammenhang oft zu kurz. Es sei angemerkt, dass die historische Entwicklung Bourbakis Themenauswahl und den hohen Grad an Abstraktion im Nachhinein rechtfertigte. Cogebras und Bigebras gehören heute beispielsweise zum Repertoire der theoretischen Physik.

Der angemessene Abstraktionsgrad bei der Kommunikation über Mathematik hängt auch davon ab, wo im Ausbildungsprozess eine Lehrveranstaltung angesiedelt ist. In einer Einführungsvorlesung wird man sicher der Bildung von Intuition einen größeren Raum einräumen. In Spezialvorlesungen, in denen Studierende die nötigen Grundlagen zu eigenständiger Forschung erwerben möchten, konzentriert man sich besser auf eine stringente Darstellung und verlagert die Intuitionsbildung auf das persönliche Gespräch.

Auf die Gesamtausbildung von Mathematikern bezogen wird man ein Spiralmodell bevorzugen, in dem Konzepte mehrfach, in zunehmender Tiefe, behandelt werden. Die Präsentation des Stoffes ersetzt natürlich in keinem Fall die aktive Auseinandersetzung mit den Themen. Darum sind Lehrbücher für Mathematik voller Übungsaufgaben.

Mode und Geschmack: Was soll man beweisen?

Nur den wenigsten Nichtmathematikern ist klar, dass es in der Mathematik Moden gibt und dass Mathematiker voneinander behaupten, der eine hätte einen guten, der andere einen schlechten oder gar keinen Geschmack. Viele Laien glauben, Mathematik sei ein fertig gestelltes Gebäude, in dem es nur um falsch oder richtig gehe. Wenn es nur einen richtigen und einen falschen Weg gibt, bedarf es keines guten Geschmacks bei der Auswahl des Weges. Wenn man sich aber einmal klar gemacht hat, dass es zum Beispiel bei allem Reichtum an mathematischer Theorie in der Regel nicht einmal möglich ist, die in der Praxis der Ingenieure auftauchenden Gleichungen zu lösen, dann ist es auch nicht schwer einzusehen, dass die Fülle der offenen Probleme eine Auswahl erzwingt.

Welche Probleme nun ausgesucht werden, hängt von vielen Faktoren ab: außermathematischen, wie einer Nachfrage nach Spezialisten auf einem bestimmten Gebiet, beispielsweise Strömungsmechanik oder Kontrolltheorie für die Entwicklung von Raketen in Kriegszeiten, und innermathematischen. Innerhalb der mathematischen Gemeinschaft kann es zum Beispiel vorkommen, dass ein Mathematiker einen überraschenden Satz beweist und dazu eine neue Querverbindung oder eine unerwartete Anwendung findet. Ein etablierter Forscher stößt dazu und propagiert die neue Erkenntnis. Es bilden sich Seminare zum Thema an den Brennpunkten der Zunft, Nachfolgearbeiten entstehen. Die Ergebnisse werden in hochangesehenen Zeitschriften veröffentlicht und dringen ins Bewusstsein einer breiteren Öffentlichkeit. Wenn die neue Erkenntnis tatsächlich bahnbrechend ist, etabliert sich die Forschungsrichtung längerfristig, wenn nicht, gerät sie nach einiger Zeit in Vergessenheit oder tritt zumindest ins zweite Glied zurück und harrt einer neuen Sternstunde.

Wenn es auch oft keine Übereinstimmung darüber gibt, was guter mathematischer Geschmack und lohnende Forschungsgebiete sind, so kann man doch die folgenden Kriterien anführen:

- Verpönt ist die Verallgemeinerung um ihrer selbst willen. Kein Herausgeber wird eine Arbeit veröffentlichen, in der eine Eigenschaft für 48-mal differenzierbare Funktionen nachgewiesen wird, die vorher nur für 49-mal differenzierbare Funktionen bewiesen war. Es sei denn, der Autor hätte einen triftigen Grund, auf die 49. Ableitung verzichten zu wollen. Und da steckt natürlich das Problem.
- Verpönt sind lange Rechnungen, wenn sie durch begriffliche Begründungen ersetzt werden können. Dahinter steckt wieder die Vorstellung, dass eine komplizierte Rechnung den Aufbau eines geeigneten internen Modells behindert. Manchmal haben solche Rechnungen aber eben doch einen ganz eigenen Erkenntniswert.
- Verpönt sind Sätze, in deren Formulierung 713 Bedingungen vorkommen. Man geht dann davon aus, dass die Sache noch nicht richtig verstanden ist, möglicherweise die richtigen Begriffe noch gar nicht gefunden sind. Vielleicht sind das aber genau die 713 Bedingungen, die in (außermathematisch) natürlicher Weise hier erfüllt sind.
- Gefragt sind Aussagen, die verschiedene Disziplinen (der Mathematik) miteinander verbinden. Vielleicht ist die Verbindung aber banal.
- Gefragt sind Beweise mit neuen Ideen, die transparent, aber nicht trivial sind. Was immer das heißt.
- Gefragt sind *schöne* Beweise. Die Schönheit der Mathematik allerdings ist nur für Eingeweihte sichtbar.

Zusammenfassung

Beweise sind Argumentationsketten für die Richtigkeit mathematischer Aussagen. Im Idealfall bestehen sie aus vollständigen logischen Schlussketten, die von den gemachten Annahmen bis zur gewünschten Aussage führen. Dieser Grad an Genauigkeit wird in der Praxis so gut wie nie erreicht, und es gibt auch grundsätzliche theoretische Probleme mit dem Konzept des Beweises, auf die hier nicht eingegangen werden kann. Dennoch besteht ein breiter Konsens über den Korpus abgesicherter mathematischer Erkenntnisse.

In der mathematischen Praxis gibt es diverse Beweistechniken, die von konstruktiven Problemlösungen bis zu reinen Existenzaussagen reichen. Ebenso gibt es eine ästhetische Wahrnehmung von Beweisen. Welche Aussagen bewiesen werden, hängt davon ab, welche Fragestellungen unter den Mathematikern als relevant betrachtet werden. Dabei unterliegen solche Einschätzungen durchaus auch Modeströmungen.

1.7 Definitionen

Die Bedeutung von Definitionen wird im Allgemeinen unterschätzt. Der Begriff wird häufig im Dreiklang *Definition - Satz - Beweis* genannt, in der Absicht, die Fantasielosigkeit in der Lehrpraxis in der Mathematik anzuprangern. Dabei gerät leicht aus dem Blick, dass es die Definitionen sind, die die Begriffe fassbar machen: In Definitionen fixiert man mathematische Konzepte, mit denen man dann in präziser Art und Weise arbeiten kann. Wie wichtig es ist, die richtigen Begriffe in Definitionen zu fassen, haben wir andeutungsweise schon in Abschnitt 1.2 gesehen, wo das Problem, Gleichungen zu lösen, auf neue Begriffe führte. Je mehr man von der Mathematik und vor allem von ihrer Entwicklungsgeschichte weiß, desto klarer erkennt man die zentrale Rolle, die die Auswahl der Begriffsbildungen spielt. Ungünstige Begriffe können den Blick auf das Wesentliche verstellen, und gut gewählte Begriffe erleichtern es, komplexe Einsichten für andere verständlich darzustellen.

Ein gutes Beispiel für die ordnende Funktion von Definitionen ist die höherdimensionale Geometrie, die in der zweiten Hälfte des 19. Jahrhunderts in verschiedenen Ausprägungen insbesondere in Italien enorme Fortschritte machte (Seite 37). Allerdings gelang es den italienischen Geometern nicht, ihre Intuitionen in klare Begriffe zu gießen, und es mutet wirklich abenteuerlich an, wie sie durch ihre komplizierten Rechnungen voller Indizes zu geometrischen Einsichten gelangen konnten. Mit den klaren Begriffen von *differenzierbaren Mannigfaltigkeiten*, *algebraischen Varietäten* und *Geradenbündeln*, die wir den Entwicklungen des 20. Jahrhunderts verdanken, wird die Struktur vieler dieser Rechnungen erst allgemein kommunizierbar.

Ein weiteres elementares Beispiel für die Bedeutung gut gewählter Definitionen ist die Definition der komplexen Zahlen. Schon 1560 hat Raphael Bombelli (1526–1572) in Untersuchungen kubischer Gleichungen formal mit Wurzeln aus negativen Zahlen gerechnet, und Leonhard Euler (1707–1783) benutzte die Wurzel aus -1 virtuos als eigenständige Größe. Die Vorstellung von der Zahlenebene (Seite 9) prägte aber erst Carl Friedrich Gauß (1777–1855), und den *Körper* der komplexen Zahlen gibt es unter diesem Namen erst seit Ende des 19. Jahrhunderts. Heute haben die komplexen Zahlen nichts Geheimnisvolles mehr. Man kann sie in der Schule unterrichten, und sie sind jedem Elektrotechniker wohlvertraut.

Das Beispiel der komplexen Zahlen zeigt auch, dass es manchmal Jahrhunderte gedauert hat, bis man die Essenz aus Beispielen und Rechnungen herausdestilliert hatte. Dieses „Auf-den-Punkt-Bringen“ von vorhandenem Wissen ist typisch für gute Definitionen. Hin und wieder erlaubt eine neue Definition sogar eine völlig neue Sicht auf altbekannte Probleme. Solche Aspekte spielen in den Abschnitten 2.5 und 2.6, in denen es um Vereinheitlichung und Perspektivenwechsel geht, eine wichtige Rolle.

In Abschnitt 1.6 mussten wir an einigen Stellen an die Anschauung appellieren, weil präzise Definitionen noch nicht eingeführt werden konnten. Mehrfach wurde zum Beispiel die Vollständigkeit der reellen Zahlen bemüht. Um den Begriff „Vollständigkeit“ sauber zu erklären, muss man weiter ausholen, was den Rahmen dieses Abschnitts gesprengt hätte. Andererseits ist die Beschreibung der reellen Zahlen über eine Reihe von Eigenschaften oft der Einstieg in das Mathematikstudium und gleichzeitig die erste Begegnung der Studierenden mit der axiomatischen Methode. Die Axiomatik der reellen Zahlen wird deshalb in Abschnitt 2.2 (algebraische Aspekte) und im Anhang (Anordbarkeit und Vollständigkeit) im Detail erklärt. Im Anhang wird auch gezeigt, wie man eine Menge, die diesen Axiomen genügt, wirklich konstruiert.

Literatur: Artmann (1999), Behrends (2003), Byers (2007), Cigler (1995), Dieudonné (1982), Ebbinghaus et al. (1992), Engelen et al. (2010), Feynman (1997), Gowers (2002), Griffith und Hilton (1970), Halmos (1972), Hawking und Mlodinow (2010), Hersh (1997), Manin (2007), Mashaal (2002), Penrose (2004), Singh (1998), Sonar (2001)

http://www.springer.com/978-3-8274-2931-5

Mathematik – ein Reiseführer
Hilgert, I.; Hilgert, J.
2012, X, 278 S. 20 Abb., Softcover
ISBN: 978-3-8274-2931-5
A product of Spektrum Akademischer Verlag