SCHRIFTEN ZUR WISSENSCHAFTLICHEN WELTAUFFASSUNG

HERAUSGEGEBEN VON

PHILIPP FRANK
o. ö. PROFESSOR AN DER UNIVERSITÄT PRAG

UND

MORITZ SCHLICK
o. ö. PROFESSOR AN DER UNIVERSITÄT WIEN

BAND 3

WAHRSCHEINLICHKEIT STATISTIK UND WAHRHEIT

EINFÜHRUNG IN DIE NEUE WAHRSCHEINLICHKEITSLEHRE UND IHRE ANWENDUNG

VON

RICHARD von MISES
PROFESSOR AN DER UNIVERSITÄT ISTANBUL

ZWEITE, NEUBEARBEITETE AUFLAGE

Springer-Verlag Wien GmbH 1936

SCHRIFTEN ZUR
WISSENSCHAFTLICHEN WELTAUFFASSUNG
HERAUSGEGEBEN VON
PHILIPP FRANK
o. ö. PROFESSOR AN DER
UNIVERSITÄT PRAG
UND
MORITZ SCHLICK
o. ö. PROFESSOR AN DER
UNIVERSITÄT WIEN
BAND 3

WAHRSCHEINLICHKEIT STATISTIK UND WAHRHEIT

EINFÜHRUNG IN DIE NEUE WAHRSCHEINLICHKEITSLEHRE
UND IHRE ANWENDUNG

VON

RICHARD VON MISES
PROFESSOR AN DER UNIVERSITÄT ISTANBUL

ZWEITE, NEUBEARBEITETE AUFLAGE

Springer-Verlag Wien GmbH 1936

Ursprünglich erschienen bei Julius Springer in Vienna 1936

ISBN 978-3-662-41724-9 ISBN 978-3-662-41863-5 (eBook)
DOI 10.1007/978-3-662-41863-5

Vorwort.

Die allseits freundliche Aufnahme, die mein Buch bei seinem ersten Erscheinen 1928 gefunden hat und die jetzt eine Neuauflage erforderlich macht, muß ich wohl in erster Linie dem gesteigerten Interesse zuschreiben, das in der letzten Zeit allen Fragen der Wahrscheinlichkeitslehre und ihrer Anwendung entgegengebracht wird.

Der wesentliche Inhalt der hier folgenden Vorträge besteht in der Darlegung einer neuen Auffassung der Wahrscheinlichkeitsrechnung. Sie ist vor mehr als zwanzig Jahren entstanden und bedeutete bei ihrer ersten Veröffentlichung 1919 einen völligen Bruch mit den bis dahin verbreiteten Anschauungen. Inzwischen ist eine große Reihe von Schriften, zum Teil durch meine Arbeiten angeregt oder durch sie beeinflußt, erschienen, in denen ein mehr oder weniger verwandter Standpunkt vertreten wird. Ich habe sorgfältig geprüft, was sich von dem von anderer Seite Vorgebrachten zur Verbesserung oder Vereinfachung meiner Theorie verwerten ließ. Der Leser wird den Niederschlag davon in dem gegenüber der ersten Auflage stark vermehrten dritten Abschnitt „Kritik der Grundlagen" finden. Auch alle anderen Teile des Buches sind nach verschiedenen Gesichtspunkten ergänzt worden; vielleicht wird die Hinzufügung eines vollständig ausgeführten Rechnungsbeispiels am Ende des zweiten Vortrages dem Anfänger willkommen erscheinen. Der letzte, den physikalischen Problemen gewidmete Teil mußte naturgemäß eine Erweiterung erfahren im Hinblick auf die neue Wendung, die der physikalischen Statistik durch die Quantenmechanik gegeben wurde, eine Wendung übrigens, die mit der von mir vertretenen Grundauffassung in bestem Einklang steht. In den Fragen, die sich mit dem allgemeinen Kausalitätsproblem berühren, konnte ich mich auf Andeutungen beschränken, da sie in dem kürzlich in dieser Sammlung erschienenen ausgezeichneten Werke von Philipp Frank eingehende Behandlung erhalten haben. So hoffe ich jedenfalls, daß man das Maß der aufgenommenen Ergänzungen berechtigt finden und daß die nicht unbedeutende Vermehrung des Umfanges dem Buche nicht zum Nachteil gereichen wird.

Nichts geändert wurde an der Form der Darstellung, die für den Nichtmathematiker bestimmt ist. Sie verzichtet nicht nur äußerlich fast vollständig auf Formeln, sondern unterläßt auch alles nähere Eingehen auf Probleme, die sich vernünftigerweise nur unter Zuhilfenahme der mathematischen Zeichensprache behandeln lassen. Dauernd habe ich mich bemüht, eindeutig und verständlich zu bleiben, selbst wenn es auf Kosten der Systematik ging, und habe mit gelegentlichen Wiederholungen nicht gespart, wo sie mir nützlich schienen, um den Leser die Hauptzüge des Gedankenganges nicht aus den Augen verlieren zu lassen. Diesem Ziel dienen auch die mehrfach eingeschalteten „Zusammenfassungen" der jeweils erreichten Ergebnisse. Die sechs Abschnitte des Buches sind wirklich als fortlaufend gesprochene Vorträge zu denken, die Zwischenüberschriften sollen nur beim Nachblättern die Übersicht erleichtern. — In den am Schluß beigefügten Anmerkungen sind neben anderen Literaturnachweisen auch meine eigenen Veröffentlichungen angeführt, die als Ergänzung zu dem hier Vorgetragenen in Betracht kommen. Insbesondere sei der mathematisch Interessierte auf meine 1931 erschienenen „Vorlesungen über Wahrscheinlichkeitsrechnung" verwiesen.

Im Vorwort zur ersten Auflage habe ich großen Nachdruck darauf gelegt, meine Ausführungen gegenüber allen metaphysischen Ansprüchen zu distanzieren, die etwa durch das Titelblatt geweckt werden könnten. Ich muß auch jetzt sagen, daß ich mich durchaus als Mathematiker und Naturwissenschaftler fühle und in der Wahrscheinlichkeitsrechnung nichts anderes sehe als eine systematische Beschreibung bestimmter beobachtbarer Erscheinungen mit den gleichen gedanklichen Hilfsmitteln, wie sie in jedem Teil der exakten Wissenschaften gebraucht werden. Doch dies soll keine Widmung „In philosophos" sein. Denn klares Denken, gewissenhaftes Prüfen aller Aussagen an der Beobachtung und Unterlassen jeglichen Wortemachens sind wohl Voraussetzungen, die heute besonders selten erfüllt werden, aber es gibt auch Philosophen, die ihnen zu genügen trachten; und mit diesen wollen wir uns gewiß verständigen.

Istanbul, im April 1936.

Der Verfasser.

Inhaltsübersicht.

Erster Vortrag.

Definition der Wahrscheinlichkeit.

Wenn ich in aller Kürze in den Sinn der Gegenüberstellung einführen soll, die in dem Titel dieses Buches zum Ausdruck gebracht ist, so darf ich vielleicht die scherzhafte Äußerung eines Engländers zitieren, der einmal gesagt hat, es gäbe drei Arten von Lügen: erstens die Notlüge, die entschuldbar sei, zweitens die gemeine Lüge, für die keine Entschuldigung gelten könne, und drittens die Statistik. Es bedeutet ungefähr dasselbe, wenn wir gesprächsweise bemerken, mit Zahlen lasse sich „alles beweisen“ oder, in Abänderung eines bekannten GOETHEwortes, mit Zahlen lassen sich „trefflich streiten, mit Zahlen ein System bereiten“. Dem allen liegt die Auffassung zugrunde, daß Schlüsse, die auf statistischen Überlegungen oder auf Wahrscheinlichkeitsrechnung aufgebaut werden, zumindest sehr unsicher, wenn nicht geradezu unbrauchbar sind. Ich will gewiß nicht bestreiten, daß vieles Unsinnige und Unhaltbare unter dem Deckmantel der Statistik auftritt. Aber es ist das Ziel, das ich meinen Ausführungen stecke, hier zu zeigen, daß man sehr wohl, von statistischen Feststellungen ausgehend, über einen geläuterten und entsprechend präzisierten Wahrscheinlichkeitsbegriff, zu Erkenntnissen und Behauptungen gelangen kann, die es an „Wahrheitsgehalt“ und Zuverlässigkeit sowie an praktischer Brauchbarkeit mit den Ergebnissen jedes Zweiges der exakten Naturwissenschaft aufnehmen können. Ich muß dazu freilich bitten, mir auf einem langen und gewiß nicht mühelosen Wege zu folgen, der uns über manche, zunächst vielleicht überflüssig scheinende, vorbereitende Gedankengänge zu unserm Ziele führen wird.

Berichtigung des Sprachgebrauchs.

„Unsere ganze Philosophie ist Berichtigung des Sprachgebrauchs“, sagt einmal LICHTENBERG; wieviel Streit und wieviel Irrtümer wären in der Geschichte der Wissenschaften vermieden worden, wenn man diesen Ausspruch immer richtig beherzigt hätte! Wir wollen es jedenfalls an Sorgfalt nicht fehlen lassen, wenn wir zunächst dem Worte Wahrscheinlichkeit nachgehen, um allmählich durch eine ganze Reihe von Überlegungen hindurch zu einem geeigneten wissenschaftlichen Begriff der Wahrscheinlichkeit aufzusteigen. Daß hierin, in der Aufsuchung einer vernünftigen Fassung des Wahrscheinlichkeitsbegriffes, der Schlüsselpunkt für die Aufklärung des Verhältnisses zwischen Statistik und Wahrheit liegt, habe ich eben angedeutet und es wird sich in der Folge, wie ich denke, ganz klar herausstellen.

Wir verwenden im gewöhnlichen Sprachgebrauch ohne Schwierigkeit und in mannigfaltiger Weise den Ausdruck „wahrscheinlich“. Wir sagen, es sei wahrscheinlich, daß es morgen regnen wird, daß heute an einem fernen Orte Schnee fällt oder daß es an einem bestimmten Tage des Vorjahres kälter gewesen ist als heute. Wir sprechen von der größeren oder geringeren Wahrscheinlichkeit, wohl auch von der Unwahrscheinlichkeit der Schuld eines Angeklagten oder der Richtigkeit einer Zeugenaussage. Wir drücken uns genauer aus, es sei wahrscheinlicher, bei dieser Verlosung den Haupttreffer zu gewinnen als bei jener auch nur den kleinsten Gewinn zu erzielen. In allen diesen Fällen geraten wir auch nicht in Verlegenheit, wenn man uns nach dem Sinne dieser Wahrscheinlichkeitsaussagen fragt, nämlich solange nicht, als sich der Fragende mit einer Umschreibung des Gesagten zufrieden gibt. Es steht uns dann eine ganze Reihe von Ausdrücken zur Verfügung, mit denen wir uns helfen, wir sprechen von Vermutung, von ungenauem, unvollständigem oder unsicherem Wissen, von Zufall, von stärkeren oder schwächeren Gründen des Fürwahrhaltens usf.

Worterklärungen.

Aber die Schwierigkeiten werden sofort recht groß, wenn man eine genauere Erklärung oder gar eine Definition von dem, was „Wahrscheinlichkeit“ heißt, geben soll. Jemand könnte auf den Gedanken verfallen, in einem deutschen Wörterbuch nachzusehen, was eigentlich das Wort „wahrscheinlich" bedeutet. Im dreizehnten Bande von JAKOB und WILHELM GRIMM finden wir spaltenlange Auskunft. Es heißt da zur Worterklärung: Probabilis, in alter Zeit „der Wahrheit gleich" oder „der Wahrheit ähnlich", dann „mit einem Schein der Wahrheit", erst seit der Mitte des 17. Jahrhunderts „wahrscheinlich"; im Englischen heute noch verylike neben probable. Es folgen eine ganze Reihe von Belegstellen, in denen der Gebrauch des Wortes, meist der bei Philosophen, erklärt wird. Ich gebe nur einiges wieder: „Weil die Wahrscheinlichkeit etwas ist, so zwischen dem Wahren und Falschen gleichsam mitten inne ist", sagt THOMASIUS 1688. „Wahrscheinlich heißt die Einsicht, wovon das Gegenteil nicht gänzlich widersprechend oder unmöglich ist", so lehrt REIMARUS in seiner Vernunftlehre. Und der große KANT: „Wahrscheinlich heißt dasjenige, was für wahr gehalten, mehr als die Hälfte der Gewißheit auf seiner Seite hat." Vielleicht hat darnach jemand noch den Wunsch, zu hören, was die heutige Philosophie als Frucht ihrer jahrhundertelangen Entwicklung zu der Sache zu sagen weiß. Ich führe daher hier wörtlich an, was ich in ROBERT EISLERS Wörterbuch der philosophischen Begriffe von 1910 gefunden habe: „Wahrscheinlichkeit ist (subjektiv) ein Grad der Gewißheit, beruhend auf starken oder überwiegenden Motiven zum Urteilen, so aber, daß diesen Motiven immerhin noch andere gegenüberstehen, die berücksichtigt werden wollen oder sollen; objektiv wahrscheinlich ist das, was, auf eine Reihe von (objektiven) Gründen gestützt, das Denken als wahr anzunehmen, zu erwarten sich mit gewisser Zuversicht berechtigt weiß. Je nach der Art und Menge der Gründe oder Instanzen, auf die sich das Wahrscheinlichkeitsurteil und der Wahrscheinlichkeitsschluß stützt, gibt es verschiedene Grade der Wahrscheinlichkeit."

Nun, über Wert oder Unwert dieser und ähnlicher „philosophischer" Erklärungen läßt sich nicht streiten. Sie bestehen stets darin, daß ein Wort durch andere, meist durch sehr viele andere, ersetzt wird. Wem die letzteren geläufiger sind als das erste, mag in der Gleichsetzung eine Erklärung sehen, ein anderer wird es jedenfalls nicht können. Wenn jemand besser versteht, was es heißt, „mehr als die Hälfte der Gewißheit auf seiner Seite haben", als was „wahrscheinlich" bedeutet, so ist das seine ganz persönliche Angelegenheit. Eine Berichtigung des Sprachgebrauchs finden wir in allen diesen Worterklärungen nicht.

Synthetische Definition.

Was aber muß man anstellen, um über das offenbar unzulängliche Wörterbuchniveau hinauszukommen? Die Art der Begriffsbildung, die sich innerhalb der exakten Wissenschaften in den letzten Jahrhunderten entwickelt hat, weist uns klar und sicher den Weg, der zum Erfolg führt, wenn anders das, was in der Mathematik und der Physik unserer Zeit erreicht worden ist, als ein Erfolg angesehen wird. Nur um die landläufigen, unexakten Denkgewohnheiten nicht allzu unvermittelt in andere Bahnen zu zwingen, will ich als Vorbereitung auf die genauen Formulierungen, die später durchgeführt werden sollen, ein Beispiel fortgeschrittener Auffassung des Prozesses der Begriffsbildung aus einem geisteswissenschaftlichen Gebiete, das dem Bereich der allgemeinen Bildung nahesteht, anführen. In seinem geistvollen Buch über den proletarischen Sozialismus sucht Werner Sombart nach einer brauchbaren Definition für den Gegenstand seiner Untersuchung. Nachdem er unter den verschiedenen, dem Sprachgebrauch entnommenen Deutungen Umschau gehalten, fährt er fort: „Bleibt also nur die Möglichkeit, den Sozialismus als Idee zu fassen, einen vernünftigen Begriff zu bilden, d. h. einen Sachverhalt zu bezeichnen, in dem wesentliche Charakterzüge (Merkmale) zu einer lebendigen Einheit sinnvoll zusammengefügt sind, einen Begriff, dessen ‚Richtigkeit' wir allein aus seiner Fruchtbarkeit als schöpferische Idee im Leben ebenso sehr wie als wissenschaftliche Hilfskonstruktion zu ermessen vermögen." In der Tat enthalten diese Worte fast alles, was das Wesen wissenschaftlicher Begriffsbildung kennzeichnet. Ich lege besonderen Nachdruck auf folgende zwei Merkmale, die ich ausdrücklich hervorhebe: Erstens der Inhalt des Begriffes wird nicht von irgendeiner Wortbedeutung abgezogen, ist also auch nicht von einem mehr oder weniger verbreiteten Sprachgebrauch abhängig, sondern er wird künstlich aufgebaut und abgegrenzt und erhält dann eine geeignete Wortbezeichnung wie ein Schildchen aufgesetzt. Und zweitens der Wert der Begriffsbildung wird nicht daran gemessen, ob Deckung mit irgendwelchen geläufigen Vorstellungskreisen erreicht ist, sondern ausschließlich darnach, was sie als Werkzeug weiterer wissenschaftlicher Untersuchungen für die Wissenschaft und dadurch mittelbar für das Leben leistet.

Unter Verwendung einer von KANT eingeführten Bezeichnung könnte ich auch sagen: Es ist nicht unsere Absicht, eine analytische Definition der Wahrscheinlichkeit zu geben, sondern eine synthetische; wobei wir die grundsätzliche Möglichkeit analytischer Definitionen dahingestellt sein lassen können.

Wahl der Bezeichnung.

Zu der erstangeführten Eigenschaft der synthetischen Definition noch eine Bemerkung: Wer in dem eben erörterten Sinn einen neuen wissenschaftlichen Begriff schafft, wird sicher die Neigung empfinden, ihm einen neuen Namen zu geben, d. h. ein Wort zu suchen, das nicht durch eine andere, womöglich nahe benachbarte Verwendung schon belegt ist. Auf diese Weise sind, da man aus verständlichen Gründen neue Worte in der eigenen Sprache nur schwer bilden kann, in großer Zahl Lehn- und Fremdwörter als Fachausdrücke in die Wissenschaft gelangt. Die an sich nicht unberechtigten Bestrebungen der Sprachreiniger können nur schädlich wirken, wenn sie diesem Tatbestand nicht Rechnung tragen, sondern durch unbedachte Verdeutschungen den Prozeß der Begriffsbildung gewissermaßen rückgängig machen. Es wäre außerordentlich nützlich, wenn man für den wissenschaftlichen Wahrscheinlichkeitsbegriff, den wir im folgenden entwickeln werden, ein sonst ungebräuchliches Wort, wie etwa „Probabilität", setzen würde. Allein dies entspricht einmal nicht dem bisherigen Gebrauch und unentbehrlich sind schließlich die Fremdworte in der Fachsprache nicht, wie man an vielen Ausdrücken der Mechanik: Kraft, Masse, Arbeit usw. erkennt; wobei ich freilich die Andeutung nicht unterdrücken kann, daß manchem, auch unter den Fachleuten, die sich berufsmäßig mit Mechanik beschäftigen, das Verständnis für diese Begriffe weit besser aufgehen würde, wenn sie durch lateinische Termini statt durch Worte der Umgangssprache bezeichnet wären. Schließlich sind auch die Forscher nur Menschen, die den größten Teil ihres Lebens hindurch die Sprache in der gewöhnlichen Weise handhaben und allen Schwierigkeiten der Sprachverwirrung ausgesetzt sind, der sie sich manchmal nur allzu willig hingeben.

Der Arbeitsbegriff der Mechanik.

Bevor ich mich jetzt der Erklärung unseres „Probabilitätsbegriffes“ zuwende, will ich noch kurz die in mancher Hinsicht ähnlichen Verhältnisse ins Gedächtnis zurückrufen, die bei fast allen Begriffsbildungen in den exakten Wissenschaften vorliegen. Ich greife z. B. den heute allen Gebildeten geläufigen Begriff der Arbeit oder Arbeitsleistung, wie er in der wissenschaftlichen Mechanik verwandt wird, heraus. Jeder von uns gebraucht das Wort „Arbeit“ in vielfältigstem Sinne. Wenn ich schon einzelne ferner stehende Bedeutungen, wie die in der Verbindung „Kapital und Arbeit“ oder „Arbeit an sich selbst“ und ähnliche, ganz außer Betracht lasse, bleibt noch genug übrig, was mit der Arbeit, wie sie die exakte Wissenschaft definiert, nichts oder nur sehr wenig zu tun hat. Die Definition lautet bekanntlich in ihrer einfachsten Form: Arbeit ist das Produkt aus Kraft und Weg; genauer: das skalare Produkt von Kraft- und Verschiebungsvektor; noch richtiger: das Linienintegral der Kraft. Auf menschliche Verrichtungen angewandt, stimmt das sehr gut — es genügt für den Nicht-Mathematiker, die erste Form der Definition zu betrachten —, sobald man an das Heben eines Gewichtes, das Drehen einer Kurbel, das Treten eines Fußhebels denkt. Die Arbeit ist um so größer, je größer das zu hebende Gewicht und je größer die Höhe ist, um die es gehoben wird.

Aber wie wenig wird diese in der wissenschaftlichen Mechanik eingebürgerte Definition schon den Fällen halbmechanischer Tätigkeit, beispielsweise dem Schreiben auf der Schreibmaschine, dem Spielen eines Musikinstruments, gerecht; man wird kaum behaupten können, daß hier noch die Produkte aus Kraft und Weg (der die Arbeit leistenden Finger) ein wesentlich richtiges Maß der Leistung ergeben. Gar nicht zu reden von der Arbeit, die der Verfasser eines Buches, ein Künstler, ein Arzt leistet. Auch ein erhebliches Gewicht mit gestrecktem Arm in Ruhe halten, ist eine schwere Arbeit, obgleich hier das Produkt aus Kraft und Weg den Wert Null hat; und es ist zumindest sehr zu bezweifeln, ob bei Bewegungen, wie sie in Spiel und Sport auftreten, die nach der mechanischen Theorie berechnete Arbeitsgröße ein auch nur annähernd brauchbares Maß für die Größe der körperlichen Anstrengung bilden würde. Kein Vernünftiger nimmt heute Anstoß an diesen Verhältnissen, da man sich längst daran gewöhnt hat, bei solchen Dingen Sprachgebrauch und wissenschaftliche Terminologie auseinanderzuhalten. Wer von „Arbeit" im Sinne der Mechanik spricht, schaltet schon automatisch die nicht dazu passenden Assoziationen aus, die das Wort sonst mit sich bringt.

Historische Bemerkung.

Daß dem nicht immer so war, sondern sich diese Klärung erst langsam als Frucht einer langen Entwicklung ergeben hat, lehrt uns zum Beispiel der Streit der Cartesianer und Leibnizianer um den Begriff der lebendigen Kraft. Ob die „Wirkung einer Kraft" durch das Produkt Masse mal Geschwindigkeit oder das Produkt Masse mal halbes Geschwindigkeitsquadrat gemessen wird, ist keine entscheidbare Frage, sondern Sache willkürlicher Namengebung. Beide Produkte haben ihre Bedeutung in der Mechanik — wenn man in dem zweiten den Faktor $^1/_2$ fortließe, wäre es auch nicht anders — und welches von ihnen man „lebendige Kraft", welches man „Bewegungsgröße" nennen will, ist gleichgültig. „Es kommt nicht darauf an, was sonst mit dem Wort ‚Kraft' bezeichnet wird," sagt Robert Meyer einmal, „sondern darauf, was wir mit diesem Worte bezeichnen wollen."

Welche Schwierigkeiten schon kleine Abweichungen zwischen dem alltäglichen und dem wissenschaftlichen Sprachgebrauch dem Anfänger bereiten, weiß jeder aus der Schulzeit. Daß auch die langsamste Bewegung eine „Geschwindigkeit" besitzt, daß die verzögerte Bewegung eine „beschleunigte" — mit negativer Beschleunigung — genannt werden muß, daß die Ruhe ein Spezialfall der Bewegung ist: das alles muß man in der Schule erst *lernen*, und in dem Erfassen dieser wissenschaftlichen Sprache liegen die Elemente der Verstandesbildung. Nur von hier aus eröffnet sich ein Zugang zum Verständnis der modernen Naturwissenschaft.

Diese Beispiele zeigen aber auch, daß die Präzisierung des Sprachgebrauchs, wie ihn die Wissenschaft mit sich bringt, sich mit der Zeit mehr und mehr verbreitet. Heute gilt schon als ungebildet, wer die Worte Viereck, Rechteck, Quadrat durcheinanderbringt und nicht die genauen Begriffsabgrenzungen kennt. Bald werden viel höhere Anforderungen in dieser Richtung selbstverständlich erscheinen.

Ziel rationeller Begriffsbildung.

Sicherlich wird man bei der Wahl des Namens für einen wissenschaftlichen Begriff gewisse Rücksichten der sprachlichen Zweckmäßigkeit und des guten Geschmacks walten lassen. Aber wesentlich ist nicht der Name, sondern der Inhalt. Wichtig ist nur, daß mit den Definitionen ein nützliches Ziel erreicht wird. Als ein solches sehen wir an das Ziel der Wissenschaft überhaupt: Ordnung und Übersicht in die Vielfältigkeit der beobachteten Erscheinungen zu bringen, den Ablauf von Erscheinungsreihen vorauszusagen und den Weg zu zeigen, auf dem man bestimmte erwünschte Vorgänge herbeiführen kann. In jeder dieser Richtungen hat sich das Begriffssystem der klassischen Physik und innerhalb dieses Systems u. a. der wissenschaftlich-physikalische Begriff „Arbeit" bewährt. Eine außerordentlich großzügige Übersicht über eine umfassende Gruppe physischer Vorgänge hat er uns in dem Gesetz von der „Erhaltung der Arbeit" vermittelt. Dasselbe Gesetz gewährt die Möglichkeit der Vorausberechnung mannigfacher Naturvorgänge und lehrt den Maschinenbauer und Elektrotechniker die Abmessungen seiner Maschinenanlagen zu bestimmen. Niemand wird heute an dem großen theoretischen und praktischen Erfolg zweifeln, den die auf exakten Begriffsbildungen sich aufbauende Mechanik erzielt hat. Nur ein Einwand, der gelegentlich gegen die Brauchbarkeit und den praktischen Nutzen der „Rationalisierung" der Begriffsbildung erhoben wurde, sei hier noch kurz besprochen.

Man kann sagen: Ja, es ist leicht, aus exakt abgegrenzten, künstlichen Begriffen eine Theorie widerspruchslos aufzubauen; aber wenn man dann die Theorie auf die Wirklichkeit anwenden will, so hat man es doch nur mit unscharfen Vorgängen zu tun, in denen allein die unbestimmteren, „natürlich gewachsenen" Begriffsbildungen Bedeutung haben. Daran ist gewiß etwas Richtiges, ja, es spricht sich hier ein tiefliegender Mangel aller theoretischen Betrachtung der Wirklichkeit aus. Die Vorgänge, die wir beobachten, die wir erleben, sind stets ungleich verwickelter, als selbst die eingehendsten und schwierigsten Theorien sie wiedergeben. In dem ganzen umfassenden Tatsachenbestand, den eine Naturerscheinung aufweist, den Zug, den die Theorie vielleicht als einzigen darstellt, auch nur zu sehen, ist schon eine große Kunst und nur dem wissenschaftlich Geschulten überhaupt möglich. Aber es wäre verkehrt, oder liegt wenigstens ganz außerhalb der Linie der bisherigen vielhundertjährigen Wissenschaftsentwicklung, wollte man, wie es eine moderne Philosophenschule, die von BERGSON, verlangt, um der Komplexität der Wirklichkeit gerecht zu werden, auf scharfe Begriffe verzichten und zu den verschwommenen Vorstellungen zurückgreifen, die man anschauliche nennt, die in Wahrheit aber nur durch einen unklaren und unsicheren Sprachgebrauch bestimmt werden.

Zugegebene Unvollkommenheit jeder Theorie.

Wenn ich zum Beispiel hier mit der Kreide eine „Gerade" an die Tafel zeichne, was ist das für ein außerordentlich zusammengesetztes Ding gegenüber dem, was man in der Geometrie als „Gerade" zu definieren pflegt! Vor allem ist es keine Linie, sondern hat eine endliche Breite, ja eigentlich ist es ein dreidimensionaler Kreidekörper, mehr noch: eine Anhäufung vieler kleiner körperlicher Kreidestückchen usf. Wer nicht von Kindheit an in der Schule den Lehrer zeichnen sah, wird nicht ohne weiteres verstehen, was dieser Kreidekörper mit *der* Geraden zu tun hat, die die kürzeste Verbindung zweier Punkte heißt. Und doch wissen wir, daß die exakten, idealisierten Begriffe der reinen Geometrie unentbehrlich sind, wenn wir uns in unserer Umgebung zurechtfinden wollen. Wir *brauchen* die exakten Begriffsbildungen, weil sie einfacher, d. h. von unseren Verstandeskräften leichter zu bewältigen sind. Man hat schon versucht, eine Geometrie aufzubauen, in der es statt „unendlich dünner" Linien nur Streifen von endlicher Breite gibt, aber man ist damit nicht weit gekommen, weil eine solche Betrachtungsweise doch nur schwieriger ist als die gewöhnliche. Auch ist ein Streifen von gegebener Breite nur wieder eine andere Abstraktion von derselben Art wie die Gerade; nur komplizierter, denn er setzt ja als Begrenzung beiderseits so etwas wie eine gerade Linie voraus.

Wir müssen ohne weiteres zugeben, daß alle theoretischen Konstruktionen, deren man sich in den verschiedenen Teilen der Physik, einschließlich der Geometrie, bedient — und auch die Wahrscheinlichkeitsrechnung zählen wir zu dieser Gruppe von exakten Wissenschaften —, nur unzulängliche Hilfsmittel für die gedankliche Nachbildung der empirisch erkannten Tatsachen sind. Allein, daß es einen anderen Weg des wissenschaftlichen Fortschrittes gibt als den, mit dem einfachsten und daher exakten theoretischen Gerüst zu beginnen und es durch allmähliche Hinzufügung ergänzender Bestandteile auszubauen, glaube ich nicht. Jedenfalls will ich für die Klasse von Erscheinungen, die die Wahrscheinlichkeitstheorie behandelt, für diesen Ausschnitt aus der Gesamtheit der physischen Vorgänge kein anderes Ziel anstreben als das, das in der Geometrie, in der Mechanik und in vielen sonstigen Teilen der Physik bereits erreicht, in anderen ebenfalls angestrebt ist: eine auf möglichst einfachen, exakten Begriffsbildungen aufgebaute „rationelle" Theorie zu liefern, die bei aller Unvollkommenheit gegenüber der Totalität des wirklichen Geschehens doch einzelne seiner wesentlichen Züge befriedigend wiedergibt.

Beschränkung des Stoffes.

Nach all diesen Vorbereitungen wollen wir uns jetzt der Aufgabe zuwenden, den Wahrscheinlichkeitsbegriff selbst näher zu umschreiben. Aus dem bisher Gesagten geht schon hervor, daß es vor allem nötig sein wird, innerhalb des Gebietes, das durch die vulgäre Wortbedeutung „wahrscheinlich" gedeckt wird, alles auszuscheiden, was nicht durch die hier zu entwickelnde Theorie berührt werden soll. Ich gebe also zunächst eine vorläufige Abgrenzung für die Geltung des Wahrscheinlichkeitsbegriffes, die durch die späteren Ausführungen präzisiert werden wird.

Mit der Frage, ob und wie wahrscheinlich es ist, daß Deutschland noch einmal Krieg mit der Republik Liberia führen wird, hat unsere Wahrscheinlichkeitstheorie nicht das mindeste zu tun. Auch wenn man von der Wahrscheinlichkeit spricht, die für eine bestimmte Lesart einer Stelle in Tacitus' Annalen besteht, befindet man sich ganz außerhalb des Bereiches unserer Untersuchung. Nicht ausdrücklich erwähnen muß ich wohl, daß uns die sogenannte „innere Wahrscheinlichkeit" eines Kunstwerkes hier nichts angeht und der schöne Goehtesche Dialog „Über Wahrheit und Wahrscheinlichkeit der Kunstwerke" nur durch einen sehr nebensächlichen sprachlichen Gleichklang mit unserem Thema verknüpft ist. Die Fragen nach der objektiven Richtigkeit der biblischen Erzählungen, für die ein hervorragender russischer Mathematiker, A. Markoff, ganz erfüllt von den Gedankengängen der liberalen Aufklärungszeit, ausdrücklich die Autorität der Wahrscheinlichkeitsrechnung in Anspruch nimmt, scheiden wir ebenfalls aus und mit ihnen fast alle Fragen der „sciences morales", über die der große Laplace in seinem „Essai philosophique" so anziehend zu sprechen weiß. Es ist im übrigen für den übertriebenen Rationalismus des 18. Jahrhunderts kennzeichnend, daß er den Geltungsbereich der exakten Wissenschaften ins Ungemessene auszudehnen suchte; in diesen Fehler wollen wir nicht verfallen.

Ungefähr an der Grenze dessen, was wir noch in unsere Betrachtungen einbeziehen, liegen die früher sehr viel behandelten Fragen über die Wahrscheinlichkeit von Zeugenaussagen und gerichtlichen Urteilen, die Poisson zum Titelproblem seines berühmten Werkes gemacht hat. Mitten in den Kern trifft man aber mit der Frage nach der Wahrscheinlichkeit eines Gewinnes innerhalb eines bestimmten, wohldefinierten, reinen Glückspiels. Ob es lohnend ist, darauf zu wetten, daß beim Spiel mit zwei Würfeln in 24 Wiederholungen mindestens einmal Doppelsechs erscheint, ob ein solches Ereignis mit „Wahrscheinlichkeit" zu erwarten ist, ja sogar mit „wie großer" Wahrscheinlichkeit, das zu beantworten, fühlen wir uns berufen. Daneben gibt es Aufgaben von größerer sozialer Bedeutung, die wir in gleicher Weise zu den unseren rechnen, so alle Fragestellungen, die mit dem Versicherungswesen zusammenhängen, die Wahrscheinlichkeit von Todes- oder Krankheitsfällen unter ganz genauen, wohlumgrenzten Voraussetzungen, die Wahrscheinlichkeit dafür, daß ein Versicherungsunternehmen mit einer Prämie von bestimmter Höhe auskommt usf. Ein drittes Gebiet, neben dem der Glückspiele und der sozialen Massenvorgänge, in dem wir mit unserem Wahrscheinlichkeitsbegriff erfolgreich arbeiten können, bilden gewisse mechanische und physikalische Erscheinungen, als deren Typus ich etwa das Herumschwirren der Moleküle in einem geschlossenen Gasbehälter oder die unter dem Mikroskope beobachtbaren Bewegungen der Teilchen eines kolloidalen Körpers nenne. Man bezeichnet bekanntlich als Kolloide jene Ansammlungen fein verteilter, in einem Medium suspendierter Teilchen, die sich dem freien Auge oder dem flüchtigen Beobachter als gleichförmige, deformable Masse darstellen.

Unbegrenzte Wiederholbarkeit.

Was ist das Gemeinsame an den zuletzt aufgeführten Fällen, was kennzeichnet sie im Gegensatz zu den anderen Anwendungen des Wortes „Wahrscheinlichkeit", die wir aus unserer Betrachtung ausgeschieden wissen wollen? Man kann über ein gemeinsames Merkmal, dem wir entscheidende Bedeutung beilegen, nicht im Zweifel sein: Es handelt sich beim Glückspiel, bei den Versicherungsproblemen, bei den Gasmolekeln um Vorgänge, die sich in großer Zahl wiederholen, um Massenerscheinungen oder Wiederholungsvorgänge. Der Wurf mit einem Paar guter Würfel ist praktisch fast unbegrenzt wiederholbar; man denkt kaum daran, daß die Würfel oder der Becher sich allmählich abnutzen und schließlich zugrunde gehen könnten. Haben wir eine Versicherungsfrage zu untersuchen, so steht uns vor Augen das große Heer der Versicherungsnehmer, die häufige Wiederholung des einzelnen Falles, der von der versichernden Gesellschaft registriert wird. Auch beim dritten Beispiel liegt das Massenhafte, die unermeßlich große Zahl der Moleküle oder der kolloidalen Teilchen schon unmittelbar in der Vorstellung.

Anderseits erkennt man ebenso deutlich, daß das Merkmal der weitgehenden Wiederholbarkeit, der Massenhaftigkeit in all den Fällen fehlt, die ich vorher ausgeschlossen hatte: Die Lage, in einen Krieg mit Liberia zu geraten, wiederholt sich nicht unbegrenzt oft, Unsicherheiten in der Lesart eines Schriftstellers sind von Fall zu Fall zu verschieden, als daß man hier von Massenerscheinungen sprechen könnte, und die Frage über die Zuverlässigkeit der biblischen Berichte ist eine geradezu einzigartige, die man unmöglich als ein Glied in einer Kette sich wiederholender gleicher Fragestellungen auffassen kann. Im Falle der Zeugenaussagen muß man mindestens im Zweifel sein, ob genügende Einheitlichkeit und Gleichförmigkeit vorhanden ist, damit an einen Massenvorgang gedacht werden kann.

So stellen wir also einmal fest: Mit dem rationellen Wahrscheinlichkeitsbegriff, der die ausschließliche Grundlage der Wahrscheinlichkeitsrechnung bildet, wollen wir nur solche Fälle erfassen, in denen es sich um einen vielfältig wiederholbaren Vorgang, um eine in großen Mengen auftretende Erscheinung, physikalisch gesprochen um eine praktisch unbegrenzte Folge von gleichartigen Beobachtungen handelt.

Das Kollektiv.

Ein schönes Beispiel für einen Massenvorgang, auf den die Wahrscheinlichkeitslehre Anwendung findet, bildet die Vererbung bestimmter Eigenschaften, etwa der Blütenfarbe, bei der Aufzucht einer großen Menge gleichartiger Pflanzen aus einem gegebenen Samenbestand. Wir können hier deutlich sehen, worauf es ankommt, wenn man von einem Wiederholungsvorgang spricht. Man hat zunächst einen Einzelvorgang vor sich, d. i. hier die Aufzucht eines Pflanzenindividuums und die Beobachtung seiner Blütenfarbe, sodann die Zusammenfassung sehr vieler solcher Vorgänge zu einer Gesamtheit. Die Einzelvorgänge unterscheiden sich untereinander durch das „Merkmal", das den eigentlichen Gegenstand der Beobachtung bildet, in unserem Beispiel: die Farbe der Blüte. Dadurch, daß wir die Gesamtheit aller Pflanzen auf dieses eine Merkmal hin untersuchen, entsteht die Zusammenfassung zu einem Sammelbegriff. Beim Würfelspiel bildet den Einzelvorgang das einmalige Ausspielen der Würfel aus dem Becher und die darauf folgende Beobachtung der geworfenen Augenzahlen. Spielt man mit einer Münze „Kopf oder Adler", so ist jeder Wurf ein Einzelvorgang und das Merkmal ist eben das auf der Oberseite erscheinende Münzenbild, der Kopf oder der Adler. Im Problem der Ablebensversicherung ist als Einzelvorgang der Lebensablauf des Versicherten anzusehen, als das zu beobachtende Merkmal etwa das Alter, in dem er aus dem Leben scheidet, allgemeiner der Zeitpunkt, in dem der Versicherungsfall eintritt. Wenn wir von einer Sterbenswahrscheinlichkeit sprechen, so kann sich das — nach der hier zu begründenden exakten Auffassung des Wahrscheinlichkeitsbegriffes — immer nur auf eine bestimmte Gesamtheit von Personen beziehen, z. B. auf die Gesamtheit der „41 jährigen versicherten männlichen Person in Deutschland, die in nicht gefährlichen Berufen stehen". Von der Sterbenswahrscheinlichkeit eines bestimmten Individuums, und mag von ihm noch so viel bekannt sein, können wir nicht nur nichts aussagen, sondern dieser Ausdruck hat für uns überhaupt keinerlei Sinn. Darin liegt eine der tiefgreifendsten Folgerungen aus unserer Auffassung des Wahrscheinlichkeitsbegriffes, auf die ich noch ausführlicher zurückkommen werde.

Es ist jetzt an der Zeit, einen Ausdruck einzuführen, der uns für die weiteren Auseinandersetzungen sehr nützlich sein wird. Wir wollen eine Gesamtheit von Vorgängen oder Erscheinungen, die sich im einzelnen durch irgendein beobachtbares Merkmal, eine Zahl, eine Farbe od. dgl. unterscheiden, kurz ein Kollektiv nennen, wobei ich mir noch vorbehalte, gewisse nähere Präzisierungen zur Definition nachzutragen. Ein Kollektiv bilden also die von einem Vererbungsforscher gezüchteten Erbsenpflanzen mit ihrer Unterscheidung in rot- und weißblühende. Ein Kollektiv ist die Folge von Würfen aus einem Becher mit der jedesmal sichtbar werdenden Augenzahl, die Gesamtheit der Gasmolekel, an deren jedem eine bestimmte Geschwindigkeit beobachtet wird, endlich die Masse der Versicherten, deren Sterbealter der Registrierung unterworfen wurde. Grundlegend für unsere Auffassung des Wahrscheinlichkeitsbegriffes ist der Satz: Zuerst muß ein Kollektiv da sein, dann erst kann von Wahrscheinlichkeit gesprochen werden. Die Definition, die wir geben werden, kennt überhaupt nur die „Wahrscheinlichkeit eines Merkmals innerhalb eines gegebenen Kollektivs".

Erster Schritt zur Definition.

Zu dieser Definition in ihrer rohesten Form zu gelangen, ist nach dem bisher Gesagten nicht mehr schwer. Gehen wir etwa von dem Fall des Spieles mit zwei Würfeln aus. Als Merkmal eines Wurfes betrachten wir die Augenzahl, die auf den beiden Würfeloberseiten sichtbar wird. Was werden wir die Wahrscheinlichkeit der Augenzahl 12, also der Doppelsechs, nennen? Wenn eine große Zahl von Spielen ausgeführt ist — sagen wir etwa, es seien zweihundert Würfe geschehen —, wird eine verhältnismäßig kleine Zahl, vielleicht fünf, die Zahl der erschienenen Doppelsechser sein. Den Quotienten $5:200 = {}^1/_{40}$ bezeichnet man dann üblicherweise als die Häufigkeit, deutlicher als die „relative Häufigkeit" des Erscheinens der Doppelsechs innerhalb der ersten 200 Spiele. Setzt man das Würfeln fort, etwa bis zu 400, dann 600 Spielen und stellt nach jedem dieser Abschnitte die relative Häufigkeit der Doppelsechs fest, indem man die Anzahl der im ganzen Spielverlauf erschienenen Doppelsechs durch 400 bzw. 600 dividiert, so wird man Quotienten erhalten, die von dem zuerst gefundenen Wert ${}^1/_{40}$ mehr oder weniger abweichen. Wenn bei beliebig langer Fortsetzung des Spielens, bei 2000, 4000 und mehr Würfen immer noch erhebliche Abweichungen aufträten, kämen wir gar nicht dazu, von einer bestimmten Wahrscheinlichkeit der Doppelsechs zu sprechen. Es ist eine für die Wahrscheinlichkeitstheorie grundlegende Erfahrungstatsache, die beim Würfelspiel sich ebenso bestätigt, wie bei den anderen früher angeführten Beispielen von Massenerscheinungen, daß die relativen Häufigkeiten der verschiedenen Merkmale sich weniger und weniger ändern, wenn man die Zahl der Beobachtungen mehr und mehr vergrößert. Setzen wir fest, daß wir die Quotienten von vornherein nur mit einer begrenzten Genauigkeit, z. B. auf drei Dezimalstellen, berechnen, dann wird je nach den Verhältnissen, von irgendeiner Zahl von Spielen an, die relative Häufigkeit der Doppelsechs überhaupt dieselbe bleiben. Diesen „Grenzwert" (was dies genauer heißt, werden wir noch besprechen) der relativen Häufigkeit, der notwendig ein echter Bruch, d. h. eine Zahl kleiner als 1, sein muß, nennen wir die Wahrscheinlichkeit für das Auftreten der Doppelsechs innerhalb des Gesamtspiels mit den beiden Würfeln. Wie groß dieser Bruch ist, ist dabei ganz gleichgültig.

Zwei verschiedene Würfelpaare.

Ich habe hier zwei Paare von Würfeln, die äußerlich ganz ähnlich, für Vorführungszwecke besonders hergerichtet sind. Bei dem einen Paar findet man durch fortgesetztes Würfeln, daß die relative Häufigkeit der Doppelsechs sich bei großer Spielzahl ungefähr dem Wert 0,028 oder $^1/_{36}$ nähert, bei dem anderen Paar ist sie nahezu viermal so groß. Daß man das eine Würfelpaar ein „richtiges", das andere „falsch" nennt, kommt jetzt nicht in Frage, unsere Wahrscheinlichkeitsdefinition ist in beiden Fällen anwendbar. So kümmert sich auch der Arzt, der eine Krankheit diagnostiziert, nicht darum, ob der Kranke ein Ehrenmann ist oder nicht. Mit jedem der beiden Paare wurden 1800 Würfe ausgeführt. Beim ersten ergab sich 48mal, beim zweiten 178mal das Resultat Doppelsechs. Die relativen Häufigkeiten waren also

$$\frac{48}{1800} = \frac{1}{37{,}5} = 0{,}027 \quad \text{und} \quad \frac{178}{1800} = \frac{1}{10{,}1} = 0{,}099.$$

Dabei haben sich diese Quotienten gegen Schluß der Versuchsreihe nur mehr wenig verändert. Nach dem 1500. Versuch betrugen sie

$$0{,}023 \quad \text{und} \quad 0{,}094$$

und zwischen dem 1500. und dem letzten Wurf wichen sie untereinander höchstens um 10 bis 15% ab.

Ich kann den Versuch des fortgesetzten Würfelns hier nicht ganz durchführen, da er geraume Zeit in Anspruch nimmt. Aber wenn ich nur einigemal mit diesem Würfelpaar werfe, so sehen Sie schon, daß fast jedesmal wenigstens *eine* Sechs fällt. Bei dem anderen Paar ist das nicht der Fall. In der Tat läßt sich feststellen, daß die Häufigkeit Einer Sechs bei diesen Würfeln, wenn man mit nur einem arbeitet, fast genau $^1/_6$, bei jenen nahezu $^1/_3$ ist. Wenn Sie den Wahrscheinlichkeitsbegriff, wie wir ihn brauchen, klar erfassen wollen, müssen Sie sich nur immer wieder dieser beiden Würfelpaare erinnern. Für jedes von ihnen gibt es eine bestimmte Wahrscheinlichkeit der Doppelsechs, aber die beiden Wahrscheinlichkeiten sind weit voneinander verschieden.

Hier haben Sie das „*Urphänomen*" der Wahrscheinlichkeitslehre in seiner einfachsten Gestalt vor sich. Wir können sagen: Die Wahrscheinlichkeit, Sechs zu zeigen, ist *eine physikalische Eigenschaft* eines Würfels, von derselben Art wie sein Gewicht, seine Wärmedurchlässigkeit, seine elektrische Leitfähigkeit usf. Ebenso entspricht dem Würfel*paar* (natürlich mit Einschluß der ganzen Anordnung, der es unterworfen wird) neben anderen eine bestimmte physikalische Größe: die Wahrscheinlichkeit der Doppelsechs. Mit den mannigfachen Beziehungen derartiger Größen untereinander beschäftigt sich die Wahrscheinlichkeitsrechnung.

Der Grenzwert der relativen Häufigkeit.

Das Wort „Grenzwert", das ich vorhin ohne Erklärung verwendet habe, ist ein Terminus der höheren Mathematik. Wir brauchen jedoch nicht viel davon zu wissen, wie er von den Mathematikern definiert wird, da wir ihn nur in einer Weise verwenden werden, die jeder Laie verstehen kann. Wir wollen die relative Häufigkeit eines Merkmales, also den Quotienten aus der Anzahl seines Auftretens durch die Gesamtzahl aller Beobachtungen, wie ich schon sagte, mit einer bestimmten Zahl von Dezimalen rechnen, ohne uns darum zu kümmern, wie die späteren Stellen ausfallen, ob sie Nullen sind oder andere Ziffern enthalten. Verfolgen wir in dieser Weise einen Wiederholungsvorgang, z. B. das Spiel „Kopf oder Adler" mit einer Münze, genügend lange und rechnen wir von Zeit zu Zeit die relative Häufigkeit der Kopfergebnisse mit der festgesetzten Genauigkeit, so kann es eintreten, daß von irgendeinem Zeitpunkt ab die Häufigkeitszahl sich nicht mehr ändert. Nimmt man nur eine Dezimale, so wird man in der Regel verhältnismäßig bald, vielleicht schon nach 500 Würfen, finden, daß keine Veränderung eintritt, daß die relative Häufigkeit, auf eine Stelle genau, etwa 0,5 beträgt. Rechnet man zwei Stellen, so muß man jedenfalls viel weiter gehen. Man wird vielleicht nach je 500 Würfen wieder den Quotienten bilden, somit die relative Häufigkeit des „Kopfes“ für die ersten 1000, 1500, 2000.... Würfe ermitteln und möglicherweise finden, daß dann von 10000 ab an der zweiten Dezimalstelle immer „0“ steht, also die relative Häufigkeit 0,50 beträgt. Natürlich kann man so einen Versuch nicht ins Endlose ausdehnen. Wenn zwei Personen geschickt zusammenarbeiten, können sie es auf etwa 1000 Beobachtungen in der Stunde bringen. Nehmen wir nun an, man habe zehn Stunden auf die Sache gewandt und bemerkt, daß in den beiden letzten Stunden die relative Häufigkeit bei 0,50 — ungeachtet der weiteren Stellen — stehengeblieben ist. Ein feinerer Beobachter wird nebenbei die dritte Stelle des Quotienten ausrechnen und vielleicht finden, daß diese in der allerletzten Stunde zwar noch Änderungen gezeigt hat, aber keine großen. Ein solches Versuchsergebnis führt den naturwissenschaftlich Denkenden auf die *Vermutung*, daß, wenn man das Spiel unter gleich bleibenden Umständen weiter und weiter fortsetzt, fortsetzen *könnte*, auch die dritte, dann die vierte und schließlich jede folgende Dezimale der relativen Häufigkeit einen festen Wert annimmt. Den Tatbestand, der dieser Vermutung entspricht, bezeichnen wir kurz mit den Worten: Die relative Häufigkeit des Merkmales „Kopf“ *strebt einem Grenzwert zu*.

Wenn wir auf einem Blatt Millimeterpapier eine Kurve aufzeichnen, die zu jeder Versuchsnummer als Abszisse die zugehörige relative Häufigkeit des „Kopf“-Resultates als Ordinate enthält, so zeigt die Kurve anfänglich starke Schwankungen nach oben und unten, nähert sich aber allmählich mehr und mehr einer horizontalen Geraden. Schließlich werden die Abweichungen von der Geraden so gering, daß sie, wie groß wir auch den Maßstab wählen, nicht mehr deutlich gemacht werden können. Welche Höhenlage die Gerade hat, ob 0,5 oder 0,6 ihre Ordinate ist, ist gleichgültig;

nur darauf kommt es an, daß es eine solche horizontale Linie gibt, in die unsere Kurve übergeht. Die Ordinate dieser Horizontalen heißt dann der Grenzwert des aufgetragenen Quotienten, also der relativen Häufigkeit des „Kopf“-Ergebnisses.

Die früher gegebene Erklärung des Kollektivs ergänzen wir jetzt durch folgende Fassung: Ein Kollektiv ist eine Massenerscheinung oder ein Wiederholungsvorgang, kurz eine lange Folge von Einzelbeobachtungen, bei der die Vermutung berechtigt erscheint, daß die relative Häufigkeit des Auftretens jedes einzelnen Beobachtungsmerkmales einem bestimmten Grenzwert zustrebt. Diesen Grenzwert nennen wir die „Wahrscheinlichkeit für das Auftreten des Merkmales innerhalb des Kollektivs“. Dabei ist es natürlich nicht notwendig, immer den ganzen schwerfälligen Ausdruck zu wiederholen; gelegentlich erlauben wir uns schon, einfach „Wahrscheinlichkeit eines Kopfwurfes“ zu sagen. Das Wesentliche ist, daß dies wirklich nur eine Abkürzung bedeutet und daß das Kollektiv, auf das sich die Wahrscheinlichkeitsaussage bezieht, genau definiert sein muß. Man sieht jetzt wohl ein, daß von einer Wahrscheinlichkeit des Ausganges einer Schlacht in unserem Sinne nicht die Rede sein kann. Hier liegt kein Kollektiv vor und unsere Art zu definieren versagt genau so, wie wenn man dem Physiker die Aufgabe stellen wollte, die „Arbeit“, die ein Schauspieler beim Vortrag seiner Rolle leistet, nach mechanischem Maß zu messen.

Die Erfahrungsgrundlage bei Glücksspielen.

Wir müssen dann noch überlegen, wie der entscheidende Versuch der Wahrscheinlichkeitsermittlung sich in den anderen vorhin angeführten Fällen, bei der Lebensversicherung, bei den Gasmolekeln gestaltet. Vorerst aber noch eine kurze Bemerkung zu den Glücksspielen. Man könnte fragen: Woher nehmen wir eigentlich die Sicherheit dafür, daß ein in der Praxis ausgeführtes Glücksspiel sich hinsichtlich der relativen Häufigkeiten der verschiedenen Ausgangsmöglichkeiten so verhält, wie eben angegeben wurde? Besitzen wir wirklich eine ausreichende Stütze für eine so weitgehende Vermutung über den tatsächlichen Ablauf einer Versuchsreihe, von der wir doch nur einen bescheidenen Anfang in jedem Fall erproben können? Nun, so gering ist das Material nicht, das unsere Unterlage bildet. Da sind vor allem die großen Spielbanken in Monte Carlo und anderwärts, die in millionenfacher Wiederholung dasselbe Spiel immer wieder ausführen. Die Banken befinden sich sehr wohl dabei, nachdem sie ihre Gewinstaussichten auf Grund der Annahme fester Grenzwerte der relativen Häufigkeiten vorausberechnet haben. Daß gelegentlich die Bank „gesprengt" wird, spricht durchaus nicht gegen die Voraussetzung. Einen Gegenbeweis gäbe es erst, wenn der seit Gründung der Bank erzielte Gesamtgewinn erheblich sinken oder gar in Verlust umschlagen würde, wenn also die Bank in Zukunft annähernd so viel oder noch mehr verlöre, als sie seit Bestand gewonnen hat. Daß dies einmal eintreten könnte, wird kein Einsichtiger vermuten.

Ganz ähnlich wie bei den Spielbanken steht es bei den Lotterieunternehmungen, die lange Jahre hindurch von einzelnen Staaten betrieben wurden und stets die beste Übereinstimmung mit der Annahme stabiler Werte der relativen Häufigkeiten ergeben haben. Dies alles berechtigt uns zu der Auffassung, wonach es wirklich realisierbare Spielvorgänge gibt, für die unsere Vermutung zutrifft, daß die relativen Häufigkeiten der verschiedenen Merkmale Grenzwerten zustreben. Und nur Vorgänge dieser Art wollen wir in unseren weiteren Ausführungen in Betracht ziehen.

Lebens- und Sterbenswahrscheinlichkeit.

Die „Sterbenswahrscheinlichkeit", mit der die Versicherungsgesellschaften rechnen, wird nach einem Verfahren ermittelt, das sich in keinem wesentlichen Punkte von demjenigen unterscheidet, das wir zur Definition der Wahrscheinlichkeit im Falle des Würfelspieles benutzt haben. Vor allem gilt es, hier für jeden Anwendungsfall ein bestimmtes Kollektiv genau abzugrenzen. Ein Beispiel liefert uns die Herstellung der „deutschen Sterblichkeitstafeln aus den Erfahrungen von 23 Versicherungsgesellschaften", die lange Zeit hindurch in allgemeiner Verwendung standen. Ungefähr 900000 Einzelbeobachtungen an Personen, die bei einer der 23 Gesellschaften auf Todesfall versichert waren, wurden durchgeführt, ehe die Tafeln mit der Angabe der Sterblichkeit aufgestellt werden konnten. Dabei umfaßte eine Beobachtung jedesmal den ganzen Zeitraum vom Eintritt einer Person in die Versicherung bis zu ihrem Austritt. Eines der betrachteten Kollektivs war folgendermaßen festgelegt: Gesamtheit der Männer, die auf Grund vollständiger ärztlicher Untersuchung mit normaler Prämie vor Erreichung des Alters von 40 Jahren versichert wurden. Das dabei zu beobachtende Merkmal war: Eintritt oder Nichteintritt des Todes innerhalb des 41. Lebensjahres. Fälle, in denen aus irgendeinem Grunde, z. B. wegen Aufgebens der Versicherung, das Merkmal nicht eindeutig festgestellt werden konnte, wurden ausgeschieden. Die Zahl der Fälle in dieser Kategorie, die bis zu Ende beobachtet werden konnten, betrug 85020; in 940 Fällen wurde der Todeseintritt festgestellt. Die relative Häufigkeit betrug also $940 : 85020 = 0{,}01106$. Diese Zahl gilt — vorbehaltlich gewisser Berichtigungen, die hier beiseitegelassen werden dürfen — als die Sterbenswahrscheinlichkeit im 41. Lebensjahre für die betrachtete Kategorie von Personen, d. h. in unserer Ausdrucksweise als Sterbenswahrscheinlichkeit innerhalb des vorher genau abgegrenzten Kollektivs.

Daß man die Zahl von rund 85000 Beobachtungen für ausreichend hielt, um die berechnete relative Häufigkeit schon als Wahrscheinlichkeit, d. h. als den bleibenden, für einen beliebig großen Kreis von Fällen sich nicht mehr ändernden Grenzwert gelten zu lassen, ist bis zu einem gewissen Grade Willkür. Niemand wird meinen, daß mehr als die ersten drei Dezimalstellen, also 0,011, stabil bleiben werden. Jedermann würde auch sehr zufrieden sein, wenn es möglich wäre, die Berechnung auf eine noch umfassendere Grundlage zu stellen; naheliegende praktische Bedenken stehen dem entgegen. Auch daß die einmal festgestellte Zahl nicht für ewige Zeiten gilt, ist selbstverständlich. So geht es auch mit anderen physikalischen Maßzahlen. Man bestimmt die Größe der Erdschwere an irgendeinem Orte und rechnet mit dem gemessenen Wert so lange, bis man durch neue Messung eine etwaige Änderung festgestellt hat; in gleicher Weise verhält man sich gegenüber örtlichen Veränderungen. In diesem Sinne begnügt man sich auch im Versicherungswesen mit dem, was eben erreichbar ist, und benutzt, solange man keinen besser begründeten Wert besitzt, die Zahl 0,011 so, als ob sie der gesuchte Grenzwert wäre. Das heißt, man rechnet bis auf weiteres damit, daß innerhalb der Gesamtheit aller künftigen Versicherungsnehmer der betrachteten Kategorie gerade 11 von Tausend im 41. Lebensjahre sterben.

Über den angegebenen Rahmen hinaus hat aber die Zahl 0,011 keinerlei Bedeutung. Völlig sinnlos wäre es, zu sagen, der jetzt 40jährige Herr N. N. habe $11^0/_{00}$ Wahrscheinlichkeit, innerhalb eines Jahres zu sterben. Denn bildet man beispielsweise ein Kollektiv aus Frauen und Männern gemeinsam, so erhält man nach Ausweis der Sterblichkeitstafeln einen etwas kleineren Quotienten für das Alter von 40 Jahren. Herr N. N. gehört aber mit demselben Recht zu der Gesamtheit der Frauen und Männer wie zu der der Männer allein und zu vielen anderen, leicht angebbaren Gesamtheiten. Meint man vielleicht, einen um so „richtigeren" Wert zu bekommen, je mehr man von den Eigenschaften des Individuums berücksichtigt, d. h. je enger man das Kollektiv abgrenzt, als dessen Glied man den Einzelfall betrachtet, so gelangt man zu keinem Ende; schließlich bleibt, durch alle seine Eigenschaften eindeutig bestimmt, das eine Individuum als einziges Glied seiner Klasse übrig. Heute pflegen z. B. die Versicherungsgesellschaften die sogenannte Selektion durch den Versicherungsabschluß zu berücksichtigen, nämlich den Umstand, daß Personen, die sich früh versichern, einen anderen Ablauf ihrer Lebensdauer aufweisen als in späterem Alter Versicherte. Die Selektions-Sterbetafeln beruhen auf einer solchen Abgrenzung des Kollektivs, bei der jedesmal eine bestimmte Dauer des Versicherungsvertrages im Zeitpunkte des Todeseintrittes vorausgesetzt wird. Es ist klar, daß man in dieser oder ähnlicher Richtung auch noch weitergehen kann, aber wenn man alle Eigenschaften des Individuums in die Definition des Kollektivs aufnehmen wollte, so bestünde es eben aus dem einen Element allein, d. h. ein Kollektiv wäre dann überhaupt nicht mehr vorhanden.

Erst das Kollektiv — dann die Wahrscheinlichkeit.

Ich muß noch einen Augenblick bei diesem Punkte verweilen, der einen charakteristischen Unterschied zwischen meiner Auffassung des Wahrscheinlichkeitsbegriffes und der bisher üblichen bildet. Vorhin formulierte ich schon kurz: Zuerst muß ein Kollektiv da sein, dann erst kann von Wahrscheinlichkeit gesprochen werden. Im strengsten Gegensatz dazu erklärt aber Johannes von Kries in seinem bekannten Buch über die Prinzipien der Wahrscheinlichkeitsrechnung (auf das ich noch zurückkommen werde) folgendes: „... So werde ich mit einer gewissen Wahrscheinlichkeit annehmen, daß Cajus, daß Sempronius oder daß Titus im Laufe eines Jahres sterben werde. Wenn dagegen von der Wahrscheinlichkeit eines allgemeinen Falles gesprochen wird, welcher eine unbestimmte Zahl einzelner Verhaltungsweisen in sich begreift, etwa der Wahrscheinlichkeit, ‚daß ein jetzt 40jähriger Mann noch 20 Jahre leben werde‘: so ist klar, daß hier das Wort in einem uneigentlichen Sinne genommen sein muß, daß wir es mit einem abgekürzten Ausdruck zu tun haben. Ein derartiger Satz hat in der Tat, wenn er mit Wahrscheinlichkeit überhaupt etwas zu tun haben soll, lediglich den Sinn, daß er eine allgemeine Bestimmung über eine beliebige Anzahl einzelner Wahrscheinlichkeiten trifft.“ Demgegenüber bin ich der Meinung, daß der „uneigentliche Sinn“, von dem hier die Rede ist, der einzig zulässige ist, der einzige, in dem von Wahrscheinlichkeit innerhalb der Wahrscheinlichkeitsrechnung gesprochen werden kann. Gerade an dem Beispiel der Sterbenswahrscheinlichkeit habe ich es eben erklärt, daß eine andere Auffassung gar nicht möglich ist, und ganz allgemein betrachte ich die Einführung des Terminus „Wahrscheinlichkeit innerhalb eines Kollektivs“ als eine wesentliche „Verbesserung des Sprachgebrauchs“. Zwei Beispiele aus dem Alltag mögen dies noch erläutern.

Wenn in einer Lotterie, die eine Million Lose umfaßt, einmal der Haupttreffer auf das Los mit der Nummer 400000 gezogen werden sollte, so würde man dies äußerst auffallend, ungewöhnlich finden, die Zeitungen würden davon als einer Merkwürdigkeit berichten und jeder wäre überzeugt, hier sei ein sehr unwahrscheinliches Ereignis eingetreten. Auf der anderen Seite gehört es zum Wesen einer Lotterie, daß stets solche Einrichtungen getroffen sind, durch die für jede überhaupt mögliche Nummer die gleiche Wahrscheinlichkeit sichergestellt wird. Es hat also das Erscheinen der Zahl 400000 genau die gleiche Wahrscheinlichkeit von ein Millionstel wie das Erscheinen etwa von 786331. Was liegt da vor? Einen anderen ähnlichen Fall erwähnt Laplace in seinem Essai philosophique: Wenn man mit Buchstabentäfelchen spielt, indem man aufs Geratewohl Buchstaben aneinanderreiht, so würde man es als ein äußerst unwahrscheinliches Ergebnis ansehen, wenn einmal beim willkürlichen Nebeneinanderlegen von 14 Täfelchen plötzlich das Wort CONSTANTINOPLE dastünde. Aber auch hier wird man annehmen dürfen, daß der Mechanismus des Spieles jeder der 26^{14} Kombinationen aus 14 Buchstaben des Alphabets die gleiche Wahrscheinlichkeit sichert. Woher kommen wir also zu der Annahme, daß dem Erscheinen des genannten Stadtnamens eine besonders geringe Wahrscheinlichkeit zuzuschreiben ist?

Die Aufklärung liegt darin, daß dem isolierten Einzelereignis: der Haupttreffer fällt auf Losnummer 400000, überhaupt keine Wahrscheinlichkeit zukommt, sondern daß erst ein Kollektiv definiert sein muß, ehe das Wort Wahrscheinlichkeit einen Sinn bekommt. Nun gibt es Ein Kollektiv, dessen Elemente die Folge der Einzelziehungen, dessen Merkmale die gezogenen Losnummern bilden. In diesem Kollektiv hat — unter den erwähnten Voraussetzungen — jede Nummer die gleiche Wahrscheinlichkeit ein Millionstel. Aber wenn wir von der „Unwahrscheinlichkeit" der Zahl 400000 sprechen, so meinen wir ein ganz anderes Kollektiv. Wir würden nämlich denselben Eindruck von Unwahrscheinlichkeit haben, wenn statt 400000 etwa 700000 oder 200000 erschienen wäre. Also, das Kollektiv, an das wir jetzt denken, weist als Merkmale des Einzelzuges nur die beiden folgenden auf: entweder hat die gezogene Zahl fünf Endnullen oder sie hat sie nicht (Alternative). In diesem Kollektiv besitzt die erstgenannte Möglichkeit die Wahrscheinlichkeit 0,00001, die zweite die Wahrscheinlichkeit 0,99999, d. i. eine fast hunderttausendmal so große. Daher ist man berechtigt zu sagen: Innerhalb der Alternative zwischen einer Zahl mit fünf Endnullen und einer Zahl anderer Art besitzt das erste Ereignis eine äußerst geringe, das zweite eine sehr große Wahrscheinlichkeit.

Genau so liegt es in dem zweiten Beispiel. Was uns an dem Erscheinen der Buchstabenfolge CONSTANTINOPLE auffällt, ist, daß die 14 Buchstaben einen sinnvollen, lesbaren Ausdruck bilden. Es gibt aber unter der ungeheuren Zahl von Kombinationen ($26^{14} \sim 10^{20}$) aus 14 Buchstaben höchstens ein paar Tausend sinnvolle. Wir denken demnach hier an ein Kollektiv, dessen Elemente die willkürlichen Zusammenstellungen von je 14 Buchstaben und dessen Merkmale „sinnvolle oder nichtsinnvolle Verbindung" sind. In diesem Kollektiv hat das zweite Merkmal („nichtsinnvoll") eine erdrückend große Wahrscheinlichkeit gegenüber dem ersten, und dies bringen wir zum Ausdruck, wenn wir sagen, daß das vorerwähnte Ereignis ein in so hohem Maße unwahrscheinliches sei.

Bei vielen Anwendungen des Wahrscheinlichkeitsbegriffes im gewöhnlichen Leben, wenn sie berechtigt sind, kann man das zugehörige Kollektiv konstruieren. Wo dies nicht möglich ist, hat der Gebrauch des Wortes mit der Wahrscheinlichkeitsrechnung nichts zu tun und eine zahlenmäßige Bestimmung der Wahrscheinlichkeit ist nicht möglich. Anderseits gibt es sicher auch Fälle, in denen die Abgrenzung des Kollektivs in verschiedener Weise vorgenommen werden kann, und das sind diejenigen, in denen sich über die Größe der betreffenden Wahrscheinlichkeit streiten läßt. Eindeutig ist nur der Begriff: Wahrscheinlichkeit innerhalb eines bestimmten Kollektivs.

Wahrscheinlichkeit in der Gastheorie.

Wir kehren zurück zu der kurzen, vorläufigen Aufzählung der Anwendungsgebiete der Wahrscheinlichkeitsrechnung, indem wir jetzt Fragen der physikalischen Statistik etwas näher betrachten. Bei der Beschäftigung mit Gasmolekeln oder einem ähnlichen Problem liegen die Verhältnisse nicht wesentlich anders als in den bisher betrachteten Fällen. Das Kollektiv besteht hier etwa aus der Gesamtheit der zwischen Zylinderwand und Kolben eingeschlossenen, sich hin und her bewegenden Moleküle. Als Merkmal eines einzelnen sehen wir beispielsweise die drei rechtwinkligen Komponenten seiner Geschwindigkeit oder den Geschwindigkeitsvektor an. Nun ist es allerdings richtig, daß niemand noch den Versuch gemacht hat, alle Geschwindigkeitsvektoren der Molekel tatsächlich zu messen, zu registrieren und darnach auszurechnen, mit welcher relativen Häufigkeit jeder einzelne Wert auftritt. Der Physiker macht vielmehr eine Annahme über die Größe der relativen Häufigkeit oder, besser gesagt, über ihren (supponierten) Grenzwert und er prüft erst an gewissen Folgerungen, die sich aus dieser Annahme ergeben, ob sie zutrifft. Man sieht, daß der Fall nicht so wesentlich anders liegt als die früheren, wenn auch hier das Experiment zur Feststellung der Wahrscheinlichkeitsgröße nicht unmittelbar ausführbar ist. Die Hauptsache bleibt, daß auch jetzt die Existenz eines Grenzwertes der relativen Häufigkeit, eines bei weiterer Vergrößerung der Gasmenge nicht mehr veränderlichen Endwertes, die Grundlage aller Überlegungen bildet.

Will man das Verhältnis zu dem früher besprochenen Fall der durch tatsächliche Zählung erhobenen Sterbenswahrscheinlichkeit mittels einer Analogie sich verständlicher machen, so denke man an folgendes: Der praktische Geometer kann sich mit einem rechtwinkligen Dreieck beschäftigen, z. B. nach dem Pythagoräischen Lehrsatz seine Hypothenuse ausrechnen, nachdem er vorher durch Messung am Objekt festgestellt hat, daß der betreffende Winkel wirklich hinreichend genau 90° beträgt; aber er kann in einer anderen Lage, ohne die Messung des Winkels vorzunehmen, zunächst voraussetzen, daß das Dreieck rechtwinklig ist, dann seine Folgerungen ziehen und deren Übereinstimmung mit der Beobachtung prüfen. In dieser Lage befindet sich der Physiker, wenn er statistische Begriffsbildungen auf Moleküle oder ähnliche Gebilde anwendet. Man pflegt in der Physik zu sagen (wobei ich von der neuesten Entwicklung in dieser Frage absehe), die Geschwindigkeit eines Molekels usw. sei eine „prinzipiell“ meßbare Größe, wenn sie auch nicht gerade mit den üblichen Meßwerkzeugen gemessen werden kann. So können wir hier erklären, die relative Häufigkeit und ihr Grenzwert, die Wahrscheinlichkeit, werde auch innerhalb der physikalisch definierten Kollektivs „prinzipiell“ in der gleichen Weise beobachtet und gemessen wie in den früher besprochenen Fällen der Glücksspiele und der sozialen oder biologischen Statistik.

Historische Zwischenbemerkung.

Das, was ich bis hierher zur Begründung des Wahrscheinlichkeitsbegriffes ausgeführt habe, widerspricht wohl in weitem Maße dem, was in den älteren Lehrbüchern der Wahrscheinlichkeitsrechnung als formelle „Definition“ der Wahrscheinlichkeit an die Spitze gestellt wird. Es steht aber durchaus in Übereinstimmung mit allen Vorstellungen, die man seit den ältesten Zeiten in der Wahrscheinlichkeitsrechnung tatsächlich benutzt. Ganz besonders deutlich kommt dies in der Darstellung zum Ausdruck, mit der Poisson sein berühmtes, von mir schon einmal erwähntes Lehrbuch über die „Wahrscheinlichkeit der gerichtlichen Urteile“ einleitet. Er beschreibt, wie auf den verschiedensten Gebieten der menschlichen Erfahrung sich die Erscheinung zeigt, die wir als die Unveränderlichkeit der relativen Häufigkeiten bei großer Wiederholungszahl der Einzelbeobachtungen kennengelernt haben.

Poisson bedient sich dabei einer Terminologie, die ich bisher vermieden habe, weil ich Verwechslung mit einem anderen Gedankengang verhüten wollte, der übrigens auch von Poisson in die Wahrscheinlichkeitstheorie eingeführt oder doch von ihm wesentlich gefördert wurde. Er bezeichnet nämlich den Tatbestand der Stabilität der relativen Häufigkeiten innerhalb genügend umfangreicher Versuchsreihen als „Gesetz der großen Zahlen" und sieht in dem Bestehen dieses Gesetzes die eigentliche Grundlage für die Möglichkeit der Wahrscheinlichkeitsrechnung. In der Durchführung seiner Untersuchungen geht er dann allerdings von der durch Laplace eingeführten formalen Definition der Wahrscheinlichkeit aus, über die wir noch zu sprechen haben werden, und leitet schließlich aus dieser mit den Methoden der Analysis einen mathematischen Satz ab, den er nun auch wieder das „Gesetz der großen Zahlen" nennt. Wir werden später sehen, daß der von Poisson abgeleitete mathematische Satz tatsächlich etwas ganz anderes besagt als das, was als allgemeine Erfahrungsgrundlage in die Anfänge der Theorie eingeht. Diese doppelte Verwendung der gleichen Bezeichnung für ganz verschiedene Dinge, die schon sehr schwere Verwirrung gestiftet hat, wird uns in der Folge noch beschäftigen. Der vierte Vortrag wird diesem Gegenstand gewidmet sein. Ich werde dann auch genau die Worte anführen, mit denen Poisson die Erfahrungstatsache der Konstanz der relativen Häufigkeiten bei großen Versuchszahlen als die unumstößliche Grundlage aller Wahrscheinlichkeitsbetrachtungen festgelegt hat. Einstweilen möchte ich Sie bitten, dem Ausdruck „Gesetz der großen Zahlen" noch keinen bestimmten Sinn beizulegen.

Jetzt sei nur noch bemerkt, daß die von mir hier vertretene Auffassung, die bei Begründung des Wahrscheinlichkeitsbegriffes die Erscheinungsreihe, die Beobachtungsfolge, in den Vordergrund stellt und die Wahrscheinlichkeit auf eine relative Häufigkeit innerhalb dieser Folge zurückführt, nicht völlig neu ist. Sie wurde in mehr dialektischer Form ohne unmittelbare Absicht auf einen Aufbau der Wahrscheinlichkeitsrechnung von dem Engländer John Venn in einem Buche „Logic of chance" 1866 ausführlich dargelegt. Die moderne Entwicklung der sogenannten Kollektivmaßlehre durch Theodor Fechner und Heinrich Bruns steht der „Häufigkeitstheorie" der Wahrscheinlichkeit nahe. Am deutlichsten hat sich in dieser Richtung Georg Helm, einer der Mit-begründer des Energieprinzips, ausgesprochen, in einer 1902 erschienen Abhandlung „Die Wahrscheinlichkeitslehre als Theorie der Kollektivbegriffe". Alle diese und viele weitere, der Kürze halber unerwähnt gelassene Ansätze haben jedoch bisher zu einer vollständigen Wahrscheinlichkeitstheorie nicht geführt und nicht führen können, weil sie ein entscheidendes Merkmal des Kollektivbegriffes nicht erfaßt haben, zu dessen Besprechung wir uns jetzt wenden.

Die Regellosigkeit innerhalb des Kollektivs.

Nicht alles, was wir über ein Kollektiv zu sagen haben, über eine Folge von Einzelbeobachtungen, eine Massenerscheinung oder einen Wiederholungsvorgang, der den Gegenstand einer Wahrscheinlichkeitsuntersuchung bilden kann, nicht alles ist in der einen Forderung eingeschlossen, daß die relativen Häufigkeiten der verschiedenen Merkmale feste Grenzwerte haben müssen. Man kann leicht Fälle angeben, in denen Grenzwerte der relativen Häufigkeiten ohne weiteres feststellbar sind, in denen es aber wenig Sinn hätte, von Wahrscheinlichkeiten zu sprechen. Gehen wir etwa eine ausgedehnte Landstraße entlang, an der alle hundert Meter ein kleiner, alle tausend Meter ein größerer Markstein aufgestellt ist, so wird sicher, wenn wir nur ein genügend langes Stück des Weges ins Auge fassen, fast genau ein Zehntel der durchlaufenen Marken das Merkmal „groß" aufweisen. Genau $^1/_{10}$ ist die relative Häufigkeit nur, wenn Anfangs- und Endpunkt der betrachteten Strecke zu den nächsten Kilometersteinen gleich gelegen sind. Aber je weiter wir wandern, desto kleiner wird, wie man ohne weiteres einsieht, die äußerste Abweichung zwischen der relativen Häufigkeit und der Zahl 0,1; mit anderen Worten: ihr Grenzwert ist sicher 0,1. Man könnte gewiß auch hier den Ausdruck anwenden, die „Wahrscheinlichkeit" dafür, einen großen Markstein in der fortlaufenden Reihe anzutreffen, sei 0,1. Daß die Namengebung für wissenschaftliche Begriffe einen gewissen Spielraum freiläßt, ist von mir in der Einleitung ausführlich dargelegt worden.

Es erscheint aber doch zweckmäßiger, aus Gründen, die wir gleich erkennen werden, das Besondere, das den jetzt betrachteten Fall gegenüber den früheren kennzeichnet, nicht unbeachtet zu lassen und die Definition des Begriffes „Kollektiv“ so einzuschränken, daß der neue Fall ausgeschlossen bleibt. Wenn wir als die Folge der Einzelbeobachtungen die fortlaufenden Feststellungen, ob ein Markstein, der beim Zurücklegen des Weges angetroffen wird, ein „großer“ oder ein „kleiner“ ist, ansehen, so unterscheidet sich diese Beobachtungsfolge von der Aufeinanderfolge der Augenzahlen bei einem Würfelspiel dadurch, daß sich leicht ihre Gesetzmäßigkeit angeben läßt. Genau jede zehnte Beobachtung führt auf das Merkmal „groß“, jede andere auf „klein“. Ist man gerade an einem großen Stein vorübergegangen, so fragt man gar nicht, ob der nächste etwa auch wieder ein großer sein kann. Hat man aber mit zwei Würfeln Doppelsechs geworfen, so folgt daraus nichts über das etwaige Ergebnis des nächsten Wurfes und ebensowenig weiß eine Versicherungsgesellschaft, wenn ein Versicherungsnehmer im 41. Lebensjahre gestorben ist, wie es sich mit dem in der Liste zunächst stehenden verhalten wird, mag die Liste nach welchem Gesichtspunkte immer angelegt sein. Es ist ein rein erfahrungsmäßiger Unterschied, der zwischen der einen und der anderen Art von Beobachtungsfolgen zutage tritt. Wir wollen nun grundsätzlich daran festhalten, daß nur solche Folgen, die die Eigenschaft der Gesetzlosigkeit oder „Regellosigkeit“ aufweisen, zu den Kollektivs gerechnet werden sollen. Es fragt sich nur, wie man diese Eigenschaft hinreichend präzise erfassen kann, um die Definition nicht allzu unbestimmt zu lassen.

Formulierung der Regellosigkeit. Stellenauswahl.

Eine geeignete Formulierung für die „Regellosigkeit“ einer Folge zu finden, ist nach dem Gesagten nicht schwer. Der wesentliche Unterschied zwischen der Folge der Augenzahlen beim Würfelspiel und der regelmäßigen Beobachtungsfolge der Marksteine am Wege ist offenbar der, daß wir im letzten Fall leicht ein Auswahlverfahren angeben können, durch das die Häufigkeitszahlen verändert werden. Fangen wir z. B. bei einem großen Stein an und beachten nur jeden zweiten Stein, dem wir begegnen, so zeigt sich sofort, daß die relative Häufigkeit der großen Steine nicht mehr ein Zehntel, sondern ein Fünftel beträgt (weil keiner von den großen, aber jeder zweite von den kleinen Steinen aus-gelassen wird). Wenn wir aber beim Würfelspiel versuchen, nur jeden zweiten Wurf gelten zu lassen oder irgendein anderes, noch so verwickeltes Auswahlverfahren zu benutzen, so zeigt sich doch immer, daß die Häufigkeit der Doppelsechs, sobald nur lange genug gespielt wird, keine andere wird als sie bei dem ursprünglichen, keinen Wurf auslassenden Spiel mit demselben Würfelpaar gewesen ist. Diese Unmöglichkeit, die Gewinstaussichten beim Spiel durch ein Auswahlsystem zu beeinflussen, die Unmöglichkeit des Spielsystems, ist die charakteristische und ausschlaggebende Eigenschaft derjenigen Beobachtungsfolgen oder Massenerscheinungen, die den eigentlichen Gegenstand der Wahrscheinlichkeitsrechnung bilden.

Also gelangen wir zu folgender Formulierung. Wir verlangen von einem Kollektiv, auf das die Wahrscheinlichkeitsrechnung anwendbar sein soll, die Erfüllung zweier Forderungen: Erstens muß die relative Häufigkeit, mit der ein bestimmtes Merkmal in der Folge auftritt, einen Grenzwert besitzen; und zweitens: dieser Grenzwert muß unverändert bleiben, wenn man aus der Gesamtfolge irgendeine Teilfolge willkürlich heraushebt und nur diese betrachtet. Selbstverständlich muß die Teilfolge unbeschränkt fortsetzbar sein wie die ursprüngliche Folge selbst, also etwa so, daß man jedes zweite oder jedes zehnte Glied der Gesamtheit beibehält oder jedes, dessen Nummer eine Quadratzahl oder eine Primzahl ist usf. Im übrigen ist der Phantasie bei der Herstellung der Auswahl jede Freiheit gelassen. Man muß nur über die Zugehörigkeit oder Nichtzugehörigkeit eines Spieles zur Teilfolge unabhängig von dem Ergebnis dieses Spieles entscheiden, also durch eine Verfügung über die Nummer oder die Stellenzahl des Spieles, die man trifft, ehe man den Spielausgang kennt. Wir nennen daher eine Auswahl, wie sie hier gemeint ist, kurz eine Stellenauswahl und verlangen somit, daß die Grenzwerte der relativen Häufigkeiten innerhalb eines Kollektivs gegen Stellenauswahl unempfindlich sein sollen. Dabei heißt Stellenauswahl die Herstellung einer Teilfolge in der Art, daß über die Zugehörigkeit oder Nichtzugehörigkeit eines Elements entschieden wird, ohne daß dabei das Merkmal des Elements, d. i. der Spielausgang, benutzt wird.

Das Prinzip vom ausgeschlossenen Spielsystem.

Es erhebt sich nun die Frage, ähnlich wie früher: Woher wissen wir eigentlich, daß es Kollektivs gibt, die dieser neuen, weit gesteckten Forderung genügen? Auch hier können wir uns nur auf ein — allerdings sehr reiches — Beobachtungsmaterial stützen. Jedem, der einmal in Monte Carlo war oder eine Beschreibung der Verhältnisse in einer Spielbank gelesen hat, ist wohl bekannt, welch verwickelte „todsichere" Systeme von gewissen Spielnarren ausgeheckt und erprobt worden sind, auch immer wieder von neuem erprobt werden. Daß sie nicht zum gewünschten Ziele führen, nämlich zu einer Verbesserung der Spielchancen, also zu einer Veränderung der relativen Häufigkeiten, mit der die einzelnen Spielausgänge innerhalb der systematisch ausgewählten Spielfolge auftreten, das ist die traurige Erfahrung, die über kurz oder lang alle Systemspieler machen müssen. Auf diese Erfahrungen stützen wir uns bei unserer Definition der Wahrscheinlichkeit.

Es drängt sich hier ein Vergleich auf, dessen kurzer Besprechung ich nicht aus dem Weg gehen will. Die fanatischen Erfinder und Erprober der Spielsysteme in Monte Carlo und anderwärts weisen eine unverkennbare Ähnlichkeit mit einer anderen Klasse von „Erfindern" auf, deren Tätigkeit wir heute nur mit einem gewissen Mitleid zu betrachten gewohnt sind, mit den seit Jahrhunderten bekannten, niemals aussterbenden Konstrukteuren eines „Perpetuum mobile". Es lohnt, diese Analogie, die nicht nur eine menschlich-psychologische ist, ein wenig weiter zu verfolgen. Warum lächelt heute der Gebildete über jeden, der ein Perpetuum mobile zu konstruieren versucht? Nun, wird er sagen, weil wir von dem Prinzip der Erhaltung der Energie her wissen, daß eine solche Konstruktion unmöglich ist. Was aber ist das Energieprinzip anderes als ein sehr stark verallgemeinerter, vielfältig verankerter Erfahrungssatz, der in letzter Linie durch das Fehlschlagen der unzähligen Versuche, ein Perpetuum mobile zu bauen, begründet wurde? Man pflegt heute oft in der theoretischen Physik das Energieprinzip mit seinen verschiedenen Auswirkungen geradezu als das „Prinzip vom ausgeschlossenen Perpetuum mobile" zu bezeichnen. Davon, daß das Energieprinzip sich irgendwie „beweisen" lasse, kann keine Rede sein, wenn man unter „beweisen“ nicht eben die Übereinstimmung mit allen bisherigen Erfahrungen versteht; die außerordentliche Evidenz, die es für uns besitzt, verdankt es nur der ungeheuren Fülle von Erfahrungsmaterial, das, abgesehen von den heute nur historisch wichtigen Perpetuum-mobile-Versuchen, eigentlich in der Gesamtheit aller in der Technik gebräuchlichen Energieumformungen besteht.

Wenn wir nun aus den Erfahrungen der Spielbanken ein „Prinzip vom ausgeschlossenen Spielsystem“ abstrahieren und in die Grundlagen der Wahrscheinlichkeitsrechnung aufnehmen, verfahren wir ganz ähnlich, wie es die Physik im Falle des Energieprinzips tut. Auch hier gilt, daß außer jenen zumeist laienhaften Versuchen der Glücksritter vor allem die Erfahrungen der Versicherungsgesellschaften und ähnlicher Unternehmungen das Prinzip stützen: Wenn es eine Möglichkeit gäbe, die relative Häufigkeit, mit der der Versicherungsfall eintritt, im großen abzuändern dadurch, daß man etwa jeden zehnten Versicherungsnehmer oder dergleichen ausscheidet, so wäre die ganze finanzielle Basis der Unternehmen in Frage gestellt. Was das Energieprinzip für das elektrische Kraftwerk, das bedeutet unser Satz vom ausgeschlossenen Spielsystem für das Versicherungswesen: die unumstößliche Grundlage aller Berechnungen und aller Maßnahmen. Wie von jedem weittragenden Naturgesetz können wir von diesen beiden Sätzen sagen: Es sind Einschränkungen, die wir auf Grund der Erfahrung unserer Erwartung über den künftigen Ablauf von Naturvorgängen auferlegen. Aus der Tatsache, daß die Voraussagen sich als richtig erweisen, dürfen wir den Schluß ziehen, daß es jedenfalls solche Massenerscheinungen oder Wiederholungsvorgänge gibt, auf die das Prinzip vom ausgeschlossenen Spielsystem anwendbar ist, und nur auf diese allein werden sich alle unsere weiteren Ausführungen zur Wahrscheinlichkeitsrechnung beziehen.

Beispiel für Regellosigkeit.

Um Ihnen noch anschaulicher zu machen, wie ein Kollektiv in seiner „Regellosigkeit“ aussieht, will ich Ihnen hier einen kleinen Versuch vorführen. Daß es wieder ein Experiment aus dem Gebiete der Glücksspiele ist, liegt nur daran, daß alle anderen

Möglichkeiten der Anwendung unserer Theorie zu umfangreiche und zeitraubende Einrichtungen erfordern. Ich ziehe also aus einem Säckchen, das 90 runde Täfelchen mit den Nummern 1 bis 90 enthält, blindlings ein Stück und notiere, ob eine gerade oder eine ungerade Zahl erschienen ist. Für eine gerade Zahl will ich eine Null anschreiben, für eine ungerade eine Eins; ich ziehe 100mal und setze die Ergebnisse der Reihe nach in zehn Zeilen nebeneinander:

1	1	0	0	0	1	1	1	0	1
0	0	1	1	0	0	1	1	1	1
0	1	0	1	0	0	1	0	0	0
0	1	0	0	1	0	0	1	1	1
0	0	1	1	0	0	0	0	1	1
0	1	1	1	1	0	1	0	1	0
1	0	1	1	1	1	0	0	1	1
0	0	1	1	0	1	1	1	0	1
0	0	1	1	0	0	1	1	0	1
0	1	1	0	0	0	1	0	0	0

Unter den 100 Versuchen sind hier 51 Einser aufgetreten, d. h. die relative Häufigkeit der Eins beträgt $^{51}/_{100}$. Beachtet man nur jeden zweiten Versuch, also nur die in der ersten, dritten, fünften .... Spalte stehenden Zeichen, so findet man unter den 50 Versuchen 24 Einser, also die relative Häufigkeit $^{48}/_{100}$. Zählt man nur die Zeichen in der ersten, dritten, fünften .... Zeile, so erhält man die Häufigkeit $^{50}/_{100}$ für die Eins. Berücksichtigt man nur diejenigen 26 Versuche, deren Nummer eine Primzahl ist, also den 1., 2., 3., 5., 7., 11., 13., 17., 19., 23., 29., 31., 37., 41., 43., 47., 53., 59., 61., 67., 71., 73., 79., 83., 89. und 97., so findet man genau 13 Einser, also wieder die Häufigkeit $^{50}/_{100}$. Schließlich betrachten wir noch diejenigen 51 Versuche, denen eine Eins vorausgegangen ist — denn ein Systemspieler könnte auch auf den Gedanken kommen, auf Null zu setzen, unmittelbar nachdem sich eine Eins gezeigt hat — und finden jetzt, daß hier 27 Einser stehen, also eine relative Häufigkeit gleich $^{27}/_{51}$ oder $^{53}/_{100}$. Der Versuch zeigt uns, daß bei den verschiedenartigen Auswahlen, die wir angewendet haben, die Einser stets mit der ungefähr gleichen Häufigkeit von etwa $^{1}/_{2}$ auftreten. Sie werden selbst das Gefühl haben, daß man bei einer ausgedehnteren

Versuchsreihe, die ich aus Zeitmangel hier nicht ausführen kann, die Erscheinung der „Regellosigkeit" noch in weit vollkommenerer Weise würde nachweisen können. Natürlich läßt sich, wenn die 100 Versuchsergebnisse einmal aufgezeichnet vorliegen, hinterher ohne weiteres eine Stellenauswahl angeben, bei der nur Einser oder nur Nullen auftreten, oder die Einser in irgendeiner beliebigen, zwischen 0 und 1 gelegenen Häufigkeit. Auch kann es leicht sein, daß bei einer anderen Gruppe von 100 Versuchen eine der hier von uns vorgenommenen Stellenauswahlen ein ganz abweichendes Ergebnis liefert. Das Prinzip der Regellosigkeit besagt ja nur, daß, wenn man die Versuche genügend weit fortführt und das einmal festgelegte Auswahlverfahren beibehält, schließlich doch wieder die relative Häufigkeit 0,50 mehr oder weniger genau herauskommt.

Zusammenfassung der Definition.

Auf Einzelheiten mathematischer Natur, auf Überlegungen, die nötig sind, um die Definitionen und Formulierungen in mathematischer Hinsicht zu sichern, brauche ich hier nicht einzugehen. Wer sich dafür interessiert, sei auf meine erste Veröffentlichung im vierten Bande der Mathematischen Zeitschrift von 1918 verwiesen oder auf das als erster Band meiner „Vorlesungen über angewandte Mathematik" erschienene Lehrbuch der Wahrscheinlichkeitsrechnung, das die Theorie in vereinfachter und, wie ich denke, verbesserter Gestalt darstellt. Auf verschiedene Fragen grundsätzlicher Art, auf die gegen meine Definition vorgebrachten Einwände und ihre Widerlegung werde ich erst im dritten Vortrag näher eingehen. Dort hoffe ich noch manches Bedenken, das vielleicht bei Ihnen entstanden ist, zu zerstreuen und manchen Punkt näher aufklären zu können. Jetzt sei mir nur gestattet, in aller Kürze noch einmal die wenigen Sätze zusammenzufassen, die wir als Grundlage für alles Folgende gewonnen haben und in denen wir die für uns allein maßgebende Definition der (mathematischen) Wahrscheinlichkeit erblicken. Es sind diese:

1. Von Wahrscheinlichkeit kann erst gesprochen werden, wenn ein wohlbestimmtes, genau umgrenztes Kollektiv vorliegt.
2. Kollektiv ist eine Massenerscheinung oder ein Wiederholungs-

vorgang, der zwei Forderungen erfüllt, nämlich: es müssen die relativen Häufigkeiten der einzelnen Merkmale bestimmte Grenzwerte besitzen und diese müssen ungeändert bleiben, wenn man durch willkürliche Stellenauswahl einen Teil der Elemente aus der Gesamtheit heraushebt.

3. Das Erfülltsein der letzteren Forderung bezeichnen wir auch als das Prinzip der Regellosigkeit oder Prinzip vom ausgeschlossenen Spielsystem.
4. Den gegen Stellenauswahl unempfindlichen Grenzwert der relativen Häufigkeit, mit der ein bestimmtes Merkmal auftritt, nennen wir die „Wahrscheinlichkeit für das Auftreten dieses Merkmals innerhalb des betrachteten Kollektivs". Wenn dieser Zusatz zu dem Worte „Wahrscheinlichkeit" gelegentlich fortbleibt, so geschieht es nur der Kürze halber, er muß in Gedanken immer ergänzt werden.

Zweiter Vortrag.

Elemente der Wahrscheinlichkeitsrechnung.

In dem ersten Vortrag habe ich schon gelegentlich erwähnt, daß die von mir hier entwickelte Auffassung, die den Wahrscheinlichkeitsbegriff als Grenzwert einer beobachtbaren relativen Häufigkeit definiert, auch ihre Gegner hat. Im dritten Vortrag werde ich auf die wichtigsten Einwände, die gegen diese Auffassung erhoben wurden, zu sprechen kommen. Zunächst aber will ich jetzt kurz darlegen, in welcher Weise die von mir gegebenen Grundlagen in der Welt der Tatsachen Anwendung finden, wie sie sich in praktischen Fragen auswirken, kurz, was überhaupt mit ihnen anzufangen ist. Denn in der Anwendbarkeit einer Theorie auf die Wirklichkeit sehe ich den wesentlichsten, wenn nicht einzigen Prüfstein ihres Wertes.

Die Wahrscheinlichkeitsrechnung eine normale Wissenschaft.

Ich muß hier mit einer Erklärung beginnen, die den unmittelbarsten Widerspruch bei den Vertretern jener Auffassung hervorrufen wird, die da meinen, es handle sich in der Wahrscheinlichkeitslehre um eine besondere, von allen anderen grundsätzlich verschiedene Art der Wissenschaft, die dadurch gekennzeichnet wird, daß in ihr eine andere Art von Logik zur Geltung kommt. Es wird behauptet — dies ist keine Übertreibung —, daß, während man in den übrigen Wissenschaften Schlüsse aus dem ziehe, was man weiß, hier aus dem Nichtwissen die wertvollsten Schlußfolgerungen zu gewinnen wären. „Absolutes Nichtwissen über die Bedingungen", heißt es nach Czuber, unter denen ein Würfel fällt, führt zu der Schlußfolgerung, daß jeder der sechs Würfelseiten die Wahrscheinlichkeit $^1/_6$ zukomme. Nun, es scheint mir etwas rätselhaft, wie man bei absolutem Nichtwissen zwischen den beiden Würfelpaaren unterscheidet, die ich Ihnen im Verlaufe des ersten Vortrages vorgeführt habe, und von denen das eine jedenfalls einen sehr stark von $^1/_6$ abweichenden Wahrscheinlichkeitswert für die Sechs aufweist, wenigstens nach unserer Definition der Wahrscheinlichkeit.

In der Tat können wir mit derartigen phantastischen Vorstellungen von einer besonderen Logik u. dgl. nichts anfangen. Daß zweimal zwei vier ist, daß aus einem richtigen Vordersatz A nicht sowohl B wie das Gegenteil von B gefolgert werden kann, das alles bleibt unverändert gültig, auch wenn wir Wahrscheinlichkeitsrechnung betreiben. Ebenso bleibt richtig, daß aus Nichts nichts folgt. Wie jede andere Realwissenschaft geht die Wahrscheinlichkeitslehre von Beobachtungen aus, ordnet, klassifiziert sie, abstrahiert aus ihnen Grundbegriffe und Grundsätze, aus denen dann — mit Benutzung der normalen, überall gültigen Logik — Folgerungen gezogen werden, die sich an der Beobachtung prüfen lassen. Mit einem Worte: für uns ist die Wahrscheinlichkeitslehre eine ganz normale Wissenschaft, gekennzeichnet durch den besonderen Gegenstand, mit dem sie sich beschäftigt, aber nicht durch eine besonders geartete Denkmethode.

Die Aufgabe der Wahrscheinlichkeitsrechnung.

Stellt man sich auf diesen nüchternen, naturwissenschaftlichen Standpunkt, der auch in der Wahrscheinlichkeitslehre dieselben Denkgesetzte, dieselben Schlußweisen, dieselben methodischen Grundsätze gelten läßt wie überall sonst, so läßt sich die Aufgabe, die die Wahrscheinlichkeitsrechnung zu lösen vermag, wie folgt umschreiben.

Es gibt ersichtlich Kollektivs, die miteinander zusammenhängen, die einander nicht „fremd“ sind, z. B. das Spiel mit dem ersten dieser beiden Würfel, das Spiel mit dem zweiten und drittens das Spiel mit beiden Würfeln. Durch die beiden ersten dieser Kollektivs ist das dritte bestimmt. Die Aufgabe der Wahrscheinlichkeitsrechnung, und zwar die einzige, die ihr in dem vorliegenden Falle zukommt, ist es nun, aus den Wahrscheinlichkeiten, die innerhalb der beiden erstgenannten Kollektivs bestehen, die Wahrscheinlichkeiten innerhalb des abgeleiteten, dritten Kollektivs zu berechnen. Als gegebene Größen gehen in die Rechnung ein: die sechs Wahrscheinlichkeiten, mit dem ersten Würfel eine Eins, eine Zwei, eine Drei ...., eine Sechs zu werfen, sowie die sechs analogen Werte für den zweiten Würfel; berechnet werden kann daraus z. B. die Wahrscheinlichkeit dafür, mit den beiden Würfeln zusammen die Summe 10 zu werfen.

Es ist genau so wie in der Geometrie, die uns lehrt, aus zwei Seiten eines Dreiecks und dem eingeschlossenen Winkel die Länge der dritten Seite zu rechnen. Woher wir die Daten der Aufgabe, die Längen der beiden Seiten und den Winkel kennen, darnach fragt der Geometer nicht oder er sieht wenigstens die Feststellung dieser Ausgangswerte seiner Rechnung nicht als eine Aufgabe der Geometrie an. Mit einer solchen Feststellung beschäftigt sich etwa der Landmesser, der Geodät, der bei seiner Arbeit manche geometrische Überlegung verwerten muß — auch dazu werden wir später in der Statistik die genaue Analogie kennenlernen. Die Geometrie im engeren Sinne lehrt nur, wie man aus gewissen als gegeben vorausgesetzten Größen andere, unbekannte bestimmen kann, und zwar gleichgültig, welche Werte die ersteren haben und unabhängig davon, wie diese Werte gefunden wurden. So liefert die richtig aufgefaßte Wahrscheinlichkeitsrechnung eine Formel für die Wahrscheinlichkeit des Erscheinens der Summe 10 oder der Doppelsechs, die in allen Fällen gültig sein muß, ob ich nun das eine oder andere Paar meiner Würfel benutze oder aus den vieren ein neues Paar kombiniere usf. Wie groß die je sechs Einzelwahrscheinlichkeiten für den ersten und den zweiten der benutzten Würfel sind und woher ich sie kenne, das ist für die Rechnung genau so gleichgültig, wie Werte und Herkunft der Daten in einer geometrischen Aufgabe.

Mit einem Schlage werden unzählige, mehr oder minder ernst gemeinte „populäre“ Einwendungen gegen die Wahrscheinlichkeitsrechnung hinfällig, sobald man als ihre ausschließliche Aufgabe erkennt, aus gegebenen Wahrscheinlichkeiten innerhalb gewisser Ausgangskollektivs die Wahrscheinlichkeiten innerhalb eines aus jenen abgeleiteten Kollektivs zu berechnen. Die Frage, mit der man einen Mathematiker zu hänseln pflegt: Können Sie berechnen, wie groß die Wahrscheinlichkeit dafür ist, daß ich den nächsten Zug nicht versäume? ist in dem gleichen Sinne abzuweisen wie die Frage: Können Sie die Entfernung dieser beiden Bergspitzen voneinander berechnen? Eine Entfernung kann man nur rechnen, wenn andere geeignete Entfernungen und Winkel gegeben sind, und eine Wahrscheinlichkeit nur, sobald man andere Wahrscheinlichkeiten, die jene bestimmen, kennt. Der allgemeinen Schulbildung, die seit langem schon gewisse Elemente der Geometrie aufgenommen hat, verdanken wir es, daß der halbwegs Gebildete die praktische Aufgabe des Landvermessers und die theoretische Fragestellung des Geometers (Mathematikers) auseinanderhält; hinsichtlich der Wahrscheinlichkeitsrechnung und Statistik ist diese notwendige Aufklärung noch nicht durchgedrungen.

Anfang und Ende jeder Aufgabe: Wahrscheinlichkeiten.

Bei jeder mathematischen Aufgabe gibt es, das ist wohl bekannt, einerseits gegebene Größen, die Daten der Aufgabe, anderseits die gesuchten Größen, die aus den gegebenen abgeleitet werden und die man dann die Resultate nennt. Was ich eben ausführte, läßt sich in den Satz fassen: Sowohl die Daten wie die Resultate einer Aufgabe der Wahrscheinlichkeitsrechnung sind Wahrscheinlichkeiten. Vorhin habe ich Nachdruck auf den ersten Teil der Aussage gelegt, der sich auf den Ausgangspunkt der Rechnung bezieht, jetzt will ich noch zu dem zweiten Teil ein Wort sagen.

Das Ergebnis irgendeiner Berechnung, die in das Gebiet der Wahrscheinlichkeitsrechnung fällt, kann bei uns immer nur eine Wahrscheinlichkeit sein. Das bedeutet, sobald wir die Definition einsetzen: die Angabe einer relativen Häufigkeit in einer genügend langen (theoretisch unendlichen) Beobachtungsfolge. Niemals kann eine Aussage, die etwas Bestimmtes über ein Einzelereignis aussagt, aus der Wahrscheinlichkeitstheorie gefolgert werden. Einzig und allein darüber, was im Durchschnitt einer sehr langen Versuchsreihe zu erwarten ist, belehrt die Wahrscheinlichkeitsrechnung. Es ist wichtig zu bemerken, daß dies auch dann gilt, wenn die errechnete Wahrscheinlichkeit einen der beiden extremen Werte 0 oder 1 besitzt.

Nach der klassischen Theorie (und auch in mancher neueren Auffassung) bedeutet der Wahrscheinlichkeitswert 1, daß das betreffende Ereignis *sicher* eintritt. In diesem Falle ließe sich also, wenn wir dieser Auffassung folgten, aus der Kenntnis einer Wahrscheinlichkeit ein sicherer Schluß über das bei einem beliebigen Versuch Eintretende ziehn. Definiert man aber, wie wir es tun, die Wahrscheinlichkeit als den Grenzwert der relativen Häufigkeit, so kann man aus der Tatsache, daß er gleich eins ist, unmöglich folgern, daß das betreffende Merkmal *ausnahmslos* bei jedem Versuch auftritt. Folgendes Beispiel wird das klarmachen. Stellen wir uns eine unendliche Folge von zwei verschiedenen Merkmalen A und B vor, die so beschaffen ist: Zuerst kommt ein A, dann nach *einem* B wieder ein A, hierauf erst nach zweimaligem B ein A, dann nach einer Gruppe von drei B ein A, sodann nach vier B eines usf. Die Folge sieht also in ihrem Anfang so aus:

ABABBABBBABBBBABBBBBA....

Es ist das eine regelmäßige, durch eine mathematische Formel darstellbare Zeichenfolge und es ist leicht auszurechnen, daß hier die relative Häufigkeit des Merkmals A allmählich gegen Null, die des Merkmals B gegen Eins geht, wenn man in der Reihe weiter und weiter geht. Ein solches Verhalten kann natürlich auch bei einer regellosen Folge eintreten. Wenn das Merkmal nur selten genug erscheint, so kann es durchaus vorkommen, daß seine relative Häufigkeit, auch wenn sie an keiner endlichen Stelle den Wert Null hat, doch mehr und mehr abnimmt, wenn die Versuchszahl wächst, so daß ihr Grenzwert gleich Null ist. Wir sehen also: Wahrscheinlichkeit Null bedeutet nur *sehr seltenes*, man könnte sagen unendlich seltenes Auftreten des betreffenden Merkmals, Wahrscheinlichkeit Eins bedeutet, daß das betreffende Merkmal *fast immer*, aber nicht notwendig immer auftritt. In diesem Sinne bleibt bei unserer Auffassung des Wahrscheinlichkeitsbegriffes der indeterministische Charakter jeder Aussage der Wahrscheinlichkeitsrechnung in allen Fällen gewahrt.

Wir müssen nun noch genauer überlegen, worin die hier mehrfach genannte „Ableitung" eines Kollektivs aus anderen besteht. Nur wenn dieser Begriff ganz konkret gefaßt wird, ist wirkliche Klarheit über die Aufgabe der Wahrscheinlichkeitsrechnung zu erreichen. Vorher aber will ich einen einfachen und naheliegenden Ausdruck einführen, der uns gestatten wird, unsere Aussagen etwas übersichtlicher zu fassen. Es handelt sich dabei zunächst nur um eine sprachliche Abkürzung, sodann um eine kleine Ausdehnung des bisher verwendeten Begriffs des Kollektivs.

Die Verteilung innerhalb eines Kollektivs.

Die Glieder oder Elemente eines Kollektivs unterscheiden sich durch ihre Merkmale, die Zahlen sein können wie beim Würfelspiel, Farben wie beim Roulettespiel (Rouge et noir) oder irgendwelche durch Beobachtung feststellbare Eigenschaften. Mindestens gibt es in einem Kollektiv zwei Merkmale, sind ihrer nicht mehr, so nennen wir es eine „einfache Alternative". Innerhalb einer solchen haben wir zwei Wahrscheinlichkeiten, die natürlich die Summe 1 geben. Z. B.: Das Kopf- oder Adlerspiel mit einer Münze stellt eine Alternative dar mit den beiden Merkmalen „Kopf" und „Adler", von denen jedes — normale Verhältnisse vorausgesetzt — mit der Wahrscheinlichkeit $^1/_2$ auftritt. Auch bei der Lebens- oder Sterbenswahrscheinlichkeit liegt eine einfache Alternative vor. Die beiden Merkmale sind z. B.: Eintritt des Todes zwischen dem 40. und 41. Geburtstag oder Überleben des letzteren Datums; die Wahrscheinlichkeit des ersten Merkmales war in dem früher betrachteten Spezialfall 0,011, daher ist die des zweiten, die Überlebenswahrscheinlichkeit 0,989. Bei anderen Kollektivs ist die Zahl der Merkmale größer als zwei, beispielsweise beim Spiel mit einem Würfel. Hier bestehen sechs verschiedene Ergebnismöglichkeiten für einen Versuch, je nachdem, welche der sechs Würfelseiten obenauf fällt. Demgemäß gibt es sechs Wahrscheinlichkeiten und ihre Summe muß wieder 1 sein. Sind alle sechs Wahrscheinlichkeiten untereinander gleich, so hat jede den Wert $^1/_6$ und wir nennen in diesem Falle den

Würfel einen „richtigen". Aber gleichgültig, ob wir einen richtigen oder falschen Würfel vor uns haben, die sechs Zahlenwerte der Wahrscheinlichkeiten für seine sechs Seiten, diese sechs echten Brüche mit der Gesamtsumme 1, müssen bekannt sein, wenn wir das Kollektiv als ein gegebenes ansehen sollen.

Es empfiehlt sich nun, für den Inbegriff der sechs Größen eine kurze Bezeichnung einzuführen. Wir wollen die Gesamtheit der Wahrscheinlichkeiten, die für das Auftreten der einzelnen Merkmale innerhalb eines bestimmten Kollektivs bestehen, die Verteilung dieses Kollektivs nennen. Die Wahl dieses Ausdrucks wird verständlich, wenn man an die Verteilung der Chancen oder der Gewinnaussichten bei einem Glücksspiel denkt. Wenn etwa von sechs Spielern jeder auf eine Würfelseite setzt, so „verteilen" sich die Chancen auf die einzelnen Spieler so, daß der verhältnismäßige Anteil eines jeden gerade gleich der Wahrscheinlichkeit der betreffenden Würfelseite ist; bei einem richtigen Würfel sind die Chancen gleichmäßig oder gleich „verteilt", bei einem falschen ungleich. Bei einer einfachen Alternative besteht die ganze Verteilung nur aus zwei Zahlen, die sich zur Summe 1 ergänzen. Man kann aber auch, um sich das Wort „Verteilung" anschaulich zu machen, daran denken, wie die verschiedenen Merkmale des Kollektivs sich in der unendlich ausgedehnten Masse der Elemente des Kollektivs „verteilen". Bilden die Zahlen $^1/_5$, $^3/_5$, $^1/_5$ die Verteilung eines Kollektivs mit drei Merkmalen, ist also $^1/_5$ die Wahrscheinlichkeit des ersten und dritten, $^3/_5$ die des zweiten Merkmals, so heißt das ja nichts anderes, als daß in der genügend weit fortgeführten Beobachtungsfolge die Merkmale so „verteilt" sind, daß das erste und dritte je $^1/_5$ aller Fälle umfaßt, das dritte die restlichen $^3/_5$ usf.

Treffer-Wahrscheinlichkeit, stetige Verteilung.

Der so erklärte Verteilungsbegriff legt es nahe, unsere Betrachtungen auch auf Fälle auszudehnen, die bisher nicht in den Kreis der Überlegungen gezogen waren. Denken wir an einen Schützen, der gegen eine passend aufgestellte Scheibe fortgesetzt Schüsse abgibt. Wir haben hier einen Wiederholungsvorgang, den wir uns unbeschränkt fortsetzbar denken können. Wenn wir die Scheibe in der üblichen Weise in eine Anzahl von Kreisringen geteilt annehmen und jeden Kreisring, einschließlich des inneren

Vollkreises und der Außenfläche um den größten Kreis herum, mit einer Nummer versehen, so kann als Merkmal eines Schusses die Nummer des getroffenen Bereiches gelten. So weit wäre das noch nichts Neues gegen früher; die Zahl der Merkmale oder der Wahrscheinlichkeiten, die zusammen die „Verteilung" bilden, wäre gleich der Anzahl der bezifferten Bereiche. Die beiden Voraussetzungen der Existenz der Grenzwerte und der Regellosigkeit sehen wir natürlich als erfüllt an, wenn wir überhaupt von Wahrscheinlichkeit sprechen.

Nun läßt sich aber das Experiment mit dem Schützen und der Scheibe noch etwas anders behandeln. Sobald ein Einschuß erfolgt ist, kann man seine Entfernung vom Mittelpunkt messen und diese Entfernung (statt der Nummer des Ringes) als das Merkmal des Schusses, als das einzelne Beobachtungsergebnis ansehen. Eine Entfernung ist in letzter Linie eine Anzahl von Maßeinheiten, z. B. von Zentimetern, und wenn wir uns wirklich darauf beschränken, nur in ganzen Zentimetern zu messen, so haben wir einen nur wenig gegen früher veränderten Fall vor uns: An die Stelle der Nummern des jeweils getroffenen Bereiches ist jetzt eine andere Nummer, nämlich die Zentimeterzahl des Abstandes vom Mittelpunkt getreten. Wenn etwa die äußerste noch in Frage kommende Entfernung 1 m beträgt, so haben wir 101 verschiedene Merkmale, nämlich die Zahlen 0 bis 100, ebenso viele Wahrscheinlichkeiten, und die „Verteilung" besteht aus 101 echten Brüchen mit der Summe 1. Aber man fühlt wohl, daß man dem Begriff der „Entfernung" nicht voll gerecht wird, wenn man ihn einfach durch eine Anzahl von Zentimetern ersetzt: Es gibt zwischen 0 und 1 m mehr verschiedene Entfernungen als 100 oder 101. Der Geometer sagt, die Entfernung ist eine stetige Veränderliche, sie kann jeden Wert zwischen 0 und 100 annehmen, nicht nur die endliche Anzahl ganzzahliger Werte, sondern auch unendlich viele nicht-ganzzahlige. So gelangt man zu der Vorstellung, daß man hier ein Kollektiv mit unendlich viel verschiedenen Merkmalen vor sich hat. In den klassischen Lehrbüchern spricht man in diesem Fall von einer „geometrischen" Wahrscheinlichkeit, im Gegensatz zur „arithmetischen" bei nur endlich vielen Merkmalen. (Über die Berechtigung oder Zweckmäßigkeit dieser Bezeichnungsweise soll nicht gestritten werden.) Nun kann man gewiß nicht unendlich viel Zahlen angeben, wie es jetzt nötig wäre, wenn in dem früheren Sinn die Verteilung innerhalb der Kollektivs durch tatsächliche Aufführung der Wahrscheinlichkeiten aller Merkmale bestimmt werden sollte. Glücklicherweise ist aber die Schwierigkeit, die hier auftritt, eine von den Mathematikern längst gelöste. Wer nur das Allerelementarste von höherer Mathematik gelernt hat, weiß, wie man sich hier hilft.

Wahrscheinlichkeitsdichte.

Eine Analogie aus einem anderen Gebiet wird, wie ich denke, es jedermann verständlich machen, in welcher Weise man eine „Verteilung", die sich auf unendlich viel Merkmale, auf eine stetige Folge von Merkmalen erstreckt, beschreiben kann. Denken wir uns, daß wir ein bestimmtes Gewicht oder eine Masse, sagen wir 1 kg, über die Punkte einer Strecke von 1 m Länge zu verteilen haben. Solange nur endlich viel Massenpunkte in Frage kommen, ist die Verteilung durch Angabe endlich vieler Zahlen, nämlich der Brüche, die, in Kilogramm ausgedrückt, die einzelnen Gewichtsgrößen bezeichnen, bestimmt. Ist aber das Gewicht von 1 kg *stetig* über die Länge eines Meters verteilt oder, wie man auch sagt, die Strecke stetig mit Masse belegt, so etwa wie ein dünner gerader Stab von 1 m Länge, der im ganzen 1 kg wiegt, so kann man nicht mehr von einzelnen, Masse tragenden Punkten sprechen, versteht aber doch sofort, was unter „Verteilung" der Masse über die Stablänge gemeint ist. An einer Stelle, an der der Stab dicker ist, drängt sich mehr Masse auf das gleiche Längenstück zusammen, man sagt, hier sei die auf die Längeneinheit bezogene Massendichte größer. Ist der Stab überall gleich dick, so sprechen wir von „Gleichverteilung" der Masse, in jedem Fall ist durch Angabe der *Dichte* an jeder Stelle die Verteilung bestimmt.

Es ist nicht schwer, diese Vorstellung auf die Wahrscheinlichkeitsverteilung in dem Beispiele des Scheibenschießens zu übertragen. Auf jedes Stück der Entfernung zwischen 0 und 100 cm entfällt eine bestimmte Wahrscheinlichkeit des Einschusses und die gesamte Verteilung ist durch Angabe der auf die Längeneinheit entfallenden Dichte an jeder Stelle bestimmt. Wir stehen gar nicht an, hier, wo es sich um ein stetig veränderliches Merkmal handelt, geradezu von der „*Wahrscheinlichkeitsdichte*" zu sprechen und zu erklären: Gibt es in einem Kollektiv nur endlich viele, diskret voneinander zu unterscheidende Merkmale, so besteht die Verteilung aus der Gesamtheit der auf die einzelnen Merkmale entfallenden Wahrscheinlichkeitsgrößen; ist aber das Merkmal eine von Versuch zu Versuch stetig veränderliche Größe, z. B. eine Entfernung zweier Punkte, so wird die Verteilung durch die an jeder Stelle des Variablenbereiches auf die Längeneinheit entfallende „Wahrscheinlichkeitsdichte" festgelegt. Bei der Schußscheibe kann z. B., wenn man voraussetzt, daß blindlings geschossen wird, die Annahme berechtigt sein, daß ein weiter nach außen gelegenes Stück des Radius größere Wahrscheinlichkeit für sich hat, als ein gleich großes, mehr innen gelegenes, weil der zugehörige Kreisring bei gleicher Breite außen größer ist. Man käme so zu dem Ansatz, daß die Wahrscheinlichkeitsdichte proportional dem Radius oder einer Potenz des Radius zunimmt.

Wir werden uns später mit bestimmten Fragen, die sich an Kollektivs mit stetig veränderlichem Merkmal anknüpfen, noch zu beschäftigen haben und dabei auch zeigen, wie sich dieser Begriff auch auf mehrdimensionale Merkmalbereiche, Flächenstücke oder Raumteile statt Linien, ausdehnen läßt. Für jetzt sollte diese Einschaltung nur dazu dienen, Ihnen den Begriff der Verteilung innerhalb eines Kollektivs, möglichst in nicht zu enger Form, anschaulich zu machen. Wir können, wenn wir uns dieses Begriffes bedienen, etwas präziser aussprechen, worin die Aufgabe der Wahrscheinlichkeitsrechnung besteht:

Es lassen sich in ganz bestimmter, noch näher anzugebender Weise aus einem oder aus mehreren wohldefinierten Kollektivs neue Kollektivs ableiten; *Aufgabe der Wahrscheinlichkeitsrechnung ist es, aus den als gegeben vorausgesetzten Verteilungen innerhalb der Ausgangskollektivs die Verteilungen innerhalb der abgeleiteten Kollektivs zu berechnen.*

Zurückführung auf vier Grundaufgaben.

In dieser Formulierung bedarf nur ein Ausdruck noch näherer Erklärung, nämlich, in welcher Weise aus gegebenen Kollektivs neue „abgeleitet" werden können. Ich sagte schon, daß es darauf ankommt, dies ganz konkret und präzise anzugeben, sonst hat alles Bisherige keinen Wert. Ich füge jetzt hinzu: es gibt in letzter Linie *vier und nur vier Verfahren der Ableitung* eines Kollektivs aus anderen und alle Probleme, die in der Wahrscheinlichkeitsrechnung behandelt werden, sind auf eine Kombination dieser vier Grundverfahren zurückführbar. Natürlich sind in einer konkreten Aufgabe meist mehrere der Grundverfahren und jedes von ihnen wiederholt durchzuführen. Wir werden jetzt zunächst die vier Verfahren genau besprechen: Für jeden der vier Fälle müssen wir die Grundaufgabe lösen, das ist die Aufgabe, die neue Verteilung aus den Ausgangsverteilungen zu berechnen. Wenn ich nun noch sage, daß ich ohne Furcht, auf zu große Schwierigkeiten des Verständnisses zu stoßen, die vier verschiedenen Lösungen hier vorführen werde, daß namentlich die beiden ersten von ganz außerordentlicher Einfachheit sind, so vermute ich, daß bei Ihnen leicht ein peinlicher Verdacht entstehen wird. Der Verdacht nämlich, daß Sie über die eigentlichen mathematischen Schwierigkeiten, die doch irgendwo in der Wahrscheinlichkeitsrechnung stecken müssen — davon zeugt schon das formelreiche Äußere jedes Lehrbuches —, hinweggetäuscht werden sollen.

Allein das ist gewiß nicht meine Absicht. Indem ich alle Gedankenoperationen, durch die in der Wahrscheinlichkeitsrechnung Kollektivs miteinander verknüpft werden, auf vier verhältnismäßig einfache und leicht erklärbare Typen zurückführe, sage ich noch nichts darüber, wie enorm verwickelt die Häufung und Kombination der Grundaufgaben innerhalb eines einzelnen praktischen Problems liegen kann. Bedenken Sie nur, daß alle Algebra bis in ihre höchsten Spitzen hinauf nur aus Gedankengängen besteht, die in letzter Linie auf die vier sogenannten Grundrechnungsarten zurückführbar sind. Wer das Addieren, Subtrahieren, Multiplizieren und Dividieren versteht, besitzt grundsätzlich den Schlüssel zum Verständnis jeder algebraischen Untersuchung. Aber der Schlüssel muß — um im Bilde zu bleiben — dabei so vielfältig gehandhabt werden, daß ohne langwierige Vorbereitung und erhebliche Anstrengung doch nicht viel zu erreichen ist. Ganz ebenso liegen die Verhältnisse in der Wahrscheinlichkeitsrechnung. Ich denke nicht daran, daß ich hier in ein paar Vorträgen von wenigen Stunden Dauer Sie instandsetzen könnte, Probleme zu lösen, die von Bernoulli und Laplace an bis in die neueste Zeit Mathematiker ersten Ranges immer wieder beschäftigt haben. Anderseits würde doch niemand von uns, auch wenn er jedem mathematischen Ehrgeiz und jeder mathematischen Tätigkeit noch so fern steht, gerne auf die Vertrautheit mit den vier Grundrechnungsarten verzichten wollen, und dies sowohl aus praktischen Gründen im Hinblick auf manche nützliche Anwendung des Rechnens im täglichen Leben, wie auch im Interesse der eigenen Geistesbildung. So hoffe ich denn, mit der kurzen Darstellung der vier Grundoperationen der Wahrscheinlichkeitsrechnung zweierlei zu erreichen: Ihnen die Lösung der einen oder anderen gelegentlich auftretenden einfachen Aufgabe zu ermöglichen, vor allem aber eine befriedigende Aufklärung über die jeden Gebildeten angehende Frage nach dem eigentlichen Sinn der Wahrscheinlichkeitsrechnung zu geben.

Erste Grundoperation: Die Auswahl.

Die erste der vier Grundoperationen, durch die ein neues Kollektiv aus einem anderen gebildet wird, bezeichnen wir als die Auswahl. Gegeben sei z. B. das Kollektiv, das aus sämtlichen Würfen mit einem bestimmten Würfel besteht, oder aus der Folge aller Spiele, die an einem bestimmten Roulettetisch stattfinden. Das neue, abgeleitete Kollektiv habe zu Elementen etwa den ersten, vierten, siebenten .... Wurf mit demselben Würfel oder das zweite, vierte, achte, sechzehnte .... Spiel am Roulettetisch. Allgemein gesprochen: es wird eine Auswahl aus den Elementen der Folge gebildet. Das Merkmal der ausgewählten Elemente im neuen Kollektiv soll das gleiche bleiben wie im ursprünglichen, also im Falle des Würfels die obenauf erscheinende Augenzahl, bei der Roulette die Farbe rot oder schwarz. Gefragt ist nach den Wahrscheinlichkeiten innerhalb des neuen Kollektivs, also z. B. nach der Wahrscheinlichkeit der Farbe rot innerhalb derjenigen Gesamtheit von Spielen, die, wenn man alle Spiele fortlaufend numeriert, zu Ordnungsnummern eine Potenz von 2 haben. Nach dem, was früher über die Eigenschaften eines Kollektivs gesagt wurde, insbesondere über die Regellosigkeit, kann kein Zweifel darüber bestehen, wie die Frage zu beantworten ist: Die Wahrscheinlichkeit bleibt beim Übergang vom alten Kollektiv zu dem durch das Verfahren der „Auswahl" gebildeten neuen unverändert. Alle sechs Wahrscheinlichkeiten der Augenzahlen eins bis sechs sind in der neuen Spielfolge die gleichen wie im Ausgangskolletiv. Genau dies war es ja, was bei der Definition hinsichtlich der Regellosigkeit gefordert wurde. Die Sache ist hier so einfach, daß alle Erklärungen fast überflüssig erscheinen. Wir wollen gleichwohl das Ergebnis in einem Satze genau formulieren:

Man kann aus einem gegebenen Kollektiv in vielfacher Weise durch „Auswahl" ein neues ableiten; die Elemente des neuen Kollektivs sind eine durch Stellenauswahl gewonnene Teilfolge der Elemente des gegebenen, die Merkmale sind unverändert.

Lösung der Aufgabe: Die Verteilung im neuen Kollektiv ist gleich der im alten.

Zweite Grundoperation: Die Mischung.

Kaum weniger einfach, vielleicht noch geläufiger ist das zweite Verfahren, ein neues Kollektiv aus einem gegebenen zu bilden, die zweite Grundoperation, für die ich die Bezeichnung „Mischung" wähle. Zunächst ein Beispiel: Wenn wir als Ausgangskollektiv wieder das Spiel mit einem Würfel nehmen, bei dem also die Elemente aus den aufeinanderfolgenden Würfen, die Merkmale aus den sechs verschiedenen Augenzahlen bestehen, die bei einem Wurf sichtbar werden können, so mag jetzt die Frage gestellt werden: Wie groß ist die Wahrscheinlichkeit, eine gerade Zahl zu werfen? Man kennt die Antwort auf diese Frage. Da es unter den Zahlen 1 bis 6 die drei geraden Zahlen 2, 4 und 6 gibt, müssen die drei Wahrscheinlichkeiten für das Auftreten der Zwei, der Vier und der Sechs addiert werden; die Summe der drei Wahrscheinlichkeiten liefert dann die gesuchte Wahrscheinlichkeit für das Auftreten einer geraden Zahl. Daß diese Rechenregel richtig ist, sieht jedermann leicht ein, auch ist es nicht schwer, sie aus der Definition der Wahrscheinlichkeiten als der Grenzwerte von relativen Häufigkeiten herzuleiten. Der allgemeine Gedanke, der dem Verfahren zugrunde liegt, läßt sich ganz leicht herausschälen. Wir haben aus dem ursprünglichen Kollektiv ein neues gebildet, das dieselben Elemente umfaßt, nämlich alle einzelnen Würfe, in dem aber den Elementen andere Merkmale zugeschrieben werden. Während vorher das Merkmal eines Wurfes eine der Zahlen 1 bis 6 war, lautet das neue Merkmal „gerad" oder „ungerad". Dabei ist es wesentlich, daß mehrere alte Merkmale durch ein neues ersetzt wurden, und nicht etwa mehrere neue an Stelle eines alten treten. In dem letzteren Falle wäre eine Berechnung der neuen Wahrscheinlichkeiten aus den gegebenen nicht möglich.

Mit der Bezeichnung „Mischung“ soll zum Ausdruck gebracht werden, daß mehrere Merkmale des Ausgangskollektivs miteinander vermengt, vermischt werden oder daß eine Vermengung, Vermischung der Elemente, die gewisse Merkmale aufweisen, stattfindet.

Ungenaue Fassung der Additionsregel.

In der älteren, Ihnen vielleicht aus dem Schulunterricht geläufigen Ausdrucksweise spricht man von einer Wahrscheinlichkeit des „Entweder-Oder“ und formuliert einen Satz über die Berechnung unbekannter Wahrscheinlichkeiten aus gegebenen etwa so: Die Wahrscheinlichkeit, entweder Zwei oder Vier oder Sechs zu würfeln, ist gleich der Summe aus den drei Wahrscheinlichkeiten der einzelnen Fälle. Diese Fassung ist aber ungenau und sie bleibt auch noch ganz unzulänglich, wenn man selbst ausdrücklich hinzufügt, daß in dieser Weise nur Wahrscheinlichkeiten von einander ausschließenden Ereignissen zu addieren sind. Die Wahrscheinlichkeit dafür, zwischen dem 40. und 41. Geburtstag zu sterben, mag 0,011 betragen, die Wahrscheinlichkeit, zwischen dem 41. und 42. Geburtstag eine Ehe einzugehen, sei 0,009. Die beiden Möglichkeiten schließen einander aus, trotzdem wäre es sinnlos, zu behaupten, die Wahrscheinlichkeit dafür, entweder zu sterben oder im nächsten Jahre zu heiraten, sei für einen Vierzigjährigen gleich der Summe $0{,}011 + 0{,}009 = 0{,}020$.

Die Aufklärung und richtige Formulierung des Satzes über die bei der „Mischung“ entstehende neue Verteilung findet man nur durch Zurückgehen auf den Begriff des Kollektivs. Der Unterschied zwischen einer richtigen und einer falschen Anwendung der Additionsregel liegt darin, daß immer nur Wahrscheinlichkeiten, die innerhalb eines und desselben Kollektivs für verschiedene Merkmale bestehen, addiert werden dürfen. Die Mischung besteht eben darin, daß Merkmale der Elemente eines Kollektivs vermischt werden. In dem eben erwähnten Beispiel handelt es sich aber um zwei ganz verschiedene Kollektivs. Das eine wird gebildet von der Gesamtheit der Vierzigjährigen, die nach dem Merkmale: Eintritt oder Nichteintritt des Todes vor Vollendung des 41. Lebensjahres eingeteilt werden. Das zweite Kollektiv hat zu Elementen die Einundvierzigjährigen und teilt sie ein nach dem Merkmal: Vollzug oder Nichtvollzug einer Eheschließung im Laufe des Jahres. Beide Kollektivs sind einfache Alternativen und die einzigen Möglichkeiten einer Mischung oder einer Addition von Wahrscheinlichkeiten wären die, ganz innerhalb des einen oder des anderen Kollektivs bleibend, die beiden Wahrscheinlichkeiten für Sterben und Überleben oder die für Eheschließung und Ledigbleiben zu addieren, wobei dann jedesmal 1 als Summe erschiene. Aber kreuzweise zu addieren ist man in keiner Weise berechtigt.

Ein anderes Beispiel, in dem die Unzulänglichkeiten des üblichen Entweder-Oder-Satzes unmittelbar zutage tritt, ist folgendes. Ein guter Tennisspieler mag 80% Wahrscheinlichkeit für sich haben, in einem bestimmten Turnier, sagen wir in Berlin, zu siegen. Seine Aussichten, in einem Turnier, das am gleichen Tag in New York beginnt, seien 70%. Die Möglichkeiten, hier und dort zu siegen, schließen einander aus. Aber für die Wahrscheinlichkeit, entweder hier oder dort zu siegen, die Summe 1,50 anzugeben, wäre völliger Unsinn. Die Aufklärung liegt, wie in dem früheren Beispiel, darin, daß es sich um zwei verschiedene Kollektivs handelt, die Addition aber nur für Wahrscheinlichkeiten *innerhalb eines Kollektivs* zulässig ist.

Fall der Gleichverteilung.

Eine sehr spezielle Form der Mischung, die allgemein geläufig ist, wird oft an die Spitze der ganzen Wahrscheinlichkeitsrechnung gestellt, ja man meint vielfach, daß in ihr die Grundlage *jeder* rechnerischen Wahrscheinlichkeitsbestimmung zu suchen sei. In unserem ersten Beispiel, in dem nach der Wahrscheinlichkeit, eine gerade Zahl zu werfen, gefragt wurde, ist es natürlich gleichgültig, wie groß die drei Wahrscheinlichkeiten der Zwei, der Vier und der Sechs sind. Es kann ein „richtiger“ Würfel in Betracht kommen, bei dem jede der sechs Seiten die Wahrscheinlichkeit $1/6$ besitzt, dann ist die Summe $1/6 + 1/6 + 1/6$ zu bilden, die $1/2$ gibt. Aber ebensogut könnte es sich um den von uns schon früher benutzten „falschen“ Würfel handeln, bei dem durch Versuche andere Werte für die drei Wahrscheinlichkeiten festgestellt wurden. Die Regel, daß die Wahrscheinlichkeit einer geraden Augenzahl gleich der Summe jener drei Größen ist, bleibt unverändert. Der besondere Fall nun, über den wir uns jetzt einen Augenblick unterhalten müssen, ist gerade der der „Gleichverteilung“ oder des „richtigen“

Würfels. Hier kann man, um zu dem richtigen Ergebnis zu gelangen, auch ein klein wenig anders schließen als wir es vorhin getan haben.

Man kann von der Überlegung ausgehen, es seien im ganzen sechs verschiedene Ergebnisse eines Wurfes möglich, nämlich die Augenzahlen 1 bis 6; jedes dieser sechs Ergebnisse sei gleich wahrscheinlich und — nun kommt die kleine Wendung, die wir dem Gedankengang geben — von den sechs Möglichkeiten seien drei dem ins Auge gefaßten Ereignis, nämlich dem Erscheinen einer geraden Zahl, „günstig". Demnach sei der Quotient $3:6 = {}^1/_2$ die gesuchte Wahrscheinlichkeit für das Erscheinen einer geraden Zahl, das ist, wie man sich auszudrücken pflegt, der Quotient aus der Zahl der „dem Ereignis günstigen" Möglichkeiten durch die Gesamtzahl aller Möglichkeiten. Man sieht, daß hier in der Tat eine allgemeine Regel aufgestellt werden kann, die alle jene Fälle der Mischung umfaßt, in denen sämtliche Merkmale innerhalb des Ausgangskollektivs gleiche Wahrscheinlichkeit besitzen. Bezeichnen wir etwa die Zahl dieser Merkmale mit dem Buchstaben n, so daß die Wahrscheinlichkeit jedes einzelnen Merkmales ${}^1/_n$ ist, und ebenso die Zahl derjenigen unter diesen Merkmalen, die im abgeleiteten Kollektiv zu einem einzigen neuen Merkmal zusammengefaßt werden, mit dem Buchstaben m, so hat man nach der Additionsregel, um die Wahrscheinlichkeit des neuen Merkmales (innerhalb des neuen Kollektivs) zu erhalten, eine Summe aus m Summanden, deren jeder die Größe ${}^1/_n$ hat, zu bilden. Dafür kann man nun einfacher sagen, die gesuchte Wahrscheinlichkeit sei $m:n$, gleich dem Quotienten aus der Zahl der „günstigen" Merkmale durch die Gesamtzahl aller Merkmale. Wir werden später darauf zurückkommen, welche mißbräuchliche Anwendung man von dieser Rechenregel gemacht hat, um darauf eine Scheindefinition der Wahrscheinlichkeit zu stützen. Für jetzt kommt es nur darauf an, einzusehen, daß das Aufsuchen von Wahrscheinlichkeiten durch Abzählen gleichwahrscheinlicher, günstiger und ungünstiger Möglichkeiten nur einen sehr spezialiellen Fall der Ableitung eines neuen Kollektivs aus einem gegebenen durch Mischung darstellt.

In meinem ersten Vortrag habe ich schon einmal von dieser speziellen Form der Mischungsregel stillschweigend Gebrauch gemacht. Es war dort die Rede von zwei Kollektivs, deren Elemente

Lotterieziehungen waren. Das eine Mal bildeten die gezogenen Losnummern die Merkmale, das andere Mal waren alle jene Nummern gemischt, die fünf Endnullen besitzen. Es gibt deren 10 unter den Nummern 1 bis 1 Million, also ist unter der Voraussetzung der Gleichverteilung im ursprünglichen Kollektiv, die Wahrscheinlichkeit der Ziehung einer runden Zahl der genannten Art gleich 10 Millionstel oder 0,00001.

Zusammenfassung und Ergänzung der Mischungsregel.

Ich will jetzt die Mischungsregel, so wie sie sich uns auf der Grundlage des Kollektivbegriffes allgemein ergeben hat, kurz in einem Satz formulieren:

Aus einem gegebenen Kollektiv, das Elemente mit mehr als zwei Merkmalen umfaßt, läßt sich in vielfacher Weise durch „Mischung" ein neues Kollektiv ableiten; *seine Elemente sind die gleichen wie die des Ausgangskollektivs*, die neuen Merkmale aber sind *Vermengungen* (Mischungen) *von Merkmalen des alten Kollektivs* (z. B. die geraden oder ungeraden Zahlen zusammen, statt der Einzelzahlen).

Lösung der Aufgabe: *Die Verteilung innerhalb des neuen Kollektivs ergibt sich aus der gegebenen, indem man die Wahrscheinlichkeiten aller alten Merkmale, die zu einem neuen zusammengefaßt sind, addiert.*

Wie diese Additionsvorschrift in einfachen Fällen anzuwenden ist, haben wir an Beispielen schon gesehen. Ich will nur noch nebenbei erwähnen, daß sich die Regel sinngemäß ausdehnen läßt auf jene, früher schon erwähnten Kollektivs, in denen es nicht nur endlich viele Merkmale gibt, sondern, wie im Beispiel des Scheibenschiebens, unendlich viele. Die höhere Mathematik lehrt, daß an Stelle des Addierens einzelner Summanden in solchen Fällen eine ähnliche, freilich etwas schwerer zu erklärende Operation tritt, die man das *Integrieren* nennt. Ist in dem Ausgangskollektiv die Wahrscheinlichkeitsdichte für jeden Wert des Abstandes vom Mittelpunkt gegeben, so findet man als Resultat einer bestimmten Mischung die Wahrscheinlichkeit dafür, daß der Abstand zwischen 0,5 m und 1,0 m liegt, der Schuß also in die äußere Hälfte der Scheibe fällt, als das *Integral* über die gegebene Dichtefunktion, erstreckt vom Werte 0,5 bis zum Wert

1,0. Diese Andeutungen werden dem in die Begriffsbildungen und die Ausdrucksweise der Analysis Eingeweihten genügen, alle anderen Zuhörer mögen überzeugt sein, daß diese Dinge wohl für das tatsächliche Lösen einzelner Aufgaben der Wahrscheinlichkeitsrechnung nützlich, ja unentbehrlich sind, für das Verständnis der allgemeinen, von uns zu erörternden Fragen aber ohne Ausschlag gebende Bedeutung.

Dritte Grundoperation: Die Teilung.

Wir haben zwei von den vier Grundoperationen der Bildung neuer Kollektivs, die Auswahl und die Mischung, kennen gelernt und kommen jetzt zur dritten, die ich die „Teilung“ nenne. Der Ausdruck wird bald verständlich werden, nebenbei soll er auch daran erinnern, daß wir es hier mit einer Division von Wahrscheinlichkeiten zu tun haben. Um die Einführung in den Begriff der „Teilung“ möglichst zu erleichtern, verwende ich wieder das gleiche Ausgangskollektiv, das uns schon in den beiden früheren Fällen, der Auswahl und der Mischung, gedient hat. Es besteht aus den fortgesetzten Würfen mit einem Würfel als Elementen und den jeweils auf der Oberseite des Würfels erscheinenden Augenzahlen als Merkmalen. Gegeben sind wieder die sechs Wahrscheinlichkeiten (mit der Summe 1) für das Auftreten der Zahlen 1 bis 6, ohne daß etwa vorausgesetzt würde, jede dieser Wahrscheinlichkeiten habe den Wert $^1/_6$. Die neue Frage, die jetzt gestellt wird und deren Beantwortung eben zur Operation der „Teilung“ führt, lautet folgendermaßen: Wie groß ist die Wahrscheinlichkeit dafür, daß ein Wurf, von dem schon bekannt ist, daß sein Ergebnis eine gerade Zahl ist, die Augenzahl 2 aufweist? Vielleicht macht diese Frage einen etwas gekünstelten Eindruck, dann kann ich sie sofort auf eine Form bringen, wie sie im täglichen Leben leicht auftritt.

Wir warten auf eine Straßenbahn an einer Haltestelle, an der die Züge von sechs verschiedenen Linien vorbeigehen; darunter seien drei Linien, die immer mit Doppelwagen befahren werden, und drei, auf denen nur Einzelwagen verkehren. Die letzteren Linien mögen die Nummern 1, 3, 5, die ersteren die Nummern 2, 4, 6 tragen. Sobald wir nun in der Ferne sehen, daß ein aus zwei Wagen bestehender Zug herannaht, wie groß ist da die Wahrscheinlichkeit dafür, daß es ein Zug einer bestimmten der drei mit

Doppelwagen befahrenen Linien ist, etwa der mit 2 bezeichneten Linie? Natürlich muß die Ausgangsverteilung gegeben sein, die hier aus den sechs Einzelwahrscheinlichkeiten (also praktisch aus den relativen Häufigkeiten), mit denen die sechs Linien befahren werden, besteht. Sind alle diese sechs Wahrscheinlichkeiten gleich, somit vom Betrage $^1/_6$, entsprechend dem Fall des „richtigen" Würfels, so wird man nicht mit der Antwort in Verlegenheit sein, daß die gesuchte Wahrscheinlichkeit den Wert $^1/_3$ besitzt. Man kann dabei etwa so argumentieren: es kommen im ganzen drei verschiedene, gleich wahrscheinliche Möglichkeiten in Betracht, von denen eine allein die „günstige" ist, also haben wir nach einer früheren Regel die Wahrscheinlichkeit $^1/_3$. Allein diese Schlußweise ist nicht in allen Fällen anwendbar, sie versagt sofort, wenn die Wahrscheinlichkeit der verschiedenen Straßenbahnlinien (oder Würfelseiten) nicht von vornherein gleich sind. Zu einer vollständigen Lösung und einer genügend allgemeinen Fassung der Aufgabe gelangen wir nur, indem wir genau untersuchen, worin eigentlich das neue, das abgeleitete Kollektiv besteht. Denn darüber sind wir uns schon klar, daß die Frage nach einer Wahrscheinlichkeit allein dann eine sinnvolle ist, wenn vorerst ein Kollektiv genau umgrenzt ist, innerhalb dessen die Wahrscheinlichkeit zu nehmen ist.

Kehren wir, der einfacheren Ausdrucksweise halber, wieder zu der Würfelaufgabe zurück, so können wir das abgeleitete Kollektiv so beschreiben: Elemente der betrachteten Folge sind nicht mehr alle Würfe mit dem Würfel, sondern nur diejenigen, deren Merkmal eine gerade Augenzahl ist; das Merkmal innerhalb des neuen Kollektivs ist das alte geblieben, die Augenzahl, nur daß jetzt nicht mehr alle Augenzahlen als Merkmale vorkommen, sondern allein die 2, 4 und 6. Wir sagen, es habe eine „Teilung" des ursprünglich gegebenen Kollektivs stattgefunden, indem die Elemente in zwei Klassen geschieden wurden. Die eine Klasse, gekennzeichnet dadurch, daß die Merkmale der Elemente gerade Zahlen sind, bildet die Elemente des abgeleiteten Kollektivs. Es ist wichtig, sich klarzumachen, daß diese „Teilung" etwas ganz anderes ist als die früher behandelte Stellenauswahl; bei einer solchen wird aus der Gesamtfolge aller Würfe eine Teilfolge durch Festlegung gewisser Ordnungsnummern der Würfe herausgehoben, ohne daß man die Merkmale, die Augenzahlen der einzelnen Würfe, dazu benutzt, um über die Zugehörigkeit zur Teilfolge zu entscheiden. Wir haben auch gesehen, daß innerhalb einer so entstandenen Teilfolge die Wahrscheinlichkeiten die gleichen sind wie innerhalb des Ausgangskollektivs. Jetzt, wo wir es mit der Teilung zu tun haben, geschieht die Wahl der zurückzubehaltenden Elemente ausdrücklich vom Gesichtspunkte der Merkmale aus und die Wahrscheinlichkeiten werden dadurch wesentlich geändert; in welcher Weise, das wollen wir gleich untersuchen.

Die Wahrscheinlichkeit nach der Teilung.

Nehmen wir an, die Wahrscheinlichkeiten innerhalb des gegebenen Kollektivs seien der Reihe nach für das Auftreten der Ergebnisse 1 bis 6 gleich 0,10, 0,20, 0,15, 0,25, 0,10 und 0,20. Die Summe dieser sechs Zahlen ist 1, im übrigen ist es gleichgültig, ob wir jetzt an das Würfelbeispiel oder die Straßenbahn denken, und ebenso gleichgültig, in welcher Weise die sechs Werte ermittelt worden sind. Addieren wir die Wahrscheinlichkeiten der Zwei, der Vier und der Sechs, also die Zahlen 0,20, 0,25 und 0,20, so erhalten wir die Summe 0,65 als Wahrscheinlichkeit für das Auftreten einer geraden Zahl (zweite Grundoperation, Mischung, Additionsregel). Nach der Bedeutung des Wahrscheinlichkeitsbegriffes haben also die geradzahligen Ergebnisse bei genügend lang fortgesetzter Beobachtung 65 % Anteil an der Gesamtheit der Elemente. Unter den ersten 10000 Elementen befinden sich somit rund 6500, deren Merkmale gerade Zahlen sind. Zu diesen Elementen gehören rund 2000 solche, die das Merkmal 2 aufweisen, da ja 0,20 die Häufigkeit der Zwei in einer genügend langen Beobachtungsreihe ist. Bilden wir nun ein neues Kollektiv, das aus dem früheren durch Ausscheiden aller Elemente mit ungeraden Merkmalen entsteht, also lediglich die Elemente mit geradzahligem Merkmal enthält, so befinden sich unter den ersten 6500 von ihnen 2000 Elemente mit dem Merkmal 2, die relative Häufigkeit dieser ist also $2000 : 6500 = 0{,}308$. Da die Rechnung, die wir hier durchgeführt haben, eigentlich erst für unendlich große Beobachtungsreihen genau gilt, haben wir in der Zahl 0,308 schon den richtigen Grenzwert der relativen Häufigkeit oder die Wahrscheinlichkeit für das Auftreten des Merkmales 2 innerhalb des neuen Kollektivs. Die allgemeine Regel, nach der diese Lösung der Aufgabe zu finden ist, läßt sich aus dem vorgeführten Beispiel leicht abziehen.

Man muß zunächst aus den gegebenen Wahrscheinlichkeiten die Summe derjenigen bilden, die den ausgewählten, bei der Teilung zurückbehaltenen Merkmalen (im Beispiel die Zwei, Vier und Sechs) entsprechen und sodann die ursprüngliche Wahrscheinlichkeit des in Frage kommenden Merkmales (im Beispiel ist es die Zwei) durch jene Summe dividieren, $0{,}20 : 0{,}65 = 0{,}308$. Es liegt also in der Tat ein Divisionsgesetz der Wahrscheinlichkeit vor.

Ausgangs- und Endwahrscheinlichkeit eines Merkmales.

Man hat das Bedürfnis, die beiden Wahrscheinlichkeiten eines und desselben Merkmales, mit denen man es hier zu tun hat, die gegebene, innerhalb des Ausgangskollektivs bestehende und die berechnete, die innerhalb des durch „Teilung“ abgeleiteten Kollektivs gilt, durch besondere Bezeichnungen zu unterscheiden. Die hierfür üblichen Namen erscheinen mir aber trotz ihrer unleugbaren Anschaulichkeit ein wenig unglücklich gewählt. Man pflegt nämlich von der „a priori“-Wahrscheinlichkeit und der „a posteriori“-Wahrscheinlichkeit zu sprechen, wenn man einmal den gegebenen und dann den abgeleiteten Wert meint. Schon der Anklang der Bezeichnung an eine gewisse allgemein-philosophische Terminologie, deren Berechtigung nicht unbestritten ist, hat seine Nachteile. Es kommt noch hinzu, daß man in der klassischen Wahrscheinlichkeitsrechnung die Ausdrücke a priori und a posteriori auch in anderem Sinne verwendet, nämlich für die — in unserer Auffassung belanglose — Unterscheidung zwischen solchen Ausgangswahrscheinlichkeiten, die empirisch ermittelt wurden, und solchen, die auf Grund irgendwelcher Hypothesen angenommen sind. Es wird daher besser sein, wenn wir etwas bescheidenere, nicht durch zu weitgehende Allgemeinbedeutung belegte Namen wählen und einfach von der Ausgangswahrscheinlichkeit und der abgeleiteten oder Endwahrscheinlichkeit sprechen, wenn wir die Wahrscheinlichkeit für das Auftreten des in Betracht gezogenen Merkmales, einmal innerhalb des Ausgangskollektivs, dann innerhalb des abgeleiteten meinen. In unserem Zahlenbeispiel besitzt das Merkmal „Zwei“ (also die mit zwei Augen versehene Würfelseite oder die zweite der fraglichen Straßenbahnlinien) die Ausgangswahrscheinlichkeit 0,20 und die Endwahrscheinlichkeit $0{,}20 : 0{,}65 = 0{,}308$. Dies heißt nichts anderes als: Die Zwei besitzt die Wahrscheinlichkeit 0,20 innerhalb aller Elemente und die Wahrscheinlichkeit 0,308 innerhalb der Gesamtheit der Elemente mit gerader Nummer.

Sogenannte Wahrscheinlichkeit von Ursachen.

Eine andere, ebenfalls unberechtigte und verwirrende Terminologie, die oft in Zusammenhang mit der Aufgabe der Teilung verwandt wird, kann ich nicht ganz unerwähnt lassen. In dem Würfelbeispiel, in dem nach der Wahrscheinlichkeit der 2 unter den geraden Zahlen 2, 4 und 6 gefragt wird, pflegt man sich oft so auszudrücken: das Erscheinen einer geraden Zahl kann drei verschiedene „Ursachen" haben oder kann drei verschiedenen „Hypothesen" zugeschrieben werden; als je eine Ursache oder Hypothese sieht man das Auftreten je einer der drei geraden Zahlen 2, 4, 6 an. Demnach heißt die von uns berechnete Endwahrscheinlichkeit 0,308 auch die Wahrscheinlichkeit dafür, daß das Ergebnis „gerade Zahl" dem Erscheinen der Ursache „Zwei" zuzuschreiben ist. Auf diese Weise entsteht scheinbar ein ganz besonderes Kapitel der Wahrscheinlichkeitstheorie, das im Gegensatz zu der sonst in Frage kommenden Wahrscheinlichkeit von „Ereignissen" die Wahrscheinlichkeit der „Ursachen von Ereignissen" oder gar von „Hypothesen über die Ereignisse" zu berechnen lehrt. Die Teilungsaufgabe wird dabei gewöhnlich in folgender Form gestellt.

Es seien drei Urnen, die mit schwarzen und weißen Kugeln gefüllt sind, auf dem Tisch aufgestellt. Das Element des Ausgangskollektivs ist ein zweifacher Beobachtungsvorgang: Man greift willkürlich oder blindlings nach einer Urne, zieht eine in ihr liegende Kugel hervor und stellt als Merkmal fest: erstens die Farbe der Kugel, zweitens die Nummer der Urne. Im ganzen gibt es also sechs verschiedene Merkmale, nämlich Farbe weiß und Nummer 1, Farbe weiß und Nummer 2, Farbe weiß und Nummer 3, dann Farbe schwarz mit Nummer 1 usw. Die entsprechenden sechs Ausgangswahrscheinlichkeiten seien gegeben. Ist nun bekannt, daß eine Beobachtung eine weiße Kugel ergeben hat, aber nicht bekannt, aus welcher Urne sie gezogen wurde, so kann man die Frage nach der Wahrscheinlichkeit dafür stellen, daß der Zug der weißen Kugel gerade aus der ersten Urne erfolgt ist, oder daß das Erscheinen der weißen Kugel den Umstand zur „Ursache" hat, daß die erste Urne von der ziehenden Hand getroffen wurde. Die Lösung ist ganz analog der früheren, man muß die Ausgangswahrscheinlichkeit für „Farbe weiß, Nummer 1" durch die Summe der drei Wahrscheinlichkeiten für „Farbe weiß, Nummer 1, 2 oder 3" dividieren. Die metaphysische Ausdrucksweise ist nur historisch dadurch zu erklären, daß die Teilungsregel, um die Mitte des 18. Jahrhunderts von dem Engländer BAYES in diesem Sinne abgeleitet wurde und seither in fast unveränderter Form in allen Lehrbüchern wiederholt wird.

Formulierung der Teilungsregel.

Wir stellen aber hier ausdrücklich fest, daß es für uns keinen Unterschied zwischen Wahrscheinlichkeit „von Ursachen oder Hypothesen" und Wahrscheinlichkeit „von Ereignissen" gibt. Den einzigen Gegenstand, mit dem wir uns beschäftigen, bilden Kollektivs, d. s. Merkmalfolgen, in denen Grenzwerte der relativen Häufigkeiten bestehen und Regellosigkeit der Anordnung herrscht. An jedem Kollektiv, was immer seine Elemente seien, wenn es nur mindestens drei verschiedene Merkmale enthält, kann man eine Teilung vornehmen. Dann gibt es für jedes Merkmal, das bei der Teilung nicht ausgeschieden, sondern zurückbehalten wurde, zwei verschiedene Werte der Wahrscheinlichkeit: eine Ausgangswahrscheinlichkeit im ursprünglichen Kollektiv, eine Endwahrscheinlichkeit im abgeleiteten. Keinerlei Raum ist hier vorhanden für die Einschiebung metaphysischer Ausdrücke. Wir wollen, wie in den früheren Fällen, bevor wir zur vierten und letzten Grundoperation übergehen, die Definition der Teilung und die Lösung der Teilungsaufgabe kurz zusammenfassen:

Aus einem gegebenen Kollektiv mit mehr als zwei Merkmalen läßt sich in vielfacher Weise ein neues Kollektiv durch „Teilung" ableiten; Elemente des neuen Kollektivs sind diejenigen Elemente des alten, deren Merkmale einer gewissen Teilmenge der ursprünglichen Merkmalmenge angehören; das Merkmal eines jeden beibehaltenen Elements bleibt ungeändert.

Lösung der Aufgabe: Die neue Verteilung erhält man, indem man die einzelnen gegebenen Wahrscheinlichkeiten der beibehaltenen Merkmale durch die Summe der Wahrscheinlichkeiten aller dieser Merkmale dividiert.

Vierte Grundoperation: Die Verbindung.

Alle drei bisher erörterten Grundoperationen, die Auswahl, die Mischung und die Teilung, hatten das Gemeinsame, daß jedesmal aus einem gegebenen Kollektiv durch eine bestimmte Vorschrift, die sich auf die Verfügung über Elemente und Merkmale erstreckte, ein neues Kollektiv abgeleitet wurde. Nun werden wir in dem letzten der zu besprechenden vier Verfahren ein solches kennenlernen, das aus zwei gegebenen Ausgangskollektivs in bestimmter Weise ein neues bildet. Zugleich werden wir auch zum erstenmal Einblick in die verschiedenen Formen der Beziehungen zwischen zwei oder mehreren Kollektivs gewinnen. Als Bezeichnung der vierten Grundoperation wähle ich das Wort „Verbindung“. Das Beispiel, an das wir die Untersuchung der Verbindung anknüpfen, soll möglichst in Anlehnung an die früheren Würfelaufgaben gebildet werden. Die beiden Ausgangskollektivs mögen je aus der Folge von Würfen bestehen, die mit einem Würfel in der üblichen Weise ausgeführt werden, so daß als Merkmale innerhalb eines jeden dieser Kollektivs die Augenzahlen 1 bis 6 mit gewissen Wahrscheinlichkeiten, die nicht in beiden Fällen gleich sein müssen, auftreten. Das Endkollektiv, das Resultat der Verbindung, hat zum Element einen Wurf mit beiden Würfeln, zum Merkmal die beiden Augenzahlen, die auf den Oberseiten erscheinen. Gefragt ist nach der Wahrscheinlichkeit, ein bestimmtes Zahlenpaar, z. B. „3“ mit dem ersten, „5“ mit dem zweiten Würfel, zu werfen. Die beiden Würfel stellen wir uns dabei wohl unterscheidbar vor, sie tragen die Nummern I und II, sind verschieden gefärbt oder besitzen sonstige Kennzeichen.

Wer in der Schule nur die ersten Anfangsgründe der Wahrscheinlichkeitsrechnung gelernt oder wer sich die Frage einmal selbst ein wenig überlegt hat, der weiß, wie man in primitiver Weise das hier aufgeworfene Problem löst. Man sagt, es handle sich um eine Wahrscheinlichkeit des „Sowohl als auch“ und man schreibt als Regel vor, die beiden gegebenen Wahrscheinlichkeiten miteinander zu multiplizieren. Ist etwa $^1/_7$ die Wahrscheinlichkeit, mit dem ersten Würfel „3“ zu werfen, $^1/_6$ die, mit dem zweiten „5“ zu erzielen, so ist $^1/_7$ mal $^1/_6$ gleich $^1/_{42}$ die Wahrscheinlichkeit des Doppelwurfes „3, 5“. Die Regel ist recht anschaulich, wenn man sich vorstellt, daß erst $^1/_7$ aller Würfe herausgehoben werden muß, damit der erste Würfel die „3" ergibt, und daß dann von diesem Siebentel nochmal $^1/_6$ zu nehmen ist, damit der zweite Würfel „5" zeigt. Aber es ist auch klar, daß diese Produktregel, ähnlich wie die Additionsregel des „Entweder-Oder", einer genaueren Formulierung und Begründung bedarf, wenn sie einwandfrei sein soll. So ist z. B. die Wahrscheinlichkeit dafür, mit zwei Würfeln sowohl die Summe „8" als auch die Differenz „2" zu würfeln, sicher nicht als das Produkt der betreffenden Einzelwahrscheinlichkeiten zu finden. Ich wende mich jetzt diesen etwas feineren Untersuchungen zu und will, um nicht allzu weitschweifig zu werden, mich hierbei vorübergehend der einfachsten Elemente der Buchstabenrechnung bedienen; ich fürchte nicht, hierdurch allzu schwer verständlich zu werden.

Auswürfelung ein neues Verfahren zur Bildung von Teilfolgen.

Es wird jetzt ein Spiel betrachtet, bei dem gleichzeitig zwei Würfel I und II ausgespielt werden. In welcher Weise dies geschieht, ob beide Würfel in dem gleichen Becher liegen, ist gleichgültig. Es kommt nur darauf an, daß in eindeutiger Weise jedem einzelnen Spielresultat an I ein korrespondierendes Resultat am Würfel II zugeordnet wird. Wenden wir unser Augenmerk zunächst nur auf die Augenzahl, die der erste der beiden Würfel zeigt, so werden wir unter den n ersten Würfen eine gewisse Anzahl, etwa n_3 solche feststellen, bei denen hier tatsächlich die Zahl „3" erscheint. Das Verhältnis $n_3:n$ ist dann die relative Häufigkeit für das Auftreten der „3" beim ersten Würfel, und der Grenzwert von $n_3:n$ bei unbeschränkt fortgesetztem Würfeln ist die gegebene Wahrscheinlichkeit der Augenzahl „3" beim Würfel I.

Nun wollen wir weiterhin, jedesmal wenn der erste Würfel „3" zeigt, die Augenzahl betrachten, die gleichzeitig, d. h. bei demselben Wurf, der zweite Würfel aufweist. In irgendwelcher Abwechslung werden die Zahlen 1 bis 6 hierbei auftreten. Ein Teil der n_3 Würfe, die wir jetzt ins Auge fassen, sagen wir etwa n_5 von ihnen, werden gerade die Zahl „5" zeigen, so daß die relative Häufigkeit der Augenzahl „5" beim zweiten Würfel $n_5:n_3$ beträgt. Dies ist aber nicht die Häufigkeit der „5" innerhalb des ursprünglich gegebenen zweiten Würfelkollektivs, das aus allen Spielen mit dem zweiten Würfel besteht, da wir ja jetzt nur eine bestimmte

Teilfolge dieser Spiele in Betracht gezogen haben, nämlich diejenigen, bei denen der erste Würfel auf „3" fiel.

Eine auf diese Art hergestellte Teilfolge der Elemente eines Kollektivs stellt gegenüber dem, was wir bisher kennengelernt hatten, wiederum etwas Neues dar. Es ist weder eine „Stellenauswahl", wie sie zuerst betrachtet wurde, bei der nach einem bestimmten arithmetischen Gesetz, unabhängig von den betreffenden Merkmalen, die Spiele, die zur Teilfolge gehören sollen, ausgewählt werden; noch eine „Teilung", bei der die Elemente, die gewisse ausgewählte Merkmale besaßen, zurückbehalten bzw. ausgeschieden wurden. Wir müssen schon eine neue Bezeichnung einführen, die wir in ganz anschaulicher Weise wählen, indem wir sagen, die Teilfolge des zweiten Kollektivs sei „mittels des ersten ausgewürfelt", genauer „durch das Merkmal 3 des ersten Kollektivs ausgewürfelt". Es ist klar, was damit gemeint ist. Aus einer beliebigen Elementenfolge kann man eine Teilfolge „auswürfeln", indem man ihre Elemente der Reihe nach den Elementen eines bestimmten anderen Kollektivs zuordnet und nur jene Elemente der ursprünglichen Folge heraushebt, deren zugeordnetes Element ein bestimmtes Merkmal aufweist. Mittels des ersten Würfels kann man aus der Folge von Würfen mit dem zweiten Würfel sechs verschiedene Teilfolgen auswürfeln: eine durch das Merkmal „1", eine durch die Augenzahl „2" usf.

Voneinander unabhängige Kollektivs.

Der von uns vorhin betrachtete Quotient $n_5 : n_3$ war die relative Häufigkeit der Augenzahl „5" innerhalb der durch die Zahl „3" des ersten Würfels aus den Ergebnissen des zweiten „ausgewürfelten" Teilfolge. Wir wissen zunächst gar nichts über die Größe dieses Quotienten, höchstens daß er ein positiver echter Bruch ist. Wir wollen annehmen, daß der Quotient bei unbeschränkt ausgedehnter Versuchsfolge einem Grenzwert zustrebt. Aber welches ist dieser Wert? Es kann sein, folgt aber aus keiner unserer bisherigen Überlegungen, daß dieser Grenzwert der gleiche ist, dem sich auch die relative Häufigkeit der „5" innerhalb aller Spiele mit dem zweiten Würfel nähert. Sei für den Augenblick angenommen, daß dies zutrifft. Eine solche Annahme ist leicht verständlich. Sie besagt eigentlich nur, daß die „Auswürfelung" einer Teilfolge auf die ganze Spielfolge mit dem zweiten

Würfel nicht anders wirkt als etwa eine Stellenauswahl. Vermutet etwa jemand, daß die Aussicht, mit dem zweiten Würfel „5" zu werfen, sich dadurch verändert, daß man nur die Spiele in Betracht zieht, bei denen der erste „3" zeigt? Wenn man die Frage einem Skeptiker vorlegt, wird er antworten: Ja, es kommt darauf an, ob die Würfe mit dem zweiten Würfel „unabhängig" von denen mit dem ersten erfolgen. Was heißt aber „unabhängig"?

Man hat gewisse Vorstellungen davon, wann zwei Kollektivs „abhängig" sein werden. Man wird z. B. das Spielen mit zwei Würfeln, die durch einen kurzen Faden miteinander verknüpft sind, nicht als Verbindung zweier unabhängiger Kollektivs bezeichnen wollen. Aber zu einer Definition der Unabhängigkeit im exakt-naturwissenschaftlichen Sinn gelangt man nur durch einen analogen Gedankengang wie den, der uns früher zum Begriff des Kollektivs und der Wahrscheinlichkeit geführt hat. Man muß diejenige Eigenschaft, diejenige Seite der Erscheinung, die für die Durchführung der Theorie die nützlichste und erfolgreichste zu sein verspricht, als konstituierendes Merkmal des Begriffes postulieren. Also sagen wir — vorbehaltlich einer kleinen Ergänzung, die noch folgen wird —: ein Kollektiv II soll von einem anderen Kollektiv I unabhängig heißen, wenn eine Auswürfelung aus II mittels eines beliebigen Merkmales von I die Grenzwerte der relativen Häufigkeiten in II, also seine Verteilung, ungeändert läßt. Setzen wir jetzt die beiden Würfel unseres Spieles in diesem Sinn als unabhängig voraus, so können wir die Aufgabe, die von vornherein gestellt war, die Wahrscheinlichkeit für das Auftreten des Zahlenpaares „3, 5" zu finden, ohne weiteres lösen.

Ableitung der Produktregel.

Wir haben im ganzen n Würfe mit den beiden Würfeln betrachtet, unter denen n_3 die Augenzahl 3 auf dem ersten und n_5 überdies die Augenzahl 5 auf dem zweiten ergeben haben. Es ist somit der Quotient $n_5:n$ die relative Häufigkeit des Auftretens des zusammengesetzten Merkmales 3, 5 bei unserem Würfelpaar. Der Grenzwert von $n_5:n$ ist das, was wir suchen. Nun gilt aber, wie jeder, der die Anfangsgründe des Buchstabenrechnens beherrscht, nachprüfen kann, die Beziehung

$$n_5 : n = \frac{n_5}{n_3} \cdot \frac{n_3}{n}$$

d. h. die relative Häufigkeit $n_5:n$ ist das Produkt der beiden früher betrachteten Häufigkeiten $n_5:n_3$ und $n_3:n$. Die letztere hat zum Grenzwert die Wahrscheinlichkeit, mit dem ersten Würfel 3 zu werfen, die wir v_3 nennen wollen. Der Quotient $n_5:n_3$ soll nach der Annahme der Unabhängigkeit der beiden Kollektivs denselben Grenzwert haben wie die relative Häufigkeit eines Fünferwurfes mit dem zweiten Würfel, wenn diese Häufigkeit in der unbeschränkten Folge aller Würfe des zweiten Würfels betrachtet wird. Dies ist nichts anderes als die Wahrscheinlichkeit, mit dem zweiten Würfel 5 zu werfen. Wir wollen für die Wahrscheinlichkeiten am Würfel II den Buchstaben w wählen, also die Wahrscheinlichkeit, mit II die Augenzahl 5 zu werfen, durch w_5 bezeichnen. Nun ist der Grenzwert der Produkte zweier Größen gleich dem Produkt ihrer beiden Grenzwerte, also der Grenzwert von $n_5:n$ das Produkt von v_3 und w_5, in Worten: Die Wahrscheinlichkeit dafür, mit dem ersten Würfel die Augenzahl „3“, mit dem zweiten zugleich „5‘, zu erzielen, ist das Produkt der beiden Wahrscheinlichkeiten der Einzelereignisse $v_3 \cdot w_5$. Nehmen wir den Buchstaben W für die Wahrscheinlichkeiten am Würfelpaar, so können wir schreiben:

$$W_{3,5} = v_3 \cdot w_5.$$

Natürlich gilt dieser Satz auch sinngemäß für jedes andere Zahlenpaar, das aus den Zahlen 1 bis 6 gebildet werden kann. Z. B. ist die Wahrscheinlichkeit, mit dem ersten Würfel 5 und dazu mit dem zweiten 3 zu treffen, $W_{5,3} = v_5 \cdot w_3$, wo v_5 den Wahrscheinlichkeitswert des Merkmales 5 innerhalb des ersten gegebenen Kollektivs bezeichnet usf.

Wenn man im Rahmen unserer allgemeinen Theorie das Multiplikationsgesetz für die Verbindung unabhängiger Kollektivs formulieren will, bedarf es noch einer Ergänzung. Wir müssen uns davon überzeugen, ob die neue Elementenfolge, bestehend aus dem Spiel mit jedesmal zwei Würfeln und den beiden Augenzahlen als Merkmal, auch den Forderungen, die an ein Kollektiv gestellt werden, genügt. Andernfalls dürfen wir ja gar nicht von einer Wahrscheinlichkeit des Zahlenpaares 3, 5 sprechen. Nun, die eine Forderung, die nach der Existenz der Grenzwerte, ist sicher erfüllt, denn wir haben ja gezeigt, wie man $W_{3,5}$ und jede andere dieser 36 Größen $W_{1,1}$, $W_{1,2}$ usf. bis $W_{6,6}$ berechnen kann. Es bleibt aber noch die Frage nach der Unempfindlichkeit der Grenzwerte gegen eine beliebige Stellenauswahl. Und da ist zu sagen, daß wir in der Tat die Bedingungen der „Unabhängigkeit" noch etwas enger fassen müssen, als dies oben geschehen ist. Man muß die Forderung ausdrücklich stellen, daß der Häufigkeitsgrenzwert für das zweite Kollektiv unverändert bleiben soll, wenn man am ersten zunächst eine beliebige Stellenauswahl vornimmt und dann die so verkürzte erste Folge zur „Auswürfelung" benutzt. Ein Beispiel wird uns nachher gleich zeigen, daß das wirklich nötig ist. Vorläufig geben wir die endgültigen Formulierungen für die Verbindung unabhängiger Kollektivs wie folgt:

1. Ein Kollektiv II heißt von dem Kollektiv I, auf das es elementweise bezogen ist, *unabhängig*, wenn die Verteilung innerhalb II *unverändert bleibt*, sobald aus den Elementen von II eine Teilfolge derart gebildet wird, daß man zuerst *an I eine Stellenauswahl trifft*, dann diese ausgewählte Folge der Elemente von I *zur Auswürfelung aus II benutzt* und schließlich *wieder eine Stellenauswahl vornimmt*.

2. Aus zwei unabhängigen Kollektivs kann als ihre „Verbindung" ein neues hergestellt werden, dessen Elemente und Merkmale *die Zusammenfassungen der Elemente und Merkmale* der ursprünglichen sind.

Lösung der Aufgabe: *Die neue Verteilung ergibt sich, indem man die Wahrscheinlichkeiten der einzelnen Merkmale innerhalb der gegebenen Kollektivs miteinander multipliziert.*

Feststellung der Unabhängigkeit.

Mit diesen Sätzen ist ein viertes, für uns im wesentlichen das letzte, Verfahren zur Bildung eines neuen Kollektivs aus gegebenen definiert. Nur hinsichtlich der Verbindung von Kollektivs, die nicht als unabhängig gelten können, werden wir noch einen Zusatz hinzufügen. Vorerst aber eine kleine Einschaltung.

Man kann, ähnlich wie früher bei der Definition der Wahrscheinlichkeit überhaupt, fragen: Woher weiß man, ob jemals zwei gegebene Kollektivs in dem Sinne der eben gegebenen Definition voneinander unabhängig sind und demnach der Multiplikationssatz angewendet werden darf? Die Antwort darauf lautet: Es liegt hier genau so wie im früheren Fall und wie bei jedem Problem einer exakten Naturwissenschaft, in der man aus abgezogenen, idealisierten Voraussetzungen Schlüsse zieht, die nur an der Erfahrung geprüft werden können. So definiert die Mechanik als einen elastischen Körper einen solchen, bei dem Spannung und Formänderung an jeder Stelle einander gegenseitig eindeutig bestimmen; sie lehrt dann weiter, wie man für einen Balken aus elastischem Material die Durchhängung aus der Größe und Anordnung der Lasten berechnen kann. Woher weiß man, daß ein bestimmter Balken im Sinne der Definition elastisch ist, so daß die Rechnung auf ihn angewandt werden darf? Kann man etwa die Definitionseigenschaft an jeder Stelle des Materials prüfen? Nein, man setzt zunächst das Zutreffen der Definition voraus, rechnet, wie die Theorie es vorschreibt, und prüft das Ergebnis der Rechnung durch einen Versuch nach. Fällt die Überprüfung günstig aus, so sieht man die Voraussetzung für den geprüften und alle analog liegenden Fälle als bewiesen an. Ein noch einfacheres Beispiel: Die Geometrie lehrt uns zahlreiche Sätze über die Kugel; woher weiß man, daß diese Sätze sich auf den Erdkörper anwenden lassen? Hat jemals irgendwer nachgeprüft, ob es einen Punkt gibt, von dem alle Punkte der Erdoberfläche gleichen Abstand haben? Dies verlangt doch die Definition der Kugel. Nein, zuerst war es eine intuitive Vermutung, daß die Erde eine Kugel sei, dann wurde die Vermutung durch Überprüfung aller möglicher Folgerungen aus ihr bestätigt — und schließlich zeigten kleine Widersprüche, daß die Annahme der Kugelgestalt doch nur mit begrenzter Genauigkeit zutrifft.

Ganz so liegt es bei der Beurteilung der Unabhängigkeit zweier Kollektivs. Sind zwei Würfel in einem Becher durch einen kurzen Faden miteinander gekoppelt, wird niemand die Unabhängigkeit der Würfe vermuten. Ist der Faden etwas länger, so wird man keine Vermutung aussprechen können, solange nicht eine genügend große Versuchsreihe, die das Multiplikationsgesetz nachprüft, ausgeführt wurde.. Legt man die Würfel lose, ungekoppelt in einen Becher, so entspricht es einer sehr alten und sehr allgemeinen Erfahrung, daß die Wahrscheinlichkeiten der Paare von Augenzahlen dem Multiplikationsgesetz folgen. Werden schließlich die beiden Würfel in je einem Becher von verschiedenen Personen, womöglich noch an entfernten Orten, gleichzeitig gespielt, so spricht schon ein instinktives Gefühl, d. i. eine noch allgemeinere, überpersönliche Erfahrung, für die Annahme der Unabhängigkeit. Ob die Annahme in einem konkreten Fall zutrifft oder nicht, kann aber nur die Beobachtung entscheiden. Diese besteht darin, daß man mit den vorliegenden Würfeln (oder mit einer Anordnung, die man der gegebenen als gleich ansieht) eine lange Versuchsserie ausführt, bei der nachgesehen wird, ob das Multiplikationsgesetz stimmt.

In manchen Fällen kann man die Unabhängigkeit zweier Kollektivs, die einer Verbindung unterworfen werden sollen, aus ihrer Definition erschließen, dann nämlich, wenn die beiden selbst schon in bestimmter Weise aus einem anderen abgeleitet sind. Beispiele dafür werde ich später geben, wenn von der Zusammensetzung von Grundoperationen die Rede sein wird.

Verbindung abhängiger Kollektivs.

Man kann, wie ich jetzt zum Schlusse kurz zeigen will, das Verfahren der „Verbindung" auch noch auf gewisse Fälle ausdehnen, in denen die Voraussetzungen der „Unabhängigkeit" nicht voll erfüllt sind. Nur darf man nicht meinen, daß die beiden zu verbindenden Kollektivs A und B gar keinen Bedingungen zu genügen brauchen; die Einschränkungen sind bloß etwas vermindert. Wir wollen zwei Kollektivs A, B „verbindbar", aber „abhängig" nennen, wenn folgendes zutrifft. Wird an den Elementen von A eine beliebige Stellenauswahl vorgenommen und die ausgewählte Folge zur „Auswürfelung" aus B benutzt, so soll die dabei entstehende Teilfolge der Elemente von B ein Kollektiv bilden, dessen Verteilung noch von dem Merkmal von A, nach dem ausgewürfelt wurde, abhängt. Also: Es soll eine gewisse Wahrscheinlichkeit geben, mit dem zweiten Würfel „5" zu werfen, wenn der erste „3" zeigt, aber diese soll nicht zugleich die Wahrscheinlichkeit dafür sein, mit dem zweiten Würfel „5" zu erzielen, wenn der erste z. B. „4" ergibt. Ein praktisches Beispiel ist etwa folgendes.

In einer Urne befinden sich drei schwarze und drei weiße Kugeln; es wird zweimal hintereinander eine Kugel gezogen, ohne daß die zuerst gezogene inzwischen zurückgelegt worden wäre. Dann erst legt man beide zurück und fängt von neuem an. Hier besteht das erste Kollektiv aus den „ersten" Zügen, die aus voller Urne geschehen. Nach der üblichen Annahme der Gleichverteilung ist die Wahrscheinlichkeit eines weißen oder eines schwarzen Zuges gleich $^1/_2$. Das zweite Kollektiv hat zu Elementen die Gesamtheit der „zweiten" Züge, die aus einer nur fünf Kugeln enthaltenden Urne erfolgen. Aus dieser Gesamtfolge lassen sich zwei Teilfolgen mittels des ersten Kollektivs „auswürfeln"; einmal, indem man alle jene „zweiten" Züge betrachtet, die auf einen weißen „ersten" Zug folgen, dann jene, denen ein schwarzer „erster" Zug vorangeht. Man sieht ohne weiteres, daß die Wahrscheinlichkeit, schwarz zu ziehen, innerhalb des ersten ausgewürfelten Kollektivs $^3/_5$, innerhalb des zweiten aber nur $^2/_5$ beträgt — immer die Gleichverteilung vorausgesetzt; denn von den fünf Kugeln in der Urne sind im ersten Falle drei, im zweiten Falle nur zwei schwarz. Es hängt hier, wie wir vorher sagten, die Verteilung innerhalb der ausgewürfelten Kollektivs von dem Merkmal ab, nach dem ausgewürfelt wurde.

Wie im Falle der bloß „verbindbaren", aber nicht unabhängigen Kollektivs die Endverteilung innerhalb der Verbindung zu rechnen ist, bedarf nach dem Vorangegangenen kaum noch der Erklärung. Man muß, um die Wahrscheinlichkeit etwa des Ergebnisses „schwarzer Zug — weißer Zug" zu erhalten, folgende zwei Faktoren miteinander multiplizieren: die Wahrscheinlichkeit $^1/_2$ eines schwarzen „ersten" Zuges und die Wahrscheinlichkeit eines weißen „zweiten" Zuges unter der Voraussetzung, daß der erste schwarz war, also $^3/_5$; Resultat: $^1/_2 \,.\, ^3/_5 = {^3/_{10}}$. Analog für jede andere Kombination.

Beispiel für nicht verbindbare Kollektivs.

Es mag schließlich nicht ganz überflüssig sein, ein Beispiel für zwei Kollektivs zu geben, die weder als „abhängige" noch als „unabhängige" gelten können, die in unserem Sinne überhaupt **nicht verbindbar** sind. Es werde etwa täglich um 8 Uhr morgens eine bestimmte meteorologische Größe, sagen wir der Feuchtigkeitsgehalt der Luft, gemessen; die Maßzahl, die eine zwischen 1 und 6 liegende ganze Zahl sei, bildet das Merkmal des Einzelversuches. Dieselbe oder eine andere Variable werde nun auch täglich um 8 Uhr abends gemessen, so daß wir zwei elementweise einander zugeordnete Folgen vor uns haben, die beide die Eigenschaften eines Kollektivs besitzen mögen: Existenz der Grenzwerte und Regellosigkeit. Es mag auch weiterhin zutreffen, daß das zweite Kollektiv gegen Auswürfelung durch das erste unempfindlich ist, d. h. daß die Abendmessungen, die z. B. auf den Frühwert 3 folgen, die gleiche Verteilung aufweisen wie sämtliche am Abend gemessenen Werte. Gleichwohl könnte es sein, daß an jedem 28. Tag, und nur an einem solchen (Vollmond) ein Vormittagswert 3, falls er vorliegt, zwangläufig die Abendbeobachtung 3 nach sich zieht. In diesem Falle würde durch die Verbindung kein Kollektiv entstehen. Denn die Stellenauswahl, die jeden 28. Tag heraushebt, würde eine Verteilung ergeben, bei der die Wahrscheinlichkeit der Kombinationen 3, 1; 3, 2; 3, 4; 3, 5 und 3, 6 Null ist, die der Kombination 3, 3 gleich 1. Innerhalb der Gesamtheit der Vor- und Nachmittagsbeobachtungen trifft dies jedoch nicht zu, da kommen alle sechs Kombinationen mit bestimmten von Null verschiedenen Häufigkeiten vor. Man sieht, daß hier aus der paarweisen Zusammenfassung der Elemente zweier Kollektivs eine Folge von Elementenpaaren entsteht, die kein Kollektiv bildet, da darin die Unabhängigkeit der Grenzwerte der relativen Häufigkeit gegenüber einer Stellenauswahl nicht besteht. Es sind zwar die einzelnen Beobachtungsreihen als zufallsartig angenommen, aber in ihrer gegenseitigen Beziehung, in ihrem Zusammenhang besteht eine Gesetzmäßigkeit, die sich nicht dem Begriff des Kollektivs einfügt. Aus diesem Grunde bezeichnen wir die beiden angegebenen Beobachtungsreihen als „zu einem Kollektiv nicht verbindbar".

Das Beispiel zeigt uns, wie unzureichend und unzuverlässig die primitive, bekannte Form der Produktregel ist, die eine Unterscheidung der einzelnen wesensverschiedenen Fälle der Beziehung zwischen zwei Kollektivs gar nicht zuläßt. Nur der rationelle Wahrscheinlichkeitsbegriff, der sich auf einer genauen Analyse der als Kollektiv bezeichneten Beobachtungsfolge aufbaut, kann hier zu einwandfreien Formulierungen führen.

Übersicht der vier Grundoperationen.

Ich will noch einmal übersichtlich, schlagwortartig die vier Grundoperationen zusammenstellen, von denen hier die Rede war.

1. Die Auswahl.	Definition:	Die Merkmale unverändert, die Elemente durch Stellenauswahl vermindert.
	Lösung:	Die Verteilung bleibt ungeändert.
2. Die Mischung.	Definition:	Die Elemente unverändert, die Merkmale durch Mischung zusammengefaßt.
	Lösung:	Additionsregel.
3. Die Teilung.	Definition:	Die Merkmale unverändert, die Elemente durch Teilung vermindert.
	Lösung:	Divisionsregel.
4. Die Verbindung.	Definition:	Merkmale und Elemente von zwei Kollektivs paarweise zusammengefaßt.
	Lösung:	Multiplikationsregel.

Mit der Aufstellung dieser Definitionen und der jedesmaligen Angabe, wie die Verteilung des abgeleiteten Kollektivs aus den Verteilungen der Ausgangskollektivs gewonnen werden kann, ist der grundsätzliche Aufbau der Wahrscheinlichkeitsrechnung erledigt. Liegt ein konkretes Problem vor, so muß man, um es in unseren Aufbau einzuordnen, folgendes leisten:

Erstens ist festzustellen, welches die Ausgangskollektivs mit den gegebenen Verteilungen sind. Zweitens muß man sich klarmachen, wie das Kollektiv beschaffen ist, innerhalb dessen die gesuchte Wahrscheinlichkeit besteht. Endlich muß man den Übergang von den erstgenannten Kollektivs zu dem neuen derart in einzelne Schritte zerlegen, daß jeder Schritt genau eine Grundoperation darstellt. Dann wird die Aufgabe durch sukzessive Anwendung der vier aufgestellten Regeln gelöst. Natürlich wird man, besonders nach einiger Übung, nicht immer pedantisch in dieser Reihenfolge verfahren müssen. Der Erfahrene übersieht leicht und unmittelbar gewisse Zusammenhänge von Kollektivs, wird z. B. oft sich wiederholende Gruppen von Grundoperationen als eine neue „zusammengesetzte Operation“ ein für allemal erledigen usf. Bei vielen, und nicht den leichtesten Aufgaben reduziert sich die ganze Vorbereitungsarbeit auf ein Minimum: Es liegt z. B. von vornherein nur eine einfache Mischung vor und die ganze Schwierigkeit ist eine rein mathematische, bestehend in der Ausrechnung einer in komplizierter Weise definierten Summe oder eines Integrals.

Ich will hier, nur zur Erläuterung des grundsätzlichen Aufbaus, ein Beispiel anderer Art kurz besprechen, nämlich ein solches, bei dem die mathematischen Schwierigkeiten zurücktreten, aber dafür Gelegenheit gegeben ist, das vielfältige Ineinandergreifen der Grundregeln zu erläutern.

Die Frage des Chevalier DE MÉRÉ.

Ich wähle als Beispiel eine Aufgabe, die vielleicht die älteste ist, die je in der Wahrscheinlichkeitsrechnung gelöst wurde und die zu betrachten für uns von mehr als einer Seite lehrreich sein wird. Zur Zeit PASCALS und FERMATS, der großen Mathematiker des 17. Jahrhunderts, lebte in Frankreich der als eifriger Spieler bekannte Chevalier DE MÉRÉ. Unter den Glücksspielen, die damals im Schwange waren, gab es folgendes Würfelspiel: Ein Würfel wurde in einen Becher getan und viermal hintereinander ausgespielt. Die eine der beiden Parteien setzte nun darauf, daß mindestens einmal die Augenzahl, die erscheint, gleich 6 wird, die andere auf das Gegenteil. Der Chevalier DE MÉRÉ hatte herausgefunden, daß es günstiger sei, auf das positive Resultat, d. h. das Erscheinen mindestens einer 6 zu setzen. Nun lieben Spieler aber die Abwechslung und so wurde gelegentlich eine Variante des Spieles eingeführt. Man nahm zwei Würfel statt eines, spielte sie gemeinsam aus, aber nicht 4mal, sondern 24mal und nun wurde darauf gesetzt, daß mindestens einmal Doppelsechs erscheint, d. h. beide Würfel gleichzeitig 6 zeigen. Der Chevalier DE MÉRÉ bemerkte nun — offenbar spielte er sehr fleißig —, daß es bei diesem veränderten Spiel vorteilhafter sei, auf das Nichterscheinen der Doppelsechs zu setzen. Er hielt das für auffallend, ja er behauptete, hier müsse ein Fehler der Arithmetik stecken. Denn, so argumentierte er, bei einem Einzelwürfel gibt es 6, beim Würfelpaar 36, also 6mal so viel verschiedene Möglichkeiten. Eine von den 6 Möglichkeiten beim einzelnen Würfel ist das Erscheinen der Augenzahl 6, eine der 36 Möglichkeiten beim Spiel mit zwei Würfeln ist die Doppelsechs. Wenn man also mit zwei Würfeln sechsmal so oft, nämlich 24mal statt 4mal spielt, so müßten die Chancen für ja und nein ebenso liegen.

Fermat, dem Méré diese Angelegenheit unterbreitete, klärte sie vollständig auf und in einem Briefe, den er an Pascal schrieb, ist uns seine Lösung erhalten. Ich will sie hier in einer anderen, viel allgemeineren Form geben, die vor allem die Einordnung des Problems in unseren grundsätzlichen Aufbau der Wahrscheinlichkeitsrechnung hervortreten läßt.

Lösung der Aufgabe.

Beginnen wir mit dem ersten Fall, dem viermaligen Ausspielen eines Würfels. Es ist klar, daß hier das Ausgangskollektiv, das gegeben sein muß, in der fortgesetzten Folge der einfachen Würfe mit einem Würfel besteht. Element des Kollektivs ist einmaliges Ausspielen des Würfels, Merkmal die Augenzahl 1, 2, 3,.... oder 6. Fermat nimmt stillschweigend an, daß die Wahrscheinlichkeiten der sechs Merkmale einander gleich, also alle gleich $^1/_6$ seien, mit anderen Worten, daß es sich um einen sogenannten richtigen Würfel handelt, und von dieser Annahme ist sein ganzer Rechnungsgang beherrscht. Nach unserer allgemeineren Auffassung muß sich die Rechnung auch ohne diese einschränkende Annahme durchführen lassen. Wir setzen also voraus, daß die sechs Wahrscheinlichkeiten w_1, w_2, ... bis w_6 der sechs Würfelseiten irgendwie gegeben seien. Ihre Summe ist natürlich 1.

Wonach ist nun gefragt? Man sucht die Wahrscheinlichkeit dafür, daß mindestens einmal in einer Gruppe von vier Würfen die Sechs fällt. Offenbar ist dies eine Wahrscheinlichkeit innerhalb des Kollektivs, das folgendermaßen beschaffen ist: Element ist eine Gruppe von vier aufeinanderfolgenden Würfen, Merkmal des Elements ist ja oder nein (einfache Alternative), und zwar ja, wenn mindestens eine der vier Augenzahlen eine 6 ist, nein, wenn alle Augenzahlen von 6 verschieden sind. Dies ist das Endkollektiv, das wir aus dem gegebenen ableiten müssen. Gefragt ist nach der Wahrscheinlichkeit des „ja" innerhalb dieses Endkollektivs.

Suchen wir nun die Grundoperationen zu bestimmen, die vom gegebenen Kollektiv C zum Endkollektiv K führen. Zunächst sieht man, daß es innerhalb C, also des fortgesetzten Spiels mit dem Würfel, auf die Unterschiede zwischen den Resultaten 1, 2 .... bis 5 nicht ankommt, da am Schlusse nur von Sechs oder

Nichtsechs die Rede ist. Also nehmen wir vor allem an C eine Mischung vor, indem wir die Merkmale 1 bis 5 zusammenwerfen, anderseits die 6 allein lassen. Es entsteht so ein verändertes Kollektiv, das C' heißen möge, mit denselben Elementen wie C, aber nur zwei verschiedenen Merkmalen, nämlich: Sechs oder Nichtsechs. Nach der Mischungsregel oder Additionsregel sind die beiden Wahrscheinlichkeiten in C' gleich w_6 einerseits und $w_1 + w_2 + w_3 + w_4 + w_5$ anderseits. Da die Summe sämtlicher w gleich 1 ist, können wir kürzer die beiden Größen bezeichnen als w_6 und $1 - w_6$.

Nunmehr führen wir an C' eine Auswahl durch, indem wir aus der Gesamtfolge aller Würfe diejenigen herausheben, deren Nummern die folgenden sind:

$$1, 5, 9, 13, 17, 21, 25, \ldots.$$

usf. in infinitum, immer drei auslassend. Das so entstandene Kollektiv mit unveränderten Merkmalen (Sechs oder Nichtsechs) und ausgewählten Elementen heiße C_1'. In ihm besteht nach der Auswahlregel die gleiche Verteilung wie in C', d. h. die Wahrscheinlichkeit der Sechs ist wieder w_6, die des gegenteiligen Falles ist $1 - w_6$.

Ein zweites Kollektiv, das C_2' heißen soll, entsteht in analoger Weise durch eine zweite Auswahl aus C', indem wir diejenigen Elemente zurückbehalten, deren Nummern

$$2, 6, 10, 14, 18, 22, 26, \ldots.$$

usf. sind. Auch innerhalb dieses C_2' bestehen die gleichen Wahrscheinlichkeiten w_6 und $1 - w_6$ für die Resultate Sechs bzw. Nichtsechs. Schließlich führen wir die beiden noch übrigbleibenden Möglichkeiten aus, nämlich eine dritte Auswahl, die die Elemente mit den Nummern

$$3, 7, 11, 15, 19, 23, 27, \ldots.$$

umfaßt und C_3' heißen soll, sowie die vierte Auswahl mit den Ordnungsnummern

$$4, 8, 12, 16, 20, 24, 28, \ldots,$$

die ein Kollektiv C_4' ergibt. Im ganzen sind so aus dem gegebenen C auf dem Weg über C' vier neue Kollektivs, C_1', C_2', C_3', C_4', gewonnen worden, jedes eine einfache Alternative mit den Merkmalen Sechs oder Nichtsechs und den Wahrscheinlichkeiten w_6 und $1 - w_6$, die bekannt sind, weil ja w_1 bis w_6 gegeben vorausgesetzt war.

Nun bleibt der letzte Schritt zu tun: Man muß eine Verbindung zwischen den vier durch Auswahl gewonnenen Kollektivs herstellen. Verbinden wir zunächst C_1' mit C_2', so heißt dies, daß wir die ersten Elemente von C_1' und C_2' zusammenfassen, also die Würfe mit den Nummern 1 und 2, sodann die zweiten Elemente, das sind die Würfe mit den Nummern 5 und 6 usf. Es entsteht ein Kollektiv, das wir C_1'' nennen und dessen Elemente Paare von Würfen sind, und zwar diejenigen Paare, die in der ursprünglichen Numerierung in C so bestimmt sind:

1 mit 2, 5 mit 6, 9 mit 10, 13 mit 14, 17 mit 18, ...

usf. Das Merkmal eines solchen Elements ist eine der vier Kombinationen von zwei Resultaten, nämlich: Sechs mit Sechs, Sechs mit Nichtsechs, Nichtsechs mit Sechs, endlich Nichtsechs mit Nichtsechs.

Es fragt sich nun, ob auf C_1'' die Regel der Verbindung unabhängiger Kollektivs angewandt werden dürfen oder nicht. Die Frage ist zu bejahen, und zwar liegt hier einer der früher erwähnten Fälle vor, in dem die Unabhängigkeit zweier Kollektivs aus ihrer Ableitung folgt. Es läßt sich mathematisch beweisen und ist auch unmittelbar leicht einzusehen, daß die vorausgesetzte Regellosigkeit im Ausgangskollektiv C zum Schluß führt, daß die aus C bzw. C' durch Auswahl gewonnenen Kollektivs C_1' und C_2' unabhängig sein müssen. Demnach ergeben sich die Wahrscheinlichkeiten der vier vorgenannten Merkmale nach der Multiplikationsregel; die erste gleich w_6^2, die zweite und dritte gleich $w_6\ (1 - w_6)$, die letzte gleich $(1 - w_6)^2$.

Genau die gleiche Verbindung läßt sich nun auch zwischen C_3' und C_4' herstellen. Das so entstehende Kollektiv C_2'' umfaßt die Paare

3 mit 4, 7 mit 8, 11 mit 12, 15 mit 16, 19 mit 20, ...

usf. Die Merkmale und Wahrscheinlichkeiten sind in C_2'' die gleichen wie in C_1''.

Noch einmal müssen wir eine Verbindung ausführen, jetzt zwischen C_1'' und C_2''. Damit werden zusammengefaßt die Paare 1, 2 und 3, 4, nämlich das erste Element von C_1'' mit dem ersten von C_2'', sodann die Paare 5, 6 und 7, 8, das sind die zweiten Elemente von C_1'' und C_2'' usf. Man sieht, daß die Elemente des jetzt entstandenen Kollektivs tatsächlich Gruppen von je vier Würfen sind, und zwar die Gruppen mit den Nummern

1 bis 4, 5 bis 8, 9 bis 12, 13 bis 16, 17 bis 20, ...

usf. Dieses Kollektiv heiße K', seine Merkmale sind die 16 Kombinationen, die man erhält, wenn man eines der vier Merkmale von C_1'' mit einem von C_2'' zusammenstellt. Die 16 Wahrscheinlichkeiten folgen aus der Multiplikationsregel, die auch hier anwendbar ist. Z. B. ist die Wahrscheinlichkeit des Merkmales Sechs-Sechs mit Sechs-Sechs (oder 4mal Sechs) gleich $w_6^2 \cdot w_6^2 = w_6^4$. Die Wahrscheinlichkeit des Resultates: 4mal Nichtsechs ist ebenso gegeben durch $(1 - w_6)^4$.

Nun kommt der allerletzte Schritt, der uns von K' zu K führt: Es interessieren uns nicht alle 16 Kombinationen, sondern nur die beiden Fälle: Keinmal 6, d. h. alle 4mal Nichtsechs, oder das Gegenteil. Also ist nochmals eine Mischung vorzunehmen. Die Wahrscheinlichkeit des Merkmales: keinmal 6 bleibt $(1 - w_6)^4$, die Summe der übrigen 15 Wahrscheinlichkeiten muß die erstere zu 1 ergänzen, also bleibt

$$W = 1 - (1 - w_6)^4$$

Dies ist somit die Wahrscheinlichkeit innerhalb des (durch Mischung aus K' hervorgegangenen) Kollektivs K dafür, daß das Gegenteil von: allemal Nichtsechs, das ist also: mindestens einmal Sechs, eintritt. Damit ist die Lösung der Aufgabe gefunden.

Diskussion der Lösung.

Zunächst können wir, ohne viel weitere Rechnung, das Resultat für den zweiten Fall der MÉRÉschen Aufgabe angeben, nämlich den Fall von 24 Würfen mit einem Würfelpaar. Nimmt man als Ausgangskollektiv C das fortgesetzte Spiel mit einem Würfelpaar und betrachtet statt des Resultats Sechs jetzt die Doppelsechs, so tritt offenbar an die Stelle des früheren w_6 eine Größe w_{66}, nämlich die Wahrscheinlichkeit dafür, innerhalb C, also des fortgesetzten Spiels mit dem Würfelpaar, die Doppelsechs zu erzielen. Die weitere Ableitung ist ganz analog der früheren, nur sind jetzt statt 4 Auswahlen deren 24 zu bilden und dann durch Verbindung zu vereinigen. Wir brauchen dies nicht im einzelnen durchzuführen, denn offenbar ist das Resultat nur dies, daß in der Endformel der Exponent 24 an Stelle von 4 tritt. Wir haben also in

$$W' = 1 - (1 - w_{66})^{24}$$

die Wahrscheinlichkeit dafür, daß innerhalb einer Gruppe von 24 Würfen mindestens einmal die Doppelsechs fällt.

Macht man die Annahme, daß das Spiel mit dem Würfelpaar als Verbindung zweier unabhängiger Kollektivs aufgefaßt werden darf, so wird das neue Ausgangskollektiv auf das frühere zurückgeführt in dem Sinne, daß $w_{66} = w_6^2$ zu setzen ist. Darnach geht die ebengenannte Formel über in

$$W' = 1 - (1 - w_6^2)^{24}.$$

Das ist offenbar die Größe, die mit dem früheren W verglichen werden muß.

Man sieht vor allen Dingen, daß die Ausdrücke für W und W' nicht übereinstimmen. Für einen beliebigen Würfel, für den w_6 irgendeinen Wert besitzt, kann also W sicher nicht gleich W' sein. So war es aber auch von Méré nicht gemeint. Er dachte sicher an einen „richtigen" Würfel, für den $w_6 = {}^1/_6$ zu nehmen ist. Setzen wir diesen Wert in unsere beiden Formeln ein, so erhalten wir:

$$W = 1 - ({}^1/_6)^4 = 0{,}516; \quad W' = 1 - ({}^1/_{36})^{24} = 0{,}491.$$

In der Tat war also die Beobachtung des Chevalier richtig, der Wert von W liegt etwas über 0,5, der von W' etwas darunter. Im ersten Spiel ist es vorteilhaft, auf „Ja" zu setzen, im zweiten nicht. Falsch war nur die Überlegung, die er anstellte und die ihn zu dem voreiligen Schluß verleitete, in beiden Fällen müßten „theoretisch" die Chancen gleich stehen.

Einige Schlußfolgerungen.

Verschiedene lehrreiche Folgerungen lassen sich aus der hier entwickelten Lösung ziehen. Vor allem sieht man, daß die Lösung einer konkreten Aufgabe der Wahrscheinlichkeitsrechnung tatsächlich etwas Bestimmtes über die Wirklichkeit aussagt. Über den Ablauf einer Erscheinungsreihe der physischen Welt wird eine Voraussage gemacht, die sich durch Beobachtungen prüfen läßt. Daß historisch eine Beobachtung der Rechnung vorausging, tut nichts zur Sache, denn die Voraussage besteht ja für alle künftigen Fälle. Sie geht auch weit über das unmittelbare Beobachtungsergebnis hinaus und deckt z. B. den Fall, in dem die Wiederholungszahl 4 bzw. 24 durch eine beliebige andere ersetzt wird oder in dem ein nicht „richtiger" Würfel mit einem von $^1/_6$ abweichenden Wert von w_6 verwandt wird usf. Auch eine charakteristische Eigenschaft der Wirklichkeitsaussagen der Wahrscheinlichkeitslehre sehen wir in Erscheinung treten: Sie beziehen sich stets und ausschließlich auf die relative Häufigkeit von Ereignissen in langen Versuchsfolgen. Eine Theorie, die nicht von vornherein den Zusammenhang zwischen der relativen Häufigkeit und dem Wahrscheinlichkeitsbegriff einführt, kann für die Deutung der Wirklichkeit nichts beisteuern.

Noch auf einen zweiten Umstand muß ich hinweisen. Man meint oft, daß man im Bereiche der Glücksspiele von altersher nur mit sozusagen a priori bekannten Wahrscheinlichkeiten zu tun gehabt habe, die sich aus der Annahme der Gleichwahrscheinlichkeit aller möglichen Fälle ergaben. Wir sehen aber hier, daß auch Spiele gespielt wurden, bei denen durchaus kein Anlaß vorlag, die Gleichheit der Chancen a priori anzunehmen. Warum gerade vermuten, daß bei genau vier Würfen mindestens einmal Sechs mit 50% Wahrscheinlichkeit fällt? Nur die Beobachtung konnte dazu führen, diese Art von Würfelspiel festzulegen.

Man kann sich den historischen Vorgang etwa so denken.

Erst hatte man durch lange, über große Zeiträume sich erstreckende Versuche in Erfahrung gebracht, wie man Würfel herstellen müsse, bei denen die Chancen für jede der sechs Augenzahlen nahezu gleich sind. Dann fand man durch Beobachtungen, daß bei denselben Würfeln, den sog. richtigen, die Aussichten auf mindestens eine Sechs in vier Würfen (oder mindestens eine Doppelsechs in 24) annähernd 1:1 stehen. Endlich zeigten noch verfeinerte oder weiter fortgeführte Versuche, daß in den letzten beiden Fällen doch gewisse Abweichungen auftreten, die einer Aufklärung bedürfen. Jetzt trat die Theorie auf den Plan, die den Zusammenhang der beiden Eigenschaften eines Würfels untersuchte, nämlich der Eigenschaft, daß seine sechs Seiten gleich häufig fallen, und anderseits der, daß in der Hälfte aller Serien von vier Würfen mindestens einmal Sechs fällt. Die Rechnung ergibt, daß die beiden Eigenschaften nicht genau zusammenstimmen, denn zu $w_6 = {}^1/_6$ gehört $W = 0{,}516$ und umgekehrt (dies ist aus der gegebenen Formel leicht auszurechnen) gehört $W = {}^1/_2$ zu einem $w_6 = 0{,}1591$, was etwas weniger als ${}^1/_6$ ist. Deutlicher kann man den empirischen Charakter der Wahrscheinlichkeitsrechnung, d. h. ihre Funktion als Beschreibung sinnlich wahrnehmbarer Tatbestände gar nicht zum Ausdruck bringen.

Allein ich bin hier schon in Überlegungen und Fragestellungen gelangt, die eigentlich erst den Gegenstand des nächsten Vortrages bilden sollen. Ich will also nicht in dieser Richtung fortfahren und unterlasse es auch, noch weitere Beispiele, weitere Aufgaben hier in den Einzelheiten vorzuführen. Sie würden uns grundsätzlich nicht viel anderes lehren können. Daß man natürlich von der großen Umständlichkeit, mit der ich vorhin die Ableitung des Endkollektivs dargelegt habe, bei einiger Übung Abstand nehmen kann, habe ich schon erwähnt. Diese Abkürzungen Ihnen vorzuführen, gehört nicht in diesen Zusammenhang. Hier will ich lieber zum Abschluß noch eine kurze Übersicht geben über die wichtigsten Punkte, die den neuen Aufbau der Wahrscheinlichkeitsrechnung kennzeichnen.

Kurzer Überblick.

1. Den Ausgangspunkt der Theorie bildet der Begriff des Kollektivs. Wahrscheinlichkeit gibt es nur mit dem Zusatz: innerhalb dieses bestimmten Kollektivs.
2. Kollektiv ist eine unendliche Folge von Beobachtungen, deren jede mit der Feststellung eines Merkmales endet. Die relative Häufigkeit, mit der ein Merkmal auftritt, besitzt einen Grenzwert, und dieser ändert sich nicht, wenn durch Stellenauswahl eine Teilfolge herausgehoben wird.
3. Der gegen Stellenauswahl unempfindliche Grenzwert der relativen Häufigkeit heißt die Wahrscheinlichkeit des Merkmales innerhalb des betreffenden Kollektivs. Die Gesamtheit der Wahrscheinlichkeiten innerhalb eines Kollektivs bildet seine Verteilung. (So weit waren wir im wesentlichen am Ende des ersten Vortrages. Nun das Neue:)
4. Eine Aufgabe der Wahrscheinlichkeitsrechnung besteht immer darin, daß ein oder mehrere Kollektivs mit ihren Verteilungen

gegeben sind und aus ihnen ein neues Kollektiv abgeleitet wird, in dem die Verteilung gesucht ist. Der Spezialfall, daß die gegebenen Verteilungen gleichförmige sind (gleich mögliche Fälle), besitzt keinerlei Ausnahmestellung.

5. Die Ableitung eines neuen Kollektivs aus gegebenen besteht stets in einer Kombination von Operationen, deren jede einzelne eine der vier Grundoperationen: Auswahl, Mischung, Teilung, Verbindung, ist.
6. Bei der Operation der Auswahl bleibt die Verteilung unverändert; bei der Mischung wird durch die Additionsregel, bei der Teilung durch die Division, bei der Verbindung durch Multiplikation die neue Verteilung bestimmt.
7. Damit, daß für jede Grundoperation der Übergang von den gegebenen zu der neuen Verteilung bekannt ist, ist jede Aufgabe der Wahrscheinlichkeitsrechnung grundsätzlich gelöst. Durch Häufung der Operationen oder durch komplizierte Definition einer Mischung u. ähnl. können gleichwohl erhebliche rein mathematische Schwierigkeiten entstehen.
8. Praktisch beruht jede Rechnung auf der Kenntnis gewisser Häufigkeiten in langen Versuchsfolgen und ihr Resultat ist eine bestimmte, an Beobachtungen prüfbare Aussage über andere derartige Häufigkeiten.

Wollte ich mich noch kürzer ausdrücken und das Wesentliche in einem einzigen Satz zusammenfassen, so würde ich sagen:

Die Wahrscheinlichkeitsrechnung hat es ausschließlich mit Häufigkeiten in langen Beobachtungsreihen zu tun, sie geht von gegebenen Aussagen über solche Häufigkeiten aus und leitet daraus andere mittels von vornherein festgelegter Rechenregeln ab.

Dritter Vortrag.

Kritik der Grundlagen.

In den beiden vorangehenden Vorträgen habe ich eine kurze Skizze von dem gegeben, was ich den Aufbau der *neuen Wahrscheinlichkeitstheorie* nenne. In der Zusammenfassung am Schlusse des zweiten Vortrages sind in acht Sätzen die Hauptgesichtspunkte der Theorie übersichtlich angeführt. Wäre es etwa meine Aufgabe, einen vollständigen Lehrgang zu entwickeln, so müßte ich jetzt zeigen, wie sich durch schrittweise Zusammensetzung der vier Grundoperationen immer verwickeltere Ableitungen neuer Kollektivs aus gegebenen gewinnen lassen und wie anderseits alle Fragen, die man in der Wahrscheinlichkeitsrechnung zu behandeln pflegt, auf derartig zusammengesetzte Operationen zurückführen. Diese Aufgaben würden uns aber in rein mathematische Auseinandersetzungen hinüberleiten, für die hier nicht der Platz ist. Ich muß den dafür Interessierten auf meine Vorlesungen über Wahrscheinlichkeitsrechnung verweisen, die seit 1931 im Druck vorliegen. Hier wollen wir uns nur mit den allgemeinen gedanklichen Grundlagen der Theorie befassen, und da erwächst mir jetzt die Aufgabe, das bisher Vorgetragene kritisch zu betrachten. In zwei Richtungen hauptsächlich werden sich meine heutigen Ausführungen erstrecken. Ich werde das Verhältnis der neuen Theorie zu der klassischen Wahrscheinlichkeitsrechnung und zu den späteren Ansätzen besprechen, die die klassische Theorie fortzubilden und tiefer zu begründen suchten. Dann aber will ich auf die mannigfachen Versuche eingehen, die seit dem Erscheinen meiner ersten Veröffentlichungen zu einer Widerlegung, Abänderung oder Ergänzung meiner Theorie unternommen wurden.

Der Umfang der praktischen Anwendbarkeit.

Vorerst einige Bemerkungen allgemeinerer Natur über Einwände, die schon die in der Einleitung des ersten Vortrages formulierte grundsätzliche Einstellung zur wissenschaftlichen Begriffsbildung überhaupt treffen. Ich hatte ausdrücklich hervorgehoben, daß eine rationelle Theorie der Wahrscheinlichkeit unmöglich jeder Anwendung des Wortes Wahrscheinlichkeit in der Umgangssprache gerecht werden kann. Gerade dies wird nun in neuerer Zeit von einem auf anderem Gebiete viel genannten Schriftsteller, dem Engländer John Maynard Keynes, sehr eindringlich als ein Mangel jeder „Häufigkeitstheorie“ hingestellt.

In einer Kritik des von mir schon erwähnten John Venn, der in seiner „Logic of chance“ ungefähr das entwickelt, was von mir in der ersten Forderung an ein Kollektiv zusammengefaßt erscheint, bemängelt Keynes, daß nicht alles, was im gewöhnlichen Sprachgebrauch „wahrscheinlich“ heißt, durch die Häufigkeitsdefinition gedeckt wird. Gerade diejenigen Fälle, auf die die Definition nicht anwendbar ist, seien die wichtigsten und wenn wir auf ihre Berücksichtigung verzichteten, so wäre „die Wahrscheinlichkeit kein Führer durchs Leben“ und wir würden „nicht vernünftig handeln, wenn wir uns von ihr leiten lassen“. Den Vorwurf, der in den letzten Sätzen steckt, müßte man, wenn er berechtigt wäre, ernst nehmen. Denn eine Wissenschaft, die sich im Leben nicht bewährt, wollen wir gewiß nicht betreiben; anderseits aber auch nicht eine solche, die über ihre Grenzen hinaus Geltung beansprucht, wo sie ihr nicht zukommt. In der Tat nimmt die neue Wahrscheinlichkeitstheorie, wie ich sie in ihren Grundlagen skizziert habe, für sich in Anspruch, auf drei großen Gebieten eine zuverlässige „Führerin durchs Leben“ zu sein.

Erstens lehrt sie, wie man sich Glücksspielen gegenüber zu verhalten hat; wie Einsätze und Gewinne zu bemessen sind, wenn man bei andauerndem Spiel ein bestimmtes Ergebnis erzielen will. Sie lehrt aber zugleich auch auf das eindringlichste, daß über den Ablauf eines Einzelspiels nichts und nochmals nichts ausgesagt werden kann, sowie daß das Ersinnen eines noch so kniffiichen „Systems“ ebenso nutzlos ist wie das Konstruieren eines Perpetuum mobile. Zweitens bewährt sich die Theorie im Versicherungswesen, wo sie Prämien und Prämienreserven in einer Weise zu berechnen gestattet, die billigen Gerechtigkeitsansprüchen genügt und eine hinreichende Sicherheit des Unternehmens gewährleistet. Hierher gehört der Methode nach auch die vielfältige Anwendung wahrscheinlichkeitstheoretischer Überlegungen auf statistische Beobachtungsreihen irgendwelcher Art — ich nenne neben der biologischen und sozialen Statistik auch die neuerdings gepflegte technische Statistik, die z. B. bei Einrichtung von Fernsprechämtern mit Selbstanschluß sich als unentbehrlich und als zuverlässig erwiesen hat — einschließlich der sogenannten Theorie der Beobachtungsfehler. Drittens hat die rationelle, auf die Häufigkeitsauffassung gestützte Wahrscheinlichkeitsrechnung in enger Verbindung mit gewissen Ideenbildungen der theoretischen Physik weitgehende Erfolge erzielt, die in ihren Auswirkungen zweifellos in das materielle Leben der Menschen eingreifen, ganz abgesehen von dem ideellen Werte der Erweiterung unserer Erkenntnis. Auf all dies kommen wir noch später zurück.

Was steht dem gegenüber?

Was aber kann man, frage ich, über das hier Angeführte hinaus leisten, wenn man den Wahrscheinlichkeitsbegriff weniger eng faßt, nicht nur als Häufigkeit innerhalb fest umgrenzter Massenerscheinungen, sondern so, wie es der Sprachgebrauch will, als einen „Grad des Fürwahrhaltens" eines Satzes, einer Vermutung? Das KEYNESsche Buch enthält wohl ein Kapitel mit der Überschrift „Die Anwendung der Wahrscheinlichkeitslehre auf die Lebensführung", aber ich habe darin auch nicht die kleinste Spur einer praktischen Anweisung finden können, vergleichbar einer der vielen konkreten Vorschriften, wie sie in jedem der früher genannten Gebiete von der auf den Häufigkeitsbegriff aufgebauten Theorie geliefert werden. Die Sache steht doch wohl so: Wenn jemand beispielsweise eine Ehe schließen und möglichst wissenschaftlich entscheiden will, ob sie „wahrscheinlich" gut verlaufen werde, dann kann ihm möglicherweise die Psychologie oder Physiologie, die Eugenik, die Rassenkunde helfen, ganz sicher aber nicht eine „Wissenschaft", die sich um die Wortbedeutung von „wahrscheinlich" herumrankt. Ich bestreite unbedingt, daß etwas praktisch Brauchbares bei einer solchen Betrachtungsweise herauskommt.

Einen ähnlichen Standpunkt wie KEYNES vertritt auch der Geophysiker HAROLD JEFFREYS in seinem Buch „Scientific inference", ja er geht noch weiter, indem er ausdrücklich erklärt, daß man jede Wahrscheinlichkeit im weitesten Sinne des Wortes durch eine Zahl ausdrücken könne. Liegt eine Behauptung B vor und sind A_1, A_2, ... die Voraussetzungen, auf die sie sich stützt, so gebe es eine von den A_1, A_2, ... und von B abhängige Zahl, eben die Wahrscheinlichkeit der Behauptung B. Man sucht aber vergebens in dem Buche eine klare Anweisung darüber, wie in einem gegebenen Fall diese Zahl zu bestimmen ist, und noch viel weniger erfährt man, was mit diesen Zahlen anzufangen ist, welche Aufgaben damit gelöst, welche praktischen Fragen beantwortet werden können.

Ich kann mir nicht vorstellen, welches Ziel man mit der Aufstellung solcher Theorien verfolgt. Jedenfalls sind alle bisherigen Erfolge wahrscheinlichkeitstheoretischer Betrachtungen mit Hilfe eines rationalisierten Wahrscheinlichkeitsbegriffes erzielt worden, der von vornherein darauf verzichtet, alles zu umfassen, was der Sprachgebrauch als „Wahrscheinlichkeit“ bezeichnet.

Der „natürliche“ Wahrscheinlichkeitsbegriff.

Gewiß soll hier nicht eine Verbotstafel aufgerichtet werden gegen jede Verwendung des Wortes „Wahrscheinlichkeit“ in einem anderen Sinn als dem eines Grenzwertes von relativen Häufigkeiten. Wir gebrauchen ja auch, um an das im ersten Vortrag Gesagte zu erinnern, das Wort „Arbeit“ in mannigfachen Bedeutungen, die nichts mit der Definition der Arbeit in der Physik zu tun haben. Nicht damit begehen wir einen Fehler, daß wir sagen, wir würden „wahrscheinlich“ heute zu Hause bleiben und es bestehe „große Wahrscheinlichkeit“ dafür, daß Besuch kommt u. dgl. Das Unzulässige, weil Verwirrende, liegt nur darin, daß man solche Aussagen in einen noch so losen oder entfernten Zusammenhang mit der Wahrscheinlichkeitsrechnung bringen will. Derartige sprachübliche „Wahrscheinlichkeiten“ lassen sich nicht durch Zahlen ausdrücken, sie sind nur in sehr beschränktem Umfang miteinander vergleichbar und unterliegen gewiß nicht den Verknüpfungen, die in den Grundoperationen zum Ausdruck kommen und nur für wohldefinierte Zahlen einen Sinn haben. Die Vermengung der Sphären, die Scheinanwendung von Formeln auf Dinge, die damit nichts zu tun haben, führt nur zur Diskreditierung der Wahrscheinlichkeitsrechnung. In der Tat gibt es kaum einen Zweig der exakten Wissenschaften, der den Laien mehr zur Spottlust anregen würde, und sobald jemand einen Mathematiker hänseln will, bringt er fast immer eine Frage, die mit dem Wort „Wahrscheinlichkeit“ verknüpft ist, aufs Tapet. Ich kann dies — leider — nicht ganz unberechtigt finden, wenn ich in einem ernsthaften, von mir schon genannten Buche eine ernstgemeinte, gründliche Auseinandersetzung darüber lese, wie groß für ein Baby, das bisher in seinem Leben nur einen blauen und einen gelben Gegenstand gesehen hat, die zahlenmäßige Wahrscheinlichkeit dafür sei, daß der nächste Gegenstand, den es sieht, wieder blau sein werde.

Man sagt mit Recht, ein intelligenter, der Sprache kundiger Mensch wisse sozusagen von Natur aus, was „wahrscheinlich“ bedeutet. Das ist ebenso richtig, wie es für die Worte „warm“ oder „kalt“ stimmt. Trotzdem muß die Wärmelehre von allem Anfang an diesen „natürlichen" Wärmebegriff durch ein physikalisches Wärmemaß ersetzen, etwa die Thermometerskala, und alle Aussagen der Wärmelehre beziehen sich ausschließlich auf diesen künstlichen, zahlenmäßig faßbaren Temperaturbegriff. Es ist sinnlos, sie auf eine beliebige sprachübliche Verwendung der Worte „warm" und „kalt" zu übertragen: Wenn wir einem Freunde in unserem Hause einen „warmen" Empfang bereiten, wenn er allmählich in unserer Gesellschaft „warm" wird, wenn wir finden, daß jemand ein „kaltes" Herz habe oder daß ein Zimmer mit blauer Tapete „kalt" wirke. Wir bezeichnen allerdings diesen Gebrauch der Worte in der Regel als „metaphorisch" oder „übertragen" und bringen damit zum Ausdruck, daß uns die Verwechslung fernliegt. Aber wir sind eben von früher Kindheit an durch den Gebrauch des Thermometers, durch häusliche und Schulbelehrung mit den Elementen des physikalischen Wärmebegriffes so vertraut geworden, daß wir die Verwendung des Wortes „warm" fast nur in diesem Sinn als legitim empfinden. Daß eine Wahrscheinlichkeit nur an einer Massenerscheinung oder einem Wiederholungsvorgang gemessen werden kann, ist noch nicht so geläufig — und daher allein kommen die vergeblichen und aussichtslosen Bemühungen, auf den „natürlichen" Begriff der Wahrscheinlichkeit eine Theorie aufzubauen.

Die Rolle der Philosophen.

Es wird hier am Platze sein, ein paar Worte darüber zu sagen, wie weit man etwa von der Seite der philosophischen Literatur her eine Aufklärung des Wahrscheinlichkeitsbegriffes erwarten darf. Zunächst ist es nicht leicht, einigermaßen abzugrenzen, was hier unter dem Namen Philosophie für uns in Frage kommt. Man könnte von unserem Standpunkt, dem der exakten Wissenschaften, aus etwa so definieren: Es gibt Untersuchungen über Gegenstände, mit denen sich auch die Einzelwissenschaften beschäftigen, die aber in der Art ihrer Fragestellung, ihrer Schlußfolgerungen, ihrer Ausdrucksweise und vor allem in der Kontrolle der Resultate durch die Erfahrung sich nicht den Regeln unterwerfen, die in den Spezialwissenschaften üblich sind. An sich ist es natürlich durchaus denkbar, daß man auch auf einem solchen, völlig andersartigen Weg zu wertvollen Erkenntnissen gelangen kann. Zweifellos sind viele Vertreter der exakten Wissenschaften selbst dieser Ansicht und sie entlehnen oft in den Vorworten und Einleitungen ihrer Bücher Gedankengänge von philosophischen Schriftstellern, allerdings ohne nachher in den weiteren Ausführungen viel davon Gebrauch zu machen. Sucht man nun in den Hauptwerken der Philosophie aller Zeiten nach, welche Stellungnahme sie dem Wahrscheinlichkeitsbegriff gegenüber vertreten, so ist man überrascht, zu sehen, daß sich sehr wenig finden läßt; bis zum Beginn des 19. Jahrhunderts fast nichts und später auch kaum etwas Bemerkenswertes. Dieser im ersten Augenblick merkwürdig erscheinende Tatbestand klärt sich leicht auf, wenn man allgemein die Rolle betrachtet, die die Philosophie den Naturwissenschaften gegenüber spielt.

Während es nämlich im großen ganzen üblich ist, daß die Naturforscher die Autorität der in ihrer Epoche maßgebenden Philosophen (der sogenannten Schulphilosophie ihrer Zeit) anerkennen und ihr in höherem oder geringerem Maße offen huldigen, verhält es sich in Wahrheit so, daß die Philosophen, soweit sie sich mit naturwissenschaftlichen Fragen befassen, kaum etwas anderes tun, als die gesichert scheinenden Ergebnisse der Naturforschung der letztvergangenen Zeit interpretieren. Jede fruchtbare naturwissenschaftliche Entdeckung durchläuft drei Stadien: Sie tritt zunächst als revolutionär, das Alte zerstörend, auf, wird umstritten, bekämpft und nur allmählich rezipiert; dann folgt eine Zeit, in der sie als vollkommen gesicherte Wahrheit gilt; schließlich machen sich neue Fortschritte geltend, die alte Wahrheit wird umgewandelt und geht in geläuterter und bescheidenerer Form in einer neuen Auffassung auf. Die Philosophen halten es nun stets mit den „gesicherten" Resultaten, sie sind auf Seiten der starken Bataillone. Sobald einmal eine naturwissenschaftliche Lehrmeinung sich innerhalb ihres Bereiches durchgesetzt hat, nehmen die Philosophen sie auf, übertreiben ihre Geltung, suchen ewigen Bestand für sie in Anspruch zu nehmen und verteidigen sie so gegen etwaige Neuerer. Ein junger Zweig der positiven Wissenschaft muß sich immer erst aus eigener Kraft durchsetzen; sobald er stark geworden ist, helfen ihm die Philosophen gegen seine Nachfolger.

Beispiele für diesen Sachverhalt finden sich in großer Zahl. Man weiß, in welcher Weise KANT die Euklidische Geometrie und die Raum- und Zeitauffassung der Newtonschen Physik dogmatisiert hat. GAUSS wagte es einfach nicht, unter diesem übermächtigen Eindruck, seine Ansätze zur nicht-Euklidischen Geometrie zu publizieren. Heute fängt die Philosophie schon langsam an, sich mit der allgemeinen Raumvorstellung und der Relativitätstheorie abzufinden. Ein anderer Fall: Als GALILEI das Trägheitsgesetz aussprach, galt es als eine widersinnige, weil der Aristotelischen Philosophie widersprechende Behauptung; für SCHOPENHAUER ist es eine unumstößliche, unausweichliche Folgerung aus dem ewig gültigen Kausalgesetz. Darnach versteht man auch das Verhalten gegenüber dem Wahrscheinlichkeitsbegriff. Es gab im 18. Jahrhundert noch keine genügend anerkannte, nach außen mit dem nötigen Autoritätsanspruch auftretende Wahrscheinlichkeitsrechnung. Erst das große Werk von LAPLACE, das zum erstenmal 1812 erschien, schuf diesen Zustand und seit der Mitte des 19. Jahrhunderts steht die gesamte Philosophie unter seinem Einfluß.

Nachdem einmal LAPLACE das Lehrgebäude der Wahrscheinlichkeitsrechnung auf der sogenannten „Gleichmöglichkeitsdefinition" aufgebaut hatte, fehlte es fast keinem Philosophen an Argumenten für diese Auffassung. Erstaunlich ist die Unbefangenheit, mit der manche Philosophen sich auf den Boden der herkömmlichen Anschauungen stellen. So berechnet z. B. EDUARD v. HARTMANN, obwohl gerade er an anderer Stelle Kritik an dem Ausdruck „gleichmöglich" übt, in der Einleitung zu seiner „Philosophie des Unbewußten" mittels mathematischer Formeln die Wahrscheinlichkeit dafür, daß es Naturvorgänge gibt, die geistigen Ursachen zuzuschreiben seien, und findet sie gleich 0,5904! So mächtig wirkte die Autorität von LAPLACE, daß der Philosoph THEODOR FECHNER, der den fruchtbaren Begriff des Kollektivgegenstandes geschaffen hat, es nicht wagte, von hier aus die Wahrscheinlichkeitsrechnung zu begründen, sondern seine „Kollektivmaßlehre" neben die Wahrscheinlichkeitsrechnung stellte.

Wir aber lernen aus alledem, daß wir nicht hoffen dürfen, wesentliche Hilfe oder Aufklärung in der philosophischen Literatur zu finden, die in Hinsicht auf Wahrscheinlichkeitsfragen nur ein Spiegel der Gedanken ist, die bei Mathematikern und Physikern entstanden sind. — Ich wende mich jetzt einer eingehenderen Besprechung dessen zu, was man als das klassische Lehrgebäude der Wahrscheinlichkeitsrechnung zu bezeichnen pflegt.

Dabei werden wir, wo es nötig ist, auf die von philosophischer Seite beigebrachten Deutungen und Begründungen von Fall zu Fall eingehen.

Die klassische Wahrscheinlichkeitsdefinition.

Die „klassische“ Wahrscheinlichkeitsdefinition in der stereotypen Form, in der sie fast ausnahmslos alle Lehrbücher der Wahrscheinlichkeitsrechnung wiedergeben, wurde von LAPLACE in folgender Weise ausgesprochen: Wahrscheinlichkeit ist der Quotient aus der Anzahl der dem Ereignis günstigen Fälle durch die Gesamtzahl aller gleichmöglichen Fälle.

In zum Teil weniger deutlicher Form, aber dem Sinne nach gleich, liegt diese Annahme auch den Arbeiten der älteren Autoren zugrunde. Anderseits muß man, um niemandem Unrecht zu tun, gleich hinzufügen, daß sehr vielen Mathematikern der späteren Zeit die Unzulänglichkeit dieser Definition durchaus bewußt war. POINCARÉ z. B. leitet sie mit den Worten ein: „Man kann kaum eine befriedigende Definition der Wahrscheinlichkeit geben; man pflegt gewöhnlich zu sagen, usf.“ Wir werden später sehen, daß eine folgerichtige, restlose Durchführung der Theorie unter Zugrundelegung der klassischen Definition niemals versucht worden ist. Man fängt mit der Gleichmöglichkeitsdefinition an, verläßt aber in einem passenden Zeitpunkt die gewählte Begriffsabgrenzung und gleitet in eine — manchmal auch explizite ausgesprochene — Häufigkeitsdefinition hinüber. Darum meine ich auch, daß die Gegensätzlichkeit der klassischen Auffassung und der von mir vertretenen nicht unüberbrückbar ist. Für die meisten Mathematiker wird es sich eigentlich nur darum handeln, eine mehr oder weniger liebgewordene Darstellungsform aufzugeben, die es gestattet hat, ein paar einfache Aufgaben zu Beginn des Lehrganges bequem zu erledigen, ohne daß man gleich schwierigere, tiefergehende Fragen zu erörtern brauchte.

Der Haupteinwand gegen die LAPLACEsche Definition muß natürlich an den darin verwendeten Ausdruck „gleichmögliche Fälle“ anknüpfen. Der gewöhnliche Sprachgebrauch kennt verschiedene Gradunterschiede der Möglichkeit oder Realisierbarkeit. Ein Vorgang heißt entweder möglich oder unmöglich, er kann aber auch schwer oder leicht möglich heißen, wenn über den Aufwand, den seine Realisierung erfordert, etwas entsprechendes bekannt ist.

Es ist nur „schwer möglich“, in gewöhnlicher Handschrift im Laufe einer Minute 40 Worte niederzuschreiben, „unmöglich“ 120 Worte zu bewältigen. Dagegen ist es „leicht möglich“, dies mit Hilfe einer Schreibmaschine oder bei Verwendung der Stenographie zu leisten. In diesem Sinne nennt man zwei Vorgänge gleich möglich, wenn sie sich mit gleicher Anstrengung, mit gleichem Aufwand ausführen lassen, und bei JACOB BERNOULLI, einem Vorgänger von LAPLACE, heißt es noch „quod pari facilitate mihi obtingere possit“. Aber dies ist es jedenfalls nicht, was die LAPLACEsche Definition meint. In anderer Bedeutung sagt man von einem Ereignis, es sei „eher möglich“ als ein anderes, wenn man damit einer Vermutung über das voraussichtliche Eintreffen Ausdruck geben will. Es kann nicht zweifelhaft sein, daß die Gleichmöglichkeit, von der die klassische Wahrscheinlichkeitsdefinition spricht, in diesem Sinne zu verstehen ist und daß sie darum nichts anderes bedeutet als gleichberechtigte Vermutung, also im Sinne des gewöhnlichen Sprachgebrauches: gleiche Wahrscheinlichkeit. Der Ausdruck „gleichmögliche Fälle“ ist vollkommen synonym mit „gleich wahrscheinliche Fälle“ und mehr als diesen einfachen Tatbestand kann auch eine umfangreiche Arbeit „Über Möglichkeit und Wahrscheinlichkeit“, wie sie A. MEINONG veröffentlicht hat, nicht ergeben. Erinnert man sich, daß wir das Auftreten untereinander gleicher Wahrscheinlichkeiten in einem Kollektiv als „Gleichverteilung“ bezeichnet haben, so kann man sagen: Die klassische Definition stellt sich, wenn man sie nicht geradezu als Zirkel ansehen will, als eine Zurückführung allgemeinerer Verteilungen auf den einfacheren Fall der Gleichverteilung dar.

Die gleichmöglichen Fälle....

Sehen wir nun zu, in welcher Weise diese Zurückführung tatsächlich bewirkt wird. Beim richtigen Würfel gibt es sechs gleichmögliche Fälle, einer davon liefert die Augenzahl drei, also ist die Wahrscheinlichkeit, drei zu werfen, gleich $^1/_6$. Enthält ein Glücksrad die Lotterienummern 1 bis 90, so hat man 90 gleichmögliche Fälle; 9 von diesen 90 entsprechen einziffrigen Nummern, 9 andere solchen zweistelligen, die durch 10 teilbar sind, die übrigen 72 den zweiziffrigen, nicht durch 10 teilbaren. Also ist die Wahrscheinlichkeit einer einstelligen oder einer durch 10 teilbaren Zahl je $^9/_{90} = {}^1/_{10}$, die der übrigen Lottonummern zusammen $^8/_{10}$. Ein drittes Beispiel: Beim Spiel mit zwei richtigen Würfeln gilt jede der 36 Zahlenkombinationen als gleichmöglicher Fall; daher ist die Wahrscheinlichkeit, Doppel-Sechs zu werfen, gleich $^1/_{36}$, die, die Summe 11 zu werfen, gleich $^1/_{18}$, weil es hier zwei „günstige" Fälle gibt, nämlich die Augenzahl 5 auf dem ersten und 6 auf dem zweiten Würfel und umgekehrt. Zu diesen drei Anwendungen der Gleichmöglichkeitsdefinition ist folgendes zu sagen.

Die erste ist eine reine Tautologie, wenn man berücksichtigt, daß gleichmöglich und gleichwahrscheinlich dasselbe heißt; nur daß die Summe der Wahrscheinlichkeiten aller überhaupt möglichen Versuchsausgänge gleich 1 ist, wird hier mit benutzt. Im zweiten Beispiel, in dem mehrere „günstige" Fälle zusammengefaßt werden, hat man nichts anderes als einen speziellen Fall der Mischung von Kollektivs, wie ich ihn schon vorher erwähnt habe; es werden die Merkmale 1 bis 9, dann die Merkmale 10, 20, 30, ..., 90, endlich die übrigen zu je einem neuen zusammengefaßt und die Addition der einzelnen Wahrscheinlichkeiten, die hier eben alle gleich $^1/_{90}$ sind, führt zu dem angeführten Ergebnis. Im dritten Beispiel, das ein Spiel mit zwei Würfeln betrifft, benutzt die klassische Theorie einen Satz, der die Verbindung unabhängiger Kollektivs in einer ebenfalls stark spezialisierten Form regelt und der besagt: Jede Kombination aus einem der gleichmöglichen Fälle des ersten und des zweiten Kollektivs liefert einen gleichmöglichen Fall des neuen Kollektivs. Da man mit Mischung und Verbindung, wie wir wissen, jedenfalls die meisten und wichtigsten Aufgaben der Wahrscheinlichkeitsrechnung beherrscht, so kann man somit im Rahmen der Gleichmöglichkeitstheorie im wesentlichen alle jene Aufgaben erledigen, in denen die Verteilungen in den Ausgangskollektivs als gleichförmige, als Gleichverteilungen gegeben sind. Dies trifft im allgemeinen stets dann zu, wenn es sich um normale Glücksspielaufgaben, um richtige Würfel, eine fehlerfreie Roulette oder dergleichen handelt.

....sind nicht immer vorhanden.

Wie aber kann man dem Fall eines „falschen" Würfels mit einer Theorie gerecht werden, die von vornherein eine Wahrscheinlichkeit nur innerhalb eines Bereiches gleichwahrscheinlicher Möglichkeiten kennt? Sicher ist, daß, wenn wir einen richtigen Würfel noch so wenig anfeilen, die Gleichwahrscheinlichkeit der sechs Würfelseiten verloren geht. Gibt es nun keine Wahrscheinlichkeit mehr, die Augenzahl drei zu werfen? Kann man jetzt nicht mehr behaupten, daß die Wahrscheinlichkeit eines geradzahligen Wurfes die Summe der Wahrscheinlichkeiten der 2, 4 und 6 ist? Der Wortlaut der klassischen Definition läßt jedenfalls keine Anwendung der früher abgeleiteten Sätze zu, denn nach ihr gibt es einfach keine Wahrscheinlichkeit, wo es keine gleichmöglichen Fälle gibt. Gleichwohl hat LAPLACE in seinem grundlegenden Werk versucht, den Fall einer Münze mit ungleichen Chancen für die Resultate „Kopf" und „Adler" zu behandeln. Er gelangt zu Schlüssen, die heute niemand anerkennt. Die späteren Lehrbücher der klassischen Theorie haben diese Betrachtungen einfach fortgelassen, sie betrachten den „falschen" Würfel überhaupt als keine beachtenswerte Aufgabe der Wahrscheinlichkeitsrechnung; nun darüber läßt sich schließlich nicht streiten.

Aber es gibt ja andere hierher gehörige Fragen, über die man unmöglich hinweggehen kann und in denen es durchaus nicht besser liegt als beim falschen Würfel, z. B. die Frage nach der Sterbenswahrscheinlichkeit. Wenn eine Tabelle angibt, daß 0,011 die Wahrscheinlichkeit dafür ist, daß ein 40jähriger Mann im Laufe eines Jahres stirbt, wo sind da die gleichmöglichen und die günstigen Fälle? Sind es 1000 Möglichkeiten, von denen 11 dem Todeseintritt günstig sind, oder 3000 mit 33 günstigen? Eine Antwort darauf oder gar eine *Erklärung darüber, was man sich hier unter den abgezählten Möglichkeiten vorstellen* soll, wird man vergeblich in den Lehrbüchern suchen. Wenn sie nämlich an die Stelle gelangt sind, an der von Sterbenswahrscheinlichkeit und ähnlichem die Rede ist, haben sie längst vergessen, daß alle Regeln und Sätze der Theorie unter Zugrundelegung einer Definition, die nur Gleichverteilung als Ausgangspunkt kennt, abgeleitet wurden. Ganz unbefangen wird jetzt behauptet, man könne Wahrscheinlichkeiten, die nicht „a priori" bekannt sind, empirisch oder „a posteriori" bestimmen, indem man in einer genügend langen Beobachtungsreihe die relativen Häufigkeiten der verschiedenen Ereignisse feststellt. Und mit der größten Kühnheit werden alle Sätze, die unter einem ganz anderen Gesichtspunkt abgeleitet worden sind, auf die neuen Wahrscheinlichkeiten übertragen. Wer etwas übriges tun will, beruft sich dabei auf das sogenannte BERNOULLIsche Theorem oder das Gesetz der großen Zahlen, das angeblich die Brücke zwischen der a priorischen und der empirischen Wahrscheinlichkeitsbestimmung bilden soll.

Daß dies durchaus nicht der Fall ist, daß hier ein ganz glatter Zirkelschluß vorliegt, werde ich später ausführlich zeigen. Aber wie dem auch sei, können wir schon jetzt feststellen, daß man ohne die Deutung der Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit niemals auskommt und niemals auskommen kann, will man nicht gerade die praktisch wichtigsten Fälle der Anwendung ausschließen. Was mag es da für einen Sinn haben, vorerst an einer Definition festzuhalten, die doch zu eng ist, um alle Fälle zu umfassen oder sich nur in ganz gezwungener Weise vielen unausweichlichen Aufgaben gefügig machen läßt? Eine einfache Analogie aus dem Gebiete der elementaren ebenen Geometrie wird uns die Verhältnisse gut veranschaulichen.

Eine geometrische Analogie.

Man könnte es sich in den Kopf setzen, eine Geometrie der von geraden Linien begrenzten Figuren in der Ebene (also der Vielecke oder Polygone) derart aufzubauen, daß man ausschließlich gleichseitige Vielecke betrachtet, und zwar solche, bei denen alle Seitenlängen überhaupt ein und dieselbe Größe besitzen. In dieser Geometrie gibt es keine Längenmessung; jedes Gebilde wird durch seine Winkel und die Anzahl der Seiten vollständig gegeben. Hat man es dann einmal mit einem Dreieck zu tun, dessen Seiten nach der gewöhnlichen Auffassung die 3-, 4- bzw. 5fache Länge der Einheit haben, so wird man dieses Dreieck als ein gleichseitiges Zwölfeck erklären, von dem erst drei Seiten unter gestreckten Winkeln aneinandergefügt sind, dann weitere vier und schließlich die letzten fünf ebenso. Sobald die Längen der Seiten ganze Vielfache der Einheitslänge sind — und bei genügend kleiner Einheit kann man mit beliebiger Annäherung jede Strecke in dieser Weise auffassen —, macht diese Zurückführung auf gleichseitige Polygone keine grundsätzlichen Schwierigkeiten. Man wird auch hier zu der Unterscheidung geführt zwischen Vielecken, deren Seitenzahl „a priori" gegeben ist (weil alle benachbarten Seiten von 180^0 verschiedene Winkel bilden), und solchen, die man erst „a posteriori“ durch Längenvergleichung mehr oder weniger genau in die Klasse der gleichseitigen Vielecke einordnen kann.

Die Durchführung einer derartigen Theorie ist sicher möglich, aber kein Vernünftiger wird behaupten, daß auf diese Weise der Begriff der Streckenlänge oder der Längenmessung wirklich aus der Geometrie ausgeschaltet werden könnte; er wird nur auf einen Umweg verwiesen. Ganz genau so verhält es sich mit der Gleichmöglichkeitstheorie der Wahrscheinlichkeit. Es ist historisch verständlich, daß man mit dem Analogon der gleichseitigen Vielecke, den gleichmöglichen Fällen, angefangen hat, da Glücksspielfragen die ersten Aufgaben bildeten, die in der Wahrscheinlichkeitsrechnung behandelt wurden. Wenn man aber heute die tatsächlich als relative Häufigkeit ermittelte Lebens- und Sterbenswahrscheinlichkeit hinterher durch Zurückführung auf rein hypothetische „gleichmögliche“ Fälle zu deuten sucht, so spielt man nur Verstecken mit der Notwendigkeit einer umfassenderen Wahrscheinlichkeitsdefinition, die sich am Ende ebensowenig entbehren läßt, wie der Begriff der Längenmessung oder Längenvergleichung in der Geometrie.

Das Erkennen der Gleichmöglichkeit.

Damit glaube ich deutlich gemacht zu haben, wie sich meine Wahrscheinlichkeitsdefinition zu der bisher vorgezogenen Darstellungsweise im großen ganzen verhält. Wenn mit der Zeit die historisch überkommenen Glücksspielaufgaben gegenüber den wichtigeren Fragen, die mit Versicherungswesen, Statistik und Fehlertheorie zusammenhängen, noch mehr zurücktreten werden, als dies schon heute der Fall ist, wird man sich, denke ich, ganz von selbst einem Aufbau der Wahrscheinlichkeitsrechnung zuwenden, der allein den Forderungen der Einfachheit und Vernünftigkeit genügt.

Andrerseits ist mit dem, was sogenannte logische oder erkenntnistheoretische Untersuchungen zur Stützung, zur Begründung, zur Vertiefung der klassischen Wahrscheinlichkeitsrechnung beigetragen haben, die neue Auffassung, die ich vertrete, in keiner Weise vereinbar. Als die grundlegende Frage in allen diesen Untersuchungen erschien nämlich etwas, was überhaupt keine Frage ist, wenn man die Bedeutung der Gleichmöglichkeitsdefinition auf ihr richtiges Maß zurückführt. Es wurden immer wieder und wieder Nachforschungen darüber angestellt: Woher wissen wir, daß die sechs Seiten des Würfels, daß die Ziehungen der 90 Nummern beim kleinen Lotto usf. lauter gleichmögliche Fälle bilden? Meine Antwort darauf ist: Wir wissen es im allgemeinen, d. h. für einen beliebigen Würfel, für eine beliebige Lotterieurne überhaupt nicht, sondern ausschließlich für eine bestimmte, durch hinreichend lange Beobachtung erprobte Anordnung und wir vermuten es dann natürlich für jede, die nach dem Muster einer erprobten genau nachgebildet ist. Von den beiden Würfelpaaren, die ich im ersten Vortrag vorführte und die äußerlich ganz gleich aussehen, zeigt das eine ein Verhalten, das der Annahme, die sechs Seiten seien gleichwahrscheinlich, gründlich widerspricht. Der Unterschied zwischen den richtigen und den falschen Würfeln läßt sich hier wohl schon durch rohes Abwägen in der Hand oder durch einen einfachen mechanischen Versuch feststellen. Ob etwas ähnliches auch bei feineren Fälschungen immer möglich ist, ob nicht manchmal der statistische Versuch der einzige Weg zur Klarstellung ist, mag vorläufig unentschieden bleiben. Wir wollen uns, bevor wir darauf eingehen, noch ein wenig mit den Quellen, aus denen die Philosophen ihre Fragestellung herleiten, beschäftigen; es dreht sich dabei, wie Sie schon bemerkt haben werden, um die Auffassung einer Wahrscheinlichkeit als „Erkenntnis a priori“.

Die Wahrscheinlichkeit „a priori“.

Wenn man aus vollkommen homogenem Material einen Körper herstellt, der geometrisch vollkommen genau die Würfelgestalt besitzt, dann, meint man, sei doch a priori evident, daß keine Seite vor der anderen bevorzugt ist, daß also das Auffallen auf jede der sechs Seiten gleichmöglich, d. h. gleichwahrscheinlich ist. Ich will das vorerst einmal zugeben, trotz der großen Schwierigkeiten, die daraus entstehen, daß es ja gar nicht allein auf den Würfelkörper ankommt, sondern auch noch auf die Beschaffenheit des Bechers, auf den ganzen Vorgang, mit dem der Würfel jedesmal erfaßt, in den Becher getan, dieser geschüttelt wird usf. Ich will auch ganz davon absehen, daß der Satz erst dann einen Inhalt gewinnt, also überhaupt eine Erkenntnis, sei es a priori oder nicht, darstellt, wenn man schon weiß, was „gleich wahrscheinlich“ heißt, also z. B. im Sinne der

Häufigkeitsdefinition dem Ausdrucke die Bedeutung beilegt, daß die sechs Seiten bei fortgesetztem Würfeln gleich oft auffallen. Nehmen wir an, es handle sich nur darum, den sechs Würfelseiten irgendwelche Zahlen zuzuordnen und es frage sich jetzt lediglich, ob hier wirklich ein von der Erfahrung unabhängiger, denknotwendiger Schluß dahin führt, daß beim homogenen Würfel diese sechs Zahlen gleich sein müssen. Überlegt man nun einmal die Voraussetzungen der Homogenität und der Symmetrie genau, so sieht man bald, wie wenig von der ganzen Aussage übrig bleibt; viel weniger jedenfalls, als auch nur die bescheidenste Anwendung auf einen konkreten Fall erfordert.

Als homogen im Sinne der logischen Bedeutung der Aussage können wir ein Material nur dann bezeichnen, wenn von keinem seiner Teile etwas Besonderes ausgesagt werden kann, wenn also alle seine Teile auch genau die gleiche Entstehungsgeschichte durchgemacht haben. Die eine Hälfte eines Elfenbeinwürfels war aber sicher im Elefantenzahn, aus dem das Material stammt, der Spitze näher gelegen als die andere; dann ist es schon nicht mehr denknotwendig, daß beide Hälften des Würfels sich ganz gleich verhalten, sondern dies beruht auf dem Erfahrungssatz, daß jene ursprüngliche Lage im Tierkörper keinen Einfluß auf den Ablauf der Würfelversuche besitzt. In der Tat benutzen wir in jedem einzelnen Falle noch viel weitergehende Erfahrungen. Wir beschreiben z. B. die sechs Seiten des Würfels mit verschiedenen Zahlzeichen und nehmen an, daß dies ohne jeden Einfluß ist; während bekanntlich primitive, d. h. im weitesten Umfang unerfahrene Volksstämme davon überzeugt sind, daß man durch Bemalen eines menschlichen Körperteiles mit derartigen Zeichen das Lebensschicksal des betreffenden Menschen wenden kann. Wir beschränken uns auch nicht auf aufgemalte Zeichen, sondern machen ein bis sechs Kerben auf die Würfeloberfläche und verändern damit die geometrische Symmetrie in sichtbarer Weise, immer auf Grund der Erfahrung, daß dies nichts ausmacht.

Wenn man einmal einen Vertreter der a priori-Auffassung zu einer deutlichen Erklärung zwingt, was er eigentlich unter der vollkommenen Homogenität versteht, so beschränkt er sich schließlich auf die Forderung, daß der Schwerpunkt des Körpers mit dem geometrischen Mittelpunkt zusammenfallen muß und — falls der Befragte über genügende Kenntnis der Mechanik verfügt

— daß die Trägheitsmomente für die zwölf Kanten als Drehachsen gleich sein sollen. Daß nun gerade diese Bedingungen für die „Gleichmöglichkeit“ der sechs Würfelseiten notwendig und hinreichend sind (und nicht z. B. noch Momente höheren Grades in Betracht kommen), wird niemand mehr für a priori evident erklären: ein erheblicher Teil der in letzter Linie auf Erfahrungssätzen aufgebauten Mechanik starrer Körper steckt in dieser Formulierung. Das Ergebnis unserer Überlegung läßt sich also dahin zusammenfassen: Aus dem, was man a priori, was man denknotwendig als gleichmöglich erkennt, gewinnt man nichts Ausreichendes zur Beurteilung eines bestimmten realen Falles; man muß immer noch mehr oder weniger allgemeine, aus Beobachtung und Erfahrung abgeleitete Kenntnisse hinzunehmen, die darüber belehren, welche Eigenschaften einer Anordnung auf den Ablauf der Versuche von Einfluß sind und welche nicht.

Es liegt hier ganz ähnlich wie bei der bekannten Anwendung des Symmetrieprinzips zur Ableitung der Gleichgewichtsbedingung am gleicharmigen Hebel. Wenn die beiden Hebelhälften völlig gleich sind, meint man, müsse die Bedingung, rein aus „Symmetriegründen“, dahin lauten, daß auch die Kräfte gleich groß seien. Aber in dieser Form ist der Satz ja viel zu eng. Abgesehen davon, daß die Gleichheit der Hälften in dem früher erörterten strengen logischen Sinn gar nicht erreichbar ist, versteht man doch unter einem gleicharmigen Hebel einen solchen, bei dem die Kräfte gleich weit von der Drehachse angreifen, ohne daß volle Symmetrie in der geometrischen Gestaltung bestehen muß. Es ist da sehr lehrreich zu sehen, wie ältere Lehrbücher der technischen Mechanik durch viele Abbildungen von verschieden gestalteten Hebeln dem Leser den Satz vom Gleichgewicht bei Gleicharmigkeit trotz Unsymmetrie vertraut zu machen suchen. Daß es eben auf die Abstände der Lasten vom Drehpunkt und nur auf diese ankommt, darin liegt die entscheidende, aus Beobachtungen geschöpfte Erkenntnis.

Sonderstellung der Gleichmöglichkeit?

Ist es nun nichts mit der „a priori-Erkenntnis“ der Gleichmöglichkeit, so bleibt noch der folgende Ausweg denkbar, um den gleichmöglichen Fällen eine Sonderstellung gegenüber allen anderen Wahrscheinlichkeitsbestimmungen zu sichern. Man könnte sagen: Wenn ein Würfel außer der geometrischen Symmetrie auch die (vorhin schon erwähnten) Eigenschaften besitzt, die man als „kinetische Symmetrie“ bezeichnen kann, daß nämlich die statischen Momente und die Trägheitsmomente für die zwölf Würfelkanten gleich sind, dann folge schon aus der Mechanik starrer Körper, daß die Wahrscheinlichkeit des Auffallens irgendeiner Seite gleich $1/_6$ ist. Sind aber für einen falschen Würfel alle mechanischen Größen, Schwerpunktslage, Trägheitsmomente usf. genau bekannt, so lasse sich in keiner Weise aus den Sätzen der Mechanik berechnen, mit welcher relativen Häufigkeit eine Seite auffällt; das einzige Mittel der Wahrscheinlichkeitsbestimmung bleibe dann der statistische Versuch, während in dem anderen Fall eine Voraussage, wenn schon nicht „a priori“, so doch auf Grund einer Erfahrungswissenschaft von deterministischem Charakter möglich sei.

Dieser Gedankengang enthält, wie ich meine, einen Fehlschluß. Ich habe schon früher bemerkt, daß nicht der Würfelkörper allein für den Ablauf einer Versuchsreihe, die aus wiederholtem Würfeln besteht, entscheidend ist. Man kann offenbar auch mit einem allen Symmetriebedingungen genügenden Würfel falsch spielen, und zwar bewußt oder auch unbewußt, indem man irgendwie beim Einlegen des Würfels in den Becher oder beim jedesmaligen Rütteln des Bechers „falsch“ vorgeht. Dabei können ganz feine psychologische oder sinnesphysiologische Erscheinungen in Frage kommen, wie man teils aus den Erfahrungen mit Taschenspielern, teils aus neueren Beobachtungen weiß, die — zum Teil unaufgeklärt — in die sogenannte Parapsychologie eingereiht werden. Selbstverständlich will ich hier nicht als Fürsprecher irgendeiner „okkulten Wissenschaft“ auftreten, aber ich glaube bestimmt, daß wir, ganz auf dem Boden der bisherigen empirischen Methoden, in vorurteilsfreier Fortführung und Verarbeitung von Beobachtungsergebnissen, noch zur Entdeckung von Zusammenhängen gelangen werden, die uns heute unbekannt sind. Wie dem auch sei, nach dem heutigen Stande unseres Wissens scheint es jedenfalls unmöglich, restlos alle Bedingungen „theoretisch“ anzugeben, durch welche das gleich häufige Auffallen der sechs Würfelseiten bei einer genügend ausgedehnten Versuchsreihe sichergestellt wird. Unter „theoretisch“ ist dabei verstanden eine Angabe, die sich zwar auf irgendwelche empirisch begründete Wissenschaftszweige stützt, aber nicht auf einen statistischen Versuch, der mit der vorliegenden oder einer ihr gleichen Anordnung ausgeführt wurde.

Es ist ein wesentlicher, wenn auch von der Grundlegung der Wahrscheinlichkeitsrechnung im engeren Sinne unabhängiger Bestandteil meiner Auffassung der statistischen Vorgänge, den ich dahin formulieren möchte: Das Bestehen der Gleichverteilung innerhalb eines Kollektivs läßt sich (ebenso wie das jeder anderen Verteilung) ausschließlich und allein durch einen genügend lang fortgeführten Wiederholungsversuch feststellen; der Versuch kann mit dem gegebenen Objekt oder mit einem anderen, das man auf Grund entsprechender Beobachtungen für gleichwertig hält, ausgeführt werden. Der Satz hat seine Bedeutung zunächst für die Verteilung innerhalb des Ausgangskollektivs, die man bei jeder Aufgabe der Wahrscheinlichkeitsrechnung als gegeben voraussetzen muß, in zweiter Linie auch für die Verteilungen innerhalb des abgeleiteten Kollektivs, sobald man den Wunsch hat, die Ergebnisse der Rechnung durch Beobachtung zu überprüfen.

Der subjektive Wahrscheinlichkeitsbegriff.

In schärfstem Gegensatz steht die eben dargelegte Auffassung der Gleichverteilung als eines speziellen, aber nicht durch prinzipielle Sonderstellung ausgezeichneten Beispieles einer allgemeinen Verteilung zu dem Standpunkt der erkenntnistheoretischen Verfechter des sogenannten subjektiven Wahrscheinlichkeitsbegriffes. Nach ihrer Meinung hängt die Wahrscheinlichkeit, die wir einem bestimmten Ereignis — oder eigentlich der Behauptung, daß dieses Ereignis eintritt — zuschreiben, ausschließlich von dem Grade unseres Wissens ab und die Annahme der Gleichwahrscheinlichkeit verschiedener Ereignisse ist eine Folge reinen Nichtwissens. Ich habe schon einmal die überaus charakteristische, knappe Formulierung von Czuber angeführt, wonach „absolutes Nichtwissen über die Bedingungen“ zu der Annahme führe, alle möglichen Fälle seien gleich wahrscheinlich. In etwas wissenschaftlicherer Aufmachung nennt man dies das „Indifferenzprinzip“. Sehr richtig bemerkt J. M. Keynes, daß aus dem Indifferenzprinzip eigentlich folgt, jeder Satz, über dessen Richtigkeit nichts bekannt ist, habe die Wahrscheinlichkeit $^1/_2$; denn er bildet ja mit seinem kontradiktorischen Gegenteil zwei gleichmögliche Fälle. Weiß man z. B. nichts von einem Buch, so hat die Behauptung, es sei rot, die Wahrscheinlichkeit $^1/_2$, aber ebenso die, es sei blau, gelb oder grün, so daß eine beliebig große Wahrscheinlichkeitssumme herausgerechnet werden kann. KEYNES gibt sich große Mühe, diese Klippe der Subjektivitätstheorie zu umschiffen, aber er hat nicht viel Glück damit. Er gibt wohl eine formale Regel, die die Anwendung des Indifferenzprinzips in dem genannten Fall ausschließt, aber ohne etwas anderes an dessen Stelle zu setzen. Auf den einfachen Gedanken, man könne, wenn man von einer Sache nichts weiß, eben auch über ihre Wahrscheinlichkeit nichts aussagen, kommt er nicht.

Der sonderbare Denkfehler der „subjektiven" Wahrscheinlichkeitsauffassung wird, wie ich glaube, durch folgende Überlegung aufgeklärt. Wenn man von der Körperlänge von sechs Personen nichts weiß, mag man die Vermutung aussprechen, sie seien alle sechs gleich groß (hier ist auch nach der KEYNESschen Regel diese Anwendung des Indifferenzprinzips zulässig). Die Vermutung kann richtig oder falsch sein, man kann sie auch — im Wortsinn des üblichen Sprachgebrauches — als mehr oder weniger wahrscheinlich bezeichnen. Ebenso läßt sich von sechs Würfelseiten, über die man nichts weiß, die Vermutung hegen, daß sie „gleichmöglich" seien; aber mehr als eine Vermutung ist das eben auch nicht, und sie kann ebensowohl richtig wie falsch sein, wie die im ersten Vortrag vorgeführten Würfelpaare es zeigen. Hier greift nun der seltsame Gedankengang der Subjektivisten ein: „Ich halte die Fälle für gleich wahrscheinlich" wird gleichgesetzt mit „Die Fälle sind gleich wahrscheinlich" — weil Wahrscheinlichkeit nur etwas Subjektives ist. Dagegen schließt man nicht, daß sechs unbekannte Personen wirklich gleiche Körperlänge haben, u. zw. deshalb, weil die Länge „objektiv meßbar" ist. Wenn man diesen Unterschied aufrechterhält, „folgt" allerdings aus dem Nichtswissen die Aussage der Gleichwahrscheinlichkeit, aber mit demselben Recht auch jede andere, z. B. die, daß die Wahrscheinlichkeiten sich wie die Quadrate der Zahlen 1 bis 6 zueinander verhalten — denn als Vermutung ist auch dies möglich.

Ich glaube wohl, daß die meisten Menschen, wenn man sie fragt, wo der Schwerpunkt eines unbekannten Würfels liegen mag, antworten werden: „Wahrscheinlich in der Mitte". Aber diese

Antwort gründet sich keineswegs auf Nichtwissen, sondern auf die nicht bezweifelbare Tatsache, daß sicher die meisten Würfel, die bisher hergestellt wurden, annähernd richtige waren. Eine eingehende psychologische Untersuchung darüber, worauf die subjektive Wahrscheinlichkeitsschätzung beruht, ist gewiß nicht undenkbar. Sie würde sich zur Wahrscheinlichkeitsrechnung so verhalten, wie eine Untersuchung des subjektiven Wärmegefühls zur physikalischen Thermodynamik. Auch die Thermodynamik hat zu ihrem Ausgangspunkt das Vorhandensein von Wärmeempfindungen, aber sie beginnt damit, daß an die Stelle der subjektiven Wärmeschätzung ein objektives, vergleichbares Maß, die Länge der Quecksilbersäule, gesetzt wird. Daß Temperaturskala und menschliche Wärmeempfindung nicht immer parallel laufen, daß ihre Beziehung durch psychologische und physiologische Einflüsse aller Art gestört wird, ist eine sehr bekannte Erscheinung. Durch sie werden Brauchbarkeit, Wert und Leistungen der physikalischen Wärmelehre nicht beeinträchtigt und niemand denkt daran, ihr zuliebe den Aufbau der Thermodynamik abzuändern. Nun, wie ich schon einmal andeutete: Wiederholungsversuch und Häufigkeitsbestimmung sind das Thermometer der Wahrscheinlichkeit.

Die Spielraumtheorie.

In eine etwas feinere, aber nicht viel widerstandsfähigere Form hat J. v. KRIES in einem oft zitierten Buche die Lehre von der Gleichmöglichkeit gebracht. Er meint, daß in den hier in Betracht kommenden Fragen unserem Urteil gewisse Spielräume offen stehen und wir ein natürliches Empfinden dafür besitzen, wie diese Spielräume einzuteilen sind, damit jeder Teil einem gleichmöglichen Fall entspricht. Er spricht von „indifferenten Spielräumen", von „freier Erwartungsbildung" und von „zwingend bestimmter, der Willkür entzogener Aufstellung gleichberechtigter Annahmen". Als Grundlage der Theorie erscheint der Satz, dessen nichtempirischen Charakter v. KRIES ausdrücklich betont, daß „Annahmen, welche gleiche und indifferente ursprüngliche Spielräume umfassen, gleich wahrscheinlich sind". Nach meiner Auffassung ist hier nur das Wort „gleichwahrscheinlich" durch den Ausdruck „gleiche Spielräume" ersetzt. Einen Sinn bekommt das erst, wenn man für jeden Fall, in dem die Wahrscheinlichkeitsrechnung zur Anwendung kommt, ein Verfahren angibt, nach dem wenigstens annähernd festgestellt werden kann, welche Spielräume gleich groß sind. Da gibt es nun wohl keine andere Möglichkeit als die Häufigkeitsbestimmung im Wiederholungsversuch und damit käme man zu meinem Standpunkt zurück.

Die Anwendung der KRIESschen Theorie hat man sich etwa so zu denken: Wenn wir z. B. ein Lotterielos kaufen, so ist unserer Erwartung der Gesamtspielraum gesetzt, daß eines der 10000 ausgegebenen Lose den Haupttreffer gewinnt; wir teilen diesen Spielraum in 10000 „gleiche" Teile und gelangen so zu dem Urteil, daß wir als Besitzer *eines* Loses die Wahrscheinlichkeit $^1/_{10\,000}$ besitzen, das große Los zu ziehen. Es mag dahingestellt bleiben, wie weit dies ein zwingender Gedankengang ist. Ganz zweifellos scheitert die Spielraumtheorie, d. h. die Auffassung, daß es *denknotwendige* Annahmen über Gleichheit der Spielräume gibt bei denjenigen Aufgaben, die BERTRAND zuerst behandelt und für die dann POINCARÉ die Bezeichnung „*BERTRANDsche Paradoxie*" eingeführt hat. Ich will an einem möglichst einfachen Fall die unüberwindlichen Schwierigkeiten darlegen, die jeder Form der klassischen Wahrscheinlichkeitstheorie, namentlich auch der KRIESschen Lehre der indifferenten Spielräume (der sogenannten logischen Theorie) hier erwachsen.

BERTRANDsche Paradoxie.

Denken wir uns, um einen möglichst einfachen Fall vor Augen zu haben, ein Glas mit einer Mischung von Wasser und Wein gefüllt. Von dem Mischungsverhältnis sei nur bekannt, daß der Mengenquotient *Wasser zu Wein* mindestens gleich 1 und höchstens gleich 2 ist, d. h. daß mindestens ebensoviel Wasser wie Wein und höchstens doppelt soviel Wasser wie Wein sich in dem Glase befindet. Der Spielraum unseres Urteils über das Mischungsverhältnis liegt also zwischen 1 und 2. Wenn nichts weiter darüber bekannt ist, so muß man nach dem Indifferenzprinzip (oder Symmetrieprinzip), nach der Spielraumtheorie, wie überhaupt nach jeder ähnlichen Art von Wahrscheinlichkeitsauffassung annehmen, daß gleich große Teilgebiete des Spielraumes gleiche Wahrscheinlichkeit haben. Es muß also 50% Wahrscheinlichkeit dafür bestehen, daß das Mischungsverhältnis zwischen 1 und 1,5 liegt und ebensoviel dafür, daß es in den Bereich 1,5 bis 2 fällt. Was aber hindert uns, an Stelle dieser Überlegung die folgende zu setzen?

Von dem Mengenverhältnis Wein zu Wasser ist uns bekannt, daß es zwischen 1 und $^1/_2$ liegt und sonst nichts. Also muß man annehmen, daß jede Hälfte dieses Spielraumes gleiche Wahrscheinlichkeit besitzt. Somit besteht 50% Wahrscheinlichkeit dafür, daß mindestens $^1/_2$ Wein zu Wasser und höchstens $^3/_4$ Wein zu Wasser in dem Glase vorhanden ist, oder was dasselbe ist, Wasser zu Wein mindestens im Verhältnis 4:3 und höchstens im Verhältnis 2:1 steht. Nach der ersten Rechnung entfiel die halbe Wahrscheinlichkeit auf den Spielraum 1,5 bis 2, nach der zweiten auf den Spielraum $^4/_3$ bis 2, was ein offenkundiger Widerspruch ist.

Ganz den gleichen Widerspruch kann man in sämtlichen Fällen aufzeigen, in denen die unterschiedlichen Merkmale (hier die Werte des Mischungsverhältnisses) nicht auf einzelne diskrete Zahlen (wie beim Würfel die ganzen Zahlen 1 bis 6) beschränkt sind, sondern einen kontinuierlichen Wertebereich (in unserem Beispiel alle Zahlen zwischen 1 und 2) erfüllen. Ich habe schon an früherer Stelle von derartigen Aufgaben kurz gesprochen und auch schon den hier üblichen, nicht sehr passenden Namen erwähnt: man spricht von „geometrischer Wahrscheinlichkeit", weil die meisten Problemstellungen dieser Art einen mehr oder weniger geometrischen Charakter tragen. Eine der ältesten und bekanntesten Aufgaben ist wohl das Nadelproblem von Buffon (1733), das darin besteht, daß eine Nadel blindlings auf den Boden geworfen wird, auf dem parallele Gerade in gleichförmigen Abständen gezogen sind, und man nach der Wahrscheinlichkeit fragt, mit der Nadel einen dieser Striche zu treffen (zu kreuzen). Der Einzelversuch ergibt als Merkmal die Lage der Nadel gegenüber dem Strichgitter, die in verschiedentlicher Art durch Zahlen, sogenannte Koordinaten, bestimmt werden kann. Gewissen Koordinatenwerten entspricht dann ein Überkreuzen eines Gitterstriches, also ein positives Versuchsergebnis, anderen ein negatives. Der Widerspruch entsteht genau so wie in unserem ersten Beispiel eben dadurch, daß es in mehrfacher Weise möglich ist, die Lage der Nadel, also den Erfolg des Einzelversuches in Zahlen auszudrücken. Das Mischungsverhältnis im Weinglas ließ sich sowohl durch den Mengenquotienten Wasser zu Wein, wie durch den reziproken, Wein zu Wasser, bestimmen; analog kann man jetzt Cartesische Koordinaten, Polarkoordinaten oder noch andere verwenden. Gleichverteilung für den einen Quotienten, bzw. den einen Koordinatenansatz, ist keine Gleichverteilung für den anderen.

An dieser Klippe muß unfehlbar jede Theorie scheitern, für die es als Ausgangswahrscheinlichkeiten nur Gleichverteilungen gibt, die irgendwie gefühlsmäßig, instinktiv, a priori oder in ähnlicher Weise erschlossen werden. Nach KEYNES, den ich schon wiederholt als konsequenten „Subjektivisten" angeführt habe, sind in den genannten Fällen tatsächlich verschiedene, gleichberechtigte Annahmen möglich, die zu verschiedenen Ergebnissen führen. Der Standpunkt der Häufigkeitstheorie ist einfach der, daß die Verteilung innerhalb der Ausgangskollektivs gegeben sein muß; woher man sie kennt, wie sie beschaffen ist, hat nichts mit der Wahrscheinlichkeitsrechnung zu tun. Will man aber in einem konkreten Fall Resultate haben, die mit der Erfahrung in Übereinstimmung stehen, so muß man die Ausgangswerte der Rechnung aus Beobachtungen entnehmen, die bei der betreffenden Anordnung gewonnen wurden. In dem Beispiel der Wasser-Wein-Mischung ist es schwer möglich, in vernünftiger Weise ein Kollektiv zu definieren; man müßte erst einen konkreten Vorgang kennen, nach dem ein Glas mit einem bestimmten Mischungsverhältnis „erfaßt" wird. Beim Nadelproblem ist der Versuchsvorgang einigermaßen gegeben; eine Nadel wird in einer noch näher zu bestimmenden Weise gegen den Boden geworfen; die Verteilung innerhalb dieses Ausgangskollektivs wird durch eine „Wahrscheinlichkeitsdichte" bestimmt, die auf irgendwelche Koordinaten bezogen sein mag. Wenn man durch eine wirklich durchgeführte Versuchsreihe die Dichtefunktion, bezogen auf irgendein beliebiges Koordinatensystem ermittelt, so ist es ganz gleichgültig, welche Koordinaten man gewählt hat, ein Widerspruch kann bei der weiteren Rechnung in keiner Weise herauskommen. Die Aufgabe selbst gehört zu dem Typus der Mischung: es werden diejenigen Koordinatenwerte zusammengefaßt, die eine Überdeckung ergeben, und diejenigen, für die sich eine zwischen den Gitterstrichen liegende Lage der Nadel ergibt. Es kann sein, daß sich solche Koordinaten wählen lassen, für die die Ausgangswahrscheinlichkeiten eine Gleichverteilung bilden, für das endgültige Ergebnis der ganzen Rechnung ist das aber ohne Belang. Auch kann es vorkommen, daß in manchen Beispielen eine bestimmte Koordinatenwahl durch irgendwelche Gesichtspunkte bevorzugt erscheint — eine denknotwendige Entscheidung gibt es niemals, solange nicht ganz bestimmte der Erfahrung entnommene Bedingungen hinzutreten.

Die angebliche Brücke zwischen Häufigkeits- und Gleichmöglichkeitsdefinition.

Die wesentlichsten Einwände, die gegen die klassische Wahrscheinlichkeitsdefinition zu erheben sind, gehen, wie wir gesehen haben, nach zwei Richtungen. Einmal ist die Definition zu eng, sie umfaßt nur einen Teil der tatsächlich vorliegenden Aufgaben und schließt gerade die praktisch wichtigsten, die die Lebensversicherung und ähnliches betreffen, aus. Zweitens räumt sie der Gleichverteilung in den Ausgangskollektivs eine Sonderstellung ein, die ihr nicht zukommt und die sich ganz besonders in den eben besprochenen Fällen geometrischer Wahrscheinlichkeiten als ganz unhaltbar erweist. Gegen den zweiten Einwand läßt sich, soweit ich sehe, nichts Erhebliches vorbringen — es scheint mir, daß man ihn mehr aus Gleichgültigkeit als aus positiven Gründen unbeachtet läßt. Aber zum ersten Punkt wird jeder, der den üblichen Lehrgang der Wahrscheinlichkeitsrechnung kennt, sofort bemerken, daß es doch innerhalb des Rahmens der klassischen Theorie eine „Brücke" zwischen den beiden Definitionen, der Gleichmöglichkeits- und der Häufigkeitsdefinition gibt, durch die zumindest für den praktischen Bedarf eine befriedigende Behandlung der anfänglich ausgeschlossenen Fälle von Lebenswahrscheinlichkeit usw. ermöglicht wird. Diese „Brücke" soll angeblich durch das auf Bernoulli und Poisson zurückgehende, von mir schon einmal flüchtig erwähnte „Gesetz der großen Zahlen" hergestellt werden. Durch einen mathematisch beweisbaren Satz soll sich zeigen lassen, daß man, mit größerer oder geringerer Annäherung, die als Quotienten der günstigen durch die gleichmöglichen Fälle gewonnenen Wahrscheinlichkeitsgrößen auch durch Bestimmung der relativen Häufigkeit in genügend großen Versuchsserien wiederfinden kann. Zwar haben schon manche Kritiker, z. B. der vorhin genannte J. v. Kries, auf die gefährlichen Schwächen dieser „Brückenkonstruktion" hingewiesen, aber mangels einer besseren wird sie immer wieder benutzt. Es läßt sich daher nicht vermeiden, daß wir später ausführlich auf diesen Gegenstand eingehen, was sich auch an sich durch das allgemeine Interesse rechtfertigt, das man dem „Gesetz der großen Zahlen“ zuerkennen muß. Ich verschiebe aber diese Betrachtung, die uns von unserer augenblicklichen Aufgabe etwas abseits führen würde, auf dennächsten Vortrag, der ihr ganz gewidmet sein wird, und will jetzt nur das Resultat vorwegnehmen, das ich dann in überzeugender Weise zu begründen hoffe: Auch das geistvolle, für alle Anwendungen unentbehrliche Gesetz der großen Zahlen in all seinen Formen und mit allen Folgerungen, die daraus fließen, enthebt uns nicht der Notwendigkeit, die Wahrscheinlichkeit in allen Fällen als Grenzwert der relativen Häufigkeit zu definieren; ja das ganze, von BERNOULLI und POISSON abgeleitete Theorem verliert den wesentlichsten Teil seines Inhaltes, eigentlich seinen Sinn überhaupt, wenn wir diese Definition nicht von vornherein annehmen. Nur versteckte Fehlschlüsse, Zirkelschlüsse u. dgl. vermögen es, dem Theorem die Rolle zuzuschreiben, daß es eine Verbindung zwischen der Häufigkeits- und der Gleichmöglichkeitsdefinition bildet.

Zusammenfassung der Kritik der Gleichmöglichkeitsdefinition.

Bevor ich mich dem zweiten Teil der heutigen Aufgabe, der Besprechung neuerer Beiträge zur Grundlegung der Wahrscheinlichkeitsrechnung zuwende, möchte ich nochmals ganz kurz zusammenstellen, was ich gegen die klassische, auf dem Gleichmöglichkeitsbegriff aufgebaute Definition der Wahrscheinlichkeit vorgebracht habe.

1. Da „gleichmöglich“ nur ein anderes Wort für „gleichwahrscheinlich“ ist, bedeutet die klassische „Definition“ höchstens die Zurückführung von Kollektivs mit beliebiger Verteilung auf solche mit Gleichverteilung.

2. Die Gleichverteilung oder die „gleichmöglichen“ Fälle als Ausgangspunkt sind nicht immer vorhanden, z. B. beim falschen Würfel, bei der Sterbenswahrscheinlichkeit usf. Hier sind die Sätze der klassischen Theorie streng genommen nicht anwendbar.

3. Die Aussage: Bei einem vollkommen homogenen Würfel ist das Auffallen der sechs Seiten gleichwahrscheinlich, ist ohne

Inhalt, solange nicht irgendwie (z. B. durch die Häufigkeitsdefinition) erklärt wird, was „gleichwahrscheinlich" heißen soll.

4. Den Begriff „vollkommen homogen" in logisch exaktem Sinn gibt es nicht. Von jedem Würfel kann, wer seine Herstellung vollständig kennt, Verschiedenes über seine einzelnen Seiten aussagen.

5. „Spielraumtheorie", Indifferenzprinzip und ähnliches sind nur Umschreibungen, durch die an den Schwierigkeiten nichts geändert wird.

6. Die Fälle stetiger Verteilung (BERTRANDsche Paradoxie) zeigen, daß Annahme der Gleichverteilung noch etwas Verschiedenes bedeutet, je nach der Wahl des Koordinatensystems. Es ist keine für alle Fälle gültige, allgemeine Vorschrift möglich, die besagen würde, welche Koordinaten die „richtigen" sind, welche Gleichverteilung also anzunehmen ist.

7. Der von BERNOULLI und POISSON mathematisch abgeleitete Satz, der das „Gesetz der großen Zahlen" genannt wird, liefert keine Vermittlung zwischen der Gleichmöglichkeitsdefinition und dem Ablauf langer Versuchsreihen. Er enthebt nicht der Notwendigkeit, die Wahrscheinlichkeit von vornherein durch die Häufigkeitsdefinition einzuführen. (Dies wird später gezeigt werden.)

Übersicht der Einwände gegen meine Theorie.

Seit dem Erscheinen meiner ersten Veröffentlichungen im Jahre 1919 hat eine lebhafte Diskussion über die Grundlagen der Wahrscheinlichkeitsrechnung eingesetzt. Es war nicht zu erwarten, daß diejenigen, die sich seit Jahrzehnten mit dem Gegenstande befassen, sich namentlich in der erfolgreichen Behandlung von Einzelproblemen verdient gemacht haben, für die Annahme einer neuen Grundlegung leicht zu gewinnen sein würden. Sieht man von dieser älteren Generation ab, so gibt es heute kaum noch einen Mathematiker, der dem Ausgangspunkt der klassischen Auffassung, der Gleichmöglichkeitsdefinition, rückhaltlos zustimmen würde. Die Mehrzahl neigt mehr oder weniger ausgesprochen zur Häufigkeitstheorie, daneben bleibt eine Gruppe, die ich die „Nihilisten" nennen möchte, weil sie den Standpunkt vertritt, daß es einer Grundlegung, die den Zusammenhang zwischen der Theorie und der Erfahrungswelt herstellt, gar nicht bedarf. Über diesen Standpunkt werde ich mich am Schlusse des Vortrages noch äußern.

Aber unter denen, die der Auffassung folgen, daß den Gegenstand der Wahrscheinlichkeitsrechnung Aussagen über Häufigkeiten bilden und daß dies auch in den grundlegenden Definitionen zum Ausdruck kommen müsse, weichen die Ansichten noch stark untereinander und teilweise von meinen eigenen Aufstellungen ab. Zunächst wäre eine Reihe von Mathematikern zu nennen, die zu Beginn des Lehrganges die Wahrscheinlichkeit durch eine Häufigkeitsdefinition einführen, dann aber in der weiteren Durchführung nicht darauf Rücksicht nehmen, sondern die Darstellung in die Gedankengänge der klassischen Theorie einmünden lassen. Zu dieser Gruppe gehört z. B. das französische Lehrbuch von Fréchet und Halbwachs (1924) und das auch ins Deutsche übersetzte Buch des Amerikaners Coolidge (1925). Eine kleine Zahl von Autoren wendet sich gegen die Annahme, daß das Kollektiv als unendliche Folge von Elementen definiert wird; sie wollen nur von Häufigkeit in langen, aber endlichen Versuchsreihen sprechen, also nicht von dem Begriff des Grenzwertes Gebrauch machen. Die Mehrzahl nimmt die erste der von mir formulierten Forderungen, die ein Kollektiv erfüllen soll, die Existenz der Grenzwerte, als Ausgangspunkt an, findet aber in der Formulierung der zweiten Forderung, die die Regellosigkeit der Anordnung verlangt, verschiedene Schwierigkeiten. Den Einwänden entsprechen in manchen Fällen auch bestimmte Vorschläge zur Abänderung. Ich werde diese Dinge jetzt der Reihe nach zur Sprache bringen.

Endliche Kollektivs.

Es besteht kein Zweifel darüber, daß die Beobachtungsreihen, auf die man die Wahrscheinlichkeitstheorie praktisch anwendet, aus endlich vielen Elementen bestehen. Ebenso wird die Punktmechanik — die sogenannte Mechanik materieller Punkte — stets auf Körper von endlicher Ausdehnung angewandt und nicht auf Punkte im mathematischen Sinn. Aber niemand verkennt den Nutzen und den hohen Wert der Abstraktion, die zu dem Begriff des materiellen Punktes geführt hat, auch wenn wir heute über mechanische Theorien verfügen, die nicht von der Betrachtung diskreter Punkte ausgehen. Wie weit übrigens noch in der

Mechanik ausgedehnter Körper die der Punktmechanik angehörenden Begriffsabstraktionen herrschend sind, kann hier unerörtert bleiben.

Sicher ist, daß man auch bei der theoretischen Untersuchung der Massenerscheinungen oder Wiederholungsvorgänge unmittelbar an die Reihen von endlichem Umfang anknüpfen könnte. Es fragt sich nur, was dabei herauskommt. Einen anderen Grund der Beschäftigung mit unendlichen Folgen als den größerer Einfachheit kenne ich nicht und habe ich nie geltend gemacht. Vor kurzem hat ein jüngerer Autor, JOHANNES BLUME, sich die Aufgabe gestellt, meine Wahrscheinlichkeitstheorie so umzuformen, daß in den grundlegenden Definitionen nur endliche Merkmalfolgen auftreten. Er geht so vor, daß er zunächst an Stelle des Grenzwertes der relativen Häufigkeiten die Forderung setzt, die tatsächlichen Werte der relativen Häufigkeiten der verschiedenen Merkmale sollen um weniger als einen vorgegebenen kleinen Betrag ε von festen Zahlen, die die Verteilung des Kollektivs bilden, abweichen. Nimmt man dieses ε nur klein genug an, so kann man an derart definierten Kollektivs gewisse Operationen ausführen und bewegt sich dabei dauernd im Gebiet einer „Annäherung“, die durch die Größe ε gemessen wird. So weit dies Verfahren durchführbar ist, läuft es, wie K. KOLMOGOROFF in einer Kritik der BLUMEschen Ansätze mit Recht bemerkt hat, auf eine bloße Umschreibung — die unter Umständen sehr nützlich sein kann — des Grenzwertbegriffes hinaus. Denn der limes oder Grenzwert wird in der Mathematik nur zur Abkürzung gewisser Aussagen über kleine Abweichungen verwendet. Anderseits ist es weder BLUME noch anderen bisher gelungen, die Gesamtheit der Eigenschaften eines Kollektivs und der Beziehungen zwischen Kollektivs, namentlich all das, was mit der Regellosigkeit zusammenhängt, in dem Rahmen der finiten Theorie restlos zum Ausdruck zu bringen. So glaube ich, daß wir nach dem bisherigen Stand der Dinge von einer vorhandenen Wahrscheinlichkeitstheorie, die mit endlichen Kollektivs arbeitet, nicht sprechen können.

Eine historische Einschaltung: Der Philosoph THEODOR FECHNER, dessen Interessen recht vielseitige waren, schuf unter dem Namen „Kollektivmaßlehre“ eine Art systematischer Beschreibung endlicher Beobachtungsfolgen, die er „Kollektiv-gegenstände“ nannte. Das Werk erschien erst nach dem Tode des Verfassers, 1897, von LIPPS herausgegeben. Daran, daß man durch Idealisierung des Kollektivgegenstandes die Grundlage für einen rationellen Wahrscheinlichkeitsbegriff gewinnen könnte, dachte FECHNER wohl nicht. Aber seine Ausführungen bildeten — wenigstens für mich — die Anregung zu der neuen Betrachtungsweise.

Und eine sachliche Bemerkung: Mit allem Nachdruck muß ich dem häufig auftretenden Mißverständnis entgegentreten, als würde unsere Theorie jedesmal, wenn man es mit einer endlichen Anzahl von Beobachtungen zu tun hat, dafür eine unendliche Folge substituieren. Das ist selbstverständlich durchaus nicht der Fall. Wir haben in dem Beispiel, das ich am Schluß des letzten Vortrages ausführte, von 24 Würfen mit einem Würfelpaar gesprochen. Eine solche Gruppe von 24 aufeinanderfolgenden Beobachtungen wird Gegenstand unserer Theorie, indem wir uns vorstellen, daß sie, als Ganzes, unendlich oft vollzogen und damit Element eines Kollektivs wird. Es fließen daraus bestimmte Wahrscheinlichkeitsaussagen, die sich auf die genaue Zahl von 24 Einzelbeobachtungen beziehen. Ebenso steht es, wenn uns etwa eine Statistik vorliegt, die die Knabenquote der Geburten in 100 verschiedenen Städten ausweist. Unsere Theorie macht dann Aussagen darüber, was bei dieser Zahl $n = 100$ von Beobachtungen im Durchschnitt zu erwarten ist und substituiert nicht etwa für 100 Versuche eine Folge von unendlich vielen. Dies wird im fünften Vortrag, der speziell den Fragen der Statistik gewidmet ist, noch eingehender dargelegt werden.

Finitistische Deutung der Wahrscheinlichkeitstheorie.

Eine ganz andere Frage als die nach dem Aufbau einer Theorie, die nur endliche Kollektivs in Betracht zieht, ist die nach einer Deutung der Wahrscheinlichkeitstheorie im Endlichen. Da wir als das einzige Ziel einer Theorie die gedankliche Nachbildung beobachtbarer Erscheinungen ansehen, so kann die Bewährung der Theorie nur daran geprüft werden, daß sie auf beobachtbare, also endliche Erscheinungsreihen anwendbar ist. Anderseits habe ich schon wiederholt hervorgehoben, daß jedes Resultat der Wahrscheinlichkeitsrechnung unmittelbar nur zu Aussagen über unendliche Folgen führt. Auch wenn wir in den statistischen

Anwendungen eine Beobachtungsreihe von bestimmten endlichem Umfang, etwa 500 Versuche, als ein Element auffassen, so besagt ein Resultat unserer Rechnung strenggenommen nur etwas über eine unendliche Folge von Wiederholungen solcher 500 Versuche. Es könnte also scheinen, als ob niemals eine Kontrolle der Theorie an der Wirklichkeit möglich wäre.

Die hier auftretende Schwierigkeit ist aber genau die gleiche wie sie sich in allen Anwendungen irgendeiner empirischen Wissenschaft zeigt. Wenn z. B. eine physikalische oder chemische Betrachtung zu dem Ergebnis führt, das spezifische Gewicht eines Stoffes sei 0,897, so können wir durch Wägung oder andere geeignete Experimente festzustellen suchen, ob das „stimmt". Aber immer wird das Experiment nur das Gewicht eines endlichen Volumens ergeben und der Schluß auf das spezifische Gewicht, d. i. auf den limes des Quotienten Gewicht durch Volumen (für unendlich kleines Volumen), bleibt ebenso unsicher wie der von der Häufigkeit in einer endlichen Reihe auf die Wahrscheinlichkeit. Ja man kann sogar behaupten, daß es ein spezifisches Gewicht in der Wirklichkeit gar nicht gibt, weil, zumindest bei atomistischer Auffassung der Materie, ein Übergang zu unendlich kleinem homogenen Volumen nicht denkbar ist. Dem steht etwa zur Seite die Tatsache, daß man mit keinem Würfel, da er sich doch abnutzt, unbeschränkt lang unter gleichbleibenden Bedingungen würfeln kann.

Man könnte einwenden, daß die physikalischen Theorien auch zu andersartigen Behauptungen führen, die sich völlig streng nachprüfen lassen, z. B. zu Aussagen über das Gewicht eines bestimmten endlichen Volumens. Allein, wenn man von einer wirklich „genauen" Nachprüfung spricht, so hängt sie, wie man weiß, von der Erfüllung einer großen Zahl verschiedener Bedingungen ab, die sich selbst gar nicht streng fassen lassen: Z. B. muß die Wägung bei einem bestimmten Luftdruck stattfinden, der Begriff „Luftdruck" beruht aber wieder auf einer Limesbildung usf. Der erfahrene Experimentator beherrscht zwar praktisch die Bedingungen, unter denen man sagen darf, daß ein Experiment „stimmt", aber es ist unmöglich, diese Bedingungen logisch vollständig aufzuzählen, etwa so wie man die Voraussetzungen eines mathematischen Satzes angibt. Die Annahme des Zutreffens einer Theorie beruht, wie H. Dubislav es sehr richtig formuliert hat, nicht auf einem logischen Schluß, sondern auf einem praktischen Entschluß.

Ich stimme durchaus der Auffassung zu, die kürzlich CARL G. HEMPEL in einem klar geschriebenen Aufsatz „Über den Gehalt von Wahrscheinlichkeitsaussagen" veröffentlicht hat. Darnach sind die Ergebnisse einer Theorie, die auf dem Begriff des unendlichen Kollektivs aufgebaut ist, in einer zwar nicht logisch abgrenzbaren, aber praktisch hinreichend genau bestimmten Weise auf endliche Beobachtungsreihen anwendbar. Das Verhältnis zwischen Theorie und Beobachtung ist dabei wesentlich das gleiche wie in jedem anderen Zweig der physikalischen Wissenschaften.

Betrachtungen über dieses Verhältnis werden oft bezeichnet als Untersuchungen des „Anwendungsproblems". Man muß sich aber vor der Auffassung hüten, als ob es neben den Beobachtungserscheinungen und der sie darstellenden Theorie noch etwas Drittes gäbe: die Anwendungstheorie. Eine besondere „Theorie" darüber, wie eine Theorie auf die Beobachtungen anzuwenden sei, d. h. ein aus Sätzen, Ableitungen, Beweisen usf. bestehendes System, das neben der Sachtheorie die Lösung des „Anwendungsproblems" bildet, gibt es nicht. Der Zusammenhang zwischen Erfahrungswelt und Theorie wird durch die Ausgangssätze, die man gewöhnlich Axiome nennt, hergestellt. Diese Bemerkung ist für uns auch deshalb wichtig, weil gelegentlich versucht worden ist, der Wahrscheinlichkeitsrechnung als solcher die Rolle der allgemeinen „Anwendungstheorie" zuzuschieben. Die Auffassung wird schon dadurch hinfällig, daß man eben für jede Aussage der Wahrscheinlichkeitsrechnung selbst auch wieder das „Anwendungsproblem" aufweisen kann.

Einwände gegen die erste Forderung an ein Kollektiv.

Von der Mehrzahl der Mathematiker wird, wie ich schon sagte, heute anerkannt, daß die Konzeption einer unendlichen Folge von Beobachtungen oder Merkmalen als Gegenstand einer rationellen Wahrscheinlichkeitsrechnung angemessen und zweckentsprechend sei. Bei vielen aber, die zum erstenmal davon hören, daß die Wahrscheinlichkeit als der Grenzwert der relativen Häufigkeit definiert wird, taucht im Augenblick ein Einwand auf, der auf einem unklaren Erinnerungsbild aus der klassischen

Theorie beruht. Ich will den Einwand hier kurz besprechen, obwohl er keinerlei ernsthafter Prüfung standhält und obwohl er in einen Zusammenhang gehört, der im folgenden Vortrag über die „Gesetze der großen Zahlen" ausführlicher behandelt werden wird.

In der Tat beruft sich der Gegenredner, den ich jetzt im Auge habe, auf den Wortlaut des schon wiederholt erwähnten Bernoulli-Poissonschen Theorems. Diesem Satz zufolge ist es „fast sicher", daß in einer sehr großen Reihe aufeinanderfolgender Würfe mit einem richtigen Würfel die relative Häufigkeit einer geraden Zahl nahe dem Wahrscheinlichkeitswert $^1/_2$ liegt; aber es besteht, wie lang auch die betrachtete Versuchsreihe sein mag, immer noch eine bestimmte, wenn auch sehr kleine, so doch von Null verschiedene Wahrscheinlichkeit dafür, daß die relative Häufigkeit etwa größer z. B. als 0,51 ist. Diese Tatsache steht angeblich in Widerspruch mit der Annahme, daß der Grenzwert der relativen Häufigkeit genau gleich $^1/_2$ sei. Unsere Annahme — so meint der Opponent — bedeute ja, es sei sicher, nicht nur „fast sicher", daß in einer genügend langen Serie die relative Häufigkeit beliebig wenig von 0,5 abweiche, es bleibe also kein Raum für die mit geringer Wahrscheinlichkeit doch vorhandene Möglichkeit einer Abweichung in dem Maße 0,51 gegenüber 0,5.

Hier liegt nichts als eine etwas unpräzise Ausdrucksweise vor, deren Berichtigung die Schwierigkeit sofort verschwinden läßt. Das genannte Theorem sagt etwas aus über die Wahrscheinlichkeit eines bestimmten Wertes der relativen Häufigkeit in einer Gruppe von n Versuchen. Dabei wird also, nach der Auffassung, die der Häufigkeitstheorie zugrunde liegt, eine Gruppe von n aufeinanderfolgenden Würfen als ein Element eines Kollektivs betrachtet — so etwa wie in dem Beispiel am Schluß des letzten Vortrages einmal 4 und einmal 24 Würfe zu einem Element zusammengefaßt wurden. Bei der jetzt gestellten Aufgabe gilt als Merkmal des Elementes die Häufigkeit, die das Resultat „gerade Zahl" innerhalb der Gruppe besitzt. Nennen wir dieses Merkmal x, so kann offenbar x einen der $(n + 1)$ Werte 0, $^1/_n$, $^2/_n$, $^3/_n$. ..., schließlich $^n/_n = 1$ annehmen. Allgemein, wenn in einer Gruppe von n Würfen m gerade Augenzahlen vorkommen, ist $x = {}^m/_n$. Für jeden dieser $(n + 1)$ verschiedenen x-Werte gibt es eine bestimmte Wahrscheinlichkeit. Sie kann z. B. in einem konkreten Fall für alle x-Werte oberhalb 0,51 zusammen die Größe 0,00001 besitzen. Dies bedeutet nach unserer Auffassung, daß, wenn man die unendliche Folge von Würfen in Gruppen zu je n abteilt, gerade in einem Hunderttausendstel dieser Gruppen mehr als 51% gerade Zahlen vorkommen. Hier handelt es sich, wohlgemerkt, immer um die Häufigkeit innerhalb einer Gruppe, also um die Anzahl der geraden Nummern innerhalb der Gruppe dividiert durch die konstante Gruppenlänge n.

Dagegen geht unsere Definition der Wahrscheinlichkeit von einer ganz anderen relativen Häufigkeit aus, nämlich: Man durchläuft die ganze Versuchsreihe, ohne sie in Gruppen zu teilen, d. h. ohne einzelne Gruppen für sich gesondert zu betrachten, und untersucht, wie oft von Beginn an das geradzahlige Resultat aufgetreten ist. Hat man, von Anfang an gerechnet, N Würfe, unter denen N_1 gerade Zahlen als Resultat ergeben haben, so behaupten wir, daß die Quotienten $\frac{N_1}{N}$, bei denen Zähler und Nenner zugleich wachsen, einem festen Grenzwert, und zwar im vorliegenden Beispiel dem Grenzwert $^1/_2$ zustreben. Zwischen den beiden Behauptungen: Existenz des Grenzwertes von $\frac{N_1}{N}$ und Vorkommen von einzelnen Gruppen mit beliebigen Werten von $\frac{m_1}{n}$, besteht kein auf den ersten Blick erkennbarer Zusammenhang, also sicher auch kein evidenter Widerspruch. Nur die zuerst angewandte, unvollständige und unpräzise Formulierung konnte den Anschein eines solchen Widerspruches hervorrufen.

Im übrigen wird es meine Aufgabe im nächsten Vortrag sein, gerade den Zusammenhang der beiden Aussagen zu untersuchen. Ich werde zeigen, daß sie nicht nur miteinander verträglich sind, sondern, daß das klassische Gesetz der großen Zahlen seinen eigentlichen Sinn und seinen vollen Wert erst erhält, wenn man von der Häufigkeitsdefinition der Wahrscheinlichkeit ausgeht.

Einwände gegen die Regellosigkeitsforderung.

Ich wende mich jetzt den Einwänden zu, die den breitesten Raum in der bisherigen Literatur einnehmen und die sich gegen die Regellosigkeitsforderung wenden. Hier befinden wir uns im Kernpunkt der Diskussion und man muß es auch als durchaus begreiflich erkennen, daß ein so neuartiger, von dem Gewohnten so stark abweichender Begriff wie der der Regellosigkeit mannigfache Kritik herausfordert. Folgendes ist der Tatbestand.

Wir nennen eine unendliche Folge von Nullen und Einsern — es genügt zunächst, sich auf eine einfache Alternative zu beschränken — regellos, wenn in ihr die relative Häufigkeit der Eins (infolgedessen auch die der Null) einen Grenzwert besitzt, der sich nicht ändert, sobald man durch Stellenauswahl einen Teil der Elemente aussondert und nur diese allein in Betracht zieht. Dabei heißt Stellenauswahl eine Vorschrift, die in gesetzmäßiger Weise eine unendliche Folge von Nummern (Stellen) festlegt, ohne daß bei der Festlegung der Merkmalwert (0 oder 1) der betreffenden Stelle benutzt wird. Zum Beispiel sind Stellenauswahlen: Jede durch 3 teilbare Nummer; jede Nummer, die um 2 vermindert, das Quadrat einer Primzahl gibt; jede Nummer, deren drittvorhergehende eine 1 als Merkmal aufwies. Mit dieser Formulierung soll eine bestimmte und wohlbekannte Eigenschaft der Glückspiele erfaßt werden: Wenn man beim Roulettespiel ständig auf „Schwarz" setzt oder nur bei jedem dritten Spiel oder nur dann, wenn das drittvorhergehende „Schwarz" ergeben hat, in allen diesen und ähnlichen Fällen sind die Chancen, bei genügend lang fortgesetztem Spiel zu gewinnen oder zu verlieren, die gleichen.

In meiner ersten Veröffentlichung von 1919 habe ich der Erläuterung des Regellosigkeitsbegriffes längere Ausführungen gewidmet. Unter anderem habe ich als Satz 5 abgeleitet: „Ein Kollektiv wird durch seine Verteilung vollkommen bestimmt; die vollständige Angabe der Zuordnung (sc. zwischen Nummern und Merkmalwerten) ist nicht möglich." Hierzu heißt es dann weiter: „Satz 5 besagt, daß man die ‚Existenz' von Kollektivs nicht durch eine analytische Konstruktion nachweisen kann, so wie man etwa die Existenz stetiger, nirgends differenzierbarer Funktionen nachweist. Wir müssen uns mit der abstrakten logischen Existenz begnügen, die allein darin liegt, daß sich mit den eingeführten Begriffen widerspruchsfrei operieren läßt." Wenn ich auch heute vielleicht manchen Ausdruck anders wählen würde, bleibt doch das Wesentliche von diesen Ausführungen bestehen: Es ist unmöglich, eine Folge von Nullen und Einsern, die der Regellosigkeitsforderung genügt, durch eine Formel oder eine Rechenvorschrift zu bestimmen, etwa derart, daß man sagt: Jede Nummer, die durch 3 teilbar ist, hat das Merkmal 1, jede andere das Merkmal 0; oder: jede Nummer, die gleich ist dem um 2 vermehrten Quadrat einer Primzahl, gibt eine 1, alle anderen geben 0 usf. Würde ein Kollektiv durch eine solche Rechenvorschrift gegeben werden, so könnte man dieselbe Vorschrift als Stellenauswahl benutzen und innerhalb der so ausgesonderten Teilfolge gäbe es dann bloß Einser. In der Teilfolge hätte mit anderen Worten die Eins immer die Grenzhäufigkeit 1, also im allgemeinen eine andere als in der ursprünglichen vollständigen Folge.

An diesen Tatbestand, die Unmöglichkeit einer expliziten Angabe der Merkmalfolge durch eine Formel, knüpft die Kritik an.

Die Frage der Existenz.

Auf die einfachste Form gebracht, lautet der Einwand, um den es sich hier handelt, so: Eine Merkmalfolge, die der Regellosigkeitsforderung genügt, existiert nicht. Dabei werden die Worte „nicht existieren" als synonym gesetzt für: sie läßt sich nicht durch eine Formel oder Rechenvorschrift darstellen.

Was bedeutet diese Nichtexistenz in mathematischem Sinn — denn nur um eine solche handelt es sich — und was folgt aus ihr? Es gibt im wesentlichen zwei verschiedene Auffassungen des Existenzbegriffes in der Mathematik. Die Mehrzahl der Mathematiker, die Formalisten, sehen mathematische Begriffe als zulässig, also ihr Objekt als existent an, wenn sich mit ihnen widerspruchsfrei operieren läßt. Die zweite Schule, die der Intuitionisten, stellt anscheinend höhere Forderungen; sie verlangt, daß die Objekte der mathematischen Untersuchung konstruierbar seien. Konstruiert wird, indem man von bestimmten, intuitiv erfaßten Grundvorstellungen ausgeht und durch Kombination von solchen neue, kompliziertere Gebilde schafft.

Um den Formalisten gerecht zu werden, müßte man den Widerspruchslosigkeitsbeweis für die Wahrscheinlichkeitstheorie erbringen. Darauf ist zu sagen, daß ein solcher expliziter Beweis noch für viele andere Teile der Mathematik aussteht, die deswegen doch nicht verworfen werden. Ich halte es für durchaus möglich, daß sich für die Wahrscheinlichkeitstheorie in der von mir gegebenen oder einer ähnlichen Form der Widerspruchslosigkeitsbeweis wird führen lassen. Vorläufig ist jedenfalls ein Widerspruch in der Anwendung der Regellosigkeitsforderung nicht nachgewiesen und viele, selbst unter den Gegnern der Theorie, behaupten, daß ein solcher Widerspruch sich gar nicht ergeben könne.

Anderseits könnte es scheinen, als ob dem intuitionistischen Standpunkt gegenüber der Regellosigkeitsbegriff ganz unhaltbar sein müsse. Das ist aber durchaus nicht der Fall. Die intuitionistische „Konstruktion“ einer Zahlenfolge ist so allgemein, daß sie ganz explizit auch solche Folgen umfaßt, die nicht durch Formel oder Rechenvorschrift „gesetzmäßig“ definiert sind (die Folgen „freien Werdens“). Ja es wird von E. J. Brouwer, dem Schöpfer der intuitionistischen Mathematik, ausdrücklich ausgesprochen, daß ein unendlicher Dezimalbruch, dessen aufeinanderfolgende Ziffern nicht anders als durch Auswürfeln festgestellt werden, zu den konstruktiv herstellbaren Gebilden gehört, die den Gegenstand mathematischer Betrachtung bilden können. Ein solcher Bruch heißt in der Brouwerschen Ausdrucksweise eine „unexakte Zahl“.

Wenn ich meine eigene Meinung andeuten darf, so geht sie dahin, daß die Frage nach der Existenz oder der Konstruierbarkeit, allgemein gestellt überhaupt nicht sinnvoll ist. Sie erhält einen Sinn nur innerhalb einer abgeschlossenen Theorie, in der man sich geeinigt hat, für welchen Tatbestand die Worte „existiert“ oder „existiert nicht“, bzw. der Ausdruck „konstruierbar“ als Abkürzung gelten sollen. Man weiß z. B., eben aus dem Streit der Formalisten und Intuitionisten, daß verschiedene Festsetzungen darüber möglich sind, wann man sagen will, es „existiere“ eine Nullstelle einer gegebenen Funktion oder sie „existiere nicht“. Es ist hier nicht der Ort, darauf näher einzugehen.

Man sollte, denke ich, einer neuen Theorie gegenüber nur fragen, ob sie sich in der Anwendung bewährt und ob es andere Theorien gibt, die denselben Problemen gegenüber vorzuziehen sind. Der Nachweis eines Widerspruches würde die Bewährung ausschließen, der Nichtnachweis kann sie natürlich noch nicht verbürgen, muß aber zunächst als Positivum gewertet werden. Im übrigen komme ich im folgenden auf einen, wie ich denke, vielversprechenden Ansatz eines Widerspruchslosigkeitsbeweises zu sprechen.

Regellosigkeitsforderung und Grenzwert.

Eine Variante des Existenzeinwandes, die gelegentlich auftritt, richtet sich gegen die Vereinigung der Regellosigkeitsforderung mit der ersten an ein Kollektiv gestellten Forderung, der nach der Existenz der Grenzwerte der relativen Häufigkeiten. Der Gedankengang ist ungefähr der folgende. Ob in einer Folge von Nullen und Einsern die relative Häufigkeit der Eins einen Grenzwert besitzt oder nicht, läßt sich nur dann entscheiden, wenn die Merkmalfolge gesetzmäßig definiert ist. Da aber eine der Regellosigkeitsforderung genügende Folge niemals durch ein Gesetz gegeben sein kann, so „darf" man hier von der Existenz eines Grenzwertes gar nicht sprechen.

Wir wollen uns einen Augenblick überlegen, welcher Sinn solchen Verbotssätzen innerhalb der Mathematik zukommt. Betrachten wir den folgenden Tatbestand. Wenn man die Quadratwurzel aus π in einen Dezimalbruch entwickelt und hinterher an Stelle jeder geraden Ziffer in der Entwicklung eine Null, an Stelle jeder ungeraden Ziffer eine Eins setzt, so hat man sicher eine gesetzmäßig definierte Folge hergestellt. Allein bis heute weiß niemand, ob in dieser Folge die Eins eine Grenzhäufigkeit besitzt oder nicht. Genau so steht es, wenn wir eine beliebige andere, in ähnlicher Weise wie π definierte irrationale Zahl nehmen, und in sehr vielen anderen Fällen. Man kann ohne Übertreibung sagen, daß von fast allen gesetzmäßig definierten Merkmalfolgen bisher nicht entschieden werden kann, ob die Grenzwerte existieren oder nicht. Allerdings glauben die Formalisten an den Satz, daß jedes mathematische Problem lösbar sei, die Intuitionisten bemerken dagegen mit Recht, daß dieser Glaube unbegründet, jedenfalls aber nicht zwingend ist.

Sicher ist, daß es nach dem bisherigen Stand der Mathematik dreierlei gesetzmäßig durch Rechenvorschrift definierte Merkmalfolgen gibt: 1. solche, für die das Bestehen der Grenzhäufigkeiten nachgewiesen ist; 2. solche, für die das Nichtbestehen der Grenzwerte bewiesen ist; 3. solche, bei denen die Frage nicht entschieden ist. Gesetzt nun, man hätte einen Satz abgeleitet, der sich auf Grenzhäufigkeiten bezieht, z. B. den folgenden: Wenn in einer Folge, die nur drei verschiedene Zeichen A, B und C enthält, das A die Grenzhäufigkeit p, das Zeichen B die Grenzhäufigkeit q

besitzt, so besteht auch für C der Grenzwert der relativen Häufigkeit, und zwar ist er gleich $1 - p - q$. Nach dem oben ausgesprochenen Verbot müßte man hier hinzufügen: Der Satz gilt nicht für Folgen der dritten Gruppe, denn bei diesen läßt sich nicht entscheiden, ob die Bedingung erfüllt ist oder nicht. Ich sehe nicht, welche Konsequenzen ein solcher Zusatz besitzt, er scheint mir völlig leerlaufend.

Ähnlich verhält es sich, wenn wir statt der Folgen, die zur Gruppe 3 gehören, solche betrachten, die als Folgen „freien Werdens" etwa durch Auswürfelung bestimmt werden. In einem gegebenen Fall läßt sich dann sicher nicht entscheiden, ob Grenzhäufigkeiten bestehen oder nicht, aber wir können Aussagen über diejenigen Folgen machen, für die die Existenz der Grenzwerte nachgewiesen ist. Solange kein Widerspruch aufgezeigt ist zwischen der Definition der Folge und einem aus der Annahme der Existenz von Grenzhäufigkeiten fließenden Satz, kann man von Folgen mit Grenzhäufigkeiten sprechen. Dem Einwand der Intuitionisten, daß bei dieser Einstellung die Mathematik ein „müßiges Spiel" wäre, begegnen wir mit dem Hinweis darauf, daß unsere Theorie auf beobachtbare Erscheinungen anwendbar ist.

Im übrigen soll hier durchaus nicht der Standpunkt vertreten werden, als ob über diese Fragen erkenntnistheoretischer Natur das letzte Wort gesprochen sei. Ich halte im Gegenteil die Diskussion erst für eröffnet. Meine Definition der Regellosigkeit stellt den ersten Versuch dar und sicher nicht den letzten, diesen für die Wahrscheinlichkeitstheorie grundlegenden Begriff zu fassen. Unter anderem werde ich im folgenden einen mir sehr sympathischen Versuch besprechen, den kürzlich K. Dörge unternommen hat, um durch eine bestimmte Modifikation meiner Theorie die in Rede stehenden Schwierigkeiten zu beseitigen.

Sogenannte Bernoullische Merkmalfolgen.

Vorerst wollen wir sehen, wie diejenigen Mathematiker, die meinen Regellosigkeitsbegriff ablehnen, aber eine auf die Häufigkeitsvorstellung gestützte Wahrscheinlichkeitsrechnung aufbauen wollen, sich zu helfen suchen. Wenn man die Merkmalfolgen, die der Betrachtung zugrundegelegt werden, keiner weiteren Einschränkung unterwirft als der Forderung nach Existenz der

Grenzhäufigkeiten, so kann man offenbar die einfachsten Fragen der Wahrscheinlichkeitstheorie nicht richtig beantworten. Nimmt man z. B. an, daß das Ergebnis eines „Kopf-oder-Adler"-Spieles durch die gesetzmäßige Merkmalfolge

$$1\,0\,1\,0\,1\,0\,1\,0\,1\ldots.$$

oder ein andermal durch die Folge

$$1\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\,0\ldots.$$

dargestellt werden kann, so besteht wohl in jedem der beiden Fälle die Grenzhäufigkeit $^1/_2$ für das Erscheinen der Eins. Fragt man aber nach der Wahrscheinlichkeit dafür, daß zweimal hintereinander Eins erscheint, so muß die Antwort im ersten Fall lauten 0, im zweiten Fall $^1/_2$ (die Hälfte der Paare, von Anfang an gebildet, zeigt das Resultat 11), während bekanntlich im wirklichen Spiel diese Wahrscheinlichkeit gleich $^1/_4$ ist. Zeigt man einen solchen speziellen Mißstand auf, wie er hier vorliegt, so besteht keine Schwierigkeit, ihm auszuweichen, indem man eine spezielle, passende Annahme über den Ausschluß gewisser Anordnungen macht. Verschiedene Autoren nehmen an Stelle der Regellosigkeitsforderung eine Bedingung auf, die genau den hier angeführten speziellen Widerspruch beseitigt (allerdings andere analoge bestehen läßt), und zwar in folgender Weise:

Es sei n eine beliebige Zahl und man denke sich die Merkmalfolge in aufeinanderfolgende Abschnitte der Länge n geteilt. Jeder solche Abschnitt enthält eine bestimmte Kombination von n Nullen und Einsern, z. B. bei $n = 2$ eine der vier Möglichkeiten 00, 01, 10, 11. Im allgemeinen gibt es 2^n verschiedene Kombinationen. Es wird nun die Forderung gestellt, daß die Häufigkeit der Abschnitte, die eine bestimmte Kombination (z. B. 01 bei $n = 2$) aufweisen, einen Grenzwert besitzt und daß dieser gleich $(^1/_2)^n$ ist, falls in der ursprünglichen Folge die Grenzhäufigkeit der Null und Eins je $^1/_2$ war. Die Wahrscheinlichkeit des Auftretens der Kombonation 11 (zwei Einser hintereinander) wird dann richtig gleich $^1/_4$. Im allgemeinen Fall, in dem an Stelle von $^1/_2$ die Wahrscheinlichkeiten p und q für die Null und Eins eintreten (mit $p + q = 1$), ist die Forderung dahin zu modifizieren, daß eine Kombination die a Nullen und b Einser umfaßt (wobei $a + b = n$), mit der Grenzhäufigkeit $p^a q^b$ auftreten soll. Die Erweiterung für den Fall von mehr als zwei voneinander verschiedenen Merkmalen ist naheliegend und braucht hier nicht erörtert zu werden.

Merkmalfolgen, die der eben dargelegten Forderung genügen, werden in der Literatur oft als BERNOULLIsche Folgen bezeichnet. Sie wurden vielleicht zum erstenmal von EMILE BOREL eingehender untersucht. E. TORNIER und unabhängig von ihm der Amerikaner A. H. COPELAND haben die Forderung definitorisch in die Elemente der Wahrscheinlichkeitsrechnung aufgenommen. Von ihnen übernahm sie E. KAMKE, der sie aber nur zur Lösung spezieller Aufgaben heranzieht, während er im übrigen den sonderbaren Standpunkt vertritt, daß es einer Einschränkung der zu behandelnden Merkmalfolgen, über die Forderung nach Existenz der Grenzwerte hinaus, nicht bedarf.

Eingeschränkte Regellosigkeitsforderung.

Man kann die Forderung, daß eine Merkmalfolge die eben dargelegte „BERNOULLIsche" Eigenschaft besitzen soll, als eine eingeschränkte Regellosigkeitsforderung betrachten. Durch die Beschränkung wird erreicht, daß Folgen, die der Forderung genügen, formelmäßig angegeben werden können. Aber was stehen dem für Nachteile gegenüber!

Gewiß können jetzt Fragen richtig beantwortet werden, wie die nach der Wahrscheinlichkeit dafür, daß innerhalb in bestimmter Weise abgeteilter Gruppen von je 5 Spielen eine Gruppe von lauter Einsern auftritt. Aber es genügt schon eine beliebig kleine Variation der Frage, um die Theorie zum Scheitern zu bringen. Bilden wir etwa die Fünfergruppen so, daß zwischen der ersten und zweiten Gruppe ein Element weggelassen wird, dann zwischen der zweiten und dritten 2 Elemente, hierauf zwischen der dritten und vierten 3 Elemente gestrichen werden, hierauf 5, dann 7, 11, 13 usf., den aufeinanderfolgenden Primzahlen entsprechend. Auch unter diesen Umständen wird in einem wirklichen Glückspiel die Häufigkeit der Gruppen, die aus lauter Einsern bestehen, den Wert $(1/2)^5$ besitzen, woran niemand zweifelt. Läßt man aber alle gesetzmäßig gebildeten Merkmalfolgen zu, die nur die BERNOULLIsche Eigenschaft aufweisen und sonst willkürlich sind, so kann die Grenzhäufigkeit der aus 5 Einsern bestehenden Gruppen bei der neuen Art der Gruppenbildung noch beliebig weit von $(1/2)^5$ abweichen. Ihre Größe bleibt überhaupt unbestimmt und ist von der willkürlichen Wahl der Formel abhängig, die die Merkmalfolge definiert. Eine Theorie, die zu solchen Resultaten führt, kann man unmöglich als brauchbar ansehen.

Es steht nicht anders, wenn man die Forderung nach beschränkter Regellosigkeit gegenüber der „BERNOULLIschen" Eigenschaft noch etwas verschärft. COPELAND verlangt von den Merkmalfolgen, die er betrachtet (er nennt sie „admissible numbers"), noch etwas mehr. Seine Definition läuft auf das hinaus, was H. REICHENBACH — und ihm folgt neuerdings K. POPPER — eine „normale" Merkmalfolge nennt. Das ist eine Folge, die die ebengenannte Eigenschaft besitzt und außerdem in gewissem Sinne „nachwirkungsfrei" ist. Dieser Ausdruck bedeutet, daß eine Stellenauswahl von der Art: Jedes Element, das auf drei Nullen folgt, oder: Jedes Element, dem die Kombination 0101 (oder eine andere) unmittelbar vorangeht, die Grenzwerte der relativen Häufigkeit unverändert läßt. Nur solche „normale" Folgen sollen den Gegenstand der Wahrscheinlichkeitsrechnung bilden. Selbstverständlich kann man diese Art, den Gegenstand der Wahrscheinlichkeitsrechnung abzugrenzen, beliebig variieren, indem man die Unveränderlichkeit der Grenzwerte der relativen Häufigkeit für eine größere oder kleinere Klasse von Stellenauswahlen fordert. Auf jeden Fall gelangt man zu einer (der meinen gegenüber) eingeschränkten Regellosigkeitsforderung.

Es sind, allgemein gesprochen, zwei Fälle eingeschränkter Regellosigkeit möglich und diese Disjunktion gilt für jede derartige Theorie. Entweder ist die Einschränkung so schwach, daß keine durch eine Rechenvorschrift darstellbare Folge die Forderung erfüllt; dann ist kein Vorteil gegenüber meiner Formulierung erreicht. Oder es lassen sich geeignete Folgen formelmäßig angeben, dann kann man sofort Aufgaben finden, denen gegenüber derselbe Widerspruch auftritt, wie er eben für den Fall der BERNOULLIschen Folgen aufgezeigt wurde. Speziell für die COPELAND-REICHENBACHschen „normalen" Folgen braucht man kein anderes Beispiel zu suchen als die erwähnten, durch Zwischenräume wachsender Länge getrennten Gruppen von je n Elementen: die Wahrscheinlichkeit für das Auftreten von n Einsern in einer derartigen Gruppe bleibt im Rahmen der Theorie der „normalen Folgen" mit der eingeschränkten Regellosigkeitsforderung völlig unbeantwortbar.

Eine Wahrscheinlichkeitstheorie, die auf einer eingeschränkten Regellosigkeitsforderung aufgebaut ist, verhält sich somit wie eine Geometrie, die die Aufgabe, ein Dreieck aus zwei Seiten und dem eingeschlossenen Winkel zu bestimmen, nur dann löst, wenn die Daten einer Einschränkung genügen, z. B. rationale Zahlen sind od. dgl.

Die Widerspruchsfreiheit der Regellosigkeitsforderung. Scheint es demnach auch ausgeschlossen zu sein, in dieser Weise eine vollständige Wahrscheinlichkeitsrechnung zu begründen, so halte ich doch keineswegs Untersuchungen für wertlos, die sich mit wohldefinierten, einer eingeschränkten Regellosigkeitsbedingung genügenden Merkmalfolgen befassen. Solche Folgen bilden in bestimmtem Sinne ein Modell des Kollektivs, ähnlich wie es im Euklidischen Raum Modelle der nicht-Euklidischen Geometrie gibt, und ihr Studium ist für das Verständnis der Widerspruchsfreiheit der allgemeinen Regellosigkeitsforderung von Bedeutung. Besonders hervorzuheben sind hier die Bemühungen von A. H. Copeland, der in zahlreichen Arbeiten die Anordnungsmöglichkeiten in Zeichenfolgen untersucht hat. Neuerdings ist es Copeland gelungen, Folgen gesetzmäßig zu konstruieren, in denen die relativen Häufigkeiten gegenüber „fast allen“ Stellenauswahlen unempfindlich bleiben. Die Worte „fast alle“ haben dabei eine mathematische wohldefinierte Bedeutung, die dem üblichen Wortsinn nahekommt.

Ich erwähnte schon, daß manche Autoren, die meine Ansätze nicht oder nicht ganz annehmen, sich in eigentümlicher Weise über deren Unangreifbarkeit äußern. So sagt z. B. E. Tornier gelegentlich: „Es liegt hier offensichtlich der sehr interessante Fall vor, daß ein praktisch durchaus sinnvoller Begriff — Auswahl ohne Berücksichtigung der Merkmalunterschiede — prinzipiell jede rein mathematische, auch axiomatische Festlegung ausschließt. ... wäre es wünschenswert, daß sich diesem Sachverhalt, der vielleicht von grundlegender Bedeutung ist, das Interesse weiter mathematischer Kreise zuwendet.“ In ähnlicher Richtung liegt die Bemerkung von E. Kamke: „Da es ... in der v. Misesschen Theorie keinerlei beweiskräftige Beispiele gibt, kann man in dieser Theorie höchstens etwas beweisen, aber niemals etwas durch Gegenbeispiele widerlegen.“ Auch K. Dörge spricht von einer „weiten Überlegenheit in den Anwendungen" der auf die uneingeschränkte Regellosigkeit gestützten Theorie, einer Überlegenheit, die doch nicht bestehen könnte, wenn die Theorie zu Widersprüchen führte. Ich glaube, daß ich zur Aufklärung dieser, wohl etwas gezwungenen, Äußerungen einen entscheidenden Beitrag liefern kann.

Die ganze Schwierigkeit, die sich der Anerkennung meiner Grundlegung der Wahrscheinlichkeitstheorie entgegenstellt, entspringt aus dem Wörtchen „alle" in der Definition der Regellosigkeit. Daß alle Stellenauswahlen oder jede Stellenauswahl den Grenzwert der relativen Häufigkeit unverändert lassen soll, ruft ein Unbehagen, das Gefühl einer Unsicherheit hervor, die sich dann in mehr psychologisch als logisch begründeten Ablehnungen äußert. Ich will nun eine Formulierung, oder wenn man will, eine Interpretation geben, die von diesem psychologischen Mangel frei ist: Die Festsetzung daß in einem Kollektiv jede Stellenauswahl die Grenzhäufigkeit unverändert läßt, besagt nichts anderes als dies: Wir verabreden, daß, wenn in einer konkreten Aufgabe ein Kollektiv einer bestimmten Stellenauswahl unterworfen wird, wir annehmen wollen, diese Stellenauswahl ändere nichts an den Grenzwerten der relativen Häufigkeiten. Nichts darüber hinaus enthält mein Regellosigkeitsaxiom.

Da nun in einer einzelnen Aufgabe niemals „alle" Auswahlen in Frage kommen, sondern deren nur wenige, so daß man jedesmal mit einer eingeschränkten, ad hoc zugeschnittenen, Regellosigkeit das Auslangen finden könnte, so kann tatsächlich nichts von dem eintreten, was ängstliche Gemüter befürchten. Darum erscheint mir auch die Beschäftigung mit formelmäßig dargestellten Folgen, die in beschränktem Sinne regellos sind, sehr nützlich; sie zeigt, wie sich innerhalb einer konkreten Aufgabe die Beziehungen gestalten. Anderseits sieht man aber auch, daß mit einer ein für allemal formulierten Einschränkung der Regellosigkeit niemals das erreicht werden kann, was die Darstellung der wirklich beobachtbaren Erscheinungen verlangt. Denn es läßt sich dann immer, wie gezeigt, ein Beispiel konstruieren, in dem die Theorie versagt.

Praktisch genommen verhält sich jeder Mathematiker, der überhaupt von der Häufigkeitsvorstellung in der Behandlung der

Wahrscheinlichkeitsprobleme ausgeht, genau so, wie ich es eben formuliert habe. Er setzt jedesmal die Zulässigkeit derjenigen Stellenauswahl voraus, die in dem betreffenden Problem eine Rolle spielt. Nur schickt er nicht den Satz voraus, daß bei allen Auswahlen die Grenzwerte unverändert bleiben. In diesem Sinne kann man sagen, daß ein ernsthafter Gegensatz, wenigstens soweit es sich um die Behandlung konkreter Probleme handelt, innerhalb der Anhänger der Häufigkeitstheorie überhaupt nicht besteht.

Der Ansatz von K. Dörge.

K. Dörge hat meiner Grundlegung der Wahrscheinlichkeitsrechnung eine neue Form gegeben, die sie gegen die vorher besprochenen Einwände sichern soll. Der Gedankengang Dörges läßt sich, wenn man von allem mathematischen Beiwerk absieht, wie folgt darstellen.

Man geht davon aus, daß zunächst nur irgendeine endliche Anzahl verschiedener Merkmalfolgen (durch ihre Formeln) gegeben seien, die sogenannten Grundkollektivs K_1, K_2, K_3, ... und ferner eine endliche Anzahl von formelmäßig definierten Stellenauswahlen A_1, A_2, A_3, ... Diese Grundelemente, Kollektivs und Auswahlen, ergänzt man nun durch Hinzufügung weiterer zu einem „Körper", ähnlich wie man aus einigen gegebenen Zahlen durch Festlegung gewisser Operationen einen „Körper" bildet. Sind z. B. die Zahlen 8, 10, 74 vorgegeben und verlangt man den „Körper", der durch Addition, Subtraktion und Multiplikation erzeugt wird, so gelangt man zu der Gesamtheit aller positiver und negativer gerader Zahlen; nämlich dadurch, daß man die genannten Operationen erst an den Grundzahlen 8, 10, 74 selbst, dann an den Ergebnissen der Rechnung, also an den Summen, Differenzen und Produkten, in jeder denkbaren Weise vollzieht. Dabei gelangt man nie zu einer anderen als einer ganzen, geraden Zahl, anderseits kann man keine gerade Zahl fortlassen, wenn keine an Zahlen des Körpers ausgeführte Operation aus dem Körper herausführen soll. In diesem Sinne bildet nun Dörge den Körper, der zu den Grundkollektivs und Grundauswahlen gehört, wobei als Operationen gelten: Vollzug einer Auswahl $A_\varkappa$ an einem Kollektiv K_λ, Vollzug einer Auswahl $A_\varkappa$ an einer anderen Auswahl A_λ (wodurch die Produktauswahl entsteht), Teilung eines Kollektivs, Auswürfelung an einem Kollektiv mittels eines anderen (diese Begriffe, Teilung und Auswürfelung, sind im zweiten Vortrag erklärt worden). Natürlich werden die Operationen immer wieder an den Resultaten wiederholt, z. B. wird an der durch A_1 aus K_4 entstandenen Teilfolge nachher die Auswahl A_2 vollzogen usf., dies alles solange, bis weitere Operationen dieser Art keine neuen Elemente mehr erzeugen.

Innerhalb einer in dieser Weise abgeschlossenen Gesamtheit von Kollektivs und von Stellenauswahlen spielt sich nun die ganze Wahrscheinlichkeitstheorie ab. Innerhalb des begrenzten Bereiches fordert Dörge das Erfülltsein der beiden Axiome: In jedem Kollektiv müssen die Grenzwerte der relativen Häufigkeit existieren und sie müssen unverändert bleiben, wenn das Kollektiv einer der Stellenauswahlen unterworfen wird. Alles weitere vollzieht sich dann genau so wie in dem von mir gegebenen Aufbau.

Was ist nun mit der Modifikation gewonnen worden? Als entscheidend ist anzusehen, daß man jetzt, sobald der Grundkörper gegeben ist, durchweg mit formelmäßig darstellbaren Merkmalfolgen zu tun hat, so daß die früher geschilderten Bedenken hinsichtlich der „Existenz" und hinsichtlich der Zulässigkeit der Annahme von Grenzwerten fortfallen. Anderseits, wenn es sich um die Anwendung handelt, kommt man vollständig auf meinen ursprünglichen Ansatz zurück. In jedem konkreten Fall muß eben angenommen werden, daß die als Daten in die Aufgabe eingehenden Ausgangskollektivs und die Stellenauswahlen, die etwa zur Ableitung der gesuchten Kollektivs benötigt werden, einem „Körper" angehören, für den die Axiome gelten. Dann verlaufen alle Rechnungen genau so, wie es in meinem zweiten Vortrag auseinandergesetzt wurde, und auch die resultierenden Kollektivs erfüllen die Bedingungen.

Die bereits angeführten neuen Resultate Copelands können dazu benutzt werden, die Dörgesche Idee zur konkreten Ausführung zu bringen. Die weitgehende Übereinstimmung der Auffassung zwischen Dörge und mir wird besonders deutlich, wenn man an meine vorhin gegebene Interpretation der Regellosigkeitsforderung denkt. Ich glaube, daß man von den Dörgeschen Untersuchungen sagen kann, sie seien ein Weg, die Widerspruchsfreiheit der auf Häufigkeitsdefinition und Regellosigkeitsbegriff aufgebauten Wahrscheinlichkeitstheorie zu beweisen.

Eine terminologische Frage.

Ich muß noch kurz auf eine Frage zu sprechen kommen, die freilich rein terminologischer Natur ist. Man bezeichnet es manchmal als einen Mangel meiner Theorie, daß sie gewisse rein mathematische Fragen, die mit Grenzwerten der relativen Häufigkeiten in gesetzmäßig definierten Zahlenreihen zusammenhängen, aus der Betrachtung ausschließt. Nun, ich schließe nichts aus und verwehre niemandem die Beschäftigung mit diesen Fragen, ja ich interessiere mich oft selbst für eine solche. Alles, was man mir vorwerfen darf, ist, daß ich es nützlich gefunden habe, für Merkmalfolgen, die der Regellosigkeitsforderung genügen, einen besonderen Namen einzuführen, den des „Kollektivs". Auch schien es mir zweckmäßig, mit dem Wert „Wahrscheinlichkeit" den Grenzwert der relativen Häufigkeit nur in dem besonderen Fall des Kollektivs zu bezeichnen. Dabei leitete mich das Bestreben, diejenigen Untersuchungen, die den Verhältnissen bei Glücksspielen und ähnlichen Erscheinungsreihen gelten, durch eine einheitliche Terminologie zusammenzufassen. Es bleibt jedem unbenommen, das Wort „Wahrscheinlichkeit" in weiterem Sinne zu verwenden und z. B. zu sagen, es bestehe die Wahrscheinlichkeit $^1/_2$ dafür, beim Durchlaufen der natürlichen Zahlenreihe auf eine gerade Zahl zu stoßen. Nur wird, wer so vorgeht, nicht darum herumkommen, zu erklären, worin er den Unterschied zwischen der natürlichen Zahlenreihe und einer durch ein Würfelspiel bestimmten Merkmalfolge erblickt. Durch Änderung der Bezeichnung schafft man die Schwierigkeit nicht aus der Welt.

Ebensowenig kann ich anerkennen, eine Theorie sei deshalb „allgemeiner" oder gar „überlegen", weil sie eine eingeschränkte Regellosigkeit zugrundelegt und daher einen größeren Kreis von Merkmalfolgen in Betracht ziehen kann. Die wesentliche Schwierigkeit bleibt immer die, diejenigen Folgen abzugrenzen, in denen man die Darstellung von Ergebnisfolgen eines Glücksspieles sieht. Daß es außerhalb dieses Gebietes auch Fragen von großem mathematischen Interesse geben kann, ist selbstverständlich und von mir unbestritten. Aber eine Theorie, die gar nicht den Versuch unternimmt, eine Abgrenzung des fraglichen Gebietes vorzunehmen, scheint mir nicht nur nicht überlegen, sondern sehr weit hinter den berechtigten Anforderungen zurückzubleiben.

Man kann, wie wir später an Beispielen sehen werden, aus Kollektivs (in meinem Sinn) durch Operationen, die nicht zu unseren vier Grundoperationen gehören und auch nicht aus solchen zusammengesetzt sind, Merkmalfolgen ableiten, die nicht mehr die Eigenschaften eines Kollektivs besitzen. Soweit solche Folgen von sachlichem Interesse sind (Folgen mit sogenannter Wahrscheinlichkeitsnachwirkung), werden sie im Rahmen meiner Theorie behandelt und ich sehe keinen besonderen Nachteil darin, daß — meiner Terminologie gemäß — die Grenzwerte der relativen Häufigkeiten, falls sie innerhalb dieser Folgen existieren, nicht mehr als Wahrscheinlichkeiten bezeichnet werden. Es ist meiner Ansicht nach durchaus bequem und dem Verständnis nur förderlich, in solchen Fällen sich mit dem Ausdruck „Grenzhäufigkeit" zu begnügen. Doch besteht natürlich kein logischer Zwang zu dieser vorsichtigen Art des Ausdruckes und ich sehe es als möglich an, daß, wenn die Häufigkeitstheorie einmal allgemein Eingang gefunden hat, man sich eine etwas größere Freiheit der Ausdrucksweise gestatten wird.

Die Nihilisten.

Es müssen schließlich noch ein paar Worte gesagt werden über diejenigen unter den zeitgenössischen Mathematikern, die mehr oder weniger ausgesprochen die Auffassung vertreten, man brauche überhaupt keine Definition oder Erklärung für den Wahrscheinlichkeitsbegriff. Was Wahrscheinlichkeit sei, wisse ohnehin jedermann, der die Umgangssprache beherrscht, und Aufgabe der Theorie sei es eben, die Größe *dieser* Wahrscheinlichkeit in jedem konkreten Fall zu berechnen. Wer so denkt, verkennt völlig das Wesen der exakten Wissenschaft, darüber ist im ersten Vortrag und auch heute, als vom „natürlichen" Wahrscheinlichkeitsbegriff der Philosophen die Rede war, schon alles nötige gesagt worden. Richtig ist nur, daß, historisch gesehen, eine derartige Auffassung am Anfang der Entwicklung stand. Alle Theorie ist ursprünglich aus dem Wunsche entstanden, zwischen Begriffen, die man als fest gegeben ansah, Beziehungen aufzufinden. Aber bei der Verfolgung dieses Weges hat sich eben gezeigt, daß das, was die Umgangssprache mit bestimmten Namen belegt, keine geeignete Grundlage für theoretische Aussagen bildet. So sind allmählich auf allen Gebieten, in denen die Forschung genügend weit fortgeschritten ist, die „künstlichen“ oder theoretischen Begriffe geschaffen worden. Wir wissen auch, daß diese Begriffsbildung ein wesentlicher Teil der Fortschritte ist, daß überall, von den abstraktesten Partien der Mathematik bis zu rein physikalischen Fragen (wofern sie überhaupt theoretisch behandelt werden), das Aufstellen von Definitionen und von Sätzen Hand in Hand geht.

Ein Beispiel aus der neueren Geschichte der Wissenschaft mag dies noch erläutern. In der gesamten theoretischen Physik galt es bis etwa an die letzte Jahrhundertwende als selbstverständlich und keiner Erklärung bedürftig, wenn von der Gleichzeitigkeit zweier Ereignisse an verschiedenen Orten gesprochen wurde. Heute ist es jedem Physiker geläufig — dies ist das wesentlichste Ergebnis der sogenannten speziellen Relativitätstheorie von Albert Einstein —, daß der Begriff der Gleichzeitigkeit einer Definition bedarf. Von dieser Definition aus entwickelt sich die ganze Theorie, die als eine der fruchtbarsten der neueren Physik anerkannt ist. Sie fällt vollständig fort für denjenigen, der sich auf den Standpunkt stellt, man wisse „ohnehin“, d. h. auf Grund der Sprachkenntnis schon, was „gleichzeitig“ heißt.

Es scheint mir daher, daß man die Widerlegung derer, die sich einer Definition der Wahrscheinlichkeit gegenüber völlig ablehnend verhalten, der fortschreitenden Entwicklung überlassen kann. Es wäre vielleicht überhaupt nicht nötig gewesen, die „Nihilisten“ zu erwähnen, wenn es nicht zwischen ihrem extremen Standpunkt und dem, der unserer Auffassung von Begriffsbildung in den exakten Wissenschaften entspricht, eine Reihe von Zwischenstufen gäbe, die nicht ganz übergangen werden können.

Beschränkung auf ein einziges Kollektiv.

Ein Standpunkt, auf den sich Mathematiker gerne zurückziehen, wird durch das schöne und sehr lesenswerte Büchlein von A. Kolmogoroff „Grundbegriffe der Wahrscheinlichkeitsrechnung“ repräsentiert. Richtet man sein Augenmerk ausschließlich auf die rein mathematische Seite dessen, was den Inhalt eines Lehrganges der Wahrscheinlichkeitsrechnung ausmacht, so wird man bemerken, daß es sich grossenteils um Rechnungen von einem und demselben Typus handelt, nämlich: Definiert ist in irgendeiner Weise die Verteilung in einem Kollektiv; gesucht wird die

Wahrscheinlichkeit, die vermöge dieser Verteilung auf einen bestimmten, in komplizierter Weise abgegrenzten Teil des Merkmalraumes entfällt. Solche Aufgaben, die nach unserer Terminologie unter den Begriff der „Mischung" fallen, können sehr kompliziert sein. Zum Beispiel:

Das gegebene Kollektiv bestehe aus der Verbindung von n einfachen Alternativen, wobei n eine sehr große Zahl ist. Das Merkmal eines Elementes wird also gebildet von einer Aufeinanderfolge von n Zeichen, die teils Nullen, teils Einser sind (oder zwei verschiedene Farben usw.). Die Wahrscheinlichkeit jedes bestimmten Gesamtergebnisses, also jeder der 2^n Kombinationen von n derartigen Zeichen wird als bekannt angesehen. Es sei nun m eine zweite feste Zahl, die ebenfalls groß, aber kleiner als n ist, ferner x eine variable Zahl, die zwischen m und n liegt, endlich $f(x)$ eine gegebene Funktion von x, z. B. die Quadratwurzel aus x. Man kann fragen: Wie groß ist die Wahrscheinlichkeit dafür, daß für alle x, die zwischen m und n liegen, die Anzahl der Einser, die sich unter den ersten x Zeichen befinden, kleiner als $f(x)$ ist? Offenbar wird durch die Formulierung dieser Frage ein bestimmter Teil der 2^n möglichen Kombinationen, ein Teil, der nur von der Zahl m und der Funktion $f(x)$ abhängt, herausgehoben und es wird nach der Summe der Wahrscheinlichkeiten dieser herausgegriffenen Kombinationen gefragt. Die Aufgabe ist genau das, was wir eine Mischung genannt haben, ihre mathematische Lösung kann sich sehr schwierig und umständlich gestalten, obwohl hier nur eine einzige Operation an dem Ausgangskollektiv (d. i. der Verbindung von n Alternativen) vorgenommen wird. In der Literatur kennt man z. B. die Lösung für den Fall, daß $f(x)$ proportional ist der Quadratwurzel aus x mal dem Logarithmus des Logarithmus von x, unter der Voraussetzung, daß m und n unendlich groß werden.

Kehren wir nun zu der allgemeinen, prinzipiellen Frage zurück, die uns hier interessiert. Es ist verständlich, daß jemand, der sich dauernd mit schwierigen Aufgaben des ebengenannten Typus beschäftigt, schließlich eine Abgrenzung seines Arbeitsgebietes sucht, die genau diesen Aufgabentypus umfaßt. So ist man zu dem Standpunkt gelangt, die Wahrscheinlichkeitsrechnung bestehe nur in der Beschäftigung mit jeweils einem einzigen Kollektiv und seiner Verteilung, an der eine gewisse Summation oder

Integration ausgeführt wird. Als „Grundlagen“ oder „Axiome“ deɪ Theorie seien daher lediglich die Einschränkungen anzusehen, die man den zugelassenen Verteilungen und Summationen auferlegt. Die „Axiomatik“ einer solchen Theorie besteht in einer Reihe funktionentheoretischer Voraussetzungen über die zulässigen Verteilungsfunktionen und über die zugelassenen Teilmengen des Merkmalraumes, für die Wahrscheinlichkeiten gesucht werden.

So beschaffen ist unter anderen das System der Grundlagen in dem erwähnten Buch von Kolmogoroff. Allerdings fügt er der Aufstellung der Axiome den Satz hinzu: „Unser Axiomensystem ist aber unvollständig: in verschiedenen Problemen der Wahrscheinlichkeitsrechnung betrachtet man verschiedene Wahrscheinlichkeitsfelder“ (Feld ist hier ein anderer Ausdruck für Verteilung). Andere Autoren, die im übrigen den gleichen „Mathematikerstandpunkt“ vertreten, sehen weniger klar die Grenzen, die dieser Auffassung gezogen sind.

Nun, unsere Darstellung der Grundlagen sucht gerade das zu bieten, was in der einseitig mathematisch-formalen Auffassung beiseite gelassen ist, und sie sieht in diesem Teil der Aufgabe die wesentliche Schwierigkeit, die zu überwinden ist.

Ein Teil der Mengenlehre? Nein!

Geht man konsequent auf dem Wege weiter, der nur ein Kollektiv kennt und nur Aufgaben eines einzigen Typus zu behandeln lehrt, so gelangt man unfehlbar zu dem Ende, daß es eine Wahrscheinlichkeitsrechnung überhaupt nicht gibt. Es gibt dann nur gewisse mathematische Aufgaben über reelle Funktionen und Punktmengen, die man natürlich in ein größeres Gebiet rein mathematischer Aufgaben einordnen kann. „Von diesem Standpunkt aus“, heißt es in einer kürzlich erschienenen Besprechung des Kolmogoroffschen Buches, „scheint die Wahrscheinlichkeitsrechnung jede selbständige Berechtigung zu verlieren, da sie ja nur ein Teilgebiet aus der Theorie der additiven Mengenfunktionen wird.“ So hat man auch schon behauptet, es gäbe keine Hydromechanik, sondern nur eine bestimmte Randwertaufgabe aus der Theorie partieller Differentialgleichungen. Und vor etwa zwanzig Jahren, als die Einsteinsche physikalische Theorie unter den Mathematikern bekannt wurde, hieß es, jetzt sei die Elektrodynamik ein Teil der Gruppentheorie geworden.

Diese Gleichsetzung zweier Dinge, die verschiedenen Kategorien angehören, dieses Gleichheitszeichen zwischen Werk und Werkzeug, ist dem logisch Denkenden unerträglich. Eine mathematische Untersuchung, so schwierig sie ist und so großen Raum sie im Rahmen einer physikalischen Theorie einnimmt, kann niemals mit dieser Theorie identisch sein, noch viel weniger kann die physikalische Theorie einen Teil eines mathematischen Aufgabenkreises bilden. Man mag sein Interesse noch so sehr auf den mathematischen, den tautologischen Teil der Gedankengänge konzentrieren und die Ausgangssätze, aus denen die mathematische Fragestellung fließt, noch so kurz abtun, nur das logische Verhältnis der beiden Begriffe darf man nicht umkehren. Um ein Gleichnis aus einem anderen Gebiete zu gebrauchen: Ein Staat ist nicht identisch mit seiner Verwaltung und ist nicht ein Teil der Verwaltungsarbeit, die irgendwo getan wird; man könnte höchstens sagen, daß der umfassendste Teil der Äußerungen staatlichen Lebens in einem bestimmten Fall die Verwaltung sei od. dgl.

So ist die Wahrscheinlichkeitstheorie niemals ein Teil der Mengenlehre, sondern sie ist die Theorie gewisser beobachtbarer Erscheinungen, die in dem Bild des „Kollektivs" idealisiert wurden, und sie bedient sich nur der Sätze der Mengenlehre, insbesondere der Integrationstheorie, um die Probleme, die durch die Definition des Kollektivs gesetzt werden, zu lösen. Es gibt daher auch nicht, wie gelegentlich behauptet wird, eine „mengentheoretische Auffassung" der Wahrscheinlichkeit, die mit der Häufigkeitsauffassung in Widerspruch steht oder mit ihr „versöhnt" werden müßte. Das einzige, was übrig bleibt, wenn man die neueste formale Entwicklung gewisser Problemstellungen charakterisieren will, ist der wenig inhaltsreiche Satz, daß innerhalb der Wahrscheinlichkeitstheorie keine anderen mathematischen Hilfsmittel zur Durchführung von Summierungen (Integrationen) verwendet werden als solche, die der allgemeinen Mengenlehre angehören.

Zusammenfassung und Schluß.

Ich bin am Ende der Ausführungen angelangt, die ich der Kritik der Grundlagen der Wahrscheinlichkeitsrechnung im Rahmen dieser Vorträge widmen wollte. Wenn ich jetzt einige

Hauptpunkte des Gesagten zusammenfassen soll, so kann ich gleich an die letzten Bemerkungen anknüpfen. Der Standpunkt, den ich vertrete, ist durch folgende Merkmale gekennzeichnet.

1. Die Wahrscheinlichkeitsrechnung (oder die Theorie der zahlenmäßig erfaßbaren Wahrscheinlichkeiten) ist die Theorie bestimmter, der Beobachtung zugänglicher Erscheinungen, der Wiederholungs- und Massenvorgänge etwa vom Typus der Glücksspiele, der Bevölkerungsbewegung, der Bewegung Brownscher Partikel usf. Das Wort „Theorie“ ist hier in demselben Sinn gemeint wie die Hydromechanik Theorie der Flüssigkeitsströmungen, die Thermodynamik Theorie der Wärmevorgänge, die Geometrie Theorie der räumlichen Erscheinungen heißt.

2. An die Spitze jeder derartigen Theorie treten gewisse Ausgangssätze, die man oft Axiome nennt. In ihnen kommen allgemeine Erfahrungsinhalte zur Verwertung, ohne daß sie unmittelbar als Erfahrungssätze angesprochen werden können. Die Axiome umgrenzen, definieren den Gegenstand der Theorie in dem Sinn, daß alle weiteren Sätze Folgerungen, d. h. tautologische Umformungen der Ausgangssätze, unter Einbeziehung der Daten der jeweiligen Aufgabe, bilden.

3. Die entscheidende Begriffsbildung in unserer Theorie ist die Definition des Kollektivs: Es gibt nur Wahrscheinlichkeiten innerhalb eines Kollektivs und alle Aufgaben bestehen darin, daß aus gegebenen Kollektivs nach bestimmten Regeln neue abgeleitet und die Verteilungen in diesen gesucht werden. Dieser Gedanke, die klare Beschränkung auf Aufsuchung von Beziehungen zwischen Verteilungen, wird in keiner anderen bisher bekannt gewordenen Theorie wirklich durchgeführt.

4. Akzessorisch und in einem gewissen Maße modifizierbar ist die genaue Festsetzung der Eigenschaften, die ein Kollektiv besitzen muß: die Existenz der Grenzwerte der relativen Häufigkeiten und die Regellosigkeit.

5. Die Einwände, die bisher gegen meine Formulierung der Regellosigkeitsforderung erhoben wurden, haben bisher zu keiner neuen geführt, die als Ersatz für sie einzutreten vermag. Die Autoren, die die allgemeine Regellosigkeit „ablehnen“ und durch eine beschränkte ersetzen, schließen entweder alle Fragen von der Beantwortung aus, die nicht der von ihnen willkürlich gesetzten Beschränkung entsprechen; oder sie nehmen in jedem konkreten

Fall die Regellosigkeit, die gerade gebraucht wird, als ein Datum der betreffenden Aufgabe an, was nur auf eine Änderung der Darstellungsform hinausläuft.

Vierter Vortrag.

Die Gesetze der großen Zahlen.

Unter den vielen schwierigen Fragen, die mit einer rationellen Grundlegung der Wahrscheinlichkeitstheorie verknüpft sind, gibt es gewiß keine, in der solche Verwirrung herrschte, wie in der Frage nach dem Inhalt und der Bedeutung des „Gesetzes der großen Zahlen“ und seiner Beziehung zur Häufigkeitstheorie der Wahrscheinlichkeit. Die meisten Autoren pendeln zwischen den beiden Behauptungen, die Erklärung der Wahrscheinlichkeit eines Ereignisses als Grenzwert der relativen Häufigkeit seines Auftretens postuliere das POISSONsche Gesetz und sie widerspreche ihm. Keines von beiden trifft zu.

Eine ausführliche Erörterung dieser Fragen, wie ich sie ja schon mehrfach angekündigt habe, muß naturgemäß im Rahmen unserer Vorträge Platz finden. Eine weitgehende Beschränkung ist mir dadurch auferlegt, daß ich keine mathematischen Kenntnisse voraussetzen und keine mathematischen Formeln benutzen will. Gleichwohl hoffe ich, das, worauf es ankommt, verständlich machen zu können. Neben der geläufigsten Aussage werde ich auch ihr klassisches Gegenstück, das sog. zweite Gesetz der großen Zahlen besprechen und dann die Ergänzungen skizzieren, die die neuere mathematische Forschung hinzugefügt hat. Zunächst also etwas über das POISSONsche Gesetz, das soviel Verwirrung in den Köpfen angerichtet hat.

Die beiden verschiedenen Aussagen von POISSON.

Die letzte Ursache der Verwirrung liegt bei POISSON selbst, der an verschiedenen Stellen seiner „Recherches sur la probabilité des jugements“, wie schon einmal erwähnt, zwei ganz verschiedene Aussagen mit dem gleichen Namen belegt und wohl auch tatsächlich für gleichbedeutend hält. In der Einleitung des Buches spricht er sich ganz deutlich nach der einen Richtung aus: „Erscheinungen verschiedenster Art sind einem allgemeinen Gesetze unterworfen, das man das ‚Gesetz der großen Zahlen' nennen kann. Es besteht darin, daß, wenn man sehr große Anzahlen von gleichartigen Ereignissen beobachtet, die von konstanten Ursachen und von solchen abhängen, die unregelmäßig, nach der einen und anderen Richtung veränderlich sind, ohne daß ihre Veränderung in einem bestimmten Sinn fortschreitet, man zwischen diesen Zahlen Verhältnisse finden wird, die nahezu unveränderlich sind. Für jede Art von Erscheinungen haben diese Verhältnisse besondere Werte, denen sie sich um so mehr nähern, je größer die Reihe der beobachteten Erscheinungen ist, und die sie in aller Strenge erreichen würden, wenn es möglich wäre, die Reihe der Beobachtungen ins Unendliche auszudehnen."

Aus diesen Worten und den anschließenden Ausführungen, die eine Fülle von Erfahrungsmaterial vor dem Leser ausbreiten, geht ganz unzweideutig hervor, daß hier mit „Gesetz der großen Zahlen" eine Beobachtungs- oder Erfahrungstatsache gemeint ist. Die Zahlen, von deren Verhältnissen die Rede ist, sind offenbar die Wiederholungszahlen der einzelnen Ereignisse oder der verschiedenen Ausgangsmöglichkeiten eines Versuches. Tritt ein Ereignis in n Versuchen m-mal ein, so nennen wir den Quotienten $m:n$ die „relative Häufigkeit" seines Auftretens. Hiernach ist der Inhalt des von Poisson in seiner Einleitung dargelegten Gesetzes völlig gleichbedeutend mit dem, was ich früher als die „erste Forderung" an ein Kollektiv formuliert habe: Die relative Häufigkeit, mit der ein Ereignis auftritt, nähert sich bei andauernder Fortsetzung der Versuche immer mehr einem festen Wert. Würde man mit „Gesetz der großen Zahlen" immer nur das meinen, was Poisson in seiner Einleitung so bezeichnet hat, so wäre es ganz richtig, zu sagen, daß dieses Gesetz die Erfahrungsgrundlage zum Ausdruck bringt, auf die sich die Definition der Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit stützen muß.

Allein ein großer Teil des Poissonschen Werkes ist der Ableitung und der Besprechung eines bestimmten mathematischen Theorems gewidmet, für das Poisson selbst wieder die Bezeichnung „Gesetz der großen Zahlen" verwendet und das heute zumeist unter diesem Namen, manchmal auch als „Poissonsches Gesetz" schlechthin angeführt wird. Dieses Theorem ist eine gewisse Verallgemeinerung eines schon von JACOB BERNOULLI herrührenden Satzes, den wir in folgender Form aussprechen können:

Wenn man einen einfachen Alternativversuch, dessen positives Ergebnis die Wahrscheinlichkeit p besitzt, n-mal wiederholt und mit ε eine beliebig kleine Zahl bezeichnet, so geht die Wahrscheinlichkeit dafür, daß der Versuch mindestens $(pn - \varepsilon n)$-mal und höchstens $(pn + \varepsilon n)$-mal positiv ausfällt, mit wachsendem n gegen Eins.

Konkreter formuliert besagt der BERNOULLIsche Satz: Wenn man 100mal mit einer Münze „Kopf oder Adler" wirft, so gibt es eine gewisse Wahrscheinlichkeit dafür, 49- bis 51mal die „Kopf"-seite zu treffen; wirft man 1000mal, so ist die Wahrscheinlichkeit, 490- bis 510mal „Kopf" zu werfen, schon größer; noch näher an Eins liegt die Wahrscheinlichkeit dafür, unter 10000 Versuchen mindestens 4900 und höchstens 5100 „Kopf"-Ergebnisse zu erzielen usf. (In diesem Beispiel ist ersichtlich $p = {}^1/_2$ und $\varepsilon = {}^1/_{100}$ gesetzt.) Die POISSONsche Erweiterung des Satzes geht nur dahin, daß die Versuchsreihe nicht mit einer Münze und auch nicht mit lauter gleichen ausgeführt werden muß; man darf jedesmal eine andere Münze nehmen, nur müssen die Münzen in ihrer Gesamtheit die Eigenschaft besitzen, daß das arithmetische Mittel aus den n Wahrscheinlichkeiten eines Kopfwurfes den Wert p, in unserem Fall ${}^1/_2$, besitzt. Eine noch allgemeinere Fassung des Satzes, die durch TSCHEBYSCHEFF in besonders einfacher Form abgeleitet wurde, bezieht sich auf den Fall, daß die einfache Alternative („Kopf oder Adler") durch einen Versuch von mehrfacher Ausgangsmöglichkeit ersetzt wird. Für unsere grundsätzliche Erörterung genügt es durchaus, die engste Form, wie sie beim gewöhnlichen Spiel mit einer Münze auftritt, ins Auge zu fassen.

Wir fragen: Wie hängt der Inhalt des bewiesenen mathematischen Satzes, den wir der Kürze halber als „POISSONsches Theorem" bezeichnen wollen, mit der Erfahrungstatsache zusammen, die POISSON an die Spitze seiner Betrachtungen gestellt hat? Kann man wirklich behaupten, daß jene Tatsache durch diesen Satz wiedergegeben wird oder daß überhaupt hier durch theoretische Überlegungen etwas abgeleitet wurde, was sich an der Beobachtung prüfen läßt und durch sie bestätigt wird?

Der Standpunkt der Gleichmöglichkeitsdefinition.

Um diese Frage zu beantworten, müssen wir davon ausgehen, was BERNOULLI und seine Nachfolger unter Wahrscheinlichkeit verstehen. Denn das „POISSONsche Theorem" enthält das Wort „Wahrscheinlichkeit", die eingangs formulierte Aussage enthält es nicht. Wir müssen also in den Wortlaut des POISSONschen Theorems für das Wort „Wahrscheinlichkeit" das einsetzen, wodurch die Wahrscheinlichkeit bei POISSON definiert wird, um so den vollständigen Inhalt dessen, was POISSON beweist, restlos zu erfassen. Nun ist, wie wir wissen, die klassische Wahrscheinlichkeitsdefinition dadurch gekennzeichnet, daß sie keinerlei Bezug nimmt auf die Häufigkeit des Auftretens eines Ereignisses, sondern rein formal erklärt: Wahrscheinlichkeit ist der Quotient aus der Anzahl der „günstigen" Fälle durch die Gesamtzahl aller „gleichmöglichen" Fälle. Bei einer normalen Münze ist das Auffallen auf die eine oder andere Seite „gleichmöglich", der eine dieser Fälle ist dem Erscheinen der Kopfseite „günstig", demnach ist die Wahrscheinlichkeit des „Kopf"-Ergebnisses $^1/_2$. Nach dieser Erklärung, die der Ableitung des POISSONschen Theorems ausschließlich zugrunde liegt, ist der Ausdruck, ein Ereignis habe eine Wahrscheinlichkeit nahe 1, gleichbedeutend mit der Aussage, daß fast alle unter den „gleichmöglichen" Fällen zu den dem Ereignis „günstigen" gehören. Führt man mit einer Münze n Würfe aus, so gibt es bei großem n außerordentlich viele Ergebnismöglichkeiten; es können die ersten zehn Würfe „Kopf", die übrigen „Adler", die ersten zwanzig und die letzten dreißig Würfe „Kopf" zeigen, die übrigen „Adler" usf., in wahrhaftig sehr reichhaltiger Abwechslung. Jedes dieser Gesamtergebnisse muß als „gleichmöglich" gelten, wenn man für die Wahrscheinlichkeit eines „Kopf"-Wurfes, also für die oben mit p bezeichnete Größe, $^1/_2$ setzt. Wählen wir wieder $\varepsilon = {}^1/_{100}$, so besagt das POISSONsche Theorem: Unter den sehr vielen Ergebnissen, die bei großem n möglich sind, haben weitaus die meisten die Eigenschaft, daß die Anzahl der in ihnen vorkommenden „Kopf"-Würfe um höchstens $^n/_{100}$ nach oben oder unten von $^n/_2$ abweicht.

Nichts als dieses hat POISSON bewiesen: Der von ihm mathematisch abgeleitete Satz steht in keinerlei Zusammenhang mit dem Ablauf einer Versuchsreihe.

Arithmetische Darstellung.

Um uns die Aussage noch anschaulicher zu machen, wollen wir uns eine Ergebnisgruppe des Spieles mit einer Münze so dargestellt denken, daß wir für die Kopfseite jedesmal eine 1, für die andere eine 0 anschreiben. Auf diese Weise entspricht jeder möglichen Serie von $n = 100$ Würfen eine bestimmte hundertstellige Zahl, mit Nullen und Einsern als einzigen Ziffern. Wenn etwaige Nullen links vom ersten Einser gestrichen werden, stellt auch jede solche Zahl von weniger als hundert Stellen eine Versuchsserie dar. Man kann im Prinzip alle diese Zahlen, von denen jede ein gleichmögliches Ergebnis repräsentiert, nach einem einfachen Schema hintereinander aufschreiben. Die ersten, der Größe nach geordnet, sind

0, 1, 10, 11, 100, 101, 110, 111, 1000, 1001 usw.

(Die Anzahl der so zu bildenden Zahlen ist bei $n = 100$ allerdings enorm groß, gleich 2 zur hundertsten Potenz, etwa eine Million Trillionen.) Dabei bedeutet z. B. 101, daß in der Versuchsreihe zuerst 97 Nullen stehen, dann eine Eins, hierauf eine Null, endlich wieder eine Eins folgt. Hätten wir $n = 1000$, so würde die Zahlenreihe, die sämtliche Ergebnismöglichkeiten umfaßt, in der gleichen Weise beginnen, nur sehr viel länger werden (nämlich 2 hoch 1000 Zahlen umfassen), da man jetzt bis zu 1000 Stellen gehen muß, und es würde 101 bedeuten, daß zuerst 997 Nullen kommen, dann der erste Einser usf. Der Poissonsche Satz ist nun nichts als eine Aussage über diese Zahlen, von denen die ersten angeschrieben wurden. Wenn wir alle aus Nullen und Einsern gebildete Zahlen bis zu den 100stelligen nehmen, so weisen etwa 16 % unter ihnen 49 bis 51 Einser auf; wenn wir bis zu den 1000stelligen gehen, so finden wir, daß ein weit größerer Prozentsatz, nämlich rund 47 %, unter ihnen 490 bis 510 Einser besitzt. Unter den 10000stelligen Kombinationen von Nullen und Einsern besitzen schon mehr als 95 % zwischen 4900 und 5100 Einser, kaum 5 % entfallen auf solche Kombinationen, bei denen die Zahl der Einser um mehr als $^1/_{100}$ von der Hälfte, d. i. von 5000, abweicht. Dieses Verhalten wird bei weiterer Vergrößerung der Stellenzahl immer ausgeprägter. Es wird bei Verwendung der klassischen Wahrscheinlichkeitsdefinition in der Form zum Ausdruck gebracht, daß man sagt: die „Wahrscheinlichkeit“ der Ergebniszahlen 49 bis 51 sei im ersten Fall 0,16, die analoge im zweiten Fall 0,47, im dritten 0,95. Der Inhalt des Theorems (für $p = {}^1/_2$), wie es bei Bernoulli und Poisson bewiesen ist, läßt sich, wenn wir $\varepsilon = 0{,}01$ wählen, wie folgt aussprechen: Schreibt man alle aus Nullen und Einsern bestehenden Zahlen der Größe nach geordnet bis einschließlich der n-stelligen auf, so bilden diejenigen unter ihnen, bei denen die Anzahl der Einser mindestens 0,49 n und höchstens 0,51 n beträgt, eine mit wachsendem n immer stärker werdende Majorität.

Diese Aussage ist rein arithmetischer Natur, sie bezieht sich auf gewisse Zahlen und stellt über deren Eigenschaften etwas fest. Mit dem, was bei der ein- oder mehrmaligen Vornahme von 1000 Würfen wirklich geschieht, d. h. welcher Anordnung von Nullen und Einsern die wirklich eintretenden Versuchsserien entsprechen werden, damit hat das Ganze nichts zu tun. Ein Schluß auf den Ablauf einer Versuchsreihe ist im Rahmen dieses Gedankenganges nicht möglich, weil nach der angenommenen Definition der Wahrscheinlichkeit diese nur etwas über das Verhältnis zwischen der Anzahl der günstigen und ungünstigen Fälle besagt, aber nichts über die Häufigkeit, mit der ein Ereignis eintritt oder ausbleibt.

Man sieht leicht ein, daß sich unsere Betrachtung grundsätzlich auf jeden anderen Fall an Stelle des „Kopf-oder-Adler“-Spieles übertragen läßt: Wenn es sich etwa um das Spiel mit einem „richtigen“ Würfel handelt, so treten nur an Stelle der aus Nullen und Einsern bestehenden Zahlen alle n-stelligen Zahlen mit den Ziffern 1 bis 6, und das Theorem besagt, daß bei großem n diejenigen, die ungefähr ${}^n/_6$ Einser enthalten, eine überwiegende Majorität bilden. Wir fassen das bisher Gesagte zusammen: Solange man mit einem Wahrscheinlichkeitsbegriff arbeitet, der nicht Bezug nimmt auf die Häufigkeit des Eintretens des Ereignisses, führt die mathematische Ableitung von Bernoulli-Poisson-Tschebyscheff zu keiner wie immer gearteten Aussage über den Verlauf einer Versuchsreihe und läßt sich daher in keinen Zusammenhang bringen mit der allgemeinen Erfahrungsgrundlage, von der Poisson ausgegangen war.

Nachträgliche Häufigkeitsdefinition.

Wie kam nun aber POISSON dazu, in seiner mathematischen Ableitung eine Bestätigung jenes erfahrungsgemäßen Verhaltens zu sehen, das er in seiner Einleitung als „Gesetz der großen Zahlen" bezeichnet hatte? Die Antwort auf diese Frage kann niemandem schwer fallen, der sie sich einmal stellt. Sie lautet: POISSON hat dem Wort „Wahrscheinlichkeit" am Ende seiner Rechnung eine andere Bedeutung beigelegt als die, die er ihm zu Anfang gegeben hatte. Die Wahrscheinlichkeit $^1/_2$ eines „Kopf"-Wurfes, die in die Rechnung eingeht, soll nur der Quotient der „günstigen" durch die „gleichmöglichen" Fälle sein, aber die Wahrscheinlichkeit nahe 1, die aus der Rechnung hervorgeht und in seinem Satze die entscheidende Rolle spielt, die soll bedeuten, daß das betreffende Ereignis, nämlich das Auftreten von $0{,}49\, n$ bis $0{,}51\, n$ „Kopf"-Würfen in einer Serie von n Versuchen, fast immer, bei fast jedem Serienversuch, zu beobachten ist. Niemand kann behaupten, daß eine solche Bedeutungsverschiebung zwischen Beginn und Ende einer Rechnung statthaft wäre. Auch ist es ganz unklar, bei welcher Wahrscheinlichkeitsgröße die Bedeutungsänderung eintreten soll. Darf schon die Wahrscheinlichkeit 0,16 bei 100 Würfen dahin gedeutet werden, daß Serien mit 49 bis 51 „Kopf"-Ergebnissen mit der relativen Häufigkeit 16 % auftreten oder gilt das erst für die Wahrscheinlichkeit 0,95, die bei der Serienlänge 10000 ausgerechnet wurde?

Wenn man die klassische Definition der Wahrscheinlichkeit unbedingt aufrecht erhalten will, so läßt sich die gewünschte Bedeutung des POISSONschen Theorems nur retten, wenn man — als deus ex machina — eine Hilfshypothese etwa folgender Art hinzunimmt: Sobald eine Rechnung für ein Ereignis einen Wahrscheinlichkeitswert, der nur wenig kleiner als 1 ist, ergeben hat, tritt dieses Ereignis bei fortgesetzten Versuchen fast jedesmal ein. Was ist das aber anderes als eine, wenn auch etwas eingeschränkte, Häufigkeitsdefinition der Wahrscheinlichkeit? Wenn ein Wahrscheinlichkeitswert von 0,999 bedeuten soll, daß das Ereignis fast immer beobachtet wird, warum nicht gleich zugeben, daß 0,50 Wahrscheinlichkeit heißt: das Ereignis trifft in der Hälfte aller Fälle ein? Freilich muß diese Festsetzung noch präzisiert werden und damit allein ist es auch noch nicht getan. Man muß erst zeigen, daß aus einer entsprechend präzisierten Häufigkeitsdefinition der Wahrscheinlichkeit sich der Poissonsche Satz ableiten läßt; der Gedankengang des klassischen Beweises wird sich dabei in entscheidenden Punkten ändern. Das Verfahren aber, nach der Ableitung eines Satzes einem darin vorkommenden Ausdruck eine neue Deutung zu geben, ist sicher nicht zulässig. Wir stellen nochmals fest: Nur, wenn die Wahrscheinlichkeit in irgendeiner Form als Häufigkeit des Ereigniseintrittes erklärt wird, läßt sich das Ergebnis der Poissonschen Rechnung überhaupt in Beziehung setzen zu dem, was Poisson in der Einleitung seines Buches als Gesetz der großen Zahlen bezeichnet.

Der Inhalt des Poissonschen Theorems.

Nun wird man aber mit Recht folgenden Einwand machen: Wenn man die Wahrscheinlichkeit eines „Kopf"-Wurfes als den Grenzwert der relativen Häufigkeit, mit der „Kopf" fällt, definieren will, muß man voraussetzen, daß es einen solchen Grenzwert gibt, mit anderen Worten, man muß das „Gesetz der großen Zahlen" der Poissonschen Einleitung von vornherein als gültig annehmen. Welchen Zweck kann es dann noch haben, durch mühsame Rechnung den Weg zu einem Theorem zu bahnen, das doch nur dasselbe aussagt, was schon als gültig vorausgesetzt wurde? Darauf antworten wir: Der Satz, der als Resultat der mathematischen Deduktion von Bernoulli-Poisson-Tschebyscheff erscheint, sagt, wenn man erst einmal die Häufigkeitsdefinition der Wahrscheinlichkeit angenommen hat, ungleich mehr aus, als die bloße Existenz eines Grenzwertes, er ist sehr viel inhaltsreicher als das in der Poissonschen Einleitung ausgesprochene „Gesetz der großen Zahlen". Sein wesentlicher Inhalt ist dann eine bestimmte Aussage über die Anordnung etwa der „Kopf"-und-„Adler"-Würfe bei unbegrenzt fortgesetztem Spiel. Man kann ohne Schwierigkeit Erscheinungsreihen angeben, für die das zutrifft, was Poisson in seiner Einleitung als kennzeichnend ausspricht, ohne daß für sie das Poissonsche Theorem Geltung hätte. Um dies einzusehen, wollen wir eine ganz bestimmte Versuchsfolge betrachten, die tatsächlich die Eigenschaft besitzt, daß die relative Häufigkeit des positiven Ergebnisses sich dem Wert $1/2$ nähert, wie dies beim „Kopf-oder-Adler"-Spiel der Fall ist, für die aber das Poissonsche Theorem nicht zu Recht besteht. Selbstverständlich ist das eine Versuchsreihe, auf die niemand, der ihre Eigenschaften kennt, die Wahrscheinlichkeitsrechnung anwenden wird. Aber unser Ziel ist ja nur, zu zeigen, daß durch das POISSONsche Theorem den Beobachtungsfolgen, die in der Wahrscheinlichkeitsrechnung betrachtet werden, d. h. den Kollektivs, eine besondere, für sie typische Eigenschaft zugeschrieben wird.

Ein Gegenbeispiel.

Wir nehmen als Versuchsobjekt eine Quadratwurzeltafel, d. i. eine Tabelle, die zu den aufeinanderfolgenden ganzen Zahlen 1, 2, 3, ... in inf. die Werte ihrer Quadratwurzeln etwa auf sechs oder mehr Dezimalstellen genau angibt. Unsere Aufmerksamkeit lenken wir ausschließlich auf die Ziffer, die jedesmal an der sechsten Stelle hinter dem Komma steht, und wollen als „positives" Versuchsergebnis, das wir auch mit einer „1" bezeichnen, den Fall ansehen, daß die sechste Ziffer eine 5, 6, 7, 8 oder 9 ist. Mit „0" oder als „negatives" Versuchsergebnis sei der Fall bezeichnet, daß die Quadratwurzel einer Zahl an sechster Stelle eine 0, 1, 2, 3 oder 4 aufweist. Auf diese Art erhalten wir als Ergebnis der ganzen Beobachtungsreihe eine Folge von Nullen und Einsern, die in bestimmter Weise abwechseln. Wenn auch jede reale Tafel naturgemäß bei irgendeiner Zahl abbrechen muß, macht es keine Schwierigkeit, sich die Folge der Nullen und Einser unbeschränkt fortgesetzt zu denken. Es ist plausibel und läßt sich mathematisch exakt beweisen, daß in dieser Folge die relative Häufigkeit sowohl der Nullen wie der Einser den Grenzwert $^1/_2$ besitzt. Ja es gilt der noch etwas allgemeinere Satz, daß an der sechsten Stelle der Quadratwurzeln jede der zehn Ziffern 0 bis 9 mit der Häufigkeit $^1/_{10}$ in der unendlichen Reihe auftritt. Nun aber wollen wir sehen, ob das Verhalten unserer Folge von Nullen und Einsern etwa auch dem entspricht, was das POISSONsche Theorem fordert. Darnach müßte, wenn man nur eine genügend große Serienlänge n wählt, in der unendlichen Folge fast jede Serie von n Ziffern ungefähr zur Hälfte aus Nullen und ungefähr zur Hälfte aus Einsern bestehen.

Der Anfang der Tafel weist in der Tat ein derartiges Bild auf, wovon man sich leicht überzeugt, wenn man etwa die Anzahl (ca. 30) der auf einer Seite stehenden Zahlen zur Serienlänge wählt und dann feststellt, wie oft eine Ziffer kleiner oder gleich 4 an sechster Stelle sich findet. Aber wie es weiter geht, wenn man sich die Tafel über das tatsächlich Gedruckte hinaus fortgesetzt denkt, das muß uns eine kleine Rechnung lehren. Wir benutzen dabei die leicht zu erweisende Formel, nach der die Quadratwurzel aus dem Ausdruck $a^2 + 1$, wenn a sehr groß gegen 1 ist, ungefähr $a + {}^1/_{2a}$ ist. Setzen wir etwa a gleich einer Million, also a^2 gleich einer Billion, so sehen wir, daß die Quadratwurzel, wenn man in der Tafel von $a^2 = 10^{12}$ um eine Zeile fortschreitet, sich nur um ${}^1/_2\, 10^{-6}$, d. i. um eine halbe Einheit der sechsten Stelle, ändert. Erst wenn man ungefähr zehn Schritte gemacht hat, erhöht sich der Wurzelwert um fünf Einheiten der sechsten Stelle. Dies bedeutet, daß in der Gegend der Tafel, in der a ungefähr eine Million beträgt, unsere Folge von Nullen und Einsern regelmäßige Iterationen etwa der Länge 10, und zwar abwechselnd solche von 10 Nullen und von 10 Einsern, aufweisen muß. Tatsächlich sieht ein Stück der Tafel in dieser Gegend so aus:

a^2	a	Fall
$10^{12} + 1237$	$10^6 + 0{,}0006185$	1
$+ 1238$	$+ 0{,}0006190$	1
$+ 1239$	6195	1
$+ 1240$	6200	0
$+ 1241$	6205	0
$+ 1242$	6210	0
....	....	....
$+ 1249$	6245	0
$+ 1250$	6255	1
....	....	....

Gegen wir noch weiter, z. B. bis a gleich hundert Millionen, so lehrt die gleiche Überlegung, daß jetzt abwechselnde Iterationen der ungefähren Länge 1000, von Nullen bzw. Einsern, auftreten werden usf.

Die Struktur der aus der Quadratwurzeltafel hervorgegangenen Folge von Nullen und Einsern ist also eine ganz andere als die, die in Übereinstimmung mit allen Erfahrungen das Poissonsche Theorem für eine solche Folge fordert, die das Ergebnis eines fortgesetzten „Kopf-oder-Adler“-Spieles darstellt. Nur in ihren ersten Anfängen zeigt sie eine scheinbare Regellosigkeit, im weiteren Verlauf besteht sie aus regelmäßigen Iterationen von langsam, aber unbeschränkt wachsender Länge. Der Widerspruch gegen das POISSONsche Theorem liegt auf der Hand. Wählen wir ein noch so großes festes n (z. B. $n = 500$) als Serienlänge, so wird, wenn man nur genügend weit fortschreitet (etwa bis $a = 100$ Millionen), die Länge der Iterationen viel größer als n sein. Es werden von da an fast alle Serien aus lauter Nullen oder lauter Einsern bestehen, während nach dem POISSONschen Theorem fast alle annähernd zur Hälfte aus Nullen und zur Hälfte aus Einsern zusammengesetzt sein sollten. Man sieht, daß hier, im Falle der Quadratwurzeln, der Häufigkeitsgrenzwert $^1/_2$ nur dadurch zustande kommt, daß immer annähernd gleich viel Serien aus lauter Nullen bzw. lauter Einsern bestehen, während im Falle eines Glücksspieles der Ausgleich zwischen Nullen und Einsern schon annähernd innerhalb jeder oder fast jeder Serie von genügender Länge erfolgt.

Unzulänglichkeiten.

Man könnte dies Beispiel nebenbei dazu benutzen, um zu zeigen, wie kritiklos auch manche Mathematiker Fragen der Wahrscheinlichkeitsrechnung behandeln oder behandelt haben. In manchem Lehrbuch auch der neuesten Zeit finden sich als Beispiele für die Anwendung der Lehrsätze der Wahrscheinlichkeitsrechnung Zahlenfolgen der eben besprochenen Art angeführt, die diesen Lehrsätzen direkt widersprechen und den Voraussetzungen einer wahrscheinlichkeitstheoretischen Betrachtung schon ihrer Herkunft nach nicht genügen können. Gar nicht zu reden von dem durch POINCARÉ zur Diskussion gestellten Beispiel der Logarithmentafel, in dem — wenn man richtig rechnet — sich zeigt, daß nicht einmal der Grenzwert der relativen Häufigkeit vorhanden ist.

Von den Klassikern der Wahrscheinlichkeitsrechnung BERNOULLI, POISSON, LAPLACE rühren auch zahlreiche rein analytische Sätze auf dem Gebiete der Reihenlehre, der Integralrechnung usf. her, deren hohen Wert im Rahmen der historischen Entwicklung der Analysis niemand anzweifeln wird. Trotzdem denkt kein Mathematiker daran, heute diese Sätze in ihrer alten Form unverändert aufrechtzuerhalten. Dies verbietet sich schon dadurch, daß in der Analysis alle grundlegenden Definitionen, die erst einen sicheren systematischen Aufbau ermöglichen, in der Zwischenzeit ganz neu geschaffen wurden. So selbstverständlich dieses Verhalten der Analytiker erscheint, so erstaunlich ist es, mit welch vollendeter Kritiklosigkeit Mathematiker ersten Ranges gelegentlich an der überlieferten Fassung des BERNOULLIschen oder POISSONschen Theorems einschließlich ihrer historischen Ableitungen festzuhalten suchen, obgleich hier dieselben Zirkelschlüsse, Vertauschungen von Grenzübergängen und ähnliche Fehler vorliegen, wie sie aus der Analysis längst ausgemerzt worden sind.

Das Beispiel der aus der Quadratwurzeltafel hervorgegangenen Null- und Einsfolge lehrt uns, daß es Erscheinungsreihen gibt, in denen die relativen Häufigkeiten der einzelnen Ergebnismöglichkeiten bestimmte Grenzwerte besitzen, ohne daß zugleich auch das POISSONsche Theorem gültig wäre. Damit wird vor allem der häufig wiederholte Einwand entkräftet, wonach mit der Annahme der Häufigkeitsdefinition der Wahrscheinlichkeit schon das Bestehen dieses Theorems postuliert würde. Noch unbegründeter ist es natürlich, zu behaupten, mit der Annahme, daß die relative Häufigkeit einen Grenzwert besitzt, sehe man das als „sicher“ an, was nach dem POISSONschen oder BERNOULLIschen Satz nur als „höchst wahrscheinlich“ gelten kann. Alle diese und ähnliche Irrtümer entstehen nur daraus, daß man von der klassischen Wahrscheinlichkeitsdefinition ausgeht, die nichts mit dem Erscheinungsablauf zu tun hat, und sich nachträglich einer Ausdrucksweise bedient, die auf diesen Ablauf Bezug nimmt („es ist mit Sicherheit zu erwarten, daß...“).

Richtige Ableitung.

Es ist jetzt klargestellt, daß das von POISSON im vierten Kapitel seines Buches abgeleitete Theorem, wenn man den Wahrscheinlichkeitsbegriff als Häufigkeitsgrenzwert auffaßt, eine wertvolle Aussage enthält, die über die Anordnung der Versuchsergebnisse, wie sie in einer lange fortgesetzten Versuchsreihe zu erwarten ist, einigen Aufschluß gibt. Die Frage ist aber noch offen, und dies ist die letzte von uns zu erörternde, ob man überhaupt das POISSONsche Theorem noch mathematisch beweisen kann, wenn man die Häufigkeitsdefinition der Wahrscheinlichkeit zugrunde legt. Dazu ist zu bemerken, daß ein Satz, der etwas über die Wirklichkeit aussagen soll, nur dann mathematisch ableitbar ist, wenn man an die Spitze der Ableitung bestimmte, der Erfahrung entnommene Ausgangssätze, sogenannte Axiome stellt. In unserem Fall besteht das eine dieser Axiome, wie aus allem bisherigen hervorgeht, in der Annahme, daß innerhalb der von der Wahrscheinlichkeitsrechnung zu behandelnden Erscheinungsreihen die relativen Häufigkeiten der einzelnen Ereignisse Grenzwerte besitzen. Dies ist nichts anderes als der allgemeine Erfahrungssatz, den POISSON in seiner Einleitung als Gesetz der großen Zahlen bezeichnet und den er — mit Unrecht — im vierten Kapitel seines Buches bewiesen zu haben glaubt. Aber mit diesem einen Axiom allein ist es noch nicht getan. Man muß die Erscheinungsreihen, auf die sich die wahrscheinlichkeitstheoretische Betrachtungsweise erstrecken soll, abgrenzen gegen diejenigen, die, wie die früher behandelte Quadratwurzeltafel, infolge einer ihnen innewohnenden Gesetzmäßigkeit sich dieser Art der Betrachtung entziehen. Dazu dient die von uns ausführlich besprochene Forderung der „Regellosigkeit" oder, wie wir es auch anschaulicher bezeichnet haben, das „Prinzip vom ausgeschlossenen Spielsystem".

Unter Zugrundelegung dieser *beiden* Eigenschaften des Kollektivs erhält man durch eine Reihe von rein mathematischen Schlüssen, auf die hier nicht näher eingegangen werden kann, das POISSONsche Theorem, und zwar gleich mit dem Sinn, den POISSON haben wollte (ohne daß *seine* Ableitung dazu führte), nämlich als eine Aussage über den tatsächlichen Ablauf der Erscheinungen. *Es erweist sich eben als eine spezielle Konsequenz aus der vollständigen Regellosigkeit einer Folge, daß bei Betrachtung genügend langer Serien ein statistischer Ausgleich schon innerhalb fast jeder Serie mit großer Annäherung stattfindet.* Dies ist der Inhalt des richtig abgeleiteten und richtig formulierten POISSONschen Theorems.

Man versteht, daß das Vorhandensein eines solchen Gesetzes die empirische Feststellung der Grenzwerte der relativen Häufigkeit außerordentlich *erleichtert*. Ja vielleicht wären wir praktisch nie zur Erkenntnis der ersten Eigenschaft des Kollektivs gelangt, wenn nicht infolge der Regellosigkeit eben das POISSONsche Theorem bestünde, wonach sich der Grenzwert annähernd in fast jeder längeren Beobachtungsserie finden muß. Aber dieser praktische Zusammenhang kann das logische Verhältnis der einzelnen Sätze nicht verwischen, das wir hier aufzuklären versucht haben und noch in folgenden kurzen Formulierungen zum Ausdruck bringen wollen.

Zusammenfassung.

1. Was POISSON in der Einleitung seines Buches als „Gesetz der großen Zahlen" bezeichnet, ist die Feststellung der Erfahrungstatsache, daß bei gewissen Erscheinungsreihen die relative Häufigkeit des Auftretens eines Ereignisses sich bei unbeschränkt fortgesetzter Beobachtung einem bestimmten Grenzwert nähert. Diese Feststellung wird unmittelbar in unserer „ersten Forderung" an ein Kollektiv zum Ausdruck gebracht.

2. Hält man an der klassischen Wahrscheinlichkeitsdefinition fest, so liefert der mathematische Satz, den POISSON im vierten Kapitel seines Buches ableitet und wieder als „Gesetz der großen Zahlen" bezeichnet, keinerlei Aussage über den Ablauf einer Erscheinungsreihe, läßt sich daher auch in keinerlei Beziehung zu der in 1. genannten Erfahrungstatsache setzen, sondern enthält lediglich eine Bemerkung rein arithmetischer Natur.

3. Nimmt man, gestützt auf den Erfahrungssatz von 1. die Definition an, Wahrscheinlichkeit eines Ereignisses sei der Grenzwert der relativen Häufigkeit seines Eintretens, so bedeutet der Satz von Kapitel 4 des POISSONschen Buches eine ganz bestimmte Aussage über den Erscheinungsablauf, die aber nicht zusammenfällt mit dem Erfahrungssatz der POISSONschen Einleitung, sondern *mehr* besagt, und zwar etwas Wesentliches über die Art der Abwechslung zwischen positiven und negativen Versuchsergebnissen.

4. Der Inhalt des eben erwähnten, mathematisch ableitbaren Satzes geht dahin, daß bei Betrachtung genügend langer Erscheinungsserien *innerhalb fast jeder von ihnen schon ein annähernder Ausgleich* der verschiedenen Ergebnismöglichkeiten stattfindet. Dagegen wäre z. B. in einer Reihe von Nullen und Einsern mit dem Bestehen des Grenzwertes $^1/_2$ für die relative Häufigkeit (aber nicht mit der Annahme der Regellosigkeit!) eine Anordnung verträglich, bei der immer länger werdende Iterationen von Nullen und Einsern miteinander abwechseln, so daß bei ge-

gebenem n innerhalb fast keiner Serie der Länge n ein annähernder Ausgleich vorhanden ist.

5. Will man das POISSONsche Theorem im Sinne der Häufigkeitsdefinition der Wahrscheinlichkeit richtig ableiten, so muß man außer auf den Erfahrungssatz von der Existenz der Grenzwerte sich auf einen weiteren stützen, der die vollständige „Regellosigkeit“ in dem Ablauf der betrachteten Erscheinungen feststellt und in unserer „zweiten Forderung“ an ein Kollektiv zum Ausdruck gebracht ist.

Nach dieser Klarstellung bleibt noch die terminologische Frage offen, was man zweckmäßigerweise als „Gesetz der großen Zahlen“ bezeichnen soll. Ich möchte vorschlagen, diesen Namen dem Theorem vorzubehalten, dessen erste Ableitung man BERNOULLI und POISSON verdankt, dagegen den in der POISSONschen Einleitung dargelegten Tatbestand als den Erfahrungssatz oder das Axiom von der Existenz der Grenzwerte zu bezeichnen.

Eine zweite „Brücke“.

Auf keinen Fall aber kann man — und dies ist das Ergebnis, um dessentwillen diese Ausführungen in erster Linie vorgebracht wurden — in dem „Gesetz der großen Zahlen“ etwas finden, was uns erspart, die Definition der Wahrscheinlichkeit auf die Häufigkeit des Ereigniseintrittes zu stützen und was einen Übergang von der Gleichmöglichkeitsdefinition zur Betrachtung wirklicher Erscheinungsabläufe gestattet. Ich muß mich jetzt noch einer zweiten, der BERNOULLI-POISSONschen ähnlichen Gedankenbildung zuwenden, deren sich die klassische Wahrscheinlichkeitsrechnung oft als Krücke zu ähnlichem Zwecke bedient. Es handelt sich da um das manchmal auch als „Umkehrung“ des BERNOULLIschen Satzes bezeichnete BAYESsche Theorem. Viele Autoren behaupten, daß dieses Theorem und nicht das BERNOULLIsche es sei, das eine Gleichsetzung der auf Grund der Gleichmöglichkeitsdefinition errechneten Wahrscheinlichkeiten mit den beobachteten Häufigkeiten legitimiert. Ich werde diese Frage nicht mit der gleichen Ausführlichkeit behandeln wie die frühere, da die grundsätzlichen Überlegungen vielfach parallel laufen und man nach dem bisher Gesagten kaum erwarten kann, daß eine „Umkehrung“ des von uns eingehend erörterten Theorems plötzlich das Wunder leisten sollte, das von jenem vergeblich erhofft wird.

Es wird für unsere Zwecke vollständig genügen, wenn ich den wesentlichen Inhalt des Bayesschen Theorems, so wie er sich nach Annahme der Häufigkeitsdefinition darstellt, erkläre. Das, was ich dabei fortlasse, ist nur der explizite Nachweis dafür, daß man von der klassischen Definition aus eben auch hier nur einen Satz von formal-arithmetischer Bedeutung, niemals aber eine Aussage über den wirklichen Erscheinungsablauf gewinnen kann.

Das Bayessche Problem.

Man gelangt am leichtesten zum Verständnis des Bayesschen Problems, wenn man von einer Kombination der Würfel- und der Lotterieaufgabe ausgeht. Wir denken uns eine Urne mit sehr vielen, kleinen Würfeln oder würfelähnlichen Körpern gefüllt, die wir kurz „Steine“ nennen wollen. Jeder solche Körper habe sechs ebene, mit den Ziffern 1 bis 6 beschriebene Seitenflächen, auf die er — aus einem Becher ausgespielt — auffallen kann. Jedesmal, wenn einer dieser Steine aus der Urne herausgeholt, in einen Becher getan und dann auf den Tisch geworfen wird, erscheint eine der sechs Zahlen 1 bis 6 als das Ergebnis des Wurfes. Dabei sollen die einzelnen Steine materiell recht verschieden beschaffen sein, so daß einige „richtige Würfel“ sich unter ihnen befinden, für die die Wahrscheinlichkeit, die Zahl 6 zu zeigen, gerade $^1/_6$ ist, während die übrigen je eine bestimmte, mehr oder weniger von $^1/_6$ abweichende Wahrscheinlichkeit des Ergebnisses „6“ aufweisen. Diese Wahrscheinlichkeit, die natürlich für jeden der Körper durch eine Versuchsreihe bestimmt werden kann, wollen wir kurz mit dem Buchstaben x bezeichnen, so daß jeder in der Urne liegende Stein durch seinen x-Wert charakterisiert wird.

Das Kollektiv, das wir nun zu betrachten haben, hat zum Element den folgenden Versuchsvorgang: Es wird aus der Urne nach gehöriger Durchmischung des Inhaltes ein Stein gezogen, in einen Becher getan und dann wird dieser Stein eine bestimmte Anzahl, sagen wir n-mal, ausgespielt; als Ergebnis dieses ganzen Versuches, als Beobachtungsresultat oder als Merkmal des jetzt definierten Kollektivs wird einerseits der x-Wert des gezogenen Steines, anderseits der Quotient angesehen, den man erhält, wenn man die Anzahl der beobachteten Sechser-Würfe (n_1) durch die Gesamtzahl n der Wiederholungen dividiert. Wir haben demnach ein zweidimensionales Kollektiv vor uns, dessen Merkmal ein

Wertepaar ist, nämlich die Zusammenfassung der Werte x und n_1/n.

Die Verteilung innerhalb dieses Kollektivs, also die Wahrscheinlichkeit dafür, einen vorgegebenen Wert von x und zugleich einen solchen von n_1/n zu bekommen, rechnet sich nach der Verbindungsregel als ein Produkt zweier Faktoren. Der erste Faktor ist die Wahrscheinlichkeit $v(x)$, die dafür besteht, beim Ziehen aus der Urne einen Stein zu ziehen, dem die Wertziffer x zukommt, der zweite ist die Wahrscheinlichkeit dafür, daß ein Ereignis, das die Einzelwahrscheinlichkeit x besitzt, in n Wiederholungen n_1-mal eintritt. Wie der zweite Faktor aus seinen Bestimmungsstücken x, n und n_1 zu rechnen ist, das lehrt die elementare Wahrscheinlichkeitsrechnung. Wir brauchen uns hier um den formelmäßigen Ausdruck, der übrigens schon NEWTON bekannt war, nicht weiter zu kümmern, es genügt für uns zu wissen, daß es eine solche Funktion — wir bezeichnen sie mit $w(x, n, n_1)$ — gibt. Nach der Regel, die wir für den Fall der Verbindung zweier Kollektivs kennengelernt haben, hat die Wahrscheinlichkeit dafür, einen Stein mit dem Kennzeichen x aus der Urne zu ziehen und dann bei n-maligem Ausspielen genau n_1 Sechser zu erzielen, die Größe des Produktes $v(x) \cdot w(x, n, n_1)$. Wenn wir die Wiederholungszahl n als fest gegeben ansehen, können wir diese Größe als eine Funktion von x und von n_1/n betrachten und für sie etwa $f(x, n_1/n)$ schreiben.

Zu der Fragestellung, die dem BAYESschen Problem entspricht, gelangen wir jetzt, indem wir auf das eben betrachtete, durch eine Verbindung entstandene Kollektiv noch die Operation der Teilung anwenden. Erinnern wir uns daran, was „Teilung“ bedeutete; in dem einfachsten Würfelbeispiel war es die Frage: Wie groß ist die Wahrscheinlichkeit des Wurfes „2“, wenn man schon weiß, daß es sich um den Wurf einer geraden Zahl handelt? Es wurden da aus der Gesamtheit aller Würfe diejenigen „abgeteilt“, deren Merkmal geradzahlig ist. Bei der jetzt in Rede stehenden Aufgabe soll folgende Teilung durchgeführt werden: Wenn man schon weiß, daß in den n Wiederholungen des Ausspielens n_1-mal die Sechs erschienen ist, wie groß ist dann die Wahrscheinlichkeit dafür, daß der aus der Urne gezogene Stein (mit dem die Würfe ausgeführt wurden) das Kennzeichen x besitzt, also ein solcher ist, für den die Wahrscheinlichkeit der Sechs gerade x beträgt? Abgeteilt und zurückbehalten werden somit diejenigen Elemente, deren zweites Merkmal n_1/n einen bestimmten Wert, sagen wir $n_1/n = a$, besitzt. Nach der Teilungsregel, die wir kennengelernt haben, muß man, um die gesuchte Endwahrscheinlichkeit zu erhalten, die (früher als Produkt zweier Faktoren gefundene) Anfangswahrscheinlichkeit $f(x, a)$ durch die Summe aller Wahrscheinlichkeiten, die den zurückbehaltenen Merkmalen entsprechen, dividieren. Zu summieren sind hierbei alle $f(x, a)$ für sämtliche x-Werte bei gegebenem a. Das Ergebnis der Summierung ist eine Funktion $F(a)$ von a allein und das schließliche Ergebnis der Division $f(x, a) : F(a)$ ist eine Funktion der beiden Veränderlichen x und a, die in der üblichen Ausdrucksweise die „a-posteriori-Wahrscheinlichkeit" der Größe x bei gegebenem a zum Ausdruck bringt.

Die Wahrscheinlichkeit der „Ursachen".

Ich habe hier, vielleicht schon zu ausführlich, den ganzen Vorgang der Kollektivableitung für die Bayessche Fragestellung beschrieben und es ist für unsere Zwecke gewiß nicht notwendig, ihn in allen Einzelheiten genau zu verfolgen. Es kam mir nur darauf an, möglichst deutlich zu zeigen, daß es sich da um eine durchaus normale Aufgabe der Wahrscheinlichkeitsrechnung handelt, die sich ohne weiteres in den Rahmen fügt, den ich an früherer Stelle ganz allgemein für das Gesamtgebiet der Wahrscheinlichkeitsrechnung abgesteckt habe: Aus gegebenen oder als gegeben angenommenen Kollektivs werden neue durch Zusammensetzung und Kombination der vier Grundoperationen abgeleitet.

Wie ich schon bei Besprechung der Teilung im zweiten Vortrage erwähnt habe, muß ich aus bestimmten Gründen besonderen Wert darauf legen, diese „Normalität" des Bayesschen Problems zu betonen. Denn fast alle älteren Lehrbücher scheinen sich darüber einig zu sein, daß in der Fragestellung von Bayes eine ganz besondere Art von Wahrscheinlichkeit gemeint sei, eine „Wahrscheinlichkeit der Ursachen oder Hypothesen", die grundsätzlich verschieden wäre von der „Wahrscheinlichkeit der Ereignisse", mit der man es sonst zu tun habe. Es ist allerdings nicht schwer, das Wort „Ursache" in die Formulierung der Aufgabe hineinzubringen. Man sagt eben, das Erscheinen von n_1

Sechsern in n Versuchen habe zur „Ursache“ das Bestehen der Wahrscheinlichkeit x für ein Sechserergebnis bei dem verwendeten Würfel. Da nach der Wahrscheinlichkeit verschiedener x-Werte gefragt wird, kann man dann sagen, es werde die Wahrscheinlichkeit verschiedener Ursachen untersucht. Aber das ist nur ein Leerlauf von Worten. Gerechnet werden kann nur die Wahrscheinlichkeit innerhalb des wohl abgegrenzten Kollektivs, dessen Elemente die früher genau beschriebenen Einzelvorgänge sind: Ziehen eines Steines aus der Urne und darauffolgendes n-maliges Ausspielen des gezogenen Steines aus einem Becher unter Abzählung der dabei auftretenden Sechserergebnisse. Diese Wahrscheinlichkeit ist in genau der gleichen Weise definiert wie die in jedem früheren Fall als der Grenzwert einer relativen Häufigkeit und sie ist die „Wahrscheinlichkeit für das Auftreten eines bestimmten Merkmales innerhalb eines bestimmten Kollektivs“. Keinerlei Grund besteht, jetzt von einer neuartigen „Wahrscheinlichkeit der Ursachen oder Hypothesen“ zu sprechen.

Das BAYESsche Theorem.

Nun komme ich erst dazu, auseinanderzusetzen, was man unter dem BAYESschen Theorem versteht. Es ist das die Feststellung einer ganz bestimmten Eigenschaft jener Verteilungsfunktion $f(x, a) : F(a)$, auf die unsere frühere Überlegung geführt hat. Natürlich übergehe ich die mathematische Ableitung, die uns diese Eigenschaft kennen lehrt und teile nur das Ergebnis mit, genau so, wie ich es beim BERNOULLI-POISSONschen Theorem gemacht habe, mit dem das jetzt zu besprechende gewisse Ähnlichkeit besitzt. Es zeigt sich, um es zunächst nur ungenau auszusprechen, daß, wenn die Zahl n sehr groß gewählt wird, die Wahrscheinlichkeiten aller x-Werte, die erheblich von a abweichen, sehr klein ausfallen und daß die x-Werte, die nahe a gelegen sind, zusammen fast die Wahrscheinlichkeit 1 für sich besitzen. Mit anderen Worten: Es ergibt sich als sehr wahrscheinlich, daß der Stein, mit dem man unter n Würfen $n_1 = an$ Sechser geworfen hat, ein solcher ist, für den die Wahrscheinlichkeit x, eine Sechs zu werfen, wenig von a verschieden ist.

In Anlehnung an die Formulierung, die vorhin für das BERNOULLIsche Theorem gegeben wurde, können wir das BAYESsche Theorem wie folgt formulieren:

Hat n-maliges Ausspielen eines Steines an-mal das Ergebnis „6" gezeitigt und bezeichnet ε eine beliebig kleine Zahl, so liegt die Wahrscheinlichkeit dafür, daß der dem Stein zugehörige x-Wert mindestens $a - \varepsilon$ und höchstens $a + \varepsilon$ beträgt, um so näher an Eins, je größer n ist.

Bedenkt man, daß $a = n_1 : n$ die relative Häufigkeit der „6" unter den n Würfen ist, so erkennt man, daß das Theorem wieder einen Zusammenhang zwischen der relativen Häufigkeit innerhalb einer Serie von n Versuchen und der Wahrscheinlichkeit herstellt. Da die Wahrscheinlichkeit der Sechs für einen bestimmten Stein (wofür wir jetzt auch Würfel sagen können) sich als relative Häufigkeit bei unbeschränkt ausgedehnter Versuchsreihe ergibt, kann man den Satz auch so aussprechen:

Hat sich bei n Versuchen mit einem Würfel die Sechs mit der relativen Häufigkeit a gezeigt, so ist, wenn n groß ist, fast mit Sicherheit zu erwarten, daß bei unbegrenzter Fortsetzung der Versuche mit dem gleichen Würfel sich eine nur wenig von a verschiedene relative Häufigkeit ergeben wird. Dabei heißt natürlich „fast mit Sicherheit zu erwarten" nur dies: Wenn man den ganzen Vorgang, Ziehen eines Steines aus der Urne, n-maliges Ausspielen und (praktisch) unbeschränkte Fortführung der Versuche, sehr oft wiederholt, dann wird in der überwiegenden Mehrheit der Fälle das Vorausgesagte eintreten, nämlich annähernde Gleichheit zwischen der nach n Versuchen und der am Ende festgestellten relativen Häufigkeit.

Präzisierung hinsichtlich der „a-priori-Verteilung".

Bevor ich die Bedeutung diese Satzes näher erörtere, muß ich etwas zu seiner mathematischen Präzisierung sagen und die Voraussetzungen, unter denen er gilt, etwas einschränken. Wir haben zunächst angenommen, daß sich in der Urne sehr viele Steine mit verschiedenen x-Werten befinden. Hierbei ist x, die Wahrscheinlichkeit, mit einem Stein die Zahl „6" zu werfen, sinngemäß eine zwischen 0 und 1 gelegene Größe. Wie das Mischungsverhältnis der verschiedenen x in der Urne beschaffen ist oder die Verteilung $v(x)$ aussieht, die die Wahrscheinlichkeit, einen Stein mit bestimmtem x zu ziehen, angibt, das ist in weitem

Maße für das Bestehen des BAYESschen Theorems gleichgültig. Nur eines müssen wir voraussetzen: Die überhaupt vorkommenden x-Werte müssen hinreichend dicht beieinander liegen, damit, wenn wir eine feste Umgebung oder eine „Nähe“ der Zahl a ins Auge fassen, immer mindestens einer dieser x-Werte hineinfällt. Bequemer für eine kurze mathematische Formulierung und auch mit den wirklichen Anwendungsfällen in bester Übereinstimmung ist es, sich vorzustellen, daß die Verteilung der x eine stetige oder, wie wir dies schon nannten, eine geometrische ist. Es bedeutet dann $v(x)$ die Wahrscheinlichkeitsdichte an der Stelle x und es ist ein wesentliches Ergebnis der mathematischen Ableitung, daß der im BAYESschen Theorem zum Ausdruck kommende Sachverhalt unabhängig davon ist, ob $v(x)$ eine Gleichverteilung darstellt oder eine beliebige andere Verteilung. Der mathematisch präzisierte Satz lautet: Wenn in n Versuchen mit einem Würfel die Sechs mit der relativen Häufigkeit a erschienen ist und ε eine beliebig kleine positive Zahl bezeichnet, so geht die Wahrscheinlichkeit dafür, daß der Grenzwert der relativen Häufigkeit der Sechs bei unbeschränkt fortgesetzten Versuchen zwischen $a - \varepsilon$ und $a + \varepsilon$ liegt mit wachsendem n gegen Eins, gleichgültig wie die (sogenannte „a-priori-“) Wahrscheinlichkeit, einen Würfel mit einer bestimmten Wahrscheinlichkeit der Sechs zu erfassen, verteilt ist.

Wie man etwa den Fall, der einer stetigen Verteilung der x-Werte in der Urne entspricht, sich realisiert vorstellen kann, ist leicht einzusehen. Man denke sich aus der Gesamtheit, sagen wir, aller bisher hergestellten Elfenbeinwürfel einer bestimmten Größe einen beliebigen herausgegriffen. Es ist dann anzunehmen, daß für ihn eine Wahrscheinlichkeit x, die Zahl „6“ zu zeigen, besteht. Dieses x wird um so näher an $^1/_6$ liegen, je „richtiger“ der Würfel ist, aber von vornherein ist x nicht auf irgendwelche ausgewählte Zahlenwerte (etwa auf ganze Vielfache von $^1/_{100}$ oder $^1/_{1000}$) beschränkt, sondern kann jeden in den Bereich 0 bis 1 fallenden Wert besitzen. Es ist dann die Vorstellung zulässig — Näheres darüber wird bei Besprechung der Fehlertheorie zu sagen sein—, daß es eine Wahrscheinlichkeitsdichte $v(x)$ gibt, durch die die Wahrscheinlichkeit jedes Teilbereiches von x-Werten bestimmt wird, nämlich die relative Häufigkeit, mit der bei genügend oft wiederholter Würfelauswahl solche Würfel auftreten, deren x in den betreffenden Teilbereich fällt. Damit sind die Voraussetzungen erfüllt, unter denen das BAYESsche Theorem streng gültig ist. Wir wollen nun einen Augenblick noch dabei verweilen, seinen Inhalt näher zu betrachten und gegenüber dem früher besprochenen Gesetz der großen Zahlen abzugrenzen.

Das Verhältnis des BAYESschen zum POISSONschen Theorem.

Ich habe schon erwähnt, daß ich nicht ausführlich darauf eingehen will, zu zeigen, wie auch das BAYESsche Theorem gleich dem BERNOULLI-POISSONschen auf eine Aussage rein arithmetischer Natur, ohne jede Beziehung auf den wirklichen Ablauf irgendwelcher Erscheinungen, zusammenschrumpft, wenn man die darin auftretenden Wahrscheinlichkeitsgrößen nicht als relative Häufigkeiten, sondern als Quotienten günstiger durch gleichmögliche Fälle deutet. Will man von der Gleichmöglichkeitsdefinition ausgehend irgend etwas aussagen, was sich auf die zu erwartenden Tatsachen oder Vorgänge bezieht, so muß man, genau wie dies früher gezeigt wurde, einmal die Hypothese einschmuggeln: Was rechnungsmäßig mit einer nahe 1 gelegenen Wahrscheinlichkeit eintreffen soll, das geschieht „fast immer", d. h. in der überwiegenden Mehrheit der Versuche. Damit hat man aber, zumindest in beschränktem Umfang, die Häufigkeitsdefinition eingeführt und dazu noch die Unzuträglichkeit in Kauf genommen, am Ende der Untersuchung einen anderen Wahrscheinlichkeitsbegriff benutzen zu müssen als zu Anfang.

Wichtiger erscheint es, den Inhalt des BAYESschen Theorems, unter Voraussetzung der von vornherein zugrunde gelegten Häufigkeitsdefinition, gegenüber dem früher betrachteten Gesetz der großen Zahlen und auch gegenüber dem Erfahrungssatz oder Axiom von der Existenz des Grenzwertes der relativen Häufigkeit genau abzugrenzen. Auf den ersten Blick scheint nichts selbstverständlicher, als daß der Satz: Wenn man bei großem n eine relative Häufigkeit a gefunden hat, so bleibt sie fast sicher annähernd unverändert bei unbegrenzter Vergrößerung von n — mit der einfachen Aussage identisch ist, daß die relative Häufigkeit mit wachsendem n sich einem bestimmten

Grenzwert nähere. Allein in den Worten „fast sicher“, die man auch anschaulicher durch „fast immer“ ersetzen könnte, liegt hier das Entscheidende. Hat man an einem aus der Urne gezogenen Stein in einer langen Serie von n Versuchen die relative Häufigkeit $n_1/n = a$ beobachtet, so weiß man zunächst nichts darüber, wie sich dieser Stein bei weiterer Fortsetzung der Versuche verhält. Solange man nur das Vorhandensein der Grenzwerte, aber nichts über die „Regellosigkeit“ voraussetzt, könnte es noch sehr gut sein, daß nach Wahl eines beliebig großen n, bei fast allen Steinen, die dasselbe $n_1/n = a$ ergeben haben, der Grenzwert der relativen Häufigkeit ein ganz anderer wird als der zuerst beobachtete Wert n_1/n. Unser Satz besagt, daß dies nicht der Fall ist, sondern daß in fast allen Fällen ein Grenzwert, der nahe dem zuerst beobachteten $n_1/n = a$ liegt, herauskommt. Es steckt sonach im Bayesschen Theorem eine ganz bestimmte, eben nur durch die mathematische Ableitung zu gewinnende Aussage, die man auch gerne, mit Rücksicht auf die Analogie mit dem Bernoulli-Poissonschen Theorem als zweites Gesetz der großen Zahlen bezeichnet. Ich will — zum Abschluß dieser Betrachtungen, die uns wohl etwas weiter, als ich es im allgemeinen gewünscht hätte, in das Gebiet abstrakter Schlüsse geführt haben — den Inhalt der beiden Gesetze und zugleich das Axiom von der Existenz der Grenzwerte noch einmal in konkreter Anwendung auf die Würfelaufgabe einander gegenüberstellen.

Die drei verschiedenen Aussagen.

Es sei in n Versuchen mit einem Würfel n_1-mal die Sechs gefallen. Dann besagt

1. das erste Axiom bzw. unsere Definition der Wahrscheinlichkeit, daß bei unbegrenzter Vergrößerung der Versuchszahl der Quotient n_1/n sich einem festen Grenzwert nähert, der eben die Wahrscheinlichkeit eines „Sechser“-Wurfes mit dem betreffenden Würfel heißt;

2. das erste Gesetz der großen Zahlen oder Bernoulli-Poissonsche Theorem, daß, wenn wir eine Versuchsreihe vom festen Umfang n mit demselben Würfel unbeschränkt oft wiederholen, falls nur n nicht zu klein ist, bei fast allen Versuchsreihen annähernd der gleiche Quotient n_1/n erscheinen wird;

3. das zweite Gesetz der großen Zahlen oder BAYESsche Theorem, daß unter sehr vielen verschiedenen Würfeln, von denen jeder einmal in n Versuchen n_1 Sechser ergeben hat, falls nur n nicht zu klein ist, fast alle Würfel die Eigenschaft besitzen, daß für sie der Grenzwert der relativen Häufigkeit der Sechs bei unbeschränkt fortgesetztem Würfeln nahe dem zuerst beobachteten Wert von n_1/n liegt.

Diese Formulierungen grenzen die drei Aussagen exakt gegeneinander ab. Man muß nur noch hinzufügen, daß die erste eine nicht beweisbare, d. h. nicht auf einfachere zurückführbare Erfahrungstatsache darstellt, während die beiden anderen aus der Voraussetzung der ersten unter Hinzunahme der „Regellosigkeit" ableitbar sind. Anderseits haben die beiden Gesetze der großen Zahlen, also die Aussagen 2. und 3., keinen vernünftigen Sinn, wenn man nicht die Existenz der Grenzwerte der relativen Häufigkeiten von vornherein voraussetzt.

Erweiterung der Gesetze der großen Zahlen.

Die beiden hier behandelten Theoreme, das von BERNOULLI und POISSON und das nach BAYES benannte, das eigentlich erst von LAPLACE ausgesprochen wurde (BAYES hatte nur das Problem gestellt und die prinzipielle Lösung angegeben), bilden die klassischen Gesetze der großen Zahlen. In neuerer Zeit sind mehrfach Ergänzungen und Erweiterungen gefunden worden. Sie bedürfen keiner ausführlichen Besprechung im Rahmen dieser Vorträge, da durch sie die grundsätzlichen Fragen der Auffassung des Wahrscheinlichkeitsbegriffes nicht berührt werden. Gleichwohl kann ich nicht ganz über diese neueren Forschungsergebnisse hinweggehen. Sie sind zum Teil an sich wichtig für die Anwendungen der Theorie, teilweise knüpft sich an sie eine Polemik, die uns hier angeht. Denn wenn auch der Tatbestand an sich klar ist, versuchen doch manche Gegner der Häufigkeitstheorie für jedes neugefundene mathematische Theorem dieser Art eine Auslegung zu finden, die irgendeinen Widerspruch zu der Häufigkeitsauffassung bildet. Dies gilt insbesondere von dem Satz, den man heute das „starke" oder das verschärfte Gesetz der großen Zahlen zu nennen pflegt.

Fragen wir uns zunächst allgemein, welche Art von Aussagen man zweckmäßigerweise unter dem Sammelnamen „Gesetze der großen Zahlen“ zusammenfassen kann. Das ist natürlich mehr oder weniger willkürlich. Aber es entspricht ungefähr dem Sinn dessen, was man seit jeher unter diesem Ausdruck verstanden hat, wenn wir folgende Abgrenzung vornehmen. Zunächst soll in die Aufgabe, die untersucht wird, eine Anzahl n von Einzelbeobachtungen oder Einzelversuchen eingehen, also in unserer Ausdrucksweise: es soll eine Verbindung von n Kollektivs gebildet werden. Dann soll nach einer Wahrscheinlichkeit W gefragt sein, die durch diese n Versuche bestimmt wird, d. h. das betrachtete Endkollektiv soll mittels der genannten Verbindung definiert sein; es wird also W von der Zahl n abhängen. Schließlich ist es das Charakteristische an dem Resultat, daß die gesuchte Wahrscheinlichkeit W sich unbeschränkt dem Wert 1 nähert, wenn n unbeschränkt ins Unendliche wächst. Eine solche Aussage kann also immer mit den Worten beginnen: „Es ist bei großem n fast sicher, daß ...“ Im BERNOULLI-POISSONschen Falle war es „fast sicher, daß“ die Häufigkeit des Ereignisses innerhalb eines (aus n Einzelversuchen bestehenden) Gesamtversuches nahe bei der Wahrscheinlichkeit des Ereignisses liegt. Im BAYESschen Falle hieß es „fast sicher, daß“ die unbekannte Grundwahrscheinlichkeit, unter bestimmten Voraussetzungen, nahe der in einer Serie gefundenen Häufigkeit liegt. — Natürlich wird in dieser abgekürzten Ausdrucksweise der Inhalt der Theoreme nicht vollständig wiedergegeben; wir haben aber vorhin die genauen Formulierungen kennengelernt.

Will man das mathematische Faktum kennzeichnen, das die verschiedenen Gesetze der großen Zahlen zum Ausdruck bringen, so wird man angemessenerweise von einer „Verdichtung“ der Verteilung zu sprechen haben. Während im allgemeinen die Wahrscheinlichkeit sich über verschiedene Merkmalwerte mehr oder weniger gleichmäßig verteilt, liegt hier der Fall vor, daß nahezu die ganze Wahrscheinlichkeit 1 sich auf einen Punkt oder seine nächste Umgebung konzentriert, und zwar um so stärker konzentriert, je größer der Parameterwert n ist, von dem die Verteilung abhängt. In weniger glücklicher Weise hat man, um den hier vorliegenden Verhältnissen gerecht zu werden, in Anlehnung an die Ausdrucksweise der Analysis, einen Begriff von „Konvergenz im Sinne der Wahrscheinlichkeitsrechnung“ zu schaffen versucht. Diese Begriffsbildung scheint mir der Aufklärung der Fragen, um die es sich handelt, wenig dienlich zu sein. Sie fällt, nebenbei bemerkt, in das Gebiet derjenigen Gedankengänge, die gerne aus der Wahrscheinlichkeitstheorie einen Teil der Mengenlehre und Funktionentheorie machen möchten und die ich am Schlusse des vorigen Vortrages gekennzeichnet habe.

Das verschärfte Gesetz der großen Zahlen.

Ein Satz, der etwas weiter geht, etwas mehr aussagt als das Bernoulli-Poissonsche Theorem ist in neuerer Zeit von verschiedenen Mathematikern, in erster Linie von Cantelli, von Pólya und einigen anderen entwickelt worden. Wir betrachten wieder wie im Bernoullischen Fall eine Gruppe von n Wiederholungen einer einfachen Alternative (Null oder Eins) als neues Element eines Kollektivs, stellen uns also vor, daß das Experiment, das aus dieser n-fachen Beobachtung besteht, unendlich oft wiederholt wird. Als das entscheidende Ergebnis eines solchen Versuches wurde früher betrachtet die Anzahl n_1 der Einser, die sich bis zum Schluß des Versuches ergeben. Bezeichnen wir kurz den Quotienten $n_1/n = x$ als die Häufigkeit der Eins, so besagt das Bernoulli-Poissonsche Gesetz: Es ist bei großem n fast sicher, daß die Häufigkeit x der Eins nahe bei dem Wahrscheinlichkeitswert p liegt.

Jetzt wollen wir die n Bestandteile des Versuches, der das Element unseres Kollektivs bildet, etwas genauer betrachten. Es sei m eine Zahl kleiner als n, so daß die Differenz $n—m$ einen positiven Wert k besitzt. Dann gibt es unter den ersten $(m + 1)$ Teilergebnissen eine gewisse Anzahl von Einsern (zwischen 0 und $m + 1$) und demgemäß eine gewisse Häufigkeit x_1 der Eins. Dabei meinen wir unter x_1 den Quotienten aus der Anzahl der bis zum $(m + 1)$sten Teilversuch erschienenen Einser durch $(m + 1)$. Ebenso gibt es eine Häufigkeit x_2 der Eins unter den ersten $(m + 2)$ Teilergebnissen, eine weitere x_3 unter den ersten $(m + 3)$ und so weiter bis zu x_k. Ist z. B. $m = 10$ und $n = 15$, also $k = 5$, so fangen wir nach dem elften Einzelversuch zu zählen an. Sind bisher 6 Einser erschienen, so ist $x_1 = {}^6/_{11}$. Ist das zwölfte Ergebnis wieder eine Eins, so ist $x_2 = {}^7/_{12}$, im entgegengesetzten Fall wäre $x_2 = {}^6/_{12}$. Die letzte Zahl x_5 ist die durch 15 dividierte Gesamtzahl der beobachteten Einser. Allgemein stimmt das letzte x_k mit der Größe überein, die vorhin einfach als x bezeichnet worden war.

Als Merkmal eines Elementes betrachten wir die Zahlen x_1 bis x_k, die alle positiv und höchstens gleich 1 sind. Wir wollen aber jetzt noch eine „Mischung" vornehmen, und zwar in folgender Weise. Es sei p die Wahrscheinlichkeit der Eins in der ursprünglichen einfachen Alternative (z. B. $p = 1/2$ beim Spiel mit einer gewöhnlichen Münze) und ε eine kleine positive Größe. Unter den k Zahlen x_1 bis x_k kann es solche geben, die in den Bereich $p-\varepsilon$ bis $p+\varepsilon$ fallen, und andere. Es soll als „Ereignis" bezeichnet werden, wenn mindestens eine der k Zahlen x_1 bis x_k außerhalb des genannten Bereiches $p-\varepsilon$ bis $p+\varepsilon$ liegt. Nur wenn sämtliche k Häufigkeiten x_1 bis x_k, die das Ergebnis des (aus n Einzelbeobachtungen zusammengesetzten) Versuches bilden, innerhalb des Bereiches liegen, sagen wir, das Ereignis sei nicht eingetreten.

Jetzt haben wir wieder eine klar definierte einfache Alternative vor uns. Die Wahrscheinlichkeit W für den Eintritt des Ereignisses in dem jetzt betrachteten Kollektiv kann man auch so bezeichnen: W ist die Wahrscheinlichkeit dafür, daß in einer Gruppe von n Einzelbeobachtungen der ursprünglichen Alternative die Häufigkeit der Eins nach dem m-ten Einzelversuch (bis zum n-ten natürlich) mindestens einmal um mehr als ε vom Festwert p abweicht. Die Berechnung dieser Wahrscheinlichkeit ist mittels wiederholter Anwendung der im zweiten Vortrag auseinandergesetzten Regeln möglich. Sie muß naturgemäß von den vier Größen n, m, p und ε abhängen. Hier kommt es aber keineswegs auf ihre genaue Größe an. Vielmehr interessiert uns eine bestimmte Eigenschaft dieser Wahrscheinlichkeit, die sehr merkwürdig ist: Ihren Wert ergibt die Rechnung unter allen Umständen (wie groß auch n und p sein mögen) kleiner als den reziproken Wert des Produktes von m und ε^2, also

$$W \text{ kleiner als } \frac{1}{m\,\varepsilon^2}.$$

Sehen wir einmal zu, was dies bedeutet! Wählt man für ε irgendeine kleine Zahl, z. B. ein Hundertstel oder ein Millionstel, auf jeden Fall wird, wenn m wächst, der Ausdruck $1/m\,\varepsilon^2$ kleiner und kleiner und nähert sich der Null, sobald m ins Unendliche geht. (Natürlich muß auch die Zahl n, die ja größer als m ist, zugleich ins Unendliche wachsen.) Wenn die Wahrscheinlichkeit des Ereignisses sich der Null nähert, muß die Wahrscheinlichkeit seines Nichteintretens gegen 1 gehen. Man kann dies auch so ausdrücken: Es ist fast sicher, daß die Häufigkeit der Eins zwischen dem m-ten und n-ten Einzelversuch sehr nahe dem Wert p bleibt, also kein einziges Mal aus dem Bereich $p-\varepsilon$ bis $p+\varepsilon$ herausfällt, wenn nur m (und damit n) eine sehr große Zahl ist. Der Unterschied gegenüber dem Bernoullischen Theorem ist klar: Dort wurde es als fast sicher bezeichnet, daß die Häufigkeit am Schlusse der Gruppe von n Versuchen beliebig nahe bei p liegt, jetzt heißt es, daß sie vom m-ten Versuch an dauernd bei p bleibt, wenn man nur m groß genug wählt.

Das Auffallende an diesem Ergebnis und das, was zu den erwähnten und unklaren Deutungen Anlaß gegeben hat, ist, daß die obere Grenze $1/m\varepsilon^2$, die wir für die Wahrscheinlichkeit des Ereignisses gefunden haben, nicht von der Zahl n, sondern nur von m abhängt. Wenn man für m eine feste Zahl, z. B. 1000 nimmt und für n hintereinander 2000, 3000, 4000 usf. setzt, so muß die Wahrscheinlichkeit für den Eintritt des von uns betrachteten Ereignisses immer größer werden, denn es sind ja zunehmend mehr Möglichkeiten für die Realisierung einer Abweichung, die ε überschreitet, vorhanden. Das angegebene Rechnungsresultat zeigt aber, daß trotz der vermehrten Möglichkeiten die Wahrscheinlichkeit nicht über eine feste Grenze, eben die Größe $1/m\varepsilon^2$ hinaus wächst. Ist z. B. $\varepsilon = 0{,}1$ und $m = 10000$, also $\frac{1}{m\varepsilon^2} = 0{,}01$, so können wir sagen: Es besteht mehr als 99% Wahrscheinlichkeit dafür, daß die Häufigkeit eines „Kopf"-Ergebnisses beim Wurf mit einer Münze nach dem zehntausendsten Versuch nicht mehr aus dem Bereich 0,4 bis 0,6 (d. i. für unsern Fall $p-\varepsilon$ und $p+\varepsilon$) heraustritt, mag die Gesamtgruppe der betrachteten Versuche noch so groß sein, eine Million oder zehn Millionen umfassen usf.

Man sieht an dieser Formulierung und allem, was zu ihrer Ableitung gesagt wurde, deutlich, daß die Fragestellung des „starken" Gesetzes der großen Zahlen sowie die Lösung dieser Aufgabe sich mühelos in den Rahmen der Häufigkeitstheorie einordnet, die auf den Begriff des Kollektivs, der Konstruktion von Kollektivs durch gewisse Operationen usw. aufgebaut ist. Nur dies zu zeigen, war meine Absicht; auf die Irrtümer und Fehldeutungen näher einzugehen, die sich an das erste Auftreten des Theorems in der Literatur geknüpft haben, liegt kein Grund vor.

Die statistischen Funktionen.

Ich will lieber zum Schlusse noch eine andere Erweiterung der Gesetze der großen Zahlen kurz besprechen, die von mehr praktischem Interesse ist und zu den Aufgaben hinüberleitet, die uns in dem nächsten Vortrag beschäftigen werden, zu den Fragen der Statistik. Zu der allgemeineren Auffassung gelangen wir, indem wir zunächst statt des einfachsten Falles des Kollektivs, der bis jetzt betrachteten Alternative „Null oder Eins", ein beliebiges Kollektiv mit mehrwertigem Material ins Auge fassen. Denken wir z. B. an die Roulette, bei der das Ergebnis eines Einzelversuches eine von 37 Zahlen (0 bis 36) ist.

Wenn eine Anzahl n, sagen wir etwa 100, Einzelversuche mit der Roulette zu einem Element eines Kollektivs vereinigt werden, so besitzt dieses Element als Merkmal zunächst eine Gruppe von 100 Zeichen, unter denen sich die Zahlen 0 bis 36 in irgendeiner Anordnung wiederholen. Man kann aber statt dieser 100 Größen eine geringere Anzahl als das kennzeichnende Ergebnis des Gruppenversuches ansehen, indem man von der Anordnung, in der die Teilresultate aufeinanderfolgten, absieht. Man sagt, daß man eine „Statistik" der Teilresultate betreibt, wenn man nur feststellt, wie oft (aber nicht an welchen Stellen des Gesamtversuches) die Null, wie oft die Eins, die Zwei usw., wie oft endlich die Zahl sechsunddreißig erschienen ist. Die Summe dieser einzelnen Anzahlen ist natürlich $n = 100$. Dividieren wir aber jede einzelne von ihnen durch die Gesamtzahl n, so bekommen wir eine Reihe von echten Brüchen, deren Summe gleich 1 ist. Die einzelnen Brüche x_1, x_2, x_3 usw. bis x_{36} geben die relativen Häufigkeiten, mit denen jedes der möglichen Resultate 0 bis 36 innerhalb des Gesamtversuches erschienen ist. Wir sagen, durch die Angabe der Größen x_1, x_2, x_3 usf., deren Summe 1 ist, werde die Aufteilung der Einzelresultate im Gruppenversuch beschrieben.

Der Übergang von dem ursprünglichen, vollständigen Ergebnis zur Angabe der Aufteilung bedeutet eine „Mischung" im Sinne der Definitionen, die im ersten Vortrag gegeben wurden. Denn ein und dieselbe Aufteilung umfaßt im allgemeinen sehr viele verschiedene Anordnungen. Schon wenn nur zwei verschiedene Einzelresultate, 0 und 1, möglich sind, ist z. B. bei $n = 10$ die Aufteilung „0,30 Nullen und 0,70 Einser" auf 120 verschiedene Arten realisierbar, d. h. es gibt 120 mögliche Anordnungen von zehn Zeichen, wenn drei von ihnen Nullen und sieben Einser sind. Kennt man die Wahrscheinlichkeit jedes überhaupt möglichen Gesamtresultates, das aus n Zeichen der gegebenen Art besteht, so findet man die Wahrscheinlichkeit einer jeden Aufteilung nach den Gesetzen der Mischung mit Hilfe einer bestimmten Summierung.

Es ist zu bemerken, daß zwei Aufteilungen auch übereineinstimmen können, die zu verschiedenen Versuchsumfängen n gehören. Hat man in 10 Beobachtungen 3 Nullen 2 Einser und 5 Zweier gezählt, so ist die Aufteilung dieselbe wie im Falle $n = 50$ bei 15 Nullen, 10 Einsern und 25 Zweiern. Wenn wir sagen, daß wir die Aufteilung eines statistischen Materials kennen, so gehört dazu nicht unbedingt die Kenntnis der Versuchszahl n.

Aber auch die Aufteilung selbst ist nicht immer das, was uns an dem Ergebnis von n Einzelversuchen interessiert. In der Regel leitet man aus der Aufteilung erst eine bestimmte Größe ab, die von ihr abhängt, und betrachtet diese als das charakteristische Merkmal. Mit anderen Worten: es wird nochmals eine Mischung vorgenommen, die alle Aufteilungen zusammenwirft, die nur in irgendeiner Eigenschaft übereinstimmen. Eine solche Größe, die von n Einzelergebnissen bestimmt wird, dabei aber nur von deren Aufteilung (nicht auch von der Anordnung und der Gesamtzahl) abhängt, nennt man eine statistische Funktion.

Der einfachste Fall einer statistischen Funktion ist der Durchschnitt der Beobachtungen. Sind in 50 Einzelversuchen 15 Nullen, 10 Einser und 25 Zweier gezählt worden, so bildet man den Durchschnitt dieser Ergebnisse, indem man die Summe der Produkte $15 \,.\, 0 + 10 \,.\, 1 + 25 \,.\, 2 = 10 + 50 = 60$ durch die Anzahl 50 dividiert, also $60 : 50 = 1{,}20$. Den Durchschnitt erhält man offenbar auch so, daß man jedes der möglichen Ergebnisse (hier 0, 1 und 2) mit der Häufigkeit seines Auftretens (hier 0,30, 0,20 und 0,50) multipliziert und dann diese Produkte addiert: $0{,}30 \,.\, 0 + 0{,}20 \,.\, 1 + 0{,}50 \,.\, 2 = 0{,}20 + 1{,}00 = 1{,}20$. Man sieht, daß in der Tat der Durchschnitt nur von den Zahlen x_1, x_2, x_3, die die Aufteilung bilden, abhängt, aber keineswegs davon, in welcher

Reihenfolge die Resultate aufgetreten sind. Er bleibt auch unverändert, wenn man die doppelte Anzahl von Beobachtungen macht und dabei jedes der früheren Einzelresultate zweimal auftritt.

Der Durchschnitt von n Größen ist also eine statistische Funktion. Im nächsten Vortrag werden wir verschiedene andere statistische Funktionen kennen lernen. Jetzt will ich nur etwas über die Gesetze der großen Zahlen sagen, denen die statistischen Funktionen allgemein unterliegen.

Das erste Gesetz der großen Zahlen für statistische Funktionen.

Besteht eine Gruppe von n Zahlen lediglich aus Nullen und Einsern, so ist ihr Durchschnitt offenbar gleich der durch n dividierten Anzahl der Einser, die sich unter ihnen finden. Daher kann man das BERNOULLIsche Theorem auch so aussprechen: Faßt man n Beobachtungen an einer Alternative zu einem Element zusammen, so ist es bei großem n fast sicher, daß der Durchschnitt der n Beobachtungen beliebig nahe an einer im voraus bekannten Zahl p liegt.

Die Erweiterung, die POISSON dem BERNOULLIschen Satz gegeben hat, geht dahin: Die n Beobachtungen müssen nicht derselben Alternative entstammen. Man kann den Durchschnitt auch aus n verschiedenen Alternativversuchen bilden (d. h. die Zahl der eingetretenen Ereignisse durch n dividieren), für das Resultat gilt immer noch die „Verdichtung“ der Wahrscheinlichkeit, die BERNOULLI für den Fall von n gleichen Alternativen festgestellt hat.

Die nächste Verallgemeinerung wird im wesentlichen TSCHEBYSCHEFF zugeschrieben. Sie besagt: Die Beobachtungen müssen nicht einfachen Alternativen entstammen, es können Kollektivs mit beliebigen mehrwertigen Verteilungen vorausgesetzt werden. Faßt man z. B. n Partien eines Roulettespieles (oder verschiedener ähnlicher Spiele) zusammen, rechnet den Durchschnitt der n Zahlen, die als Ergebnisse der einzelnen Partien erschienen sind, so ist es, wenn nur n groß genug ist, fast sicher, daß dieser Durchschnitt nahe bei einem bestimmten, im voraus bekannten Wert liegt, der nur von den Verteilungen der n Ausgangskollektivs abhängt.

Nunmehr aber können wir, auf Grund neuer Forschungen, den Satz noch mehr erweitern und ihm dadurch einen für die praktische Statistik sehr bedeutenden Wert verleihen: Die Erscheinung der „Verdichtung", die das BERNOULLIsche Theorem zum erstenmal präzisiert hat, besteht nicht nur für den Durchschnitt, sondern im wesentlichen für jede statistische Funktion von n Beobachtungen, wenn nur n sehr groß ist. Mit anderen Worten: Wird aus den Beobachtungen an n gleichen oder verschiedenen Kollektivs eine Größe abgeleitet, die nur von der Aufteilung (aber nicht von der Anordnung und Gesamtzahl) der Beobachtungen abhängt, so ist es bei großem n fast sicher, daß diese Größe in beliebiger Nähe eines festen Wertes liegt, der sich aus den Grundverteilungen der n Kollektivs im voraus berechnen läßt. Hier sind natürlich die Worte „fast sicher" usw. eine Abkürzung, die auf Grund unserer Wahrscheinlichkeitsdefinition aufzulösen ist, etwa so: Wenn man die Gruppe von n Versuchen unendlich oft wiederholt und ε eine beliebig kleine Zahl bedeutet, so wird bei großem n in der erdrückenden Mehrheit der Fälle die aus den Beobachtungen berechnete Größe sich um weniger als ε von dem „theoretischen" Festwert unterscheiden; je größer n ist, um so stärker ist, bei gleichem ε, die Majorisierung.

Das zweite Gesetz der großen Zahlen für statistische Funktionen.

Auch das früher besprochene BAYESsche Theorem läßt sich zu einem allgemeinen zweiten Gesetz der großen Zahlen für beliebige statistische Funktionen erweitern. Wir haben uns jetzt vorzustellen, daß n Beobachtungen an einem Kollektiv gemacht werden, dessen Verteilung aber unbekannt ist, z. B. n Würfe mit einem würfelähnlichen Körper, der irgendwie zufallsartig aus einer Menge derartiger Körper herausgegriffen wurde. Aus den n Beobachtungen — im Beispiel ist jede von ihnen eine der Zahlen 1 bis 6 — leiten wir eine Größe L ab, die nur von der Aufteilung, nicht von der Anordnung oder der Gesamtzahl der Beobachtungen abhängt, kurz eine statistische Funktion. Nach dem früher Gesagten gibt es einen „theoretischen" Wert von L, der L_0 heißen möge und der durch die Verteilung im Ausgangskollektiv, also in unserm Falle durch die sechs Wahrscheinlichkeiten der Würfelseiten bestimmt ist. Jetzt kennen wir L_0 nicht, denn wir nehmen an, daß der Würfelkörper, mit dem wir arbeiten, noch nicht näher untersucht wurde und nichts von ihm bekannt ist, als eben das im Werte von L zum Ausdruck kommende Resultat von n Beobachtungen.

Das zweite Gesetz der großen Zahlen in seiner Erweiterung auf beliebige statistische Funktionen besagt nun, daß bei großem n der unbekannte „theoretische" Wert fast sicher sehr nahe dem Beobachtungswert L liegen wird. Das ursprüngliche Bayessche Theorem drückte diesen Tatbestand für den Fall des Durchschnittes aus n Beobachtungen an einer Alternative aus; der zugehörige „theoretische" Wert des Durchschnittes war dabei die Grundwahrscheinlichkeit p des Ereignisses. Jetzt können wir den allgemeinen Satz wie folgt formulieren: Haben n Beobachtungen an einem sonst unbekannten Kollektiv für eine statistische Funktion der Beobachtungen den Wert L geliefert, so ist es bei großem n fast sicher, daß der theoretische Wert L_0 dieser Funktion, der sich aus der Verteilung innerhalb des Kollektivs ableitet, sehr nahe bei dem Beobachtungswert L liegt. Wie man hier die Ausdrücke „fast sicher" usf. durch Einsetzen unserer Wahrscheinlichkeitsdefinition aufzulösen hat, ist vorhin gesagt worden.

Der neue, eben ausgesprochene Satz gestattet, wie man sieht, einen Rückschluß aus einer genügend großen Versuchsreihe auf die Beschaffenheit des untersuchten Kollektivs. Er ist daher praktisch noch viel wichtiger als das erste Gesetz der großen Zahlen, ja er ist vielleicht das wesentlichste Theorem, auf das man sich in der Statistik stützt. Wir werden z. B. im folgenden Vortrag als eine in vielen Fällen wichtige statistische Funktion L den sogenannten Lexisschen Quotienten kennenlernen. Hat man in einer genügend großen Versuchsreihe für L einen bestimmten Wert, sagen wir 1,1 gefunden, so kann man aus dem zweiten Gesetz der großen Zahlen in seiner erweiterten Form schließen, daß der „theoretische", d. h. das betreffende Material charakterisierende Wert L_0 der Lexisschen Zahl nahe bei 1,1 liegen wird. — Auf die mathematischen Einzelheiten, die den Bereich der zulässigen Funktionen für diesen und den ersten Satz abgrenzen, gehe ich hier nicht ein. Ich komme zum Schluß.

Schlußbemerkung.

In diesem Vortrag habe ich versucht, soweit dies ohne Benutzung mathematischer Hilfsmittel möglich ist, den Komplex von Fragen, die mit dem populären Begriff „Gesetz der großen Zahlen" verknüpft sind, zu erörtern. Mein Ziel war ein doppeltes. Erstens wollte ich Sie mit dem Inhalt der Sätze bekannt machen, der wertvoll, ja unentbehrlich ist für jeden, der sich irgendwie mit Aufgaben der Statistik oder den Anwendungen der Wahrscheinlichkeitsrechnung auf irgendeinem Gebiet befaßt. Zweitens war es für uns besonders wichtig, die Rolle zu untersuchen, die den Gesetzen der großen Zahlen im Rahmen der von mir entwickelten Häufigkeitstheorie der Wahrscheinlichkeitsrechnung zukommt. Wie ich eingangs sagte und wie auch im vorangehenden Vortrag schon bemerkt wurde, knüpfen mancherlei kritische Einwendungen gegen die Häufigkeitstheorie gerade an die Behauptung an, daß sie mit diesen Gesetzen nicht in Einklang zu bringen sei. Ich hoffe aber zweierlei gezeigt zu haben:

1. Geht man von dem Begriff des Kollektivs und der Definition der Wahrscheinlichkeit als Grenzwert einer relativen Häufigkeit aus, so erhalten alle „Gesetze der großen Zahlen" einen klaren und eindeutigen Sinn, frei von jedem Widerspruch. Sie sagen, jedes einzeln, etwas Bestimmtes aus über den Ablauf einer sehr großen (unendlichen) Versuchsreihe, bei der jeder einzelne Versuch eine große Zahl n von Einzelbeobachtungen umfaßt.

2. Wenn man, wie es in der alten Theorie der Fall ist, die Wahrscheinlichkeit nicht auf die relative Häufigkeit in einer Versuchsreihe zurückführt, so kann kein Satz einer solchen Theorie, auch nicht ein „Gesetz der großen Zahlen", irgend etwas über den Ablauf des Geschehens in einer Versuchsfolge aussagen. Wenn gleichwohl solche Folgerungen gezogen werden, so ist es nur dadurch möglich, daß man am Schlusse der Rechnung dem Wort Wahrscheinlichkeit doch eine andere Deutung gibt als zu Anfang, und zwar eine Häufigkeitsdeutung. Ein solches Vorgehen kann natürlich Widersprüche und Unklarheiten mit sich bringen.

Schließlich muß ich noch einen Vorbehalt hinzufügen, der freilich in ganz anderer Richtung liegt. Es ist bei Vermeidung aller Formelzeichen und Ausschaltung aller nicht ganz elementarer mathematischer Begriffe nicht möglich, die Ableitungen und

Formulierungen der Sätze völlig korrekt zu gestalten. Ich hoffe, alles Wesentliche richtig dargestellt zu haben. Aber vom rein mathematischen Standpunkt aus sind mehrfache Einschränkungen, z. B. über den Kreis der zuzulassenden Funktionen usw., und Zusätze über formelle Bedingungen u. ähnl. erforderlich. Wer sich dafür interessiert und die nötigen Vorkenntnisse besitzt, muß sich in der einschlägigen mathematischen Literatur orientieren.

Fünfter Vortrag.

Anwendungen in der Statistik und Fehlertheorie.

In diesem und dem folgenden Vortrag will ich mich mit den beiden wichtigsten Anwendungsgebieten der Wahrscheinlichkeitsrechnung beschäftigen. Von Glücksspielen soll nicht mehr viel die Rede sein. Heute zunächst wollen wir uns der Betrachtung jener mannigfaltigen Erscheinungsreihen zuwenden, denen man allenthalben im praktischen Leben begegnet und deren Untersuchung man gemeiniglich mit dem Worte „Statistik" bezeichnet.

Was ist Statistik?

Nicht der sprachlichen Herkunft nach, aber dem Sinn gemäß, den das Wort im heutigen Sprachgebrauch besitzt, hat man Statistik übersetzt mit „Großzahlforschung" oder „Häufigkeitslehre". Ihr Gegenstand sind umfangreiche Reihen gleichartiger Erscheinungen, die sich irgendwie zahlenmäßig erfassen lassen. Um nur einiges anzuführen: man kann vom statistischen Standpunkt betrachten: Bevölkerungsvorgänge, Geburten, Todesfälle; soziale Erscheinungen, Eheschließungen, Selbstmorde, Einkommensfragen; biologische Probleme, Erblichkeit, Abmessungen bestimmter Organe der Lebewesen; medizinische Fragen, Wirksamkeit von Heilmitteln oder Heilverfahren; Fragen der Technik, Massenfabrikation und Massenbedarf usw.

In diesen und ähnlichen Fällen kann man zunächst ein gewisses Zahlenmaterial sammeln und dann versuchen, daraus praktisch verwertbare Schlüsse zu ziehen. Welchen Teil der Arbeit man ausschließlich oder vorzugsweise mit dem Namen „Statistik" belegen will, ist natürlich in weitem Maße willkürlich.

Wir wollen uns jedenfalls in den Streit der verschiedenen Parteigänger der „allgemeinen" und der „mathematischen", der „formalen" und der „realistischen" Richtung nicht einmengen. Es handelt sich nur darum, festzulegen, welche Fragen uns im folgenden beschäftigen werden. Dazu gehören nicht die Methoden der Planung und Durchführung statistischer Aufnahmen, etwa einer Volkszählung, und auf der anderen Seite auch nicht die Maßnahmen irgendwelcher Art, zu denen man sich auf Grund statistisch erworbener Erkenntnisse veranlaßt sieht. Natürlich ist die Isolierung eines Arbeitsgebietes niemals mit voller Schärfe möglich. Der Arzt, der einen Beinbruch einrichtet, kümmert sich wenig um den Bildungsgang oder die Zukunft des Patienten, aber er darf doch nicht völlig blind sein für etwas wesentliches, was den Kranken in irgend einer Richtung betrifft. Anderseits liegt aller Fortschritt und alle Möglichkeit gesteigerter Leistung nur darin, daß man den Blickpunkt in geeigneter Weise einstellt und von verschiedenen Zentren aus an das Studium eines Erscheinungsgebietes herangeht.

Was uns an den Wiederholungs- und Massenvorgängen interessiert, wird durch die Theorie bestimmt, die ich in den vorangehenden Vorträgen entwickelt habe. Gewiß wird eine Theorie geschaffen, weil man den Problemen, die ein bestimmtes Tatsachengebiet aufwirft, gerecht werden will. Ist aber dann die Theorie da, so zeigt sich, daß sie doch nur die eine oder die andere Seite der Wirklichkeit wiedergibt, und man muß sich damit begnügen, diese Seiten hervorzukehren.

So wird für uns der Begriff des Kollektivs, der aus der Beobachtung der Wirklichkeit abstrahierte und idealisierte Begriff der Massenerscheinung, den Ausgangspunkt aller Betrachtung bilden. Man kann niemals in zwei Worten sagen, was man später in Stunden ausführen wird. Aber andeuten will ich es, in welchem Sinn wir an die Fragen der Statistik herantreten: Wir werden an den einzelnen Beispielen von statistischen Vorgängen, die uns vor Augen kommen, untersuchen, ob und in welchem Sinne sie näherungsweise als Kollektivs aufgefaßt oder auf Kollektivs zurückgeführt werden können. Sobald wir finden, daß sich ein Zusammenhang herstellen läßt, wird es möglich, aus der auf dem Kollektivbegriff aufgebauten Theorie Schlüsse zu ziehen, die die beobachteten Erscheinungsreihen betreffen.

Abgrenzung zwischen Glücks- und Geschicklichkeitsspielen.

Zunächst sei noch eine praktische Frage über Glücksspiele, die zu den Aufgaben überleitet, deren Sinn ich eben andeutete, kurz besprochen. In den meisten Ländern wird die öffentliche Abhaltung von Glücksspielen mehr oder weniger eingeschränkt oder verboten. Dabei tritt das Bedürfnis auf, Glücksspiele gegen andere, ähnliche Spiele abzugrenzen. Früher half man sich da einfach mit der Definition, Glücksspiele seien „Spiele von aleatorischem Charakter", einer Worterklärung schlimmster Art, da sie einen gebräuchlichen Ausdruck durch einen ungebräuchlichen, aber gelehrt klingenden ersetzt. Das moderne Strafrecht sieht meist von jeder Definition ab, setzt also voraus, daß die Abgrenzung nach bekannten Gesichtspunkten möglich ist.

Eine spezielle Streitfrage, die mehrfach die Öffentlichkeit beschäftigt hat, betraf das sogenannte Bajazzospiel, bei dem eine zwischen Stiften herabfallende Kugel in einen von Hand aus gelenkten Behälter aufgefangen werden soll. Darüber, ob dies ein Glücksspiel sei oder nicht, wurde lange keine Einigung erzielt. Wenn auch nicht in diesem Falle, so liegen doch manches Mal die Verhältnisse recht schwierig. Niemand wird bezweifeln, daß die Roulette ein Glücksspiel ist, das Schachspiel keines, aber bei jedem Kartenspiel, in dem doch zumindest die Verteilung der Karten zufallsartig erfolgt, ist, wie man sagt, „Glück und Unglück mit im Spiele", ohne daß man es deswegen schon als Glücksspiel bezeichnet. Offenbar sind da alle Zwischenstufen möglich, je nachdem, wie die zufallsartigen Elemente mit den durch die Spielregeln an Geschicklichkeit und Kombinationsgabe gestellten Anforderungen sich verbinden.

Zu einer vernünftigen Definition führt ungezwungen unser Begriff des Kollektivs. Ein reines Glücksspiel ist ein solches, bei dem die Gewinstaussichten der Spieler, d. s. die relativen Häufigkeiten der Gewinne bei (unendlich) fortgesetztem Spiel, von den persönlichen Eigenschaften der Spieler nicht beeinflußt werden. Ob dieses Kriterium im Einzelfall zutrifft oder nicht, kann (wenn es sich nicht um einen ganz automatischen Vorgang handelt, an dem die Spielparteien gar nicht tätig mitwirken) in letzter Linie nur durch den statistischen Versuch, d. i. durch eine genügend lang ausgedehnte Spielreihe festgestellt werden. Eine „a-priorische" Entscheidung, d. h. eine logische Deduktion aus den Spielregeln, ist in keinem Falle möglich, eine solche Möglichkeit wird nur manchmal vorgetäuscht dadurch, daß — wie etwa bei der Roulette — allgemeinere Erfahrungen unbewußt zugrunde gelegt werden. Der Standpunkt, den wir durch unsere Untersuchungen gewonnen haben, wonach auch beim Würfel- oder Roulettespiel die Regellosigkeit der Ausgänge ein Erfahrungsergebnis ist, gestattet die völlig einheitliche Auffassung des ganzen Fragenkomplexes.

Für die praktische Aufgabe der Gesetzgebung kommt übrigens weniger die Frage des reinen Glücksspieles in Betracht, da dann wohl zu wenig ausgeschlossen würde. Eine mögliche Formulierung wäre etwa die folgende: Spielveranstaltungen, bei denen die relativen Häufigkeiten der Gewinne, die auf die einzelnen Mitspielenden nach langer Spieldauer entfallen, durch die Geschicklichkeit der Spieler gar nicht oder nur unerheblich beeinflußt werden, sind verboten (bzw. bedürfen der Genehmigung usf.). Beim Bajazzospiel erfolgt das Herabfallen der Kugel wohl zufallsartig, aber die Geschicklichkeit (Geschwindigkeit), mit der der Spieler den Auffangbecher bewegt, beeinflußt in hohem Maße, wie die Erfahrung lehrt, seine Chancen. Hingegen läge ein reines Glücksspiel vor, wenn man den Becher, bevor noch die Kugel zu rollen beginnt, in eine feste Stellung zu bringen hätte, wobei natürlich den verschiedenen Auffangstellen verschiedene Gewinstquoten zugeordnet werden müßten.

Marbes „Gleichförmigkeit in der Welt".

Ich verlasse jetzt das Gebiet der Glücksspiele und wende mich zu einer Frage, die, an der Grenze zwischen sozialer und biologischer Statistik stehend, immerhin noch eine nahe Verwandtschaft mit Glücksspielaufgaben aufweist und die seit einer Reihe von Jahren wiederholte Erörterung in der Literatur gefunden hat. Der Würzburger Philosoph Karl Marbe hat in einem umfassenden Werk „Die Gleichförmigkeit in der Welt" eine Theorie entwickelt, zu deren Kennzeichnung am besten folgendes Beispiel dient. Ein Ehemann, dem viel daran liegt, einen Sohn zu bekommen, sieht unmittelbar vor der Niederkunft seiner Frau im Standesregister nach, ob auf den letzten Seiten mehr Knaben- oder mehr Mädchengeburten eingetragen sind.

Er hält seine Aussichten für günstig, wenn das letztere der Fall ist. Während man nun im allgemeinen geneigt sein dürfte, diesen Gedankengang für töricht zu erklären, sagt Marbe: „Unsere statistischen Untersuchungen zeigen, daß (was man bisher immer übersehen hat) ein ganz, ganz kleines Körnchen Wahrheit auch in den Ansichten jenes Vaters steckt.“ Gestützt wird diese Behauptung auf die Untersuchung einer umfangreichen Statistik von etwa 200000 Geburtseintragungen in vier bayrischen Städten, bei der sich gezeigt hat, daß nur ein einziges Mal eine Iteration der Länge 17 (d. i. eine ununterbrochene Folge von 17 Geburten gleichen Geschlechtes) und niemals eine größere vorgekommen ist. Auf dem Studium der langen Iterationen beruhen auch die meisten Versuche, zu einem „Spielsystem“ zu gelangen, und Marbe gibt in seinem Buche tatsächlich an, in welcher Weise, d. h. durch welche Auswahl der Spielpartien man in Monte Carlo seine Gewinstchancen verbessern kann.

Das Marbesche Problem, über das, wie gesagt, schon viel geschrieben wurde — Marbe selbst hat seinem zweibändigen Werk von 1916 kürzlich noch einen Nachtrag folgen lassen —, ist für uns nur soweit von Interesse, als es die Grundlagen der Wahrscheinlichkeitsrechnung berührt. Der Marbesche Gedanke ist offenbar der, daß, wenn einmal 17 Eintragungen männlichen Geschlechts vorliegen, mit Sicherheit oder zumindest mit großer Wahrscheinlichkeit eine weibliche Eintragung zu erwarten steht. Faßt man die aufeinanderfolgenden Registrierungen, die etwa mit M und F (masculini und feminini) bezeichnet seien, als die Ergebnisse eines einfachen Alternativversuches wie „Kopf oder Adler“ auf, so würde aus der zweiten Eigenschaft der Kollektivs, aus der „Regellosigkeit“, folgen, daß auch dann, wenn man nur die auf 17 M unmittelbar folgenden Eintragungen allein betrachtet, unter ihnen die M und F sich im gleichen Verhältnis finden müssen wie in der Gesamtheit sämtlicher Eintragungen. Denn die Heraushebung der auf 17malige M-Eintragung folgenden Buchstaben bedeutet eine „Stellenauswahl“ der früher betrachteten Art, durch die ja die relativen Häufigkeiten der Merkmale, d. i. hier der Zeichen M und F, nicht beeinflußt werden dürfen. Somit sehen wir, daß die Frage, die durch die Behauptung Marbes über den „statistischen Ausgleich“ aufgeworfen wird, dahin mündet: Hat die Folge der Geburtseintragungen mit den Geschlechtsunterschieden als Merkmalen die Eigenschaften eines Kollektivs oder nicht?

Es ist klar, daß diese Frage nur empirisch, d. h. nur durch die Beobachtung selbst entschieden werden kann, und daß jeder Versuch, die Lehre vom statistischen Ausgleich durch „a-priorische" oder logische Überlegungen zu entkräften, sinnlos ist. Wir befinden uns hier übrigens geradezu im Zentrum der Schwierigkeit aller praktischen Statistik, da wir zu untersuchen haben, ob eine bestimmte statistische Aufnahme als Kollektiv aufgefaßt werden darf oder nicht.

Erledigung des MARBEschen Problems.

Zur Entscheidung der empirischen Frage ist nun zunächst zu sagen, daß MARBE jedenfalls trotz des ungeheuren Umfanges seiner Statistik unmittelbar keinen stichhaltigen Grund für seine Behauptung beibringen kann. Denn wenn eine Iteration der Länge 17 nur einmal vorgekommen ist, so hat er auch nur ein einziges Mal den „Versuch" gemacht, der lehren kann, was auf 17 gleiche Eintragungen folgt; mit anderen Worten, durch seine Stellenauswahl ist überhaupt nur ein einziges Element ausgewählt worden, während eine vernünftige Behauptung über die Verhältnisse in der ausgewählten Teilfolge erst möglich ist, sobald diese selbst einen angemessenen Umfang besitzt. Anderseits folgt aus diesem Tatbestand auch nichts gegen MARBE, und wenn wir nicht die Möglichkeit anderer Schlußweise hätten, könnten wir nicht viel über die Berechtigung seiner Behauptung sagen.

Aber dies ist eben der Erfolg der Wahrscheinlichkeitstheorie, daß sie solche Möglichkeiten indirekten Schließens eröffnet. Die Grundoperationen der Wahrscheinlichkeitsrechnung, die wir kennengelernt haben, gestatten es, aus dem als gegeben vorausgesetzten einfachen Kollektiv mit den Merkmalen M und F und der ungefähren Verteilung 0,5 : 0,5 ein neues abzuleiten, dessen Element eine Serie von n Versuchen ist und dessen Verteilung die Frage beantwortet: Wie groß ist die Wahrscheinlichkeit dafür, daß in einer Serie von n Versuchen genau x Iterationen der Länge m vorkommen? Setzen wir in die Formel, die als Resultat der Rechnung erscheint, z. B. $m = 17$ und $x = 1$ ein, so erhalten wir die Wahrscheinlichkeit dafür, daß genau eine Iteration von der Länge 17 auftritt, wie es im MARBEschen Fall beobachtet wurde. Anderseits können wir mit $m = 18$ und $x = 0$ berechnen, wie groß die Wahrscheinlichkeit dafür ist, daß keine 18-Iteration vorkommt. Das entscheidende Beobachtungsergebnis, das MARBE zu seinen Behauptungen führte, war ja eben, daß einmal 17 gleiche Eintragungen hintereinander vorlagen, aber keinmal 18. Wenn wir zeigen können, daß diese Verhältnisse nach der Wahrscheinlichkeitsrechnung vorauszusehen waren, so ist damit ein starkes Argument dafür gewonnen, daß die Geburtseintragungen die Voraussetzungen, denen ein Kollektiv genügen muß, erfüllen.

Rechnet man die genannten Größen tatsächlich aus für ein $n = 49152$, entsprechend den vier von MARBE untersuchten Serien, so findet man z. B. für die 17-Iterationen: Wahrscheinlichkeit einmaligen Auftretens gleich 0,16, mehr als einmaligen Auftretens 0,02, Wahrscheinlichkeit keiner 17-Iteration 0,82. Da nun von den vier Serien eine eine Iteration der Länge 17 aufwies, die anderen keine, so liegt zumindest kein Widerspruch zwischen Rechnungs- und Beobachtungsresultat vor. Für die Iterationslänge $m = 18$ ergibt die Rechnung die Wahrscheinlichkeit 0,91 dafür, daß keine vorkommt, und tatsächlich hat sich in den vier Städten keine gezeigt. Es ist hier nicht der Ort, im einzelnen durch Anführung aller rechnungsmäßigen Ergebnisse und ihre Gegenüberstellung mit der Beobachtung nachzuweisen, daß in der Tat vollständige Übereinstimmung zwischen den Folgerungen der Wahrscheinlichkeitsrechnung und den Eigenschaften der von MARBE erhobenen Statistik besteht. Nur ein paar Zahlen seien herausgegriffen. Die Rechnung ergibt als die im Mittel zu erwartende Anzahl von Iterationen der Längen 2 bzw. 6 bzw. 10 die Zahlen 6138 bzw. 385 bzw. 24,3. Tatsächlich betrug der Durchschnitt der vier von MARBE festgestellten Größen im ersten Fall 6071, im zweiten 383, im dritten 24,3. Die Einzelheiten habe ich bei früherer Gelegenheit an anderer Stelle ausgeführt. Auch die neueren, etwas subtileren Behauptungen MARBES liegen in derselben Richtung und sind durch analoge Rechnungen widerlegbar.

Jeder, der Interesse für die Grundfragen der Wahrscheinlichkeitslehre besitzt, wird MARBE Dank wissen für die Erschließung und sorgfältige Durcharbeitung des großen statistischen Materials, durch das zum erstenmal in zuverlässiger Weise ein derart umfangreiches, nicht dem Gebiet der Glücksspiele angehörendes Beispiel eines Kollektivs der wissenschaftlichen Untersuchung zugänglich gemacht wurde. Was freilich die praktischen Schlußfolgerungen betrifft, so gelangen wir zu den dem MARBEschen Standpunkt gerade entgegengesetzten. Denn die festgestellte Übereinstimmung zwischen den auf dem Kollektivbegriff aufgebauten Berechnungen und der Beobachtung kann nur dahin führen, anzunehmen, daß die Folge der Geburtseintragungen mit weitgehender Annäherung alle Eigenschaften eines Kollektivs besitzt, also auch der Forderung der Regellosigkeit genügt. Darnach würden unter allen auf männlichen Nachwuchs bedachten Vätern, die im Laufe der Zeit in die allerdings seltene Lage kommen, gerade nach 17 weiblichen Eintragungen einem Familienzuwachs entgegenzusehen, ebenso viele glückliche wie enttäuschte sich finden: ungefähr die Hälfte von ihnen wird zu einem achtzehnten „*F*" die Veranlassung geben und nur die andere Hälfte wird durch ein „*M*" die Iteration abbrechen.

Knäuelungstheorie und Gesetz der Serie.

Ein merkwürdiges Gegenstück zu der MARBEschen Lehre vom statistischen Ausgleich bildet die von einem andern Philosophen herrührende „Knäuelungstheorie" von STERZINGER, die der Biologe PAUL KAMMERER zu einer wissenschaftlichen Begründung des vielverbreiteten Glaubens an ein „Gesetz der Serie" benutzt hat. Ein Unglück komme selten allein, meint der Volksmund, und nach KAMMERER müßte das gleiche auch für die Glücksfälle gelten.

Sucht man dem „Gesetz der Serie" einen präziseren Ausdruck zu geben, so kann man nur sagen: Kurze Iterationen gleicher Versuchsergebnisse kommen unerwartet oft vor, angeblich öfter, als es der wahrscheinlichkeitstheoretischen Berechnung entspricht. STERZINGER beobachtete z. B. die Aufeinanderfolge der Zeitintervalle, in denen die Kunden einen Laden betreten. Wenn im Durchschnitt 30 Kunden auf eine Stunde entfallen, gibt es doch nur selten einen Zeitraum von zwei Minuten, in dem genau ein Kunde eintritt; viel öfter bleibt ein Zweiminutenintervall leer oder es erscheinen zwei oder mehr Kunden. Aus derartigen Beobachtungen zieht STERZINGER weitgehende Folgerungen für die gesamte Naturauffassung, vor allem sieht er in ihnen eine Widerlegung der Wahrscheinlichkeitsrechnung. Unser Standpunkt demgegenüber ergibt sich ungezwungen aus allem bisher Gesagten. Selbstverständlich gibt es auch Erscheinungen, für welche die „Zufallsgesetze“ nicht gelten, d. h. es gibt Wiederholungsvorgänge, die Kollektivs sind, und solche, die es nicht sind. Welcher von beiden Fällen vorliegt, kann nur die genaue Analyse der Beobachtung zeigen; einer oberflächlichen, rein gefühlsmäßigen Schätzung kommt da keine große Bedeutung zu. Um eine gerechtfertigte Unterscheidung treffen zu können, muß man vor allem Rechnungsergebnisse kennen, die einen Vergleich mit der Erfahrung gestatten.

Ein solches Ergebnis der rationellen Wahrscheinlichkeitstheorie, d. h. der Anwendung der von uns wiederholt betrachteten Rechenregeln, ist beispielsweise folgendes: Wenn eine Ereignisreihe zufallsartig über eine große Zeitstrecke so verteilt ist, daß auf ein Intervall der Länge a im Durchschnitt ein Ereignis entfällt, so besteht nur 0,37 Wahrscheinlichkeit dafür, daß ein beliebig herausgegriffenes Intervall von der Länge a genau ein Ereignis umfaßt, dagegen 0,63 Wahrscheinlichkeit dafür, daß entweder kein oder mehr als ein Ereignis in dem Intervall liegt. Man könnte also in dem Sterzingerschen Fall nur dann von einer Knäuelung sprechen, die der Zufallstheorie widerspricht, wenn erwiesen wäre, daß wirklich mehr als zwei Drittel der Zweiminutenstrecken eine von Eins abweichende Kundenzahl aufweist. An sich wäre das durchaus möglich und würde eben nur besagen, daß diese Erscheinungsfolge nicht unmittelbar als Kollektiv aufgefaßt werden kann. Aber die Beobachtungen von Sterzinger lassen keinen solchen Schluß zu, sie sind völlig nichtssagend. Offenbar hat Sterzinger gemeint, jedes häufigere Auftreten von Intervallen mit von 1 abweichender Zahl der Eintretenden sei bereits ein Widerspruch gegen die Wahrscheinlichkeitsrechnung.

Noch ein anderes Beispiel zum Gesetz der Serie! Nach der Selbstmordstatistik Deutschlands entfällt in einem Landesteil von etwa einer Viertelmillion Einwohner durchschnittlich ein Selbstmord auf jede Woche. Wie oft wird das betreffende Kreisblatt eine „Serie“ von drei oder mehr Selbstmorden in einer Woche zu verzeichnen haben? Die Rechnung ergibt, daß die Wahrscheinlichkeit eines mehr als zweimaligen Auftretens des

Ereignisses 0,08 beträgt. Sofern also die Folge der Selbstmorde die Eigenschaften eines Kollektivs besitzt, muß man durchschnittlich in einem Jahr (52 Wochen mal 0,08 gibt 4,16) vier bis fünf Wochen mit „Selbstmordserien" erwarten. Nur wenn der tatsächliche Verlauf diese Zahl erheblich überschreiten sollte, wäre man berechtigt, aus dieser Beobachtung auf einen inneren Zusammenhang der zeitlich benachbarten Fälle zu schließen. In dem umfangreichen Werk von KAMMERER über das „Gesetz der Serie" ist in keinem einzigen Beispiel die Unterlage gegeben, die ermöglichen würde, aus der statistischen Aufnahme (die meist ganz unzureichenden Umfang hat) zu entscheiden, ob der betreffende Wiederholungsvorgang ein Kollektiv bildet oder nicht.

Verkettete Vorgänge.

Auf der andern Seite kennt man in der Statistik mannigfache Erscheinungsreihen, die sich nicht unmittelbar als Kollektivs auffassen lassen und bei denen man sehr wohl von einem „Gesetz der Serie" sprechen könnte. Man tut es wohl nur deshalb nicht, weil diese Bezeichnung — dies ist rein historisch bedingt — einen Widerspruch gegen die Wahrscheinlichkeitsrechnung zum Ausdruck bringen soll, während die Vorgänge, die ich im Auge habe, offenkundig in bester Übereinstimmung mit ihr stehen. Ein einfaches Beispiel für solche „verkettete" Erscheinungen bilden die Todesfälle, die infolge einer bestimmten ansteckenden Krankheit, etwa der Blattern, eintreten. Es ist von vornherein klar, daß hier die Zahlen ganz anders verteilt sein müssen als etwa in der Selbstmordstatistik. Der Mathematiker G. PÓLYA hat uns gezeigt, wie man dieses Problem mit den Hilfsmitteln der Wahrscheinlichkeitsrechnung in eleganter Weise behandeln kann.

Die Todesfälle infolge einer bestimmten Todesursache ohne jede Ansteckung kann man sich etwa unter diesem Bild vorstellen: Aus einer Urne, die teils weiße, teils schwarze Kugeln in irgendeinem Mischungsverhältnis enthält, wird von Monat zu Monat für jeden Bewohner des Landes einmal gezogen; erweist sich seine Kugel als schwarz, so bedeutet dies Tod. Den verschiedenen Todesursachen entsprechen verschiedene Mischungsverhältnisse der schwarzen und weißen Kugeln in der Urne. Das Bild soll natürlich nur das besagen, daß man die Wahrscheinlichkeit für das Eintreten einer bestimmten Zahl X von Todesfällen in einem Monat, also für das Erscheinen dieser Zahl in der Statistik, so berechnen kann wie die Wahrscheinlichkeit dafür, aus einer Urne unter den gegebenen Verhältnissen gerade X schwarze Kugeln zu ziehen. Dabei ist vorausgesetzt, daß nach jedem Zug die gezogene Kugel zurückgelegt wird, so daß die Chancen unverändert bleiben.

Für eine Statistik „mit Ansteckung" trifft dieses Bild nicht mehr zu. Hier werden offenbar durch Eintritt eines Todesfalls die Aussichten, daß weitere eintreten, verstärkt. Man kann nach PÓLYA diesem Umstand in dem Bild — man sagt hierfür auch oft: in dem „Urnenschema" — dadurch Rechnung tragen, daß man annimmt: So oft eine schwarze Kugel gezogen ist, wird nicht nur diese zurückgelegt, sondern es werden die Kugeln in der Urne um eine bestimmte Zahl schwarzer Kugeln vermehrt. Auf diese Weise hat man es nicht mehr mit einem einzigen Ausgangskollektiv zu tun, sondern mit einer Reihe verschiedener, deren Verteilungen man aber genau kennt, nämlich mit Urnen von verschiedenen Füllungen. Man kann durch gewisse Grundoperationen daraus ein Endkollektiv ableiten, dessen Element der ganze, etwas verwickelte Vorgang von n aufeinanderfolgenden Ziehungen ist, wobei n für die Zahl der Einwohner gesetzt ist. PÓLYA zeigt, wie man durch eine verhältnismäßig einfache Formel die Wahrscheinlichkeiten für das Auftreten von 0, von 1, von 2, 3, . . . usf. Todesfällen in einem Monat (nämlich von ebensoviel schwarzen Zügen unter insgesamt n) darstellen kann.

Wie gut die Rechnung mit der Beobachtung stimmt, will ich an folgenden Zahlen der Schweizer Blattern-Statistik von 1877 bis 1900 klarmachen. Es wurden, bei einem Durchschnittswert von 5,5 Fällen im Monat, beobachtet: 100 Monate ohne Todesfall, 39 mit einem, 28 mit zwei, 26 mit drei, 13 mit vier, 3 mit fünfzehn Todesfällen. Die Theorie der „verketteten" Vorgänge läßt folgende Zahlen erwarten: für keinen Todesfall 100,4, für 1 bis 4 Todesfälle der Reihe nach 36,3, 23,5, 17,5 und 13,8, und für fünfzehn 3,0. Dagegen würde ohne Berücksichtigung der Verkettung, wenn man also das einfache Urnenschema mit sofortigem Zurücklegen der Kugel zugrunde legt, sich ergeben: Für keinen Todesfall die Erwartungszahl 1,2, für 1 bis 4 Fälle die Zahlen 6,5, 17,8, 32,6, 44,9, für 15 Fälle die Zahl 0,1. Man erkennt die Überlegenheit der richtigen Theorie über die primitive. Ohne Rechnung ist aber die Wirkung der Verkettung, die keineswegs bloß im Überwiegen längerer Serien sich auswirkt, nicht leicht zu übersehen.

Im nächsten Vortrag, der den Problemen der physikalischen Statistik gewidmet sein wird, werden wir noch von umfassenderen Problemen der „Verkettung" oder „Nachwirkung" zu reden haben.

Die allgemeine Aufgabe der Statistik.

Wir haben jetzt an speziellen Beispielen einiges Material gewonnen, um wieder die allgemeine Aufgabe, die uns gegenüber den statistischen Problemen erwächst, etwas näher ins Auge fassen zu können. Ich erwähnte schon vorhin, daß uns die von Marbe angeregte Fragestellung nahe an den Kernpunkt des Problems heranführt, indem sie zu der Untersuchung Veranlassung gibt, ob eine bestimmte, durch statistische Aufnahme gewonnene Zahlenreihe ein Kollektiv bildet oder nicht. Die allgemeine Aufgabe ist im Anschluß an meine einleitenden Bemerkungen etwa so zu formulieren: Es liegt ein durch statistische Aufnahmen beschafftes Zahlenmaterial vor, z. B. die Anzahl der Eheschließungen in verschiedenen Teilen Deutschlands in einer Reihe von Jahren; man stellt sich die Aufgabe, diese Zahlen entweder unmittelbar als ein Kollektiv zu erkennen — genauer als den endlichen Abschnitt einer unendlichen Zahlenfolge, die die Eigenschaften eines Kollektivs besitzt — oder sie irgendwie auf Kollektivs zurückzuführen, sie als aus Kollektivs entstanden zu deuten. Ein lehrreiches Beispiel hierfür haben wir soeben in der Statistik der verketteten Todesfälle kennengelernt, wo es möglich war, eine beobachtete Zahlenreihe, die an sich keine volle „Regellosigkeit" besitzt, also kein Kollektiv bildet, durch Einführung eines gewissen Systems von Kollektivs aufzuklären. An anderen Aufgaben werde ich gleich weiter zeigen, worin die Leistung der Wahrscheinlichkeitstheorie in diesem Sinne besteht. Gestatten Sie mir nur vorher noch, die eben gegebene Abgrenzung des Problems der theoretischen Statistik gegen einen naheliegenden Einwand in Schutz zu nehmen.

Bei flüchtigem Zusehen mag man einen Widerspruch darin finden, daß zuerst der Begriff des Kollektivs aus der Betrachtung statistischer Reihen geschöpft wurde — habe ich doch früher

z. B. für die Definition der Sterbenswahrscheinlichkeit die Todesfallstatistik herangezogen — und daß hinterher ein Problem daraus gemacht wird, zu erforschen, ob eine statistische Reihe ein Kollektiv bildet, bzw. in welcher Weise sie mit Kollektivs zusammenhängt. Aber auch hier wird, wer mit der Methodik der Naturwissenschaften nur einigermaßen vertraut ist, den scheinbaren Widerspruch rasch gelöst sehen.

Wir wissen, daß die NEWTONsche Gravitationstheorie zu der Behauptung führt, die Bahnen der Planeten um die Sonne seien Ellipsen, und daß gerade die KEPLERschen Beobachtungen der Ellipsengestalt der Bahnen einen der Ausgangspunkte der Theorie gebildet haben. Wir wissen aber auch, daß keiner der acht großen oder der tausend kleinen Planeten bei seinem Umlauf um die Sonne eine genaue Ellipse beschreibt und daß es eben die wesentlichste Aufgabe der rechnenden Astronomie ist, die beobachteten, von der Ellipsenform abweichenden Bahnen auf die Gesetzmäßigkeiten zurückzuführen, die die NEWTONsche Gravitationstheorie postuliert. Ebenso besitzen wir in der MAXWELLschen Theorie der Elektrizität einen allgemeinen Rahmen, von dem wir überzeugt sind, daß er alle elektrischen und magnetischen Erscheinungen (wenn man von denen in sehr rasch bewegten Medien absieht) umschließt. Allein es ist eine große, niemals sich erschöpfende Aufgabe, die einzelnen konkreten Vorgänge, wie sie sich z. B. in einer elektrischen Maschine zeigen, als Auswirkungen der einfachen Beziehungen zu deuten, die in den MAXWELLschen Gleichungen zum Ausdruck kommen. Es bedeutet keinen Einwand gegen MAXWELL, daß die sogenannten Gedankenexperimente, die seinen Begriffsbildungen und seinen Gleichungsansätzen zugrunde liegen, keinem in der Natur wirklich nachweisbaren Vorgang unmittelbar entsprechen und sich nur in seltenen Fällen, und dann nur mit einiger Annäherung realisiert finden.

So glaube ich auch den Einwänden begegnen zu können, die dahin gehen, daß die konkrete Statistik niemals auf Zahlenreihen führt, die unmittelbar und exakt als Kollektivs erscheinen. Nicht darauf kommt es an, sondern einzig und allein darauf, was sich mit der auf dem abstrakten Kollektivbegriff aufgebauten Wahrscheinlichkeitstheorie im Bereiche der Anwendungen beginnen läßt. Dazu jetzt noch einige weitere Beispiele.

Der Gedanke der Lexisschen Dispersionstheorie.

Der bekannte Göttinger Nationalökonom W. Lexis hat die Statistik durch einen Gedanken bereichert, der in sehr bequemer Weise dazu dienen kann, den Vergleich zwischen einer statistisch erhobenen Zahlenreihe und einem bestimmten Kollektiv durchzuführen. Ich will den Lexisschen Gedanken an Hand einer kleinen statistischen Aufnahme erläutern, wie Sie sie selbst jederzeit mit geringer Mühe ausführen können.

Es werde untersucht, wie oft der Buchstabe „a" in einem lateinischen Text, z. B. im Anfang von Caesars „Bellum Gallicum" vorkommt. Nimmt man die ersten 20 Gruppen von je 100 Buchstaben zum Gegenstand der Untersuchung, so finden sich, wenn man von der Reihenfolge absieht, zwei Gruppen mit fünf „a", je drei Gruppen mit sechs, sieben und acht, wieder zwei Gruppen mit neun, dann eine Gruppe mit zehn, je zwei mit elf und zwölf, schließlich je eine mit dreizehn und vierzehn „a". Im ganzen sind es 174 „a", also durchschnittlich $174 : 20 = 8{,}7$ in einer Gruppe. Die Frage, die sich hier aufdrängt, ist die: Wie verhält sich das Ergebnis, das die statistische Aufnahme gezeitigt hat, zu dem, was zu erwarten wäre, wenn man aus einer Urne, die Lose mit den 25 Buchstaben des Alphabets in einem geeigneten Mischungsverhältnis enthält, 20mal je 100 Ziehungen vollzogen und die in jeder Gruppe von Ziehungen zutage geförderten „a" gezählt hätte? Dabei wird man, da sich unter 100 Buchstaben durchschnittlich 8,7 „a" gefunden haben, vernünftigerweise annehmen, daß 8,7 v. H. der Lose den Buchstaben „a" tragen.

Man könnte, um die Frage zu beantworten, etwa so vorgehen, daß man zunächst für das Kollektiv, dessen Element hundertmaliges Ziehen aus der Urne ist, die Verteilung berechnet, d. h. die Wahrscheinlichkeiten dafür, daß unter 100 Buchstaben kein „a", oder deren eines, zwei drei ... usf. vorkommen; sodann diese Wahrscheinlichkeiten einzeln mit den tatsächlichen Häufigkeiten der verschiedenen, beobachteten Wiederholungszahlen des „a" vergleicht. Es ergibt sich z. B. die Wahrscheinlichkeit 0,10 für sechsmaliges, 0,14 für achtmaliges Auftreten des „a", so daß bei 20 Versuchen $20 \cdot 0{,}10 = 2$mal die Wiederholungszahl 6 und $20 \cdot 0{,}14 = 2{,}8$, also 2- bis 3mal die Wiederholungszahl 8 „zu erwarten" wäre. Tatsächlich ist die Sechs sowohl wie die Acht dreimal beobachtet worden. Es ist nicht leicht, hier von Übereinstimmung oder von Widerspruch zu reden, besonders da die Gesamtheit aller möglichen Wiederholungszahlen (die von 0 bis 100 laufen) zugleich berücksichtigt werden müßte. Man empfindet das Bedürfnis nach einem eindeutigen Maßstab zur Beurteilung der Übereinstimmung zwischen Rechnung und Beobachtung, nach einem Maßstab, der in gewissem Sinne einen Durchschnitt aus dem bildet, was sich einzeln für die verschiedenen Wiederholungszahlen ergibt. Die Lexissche Theorie besteht gerade darin, einen solchen Maßstab zu liefern; sie gibt für das vorliegende Problem in Gestalt einer einzigen Zahl ein Kriterium für den Grad der Übereinstimmung bzw. Nicht-Übereinstimmung der Beobachtung mit dem, was der Vergleich mit dem Kollektiv erwarten läßt.

Durchschnitt und Streuung.

Um zu dem Lexisschen Kriterium zu gelangen, muß ich zunächst erklären, was man unter der „Streuung" einer Zahlenreihe versteht. Es ist dies eine der statistischen Funktionen, von denen am Schlusse des vorigen Vortrages die Rede war, und zwar eine der allerwichtigsten.

Daß man durch Addition der sämtlichen Einzelwerte und Division der Summe durch die Anzahl der Summanden den „Durchschnittswert" erhält, kann ich als bekannt voraussetzen. Wir sprachen auch schon davon, daß der Durchschnitt die einfachste statistische Funktion einer Zahlenreihe bildet. In unserem Beispiel lagen die 20 Einzelwerte zwischen 5 und 14, der Durchschnitt berechnete sich zu 8,7. Wenn wir nun die Differenz ausrechnen, die zwischen je einem Einzelwert und dem Durchschnitt besteht, so bekommen wir eine neue Zahlenreihe, und zwar eine solche von teils positiven, teils negativen Zahlen. Dem dreimaligen Auftreten der Anzahl 7 z. B. entsprechen jetzt drei negative Differenzbeträge —1,7, dem zweimaligen Auftreten von 11 die zweimaligen positiven Differenzen 2,3 usf. Es folgt aus der Art, wie der Durchschnitt berechnet wurde, daß die Summe aller dieser positiven und negativen Differenzen Null ergibt, d. h. daß die positiven und negativen sich aufheben. Nach der Zusammenstellung, die ich oben gegeben habe, liegen vor:

2mal	die	Differenzen		—3,7
je 3	,,	,,	,,	—2,7, —1,7, —0,7
,, 2	,,	,,	,,	0,3
1	,,	,,	,,	1,3
je 2	,,	,,	,,	2,3 und 3,3
1	,,	,,	,,	4,3 ,, 5,3

Die Summe der negativen Differenzbeträge ergibt sich, wie man sieht zu $2 \,.\, 3{,}7 + 3 \,.\, (2{,}7 + 1{,}7 + 0{,}7) = 7{,}4 + 15{,}3 = 22{,}7$ und ebenso groß ist die der positiven. Man erkennt daraus, daß es nicht möglich wäre, durch Summieren der Abweichungen der Einzelwerte vom Durchschnittswert ein Maß der „Streuung“, der Ungleichförmigkeit der Ergebnisse, zu erhalten. Ein solches kann man aber in einfacher Weise finden, wenn man die einzelnen, positiven oder negativen Differenzen ins Quadrat erhebt und den Durchschnitt dieser, jedenfalls positiven, Quadratzahlen bildet. So wird in der Tat als das *arithmetische Mittel aus den ins Quadrat erhobenen Abweichungen* die „*Streuung*“ einer Zahlenreihe üblicherweise definiert. Es ist klar, daß diese Größe um so kleiner ist, je enger die Einzelwerte beim Durchschnittswert liegen und daß sie dann, und nur dann, gleich Null wird wenn alle Einzelwerte untereinander und daher auch mit dem Durchschnittswert zusammenfallen. In unserem Beispiel haben wir zweimal das Quadrat von —3,7, also $2 \,.\, 13{,}69 = 27{,}38$, je dreimal die Quadrate von —2,7, —1,7 und —0,7, also $3 \,.\, (7{,}29 + 2{,}89 + 0{,}49) = 3 \,.\, 10{,}67 = 32{,}01$ zu nehmen. Je zweimal sind die Quadrate von 0,3, von 2,3 und von 3,3 zu nehmen, das gibt $2 \,.\, 16{,}27 = 32{,}54$, schließlich je einmal das Quadrat von 1,3, von 4,3 und von 5,3, zusammen 48,27. Im ganzen ergibt sich die Quadratsumme zu 140,20 und nach Division durch 20 ihr Durchschnittswert, also die Streuung, zu 7,01.

Es ist leicht zu sehen, daß die hier definierte Streuungsgröße die Eigenschaft einer statistischen Funktion besitzt. Denn die Anordnung, die Reihenfolge, in der die Einzelwerte auftreten, spielt keinerlei Rolle in der Definition. Wird aber die Anzahl aller Beobachtungen verdoppelt und tritt dabei jede der erstmaligen Beobachtungen zweimal auf, so multiplizieren sich Zähler und Nenner des Bruches (Quadratsumme der Differenzen durch Versuchszahl) gleicherweise mit 2, so daß die Streuung unverändert bleibt.

Vergleich der tatsächlichen und der zu erwartenden Streuung.

Lexis vergleicht nun die in der angegebenen Weise berechnete Streuung der empirisch gefundenen Zahlenreihe mit einer bestimmten Größe, zu der die wahrscheinlichkeitstheoretische Untersuchung, ausgehend von dem schon vorhin erwähnten Kollektiv, führt. Denkt man sich aus einer Urne, die unter 1000 Losen 87 mit „a" bezeichnete enthält, je 100 Ziehungen ausgeführt, so kann man die Zahl der gezogenen „a" als Merkmal eines (aus den 100 Ziehungen bestehenden) Elementes ansehen. Man kann aber dann weiter durch die Operation der „Verbindung" 20 solcher Elemente zum Element eines neuen Kollektivs zusammenfassen. Als Merkmal des Elementes innerhalb dieses durch 20fache Verbindung entstandenen Kollektivs kann jede Größe gewählt werden, die durch die 20 Gruppen von Ziehungen bestimmt wird, also z. B. auch die Streuung der 20 Anzahlen von „a"-Ergebnissen. Durch wiederholte Anwendung der Rechenregeln, die wir früher betrachtet haben, gelingt es, die Verteilung innerhalb des neuen Kollektivs, also die Wahrscheinlichkeit jedes möglichen Streuungswertes, zu ermitteln. So könnte man z. B. ausrechnen, wie groß die Wahrscheinlichkeit ist, gerade die Streuung 7,01 zu erhalten. Angemessener aber erscheint folgender Gedankengang.

Wenn man innerhalb eines Kollektivs mit mehrfachen Ausgangsmöglichkeiten oder Merkmalwerten jeden dieser Werte mit der Wahrscheinlichkeit seines Eintreffens multipliziert und die Produkte addiert, so erhält man eine Zahl, die man mit Recht als den „Erwartungswert" des Merkmals, manchmal auch mit einem etwas romantischen Namen als „mathematische Hoffnung" bezeichnet. Wer etwa in einem Glücksspiel 20 % Wahrscheinlichkeit hat, 10 M zu gewinnen, 30 % Wahrscheinlichkeit, 20 M zu erhalten, während er in 50 % der Fälle auf keinen Gewinn rechnen kann, der wird als „Erwartungswert" ansehen: $0{,}20 \,.\, 10 + 0{,}30 \,.\, 20 = 8$ M. Die Bedeutung des Erwartungswertes ist natürlich die, daß er den Gewinn (bzw. den Merkmalwert) darstellt, der bei unbeschränkt fortgesetzter Spielfolge im Durchschnitt auf das einzelne Spiel entfällt.

Die Algebra lehrt, wie man in dem von uns vorhin ins Auge gefaßten Fall eines durch wiederholte Verbindung gebildeten Kollektivs den Erwartungswert des Merkmals in einfacher Weise bestimmen kann. Bezeichnet n die Anzahl der Gruppen (in unserem Fall 20), z die Zahl der Spiele in einer Gruppe (in unserem Fall 100), endlich p die Wahrscheinlichkeit des beobachteten Ergebnisses innerhalb eines Spieles (in unserem Fall $87 : 1000 = 0{,}087$), so ist der Erwartungswert der Streuung durch die Formel gegeben

$$\frac{n-1}{n} z p (1-p).$$

Wie man im einzelnen diese Formel ableitet, kann hier unerörtert bleiben; es ist das eine rein formal-mathematische Angelegenheit, die mit den begrifflichen Grundlagen der Theorie nichts zu tun hat. Setzt man $n = 20$, $z = 100$, $p = 0{,}087$ in die Formel ein, so erhält man als den Erwartungswert der Streuung

$$\frac{19}{20} \cdot 100 \cdot 0{,}087 \cdot (1 - 0{,}087) = \frac{19}{20} \cdot 8{,}7 \cdot 0{,}913 = 7{,}55.$$

Diese Zahl gibt nach dem eben Gesagten an, wie groß im Durchschnitt die Streuung ist, wenn man unendlich oft je 20 Gruppen von 100 Ziehungen ausführt. Lexis stellt diesen theoretischen Wert der wirklich beobachteten Streuung gegenüber (in unserem Fall 7,01) und sieht in dem Quotienten der zweiten Zahl durch die erste den gesuchten Maßstab für die Beurteilung der Übereinstimmung zwischen Theorie und Beobachtung. In unserem Fall ist $7{,}01 : 7{,}55 = 0{,}928$ nicht viel von 1 verschieden. Darin erblickt man nach Lexis eine Bestätigung der Annahme, daß das Auftreten des Buchstaben „a“ im laufenden Prosatext des lateinischen Schriftstellers sich im großen ganzen „zufallsartig“ verhält, den Gesetzmäßigkeiten eines Kollektivs annähernd folgt.

Begründung durch die Gesetze der großen Zahlen.

Eine nähere Begründung erfährt die Lexissche Theorie durch die Heranziehung der allgemeinen Gesetze der großen Zahlen für statistische Funktionen, wie sie am Ende des vorigen Vortrags dargelegt wurden. Ich kann dies nur ungefähr andeuten, da diese Fragen schon stark ins Mathematische übergreifen.

Wir haben es jetzt zunächst mit zwei statistischen Funktionen der n Beobachtungen zu tun, dem Durchschnitt D und der Streuung S. Den Durchschnitt erhält man, indem man die n Beobachtungszahlen addiert und dann die Summe durch die Anzahl der n Summanden addiert, also das arithmetische Mittel bildet. In unserem Beispiel war $D = 174 : 20 = 8{,}7$. Zur Streuung S gelangen wir in der angegebenen Weise, indem wir zunächst von jeder Beobachtungszahl den Wert D abziehen, die Differenzen ins Quadrat erheben und endlich deren arithmetisches Mittel aufsuchen. Im Beispiel hatten wir $S = 140{,}20 : 20 = 7{,}01$. Mit z hatten wir die Anzahl der Einzelversuche bezeichnet, aus denen jede der n Beobachtungen sich zusammensetzte. Bilden wir den Quotienten $D : z = d$, im Beispiel $8{,}7 : 100 = 0{,}087 = d$, so kann man d den Durchschnitt der n Häufigkeiten nennen, während D der Durchschnitt der n tatsächlichen Beobachtungen selbst war. Die Zahl 0,087 hatten wir vorhin der Wahrscheinlichkeit p in dem zum Vergleich herangezogenen Kollektiv gleichgesetzt und dann in dem Ausdruck für den Erwartungswert der Streuung für das $z\,p\,(1-p)$ den Wert $z\,d\,(1-d)$ eingeführt.

Der tiefere Gedanke der Lexisschen Theorie besteht darin, daß der Quotient von S durch das Produkt $z\,d\,(1-d)$ untersucht wird, nämlich die Lexissche Zahl

$$L = \frac{S}{z\,d\,(1-d)}.$$

Diese ist natürlich eine (zusammengesetzte) statistische Funktion, da die einzelnen Bestandteile S und d solche Funktionen sind, während z eine feste Zahl bezeichnet, die von den Ergebnissen der Beobachtung in keiner Weise abhängt. Aus den Gesetzen der großen Zahlen folgt nun zweierlei über die Funktion L, vorausgesetzt, daß n eine genügend große Zahl ist.

Erstens: Handelt es sich um n Versuche an einem mit seiner Verteilung bekannten Kollektiv, so ist mit großer Sicherheit zu erwarten, daß ein Wert für L beobachtet werden wird, der sehr nahe dem aus der Verteilung errechneten „theoretischen" Wert von L liegt. Nun zeigt sich — und das steht in Übereinstimmung mit dem früher Gesagten —, daß der theoretische Wert von L gleich 1 ist, wenn das Element des betrachteten Kollektivs in der Beobachtung von z einfachen, untereinander gleichen und voneinander unabhängigen Alternativen besteht. In diesem Falle ist also bei großem n ein L-Wert nahe 1 zu erwarten.

Zweitens: Wenn man das Kollektiv, um das es sich handelt, nicht vollständig kennt, so darf man, bei großem n, schließen, daß der unbekannte, das Kollektiv charakterisierende „theoretische" Wert von L nahe dem beobachteten liegt. Hat man also einen Wert nahe 1 beobachtet, so wird, wenn nur n groß genug ist, die Vermutung nahegelegt, daß ein Kollektiv vorliegt, das die eben angeführte Beschaffenheit wenigstens annähernd besitzt. Diese, aus dem zweiten Gesetz der großen Zahlen folgende Schlußweise ist es, die man bei der praktischen Anwendung der Lexisschen Theorie tatsächlich benutzt.

Es wird dem aufmerksamen Zuhörer vielleicht nicht entgangen sein, daß wir vorhin, bei der ersten Darstellung des Lexisschen Gedankenganges die Größen S und

$$\frac{n-1}{n} \cdot z\,d\,(1-d)$$

einander gegenübergestellt haben, während jetzt im Zähler und Nenner der Lexisschen Zahl die Größen S und

$$z\,d\,(1-d)$$

auftreten. Der Unterschied ist dadurch zu erklären, daß wir jetzt nur Schlußfolgerungen für sehr große (streng genommen unendlich große) n betrachten. Es ist aber klar, daß für große n der Bruch $(n-1)/n$ sich unbeschränkt der Einheit nähert, also als Faktor unterdrückt werden darf.

Normale und nichtnormale Dispersion.

In dem Beispiel einer Statistik über das Vorkommen des Buchstaben „a" in einem lateinischen Text hatten wir ziemlich gute Übereinstimmung zwischen der beobachteten Streuung und ihrem Erwartungswert (7,01 gegenüber 7,55) gefunden. Mit anderen Worten heißt das, der Lexissche Quotient sei nicht allzuweit von 1 entfernt, die Annahme, daß es sich bei jeder der Beobachtungen um z gleiche, unabhängige Alternativen handle, ist einigermaßen berechtigt.

Die Lexissche Untersuchungsweise einer statistischen Aufnahme ist immer dann anwendbar, wenn jede der n in Frage stehenden Zahlen auf einer Beobachtung von irgendwelchen z Alternativen beruht. Z. B. gehört hierher die Statistik der Todesfälle in einer Bevölkerung, die aus z Individuen besteht, oder die Statistik der Knabenanzahl in einer Reihe von Geburtengruppen, etwa den Geburten eines Landes in verschiedenen Jahren oder verschiedener Länder im selben Jahre. Dabei wird allerdings im allgemeinen die Zahl z von Beobachtung zu Beobachtung mehr oder weniger schwanken. Doch wollen wir der Einfachheit wegen davon absehen und annehmen, daß es sich immer um eine Reihe von Beobachtungen annähernd gleichen Umfangs handle.

Aber man darf nicht etwa erwarten, daß das Resultat der Untersuchung immer so ausfällt, wie in dem ersten Beispiel. Findet man Übereinstimmung zwischen Streuung und Erwartungswert oder eine Lexissche Zahl nahe 1, so sagt man, die Statistik weise „normale Dispersion“ auf. Wir haben gesehen, daß dies einen ganz bestimmten Schluß auf das statistische Material zuläßt hinsichtlich des Kollektivs, das ihm zugrunde liegt.

Jetzt will ich einige andere, der statistischen Praxis entstammende Fälle besprechen, in denen die Verhältnisse anders liegen, d. h. die Lexissche Zahl nach oben oder unten von der Einheit stärker abweicht. Wir sprechen dann von übernormaler bzw. unternormaler Dispersion. Unsere Aufgabe wird es in diesen Fällen sein, für diese Abweichungen eine Erklärung zu geben, d. h. Kollektivs zu suchen, denen ein von 1 verschiedener Wert von L zugehört.

Geschlechtsverhältnis der Neugeborenen.

Einen viel durchforschten Gegenstand statistischer Untersuchung bildet das Geschlechtsverhältnis der neugeborenen Kinder. In den 24 Monaten der Jahre 1908 und 1909 sind in Wien insgesamt 93661 Kinder lebend zur Welt gekommen, also durchschnittlich $93661 : 24 = 3903$ in einem Monat. Hiervon waren 48172 oder durchschnittlich 2007 männlichen Geschlechtes, so daß der Knabenanteil oder die „Knabenquote“ für den ganzen Zeitraum $48172 : 93661 = 0{,}51432$ beträgt. Bildet man für die einzelnen Monate die Quotienten: Knabenzahl durch gesamte Geburtenzahl, so schwanken die Zahlen zwischen 0,4990 im März 1909 und 0,5275 im August des gleichen Jahres. Der Reihe nach betrugen die Knabenquoten:

0,5223, 0,5125, 0,5141, 0,5246, 0,5126, 0,5136, 0,5187, 0,5213, 0,5105, 0,5203, 0,5124, 0,5141; 0,5143, 0,5093, 0,4990, 0,5097, 0,5140, 0,5089, 0,5129, 0,5275, 0,5178, 0,5130, 0,5177, 0,5027.

Der Durchschnitt dieser Zahlen ist, wie man leicht ausrechnet, 0,51433, ihre Streuung, die in der vorhin angegebenen Weise zu berechnen ist, beträgt 0,0000533.

Im Sinne der Lexisschen Theorie fragen wir nun nach dem Erwartungswert der Streuung, die als Merkmal innerhalb eines Kollektivs mit folgendem Element erscheint: 24malige Wiederholung einer Gruppe von 3903 Ziehungen aus einer Urne, die unter 1000 Losen rund 514 mit „M“ (masculinum) bezeichnete enthält. Die Formel, die hier gilt, ist von der eben angeführten nur deshalb etwas verschieden, weil wir jetzt nicht die Ereigniszahlen (d. i. die Anzahlen der Knabengeburten) selbst, bzw. ihre Streuung, sondern die Relativzahlen (Knabenquoten) als Merkmale verwandt haben. Dies hat zur Folge, daß der Faktor z in der Formel aus dem Zähler in den Nenner rückt, so daß als Erwartungswert der Streuung

$$\frac{n-1}{n} \cdot \frac{1}{z} p\,(1-p)$$

zu nehmen ist, mit $n = 24$, $z = 3909$, $p = 0{,}514$. Die Ausrechnung ergibt

$$\frac{23}{24} \frac{0{,}514 \cdot 0{,}486}{3903} = 0{,}0000613.$$

Diese Zahl ist, wie man sieht, im Verhältnis $613:533 = 1{,}15$ zu groß gegenüber dem tatsächlichen Ergebnis der Statistik. Die Lexissche Zahl wäre, wenn man von dem Faktor 23/24 absieht, gleich $533:637 = 0{,}835$. Im allgemeinen findet man bei Untersuchungen über die Knabenquote bessere Übereinstimmung und man wird daher hier nach einer Erklärung der etwas „unternormalen“ Streuung suchen. Eine solche findet sich in der Tat leicht unter Zuhilfenahme weiterer wahrscheinlichkeitstheoretischer Untersuchungen. Legt man nämlich der Rechnung statt des bisher betrachteten Kollektivs, dessen Elemente je 24 Gruppen von je 3903 Ziehungen aus einer und derselben Urne (d. h. so viele Geburtsfälle unter gleichbleibenden Verhältnissen) waren, ein etwas komplizierteres zugrunde, bei dem für jede Gruppe von 3903 Ziehungen (bzw. ebensoviel Geburten) mehrere Urnen mit verschiedenen Mischungsverhältnissen (Bevölkerungsteile mit verschiedenen Knabenquoten) zur Verfügung stehen, so zeigt sich — auf Grund eines bekannten algebraischen Satzes —, daß der Erwartungswert der Streuung dadurch *kleiner* wird. Es liegt nun nahe anzunehmen, daß die Knabenquote der Lebendgeborenen eine Eigenschaft der Rasse oder des Volksstammes ist und daß demgemäß in einem Bevölkerungskreis, der eine gewisse *Rassenmischung* aufweist, eine geringere Streuung erwartet werden muß als in einer völlig ungemischten Bevölkerung. Man wird es nicht unplausibel finden, daß diese Erklärung für das Beispiel der Geburtenverhältnisse in Wien angenommen wird.

Todesfallsstatistik mit übernormaler Streuung.

Viel häufiger als mit solchen Fällen einer gegenüber der theoretisch berechneten zu geringen Streuung hat man es in der praktischen Statistik mit Zahlenreihen zu tun, die das Maß der „erwartungsmäßigen“ Streuung beiweitem übertreffen. Sehen wir uns z. B. die Todesfallsstatistik des Deutschen Reiches in dem zehnjährigen Zeitraum von 1877 bis 1886 an, in dem kaum irgendwelche äußere Ereignisse die Gleichförmigkeit der Lebenshaltung störten. Die folgende kleine Tafel gibt in der zweiten Spalte die Zahl der in jedem Jahr eingetretenen Todesfälle, daneben in der letzten die auf 1000 Einwohner entfallende Anzahl. Es starben

im	Jahre	1877	insgesamt	1223156,	auf	je	1000	Einwohner	28,0
,,	,,	1878	,,	1228607,	,,	,,	1000	,,	27,8
,,	,,	1879	,,	1214643,	,,	,,	1000	,,	27,2
,,	,,	1880	,,	1241126,	,,	,,	1000	,,	27,5
,,	,,	1881	,,	1222928,	,,	,,	1000	,,	26,9
,,	,,	1882	,,	1244006,	,,	,,	1000	,,	27,2
,,	,,	1883	,,	1256177,	,,	,,	1000	,,	27,3
,,	,,	1884	,,	1271859,	,,	,,	1000	,,	27,4
,,	,,	1885	,,	1268452,	,,	,,	1000	,,	27,2
,,	,,	1886	,,	1302103,	,,	,,	1000	,,	27,6

Wer ohne alle Kenntnisse in der mathematischen Statistik diese Zahlen, namentlich die letzte Spalte der relativen Anzahlen ansieht, wird sich gewiß über ihre Gleichförmigkeit wundern. Ältere Schriftsteller wissen sich gar nicht genug vor Staunen darüber, wie außerordentlich „stabil" die menschlichen Verhältnisse in der Statistik erscheinen. Rechnet man aber nach, wie große hier die Streuung tatsächlich ist und vergleicht sie im Sinne der Lexisschen Theorie mit ihrem Erwartungswert, so kommt man zu einem ganz anderen Ergebnis. Das arithmetische Mittel der zehn Sterblichkeitszahlen findet sich gleich 27,41 auf 1000, oder als Dezimalbruch geschrieben gleich 0,02741. Bestimmt man jetzt die Abweichungen der einzelnen Beobachtungswerte vom Mittelwert und quadriert sie, also beispielsweise für das erste Jahr: $0{,}028 - 0{,}02741 = 0{,}00059$, $0{,}00059^2 = 0{,}0000003481$ und bildet dann das arithmetische Mittel der so gefundenen zehn Quadratzahlen, so findet man die Streuung gleich 0,0000000949. Diese Zahl scheint ja nun freilich recht klein zu sein.

Aber die Formel von S. 184 für den Erwartungswert der Streuung gibt noch viel weniger, da die im Nenner stehende Zahl z jetzt so groß ist. Offenbar hat man für n den Wert 10, für p die Größe 0,02741 einzusetzen und für z die Zahl der „Versuche" innerhalb jeder der zehn Gruppen von Beobachtungen, also die Bevölkerungszahl des Deutschen Reiches, die in den Jahren 1877 bis 1886 rund 45 Millionen betrug. Die Ausrechnung ergibt

$$\frac{9}{10}\,\frac{0{,}02741 \,.\, 0{,}97259}{45\,000\,000} = 0{,}000\,000\,000\,533,$$

d. i. eine im Verhältnis $949 : 5{,}33 = 177$mal kleinere Zahl als die beobachtete. Die tatsächliche Streuung der Sterblichkeitszahlen übertrifft demnach ihren Erwartungswert beinahe um das Zweihundertfache! Wie ist dies zu erklären?

Solidarität der Fälle.

Überlegen wir einmal, was für ein Kollektiv es ist, mit dem die Lexissche Theorie den Ablauf der Jahressterblichkeit vergleicht. Sein Element besteht aus der zehnmaligen Wiederholung einer Gruppe von 45 Millionen Ziehungen aus einer Urne, die, wie wir sagen können, unter 10000 Losen 274 „schwarze" und $10000 - 274 = 9726$ „weiße" Lose enthält. Wenn zu Beginn jedes Jahres jeder Einwohner Deutschlands aus einer Urne dieser Art sein Lebens- oder Todeslos zu ziehen hätte (selbstverständlich ist angenommen, daß nach jedem Zug das Los zurückgelegt wird, damit das Mischungsverhältnis unverändert bleibt), so müßte man erwarten, daß die Sterblichkeit in zehn Jahren eine 177mal kleinere Streuung aufweist, als sie tatsächlich beobachtet wurde. Nun ist es klar, daß dieses Bild das Spiel von Leben und Tod der Menschen nur sehr unvollkommen wiedergibt.

Schon die einfachste Erfahrung lehrt, daß es viele Erscheinungen gibt, die solidarisch als Todesursachen für eine Vielheit von Menschen wirken, z. B. schlechter Witterungsverlauf in einem Winter- oder Sommermonat, ungünstige Gestaltung der Wirtschaftslage in einem Teilgebiet des Landes, eine endemische Erkrankung usf. Man wird also den wirklichen Verhältnissen viel näher kommen, wenn man annimmt, daß nicht jeder einzelne von den 45 Millionen Einwohnern, unabhängig von allen anderen, sein Los zieht, sondern eine geringere Anzahl von „Repräsentanten" je für eine ganze Gruppe an die Urne herantritt und das Schicksal befragt. Nach der Formel von S. 184 wird der Erwartungswert der Streuung in demselben Maß vergrößert, in dem die Zahl z der unabhängigen Einzelfälle innerhalb einer Gruppe vermindert wird. Wir bekämen somit volle Übereinstimmung zwischen Beobachtung und Erwartung, wenn wir annehmen wollten, daß für je 177 Bewohner des Deutschen Reiches ein gemeinsames Los gezogen wird, das über Leben oder Tod der sämtlichen 177 entscheidet.

Wesentliche Schwankungskomponente.

Ob man diese Erklärung stark übernormaler Streuung durch „Solidarität" der Ereignisse in einem konkreten Fall als zutreffend gelten lassen kann, hängt noch von weiteren Untersuchungen ab. Vor allem wird man verlangen müssen, daß der einmal aus einer Beobachtungsreihe errechnete Umfang der Solidaritätsgruppe einigermaßen erhalten bleibt, wenn man andere analoge Reihen, etwa die Sterblichkeitszahlen eines zweiten Jahrzehnts ins Auge faßt. Das stimmt nun gerade bei dem hier von mir herangezogenen Beispiel der Gesamtsterblichkeit aller Einwohner nur sehr schlecht. Ich will mich nicht mit der Besprechung anderer Zahlenreihen, für die die Erklärung besser zutrifft, aufhalten, sondern lieber die Frage erörtern, in welcher Richtung man sonst noch eine theoretische Begründung für das Auftreten einer wesentlich das Erwartungsmaß übersteigenden

Streuung suchen kann. Man verdankt den Arbeiten von L. v. Bortkiewicz wertvolle Aufklärung über diese Verhältnisse.

Wenn man die Todesfallsstatistik für eine Reihe von Jahren betrachtet, so liegt es nahe anzunehmen, daß die Sterbens-wahrscheinlichkeit von Jahr zu Jahr wechselt, daß also, um im früheren Bilde zu bleiben, das Mischungsverhältnis der schwarzen und weißen Lose in der Urne jedes Jahr ein anderes ist. Nimmt man etwa an, daß im Jahre 1877 die Sterbenswahrscheinlichkeit 0,0280, im darauffolgenden 0,0278 usw. bestand, so kommt man zu dem Schluß, daß in jedem Jahr genau das geschah, was erwartet werden mußte. Allein damit gewinnt man keinerlei vernünftige Einsicht in die Dinge: allzu große Freiheit in den Prämissen nimmt den Schlußfolgerungen jeden Wert, beraubt die Theorie jeglichen Inhalts. Erst wenn wir von einer anderen Seite her auf dieselben Schwankungen der Sterblichkeitswahrscheinlichkeit geführt würden, könnten wir in ihrer Annahme eine Erklärung des Tatbestandes sehen.

Es läßt sich nun wirklich in dieser Richtung etwas erreichen, wenn man nur die wahrscheinlichkeitstheoretische Untersuchung einen Schritt weiter treibt. Betrachtet man nämlich das Kollektiv, dessen Element, wie früher, die Zusammenfassung von n Gruppen von je z Ziehungen ist, wobei aber jetzt für jede Gruppe eine neue Urne mit einem neuen Mischungsverhältnis benutzt wird, so ergibt sich rechnungsgemäß der Erwartungswert der Streuung derart, daß zu der Formel von S. 184 noch ein bestimmtes positives Zusatzglied, noch ein Summand, hinzutritt. Dieser Summand, der oft als die „wesentliche Schwankungskomponente" bezeichnet wird, ist — wie die Rechnung zeigt — nichts anderes als die Streuungsgröße, die man aus den n Mischungsverhältnissen der einzelnen Urnen unmittelbar bilden kann, indem man nämlich zu diesen, untereinander verschiedenen Zahlen den Mittelwert rechnet, dann die Abweichungen der Einzelwerte vom Mittelwert und deren Quadrate bildet, schließlich das Mittel dieser Quadratzahlen nimmt. Die Einzelheiten sind dabei gar nicht wichtig, entscheidend ist nur, daß, was zu dem früher nach der Formel von S. 184 errechneten Ausdruck hinzukommt, ein Betrag ist, der nicht von der Zahl z der Einzelbeobachtungen, also der Ziehungen innerhalb einer Gruppe (bzw. der Bevölkerungszahl), abhängt, sondern nur allein von den Ungleichheiten, von den Schwankungen der Wahrscheinlichkeit von Gruppe zu Gruppe (von Jahr zu Jahr). Damit gewinnt man eine Handhabe, um die Richtigkeit der erweiterten Lexisschen Theorie, d. h. die Zulässigkeit des Vergleiches der untersuchten statistischen Reihe mit dem Kollektiv, das von den Ziehungen aus n verschiedenen Urnen gebildet wird, zu überprüfen.

Man wird vernünftigerweise annehmen müssen, daß, wenn die Sterbenswahrscheinlichkeit im ganzen Reich von Jahr zu Jahr schwankt, etwa infolge wechselnder wirtschaftlicher oder klimatischer Verhältnisse, diese Schwankungen ungefähr in gleichem Maße in den einzelnen größeren Landesteilen zur Geltung kommen. Da nun, wie wir sagten, das Zusatzglied zu dem Erwartungswert der Streuung von der Bevölkerungszahl des Landes unabhängig ist, muß es in gleicher Größe bei der Statistik des ganzen Reiches wie der eines Teiles in Erscheinung treten. Dies liefert eine Kontrolle der Theorie. Man braucht nur den für das Gesamtreich gültigen Zahlen die für einen Teilbereich, z. B. einen Bundesstaat, ermittelten gegenüberzustellen. Es muß dann, wenn die Annahme richtig war, daß die „übernormale Dispersion“ von den Veränderungen der Sterbenswahrscheinlichkeit herrührt, der Überschuß der beobachteten Streuung über die nach der Formel von S. 184 errechnete in beiden Fällen, für das ganze Reich wie für den Einzelstaat, ungefähr gleich groß sein.

Ausdrücklich hebe ich noch hervor, daß bei der jetzt zur Geltung gebrachten Auffassung die Ziffern der gegebenen statistischen Reihe keinesfalls mehr als die aufeinanderfolgenden Merkmalwerte eines Kollektivs aufgefaßt werden; die statistische Aufnahme wird in viel allgemeinerer Weise auf eine Reihe von Kollektivs zurückgeführt.

Selbstmordstatistik.

Die Überprüfung dieser Theorie an dem Material der allgemeinen Sterblichkeit würde Schwierigkeiten begegnen, da hier noch andere Elemente von nichtzufallsartigem Charakter, wie die fortschreitende Verbesserung der Sterblichkeit infolge Hebung der sozialen Lage, mitspielen. Ich will daher ein anderes Beispiel heranziehen, das L. v. Bortkiewicz sorgfältig durchgearbeitet hat, die Selbstmordstatistik Deutschlands für das Jahrzehnt 1902 bis 1911.

In diesen Jahren entfielen im Deutschen Reich auf je eine Million Einwohner durchschnittlich 213,9 Selbstmorde jährlich. Im einzelnen lagen die Zahlen zwischen 204 und 223 pro Million, also in Form von Dezimalbrüchen zwischen 0,000204 und 0,000223. Die Streuung, die sich aus den zehn Zahlen in der schon wiederholt beschriebenen Weise, durch Aufsuchung des Mittels, der Einzelabweichungen und ihrer Quadrate errechnet, beträgt 30,9 Billionstel oder, wie wir zur Abkürzung, um die zehn Nullen hinter dem Komma zu sparen, schreiben wollen, $30{,}9 \cdot 10^{-12}$. Der Erwartungswert der Streuung findet sich dagegen nach der Formel, wenn für z die Einwohnerzahl des Reiches mit 61,6 Millionen, für n der Wert 10, für p der eben angegebene Durchschnittswert 0,0002139 eingesetzt wird, nur zu $3{,}12 \cdot 10^{-12}$, also fast zehnmal kleiner als der beobachtete Wert. Wir stellen nun diesem Tatbestand die Zahlen gegenüber, die für das Großherzogtum Baden ermittelt wurden, ein Teilgebiet des ganzen Reiches, in dem die Verhältnisse hinsichtlich der Selbstmorde denen im Reich sehr ähnlich waren. Es lagen hier die jährlichen Anzahlen zwischen 193 und 225 pro Million, ihr Mittelwert betrug 214,9 und die tatsächliche Streuung rechnet sich zu $126{,}3 \cdot 10^{-12}$, mehr als viermal soviel als für das Gesamtreich. Anderseits ergibt die Formel für den Erwartungswert der Streuung, da die Einwohnerzahl Badens mit 2,04 Millionen nur etwa ein Dreißigstel der des Reiches ausmachte, einen ungefähr dreißigmal so großen Wert, genau (wenn $n = 10$, $z = 2{,}04 \cdot 10^6$ und $p = 0{,}0002149$ eingesetzt wird) $94{,}8 \cdot 10^{-12}$. Hier beträgt somit die beobachtete Streuung nur etwa $^4/_3$mal soviel als die erwartungsmäßige.

Bilden wir aber die beiden Differenzen zwischen tatsächlicher und berechneter Streuung, so haben wir im ersten Fall, für das ganze Reich 30,9 — 3,12 Billionstel, also rund $28 \cdot 10^{-12}$, im zweiten für Baden 126,3 — 94,8 Billionstel, also rund $31 \cdot 10^{-12}$. Man sieht, daß die letzteren Zahlen sehr nahe beieinanderliegen. Damit ist ein Anhaltspunkt dafür gewonnen, daß der große Streuungsüberschuß, den die Selbstmordstatistik des Reiches aufweist, ebenso wie der viel kleinere, der für Baden festgestellt wurde, auf eine „wesentliche Schwankungskomponente“ zurückzuführen ist, d. h. auf eine von Jahr zu Jahr sich verändernde Disposition zum Selbstmord oder, wie wir auch sagen können, eine von Jahr zu Jahr wechselnde Selbstmord-Wahrscheinlichkeit.

Ich will nicht verschweigen, daß die Ergebnisse nicht so günstig liegen, wenn man beliebige andere Teile des Reiches untersucht und daß überhaupt erst ein viel tieferes Eindringen in den besonderen Gegenstand der statistischen Aufnahme und in die mathematische Theorie zu voller Übersicht führt. Nur das eine mag hier noch festgestellt werden, da darüber sehr unrichtige Ansichten verbreitet sind, daß die Selbstmord-Häufigkeit keineswegs, im großen gesehen, einer so starken eindeutig fortschreitenden Veränderung unterworfen ist, wie die allgemeine Sterblichkeit. So ist die für das Jahrzehnt 1902 bis 1911 gültige Zahl von rund 214 Selbstmorden pro Million fast genau im Jahre 1920 wieder verwirklicht und der Durchschnitt im Jahrzehnt 1924 bis 1933 beträgt etwa 265 gegenüber 211 im Jahrfünft 1881 bis 1885. Demgegenüber ist die Gesamt-Sterbeziffer von dem früher besprochenen Wert 274 auf 10.000 im Jahrzent 1877 bis 1886 auf 123 im Jahrzehnt 1922 bis 1931 gesunken, u. zw. von den Kriegsjahren abgesehen, fast ohne Schwankungen.

Soziale und biologische Statistik.

Die Bereiche, in denen statistische Erwägungen angestellt zu werden pflegen, dehnen sich mit der Zeit mehr und mehr aus. Das erste Anwendungsgebiet, das der Statistik in größerem Umfang erschlossen wurde, war wohl das der sozialen Erscheinungen und noch heute gibt es Vertreter der Ansicht, daß nur die Staats- und Wirtschaftswissenschaften „legitimer" Ort für die Anwendung der statistischen Methode seien. Die ursprüngliche Bedeutung des Wortes „Statistik" ist wohl auch „Staatslehre". Sicher hat dieser Teil der Statistik schon lange eine feste Stellung im Kreise der Wissenschaften gewonnen. Wenn nun gerade unter den Vertretern des staatswissenschaftlichen Anwendungsgebietes immer wieder Stimmen laut werden, die meinen, Statistik hätte nichts mit Wahrscheinlichkeitsrechnung zu tun, so darf man das nicht allzu schwer nehmen. Es ist ungefähr dasselbe, wie wenn jemand behauptet, man könne Politiker sein, ohne etwas von der Geschichte seines Landes zu verstehen, oder man könne Brücken bauen, ohne die Gesetze der Statik zu beherrschen. Die klassischen Leistungen auf dem Gebiete der Sozialstatistik sind jedenfalls in anderem Geiste entstanden.

So hat einer ihrer Begründer, Adolphe Quetelet, 1869 seinem grundlegenden Werk, der „Sozialen Physik“, in dem die berühmte Konzeption des „mittleren Menschen“ enthalten ist und das befruchtend auf zwei Generationen von Forschern gewirkt hat, eine ausführliche Abhandlung des Astronomen HERRSCHEL über Wahrscheinlichkeitsrechnung vorangestellt. Stofflich greift QUETELET mehrfach in das Gebiet der Biologie des Menschen, auch mittelbar in das medizinische über. Heute nimmt, wie schon erwähnt, die sogenannte Biometrie, die Lehre von der statistischen Erfassung biologischer Zusammenhänge, namentlich in England einen breiten Raum ein. Es verbinden sich mit diesen Forschungen Hoffnungen ungewöhnlicher Art auf eine Höherzüchtung, eine Veredlung der menschlichen Rasse, wie sie in der Bezeichnung „Eugenik“ zur Andeutung kommen; die weitgespannten Ziele haben die Wissenschaftlichkeit der einschlägigen Untersuchungen nicht immer gefördert. Viel günstiger steht es in dieser Richtung mit einem nahe verwandten Forschungsgebiet, der Vererbungslehre, deren statistische Grundlagen von dem Augustinermönch GREGOR MENDEL um 1870 geschaffen wurden.

MENDELsche Vererbungslehre.

MENDEL erkannte, daß das Auftreten gewisser biologischer Merkmale innerhalb der Individuen einer Generation und auch der Vererbungsvorgang von Generation zu Generation als Kollektiv aufgefaßt werden kann. Hinsichtlich eines jeden, wie man heute sagt, „mendelnden“ Merkmales besitzt das einzelne Individuum zwei Anlagen. Z. B. ist die Blütenfarbe der Erbse „mendelnd“, und zwar eine mendelnde Alternative: rot-weiß; ebenso die Körnerfarbe eine Alternative: grün-gelb. Jede Erbsenpflanze gehört demgemäß hinsichtlich der Blütenfarbe einem der drei Typen rot-rot, rot-weiß oder weiß-weiß an. Bei jedem Fortpflanzungsakt wird von jedem der beiden Elternteile die eine oder andere seiner Anlagen auf den Nachkommen übertragen, wobei jeder der beiden Übertragungsmöglichkeiten die Wahrscheinlichkeit $^1/_2$ zukommt.

Gehören z. B. beide Elternteile dem Mischtypus rot-weiß an, so besteht je die Wahrscheinlichkeit $^1/_4$ für das Auftreten der Kombinationen rot-rot, rot-weiß, weiß-rot, weiß-weiß in der Nachkommenschaft. Dabei geben die beiden mittleren Kombinationen den gleichen Typus, so daß auf diesen 0,50 Wahrscheinlichkeit entfällt. Mitunter ist der Mischtypus äußerlich nicht von dem einen der reinrassigen zu unterscheiden (dominierendes Merkmal), so daß eine theoretische Verteilung $^1/_4 : {^3/_4} = 1 : 3$ entsteht. Tatsächlich beobachtete MENDEL an 8023 Erbsensamen 2001 grüne und 6022 gelbe Körner, BATESON fand unter 15806 Samen 3903 grüne und 11903 gelbe, beides also in bester Übereinstimmung mit der Theorie: Die beobachteten Zahlen stehen einmal im Verhältnis 1:3,01, das anderemal im Verhältnis 1:3,05.

Aus den einfachen Annahmen der MENDELschen Vererbungslehre, die natürlich nicht immer unmittelbar nachprüfbar sind, erwachsen in Verbindung mit den Grundgesetzen der Wahrscheinlichkeitsrechnung die weittragendsten Folgerungen, deren ausgezeichnete Übereinstimmung mit den Beobachtungsergebnissen der Pflanzen- und Tierzüchter eine der schönsten Bestätigungen für die Brauchbarkeit der auf dem Kollektivbegriff aufgebauten Wahrscheinlichkeitstheorie bildet. Die Vererbungsstatistik ist heute eine weit ausgebaute, in ihren Anwendungen überaus erfolgreiche Wissenschaft. Sie bietet der Wahrscheinlichkeitsrechnung viele hübsche, gelegentlich auch recht schwierige Aufgaben, aber keine, die von prinzipieller Bedeutung für deren Grundlegung wären.

Statistik in der Technik.

Ein noch ziemlich neues und wenig bekanntes Anwendungsgebiet wahrscheinlichkeitstheoretischer Untersuchung, das ebenfalls keine grundsätzlichen Schwierigkeiten bietet, findet man in gewissen Aufgaben der industriellen Technik unserer Zeit. Zum Teil handelt es sich da um Fragen, die in den Bereich der Fehlertheorie fallen (von der gleich im folgenden ausführlich die Rede sein soll), beispielsweise wenn die Erzeugnisse einer Fabrik von Stahlkugeln (die für die Verwendung in Kugellagern mit einer Genauigkeit von $^1/_{100}$ mm herzustellen sind) auf ihre Gleichmäßigkeit geprüft werden sollen. Aber auch neuartige Fragestellungen treten auf, die theoretisch und praktisch von großem Interesse sind und die ich etwa als „Probleme der Verkehrsdichte“ kennzeichnen darf.

Ein Elektrizitätswerk muß seine Anlagen nach statistischen Beobachtungen über die Häufigkeit und Intensität der Benutzung durch die angeschlossenen Teilnehmer bemessen. Dabei erscheint die Gesamtheit der Teilnehmer als Kollektiv, etwa mit dem Zeitpunkt und der Dauer des täglichen Stromverbrauches als Merkmal. Vielfältiger und interessanter sind in dieser Richtung die Aufgaben, die bei der Errichtung von Selbstanschlußämtern in Fernsprechnetzen zur Sprache kommen, in einfacherer Gestalt auch schon als „Schalterprobleme". Es wäre natürlich unmöglich oder im höchsten Maße unwirtschaftlich, eine Telephonanlage so zu erstellen, daß sämtliche Verbindungen zwischen je zwei Teilnehmern gleichzeitig bewirkt werden können. Man muß vielmehr auf Grund gewisser Wahrscheinlichkeitsannahmen, die an der Erfahrung geprüft werden, berechnen, welche Gruppen von Verbindungen gerade noch mit einer nicht zu geringen Wahrscheinlichkeit zu erwarten sind. Eine Kombination, für die ein sehr kleiner Wahrscheinlichkeitswert errechnet wird, kann natürlich auch einmal eintreten, aber sie tritt eben sehr selten ein. In ganz unmittelbarer Form erscheint hier die Wahrscheinlichkeit als ein Näherungswert für die relative Häufigkeit — man verläßt sich vollständig darauf und riskiert es, daß in diesen seltenen Augenblicken die Einrichtung versagt.

Ein Beispiel fehlerhafter Statistik.

Das Bild, das ich Ihnen hier von den Anwendungsmöglichkeiten und den Ergebnissen statistischer Forschung zu entwerfen versuche, wäre unvollständig, wollte ich nicht auch der mannigfachen fehlerhaften, ja zum Teil unsinnigen Theorien Erwähnung tun, wie sie, leider gar nicht selten, in der Literatur auftreten. Für uns sind solche Verirrungen, die man an vielen Stellen, besonders häufig auch in medizinischen Werken findet, deshalb lehrreich, weil sie zeigen, wie jedes Verlassen der sicheren Grundlagen, auf denen wir den Wahrscheinlichkeitsbegriff aufgebaut haben, sich rächt. In der Tat kann man fast jeden Fehlschlag und seine Herkunft klar durchschauen, sobald man nur den Kollektivs ein wenig nachgeht, um deren Verteilungen es sich bei den in Rede stehenden Wahrscheinlichkeiten handelt. Ich will hier ein Beispiel etwas genauer besprechen, hinter dem der Name eines hervorragenden Psychiaters, aber auch der eines bekannten Mathematikers steht.

Einem Patienten war beim Rezitieren eines lateinischen

Verses das Wort „aliquis“ entfallen. Aufgefordert, die nächstliegende Assoziation zu „aliquis“ anzugeben, nennt er die Teilung a-liquis. Als nächste Assoziationen zu „liquis“ bringt er: Reliquie, Flüssigkeit, dann hieran anschließend sieben weitere Hinweise, die man einem Vorstellungskomplex, der durch die Worte: Blut—Flüssigkeit angedeutet wird, einigermaßen einordnen kann, schließlich eine zehnte von unbestimmtem Charakter. Aus diesem „statistischen“ Tatbestand, nämlich aus dem Auftreten der Häufigkeit $^9/_{10}$ hinsichtlich der dem Komplex angehörigen Assoziationen, wird der folgende Schluß gezogen: Die Annahme der FREUDschen psychoanalytischen Theorie, daß das Vergessen des Wortes „aliquis“ auf eine Verdrängung durch den genannten Vorstellungskomplex zurückzuführen ist, besitzt eine Wahrscheinlichkeit, die sich von 1 nur um einen Bruch unterscheidet, der hinter dem Komma 25 Nullen besitzt (der also verschwindend klein ist)! Eigentlich wird noch viel mehr gefolgert, nämlich, daß der FREUDschen Theorie an sich diese immense Wahrscheinlichkeit zuzuschreiben wäre. Aber wir wollen dies als ein Versehen gelten lassen, da doch evident ist, daß selbst die volle Sicherheit des Zutreffens einer theoretischen Behauptung in einem Falle statistisch noch nicht das geringste für die Theorie besagt.

Zu dem angeführten Zahlenwert gelangt der betreffende Verfasser in folgender Weise. Er schätzt, daß unter allen „Ideen eines gebildeten Mannes“ durchschnittlich jede tausendste sich mit dem vorhin angedeuteten Vorstellungskomplex in Verbindung bringen läßt. Nun löst er die Aufgabe, die in den Bereich des BERNOULLIschen Problems fällt: Wie groß ist die Wahrscheinlichkeit dafür, aus einer Urne, die auf tausend Kugeln eine schwarze enthält, bei zehnmaligem Ziehen neun schwarze zu erhalten? Dafür ergibt sich in der Tat etwa 0,1 zur 26. Potenz, d. h. ein Dezimalbruch mit 25 Nullen hinter dem Komma. Aber was in aller Welt hat diese ganz richtig gelöste Aufgabe mit der ursprünglichen Frage zu tun?

Richtigstellung.

Offenbar will der Autor das Ergebnis seines statistischen Experimentes darin sehen, daß die Versuchsperson in größerer Zahl Assoziationen, die dem Komplex angehören, gebildet hat als der durchschnittliche „gebildete Mann“. Dann aber hätte er zunächst feststellen müssen, wie häufig im Durchschnitt unter den Assoziationen, die zu dem Wort „liquis“ angegeben werden, solche vorkommen, die in den kritischen Komplex fallen. Es ist kaum anzunehmen, daß ein gebildeter Mensch zu dem Klang „liquis“ nicht sehr bald Reliquie und liquid = flüssig assoziiert und dann eine Weile in diesem Vorstellungskreis weiterspinnt. Der Psychiater verrät sogar, daß er selbst so denkt, denn er bemerkt, daß schon die Trennung in „a“ und „liquis“ auf den kritischen Komplex hinweist. Jedenfalls, so können wir uns ausdrücken, ist schon die Wahl des Ausgangskollektivs eine verkehrte: Nicht darauf kommt es an, wie häufig die kritischen Vorstellungen unter allen möglichen vorkommen, sondern darauf, wie oft sie sich unter denen finden, die von einem unbefangenen Durchschnittsindividuum zu „liquis“ assoziiert werden.

Aber auch die Ableitung des Endkollektivs aus dem als gegeben vorausgesetzten Ausgangskollektiv ist verfehlt. Wenn wir die Ausdrucksweise des Urnenschemas beibehalten, steht es doch so: Beobachtet wurde unter $n = 10$ Ziehungen die Häufigkeit $n_1 : n = a = 0{,}9$ eines schwarzen Zuges; gesetzt, die Wahrscheinlichkeit eines schwarzen Zuges unter normalen Verhältnissen wäre 0,5, so wird z. B. gefragt: Wie groß ist die Wahrscheinlichkeit dafür, daß die Urne, aus der gezogen wurde, ein Füllungsverhältnis x besitzt, das größer als 0,5 ist? Oder in die Sprache des Versuches übersetzt: Wie groß ist die Wahrscheinlichkeit dafür, daß das untersuchte Individuum eine mehr als normale Neigung zeigte, die betreffende Assoziation zu bilden? Die Frage gehört nicht dem Bernoullischen, sondern dem Bayesschen Problemkreis an (d. h. dem Problemkreis, der zum zweiten, nicht zum ersten Gesetz der großen Zahlen führt) und wird beantwortet durch die Bayessche Regel, die etwa 0,95 für die gesuchte Wahrscheinlichkeitsgröße ergibt. Dabei ist angenommen, daß die sogenannte a-priori-Wahrscheinlichkeit, eine Urne mit irgendeinem Füllungsverhältnis x zu ergreifen, für alle x-Werte die gleiche ist. Nur wenn die Versuchszahl n wesentlich größer als 10 wäre, würde der bei der Besprechung des zweiten Gesetzes der großen Zahlen erwähnte Satz, daß das Ergebnis der Rechnung von der sogenannten a-priori-Verteilung unabhängig wird, in Geltung treten. Abgesehen von dieser Unsicherheit liegt natürlich eine weitere in der Annahme 0,5 für die Häufigkeit einer in das kritische Gebiet fallenden Assoziation zu „liquis" beim normalen, unbeeinflußten Menschen. Jedenfalls aber hält sich das auf diese Weise gefundene Ergebnis, das der Zulässigkeit der FREUDschen Annahme für den vorliegenden Fall die Wahrscheinlichkeit 0,95 zuerkennt, in vernünftigen Grenzen gegenüber der geradezu abstrusen Behauptung eines Wertes $1 - 10^{-26}$.

Zusammenfassend ist zu sagen: Wenn überhaupt aus der angeführten Statistik ein Wahrscheinlichkeitsschluß gezogen werden sollte, so war 1. ein anderes Ausgangskollektiv zu wählen; 2. eine einigermaßen begründete Schätzung für die Wahrscheinlichkeit, die wir 0,5 gesetzt haben, zu geben; 3. die BAYESsche statt der BERNOULLIschen Formel zu verwenden; 4. die Zahl der Versuche mit Rücksicht auf die unbekannte „a-priori-Verteilung" bedeutend größer zu machen. Im übrigen wäre es vermutlich richtiger, die Rechnung gar nicht auf die zehn Beobachtungen aufzubauen, sondern als erwiesen anzunehmen, daß das Wort „aliquis" für alle Menschen eine klangliche Verknüpfung mit dem fraglichen Komplex besitzt. Es käme dann darauf an, zu untersuchen oder abzuschätzen, mit welcher Häufigkeit durchschnittlich ein Vergessen des Wortes „aliquis" bei der einen oder anderen Gruppe von Menschen statthat. Hält man eine derartige Feststellung nicht für ausführbar, so muß man eben auf einen statistischen Schluß verzichten, darf ihn aber nicht auf fehlerhaften Wegen zu erzwingen suchen.

Zusammenstellung einiger Ergebnisse.

Ich will die Zahl der Beispiele zur Anwendung des Wahrscheinlichkeitsbegriffes in der Statistik nicht weiter vermehren. Was an Grundsätzlichem zu zeigen war, geht aus dem bisher Erörterten hervor, die systematische Behandlung der verschiedenen Aufgaben ist hier nicht unsere Sache. Bestimmte Fragen, namentlich solche, die mit der sogenannten Wahrscheinlichkeitsnachwirkung zusammenhängen, werden noch im nächsten Vortrag im Rahmen der physikalischen Theorien zur Sprache kommen. Dem Verfahren entsprechend, das in den früheren Vorträgen befolgt wurde, sollen, ehe ich zu einem andern Punkte übergehe, nur noch einige der Grundgedanken, die zur Darstellung gebracht wurden, hier kurz zusammengefaßt werden.

1. Die Zahlenfolge einer statistischen Aufnahme kann in

manchen Fällen unmittelbar als Beobachtungs- (Merkmal-) folge an einem Kollektiv aufgefaßt, in anderen Fällen durch geeignete Überlegungen auf verschiedene Kollektivs zurückgeführt werden.

2. Widersprüche gegenüber der auf dem Kollektivbegriff aufgebauten Wahrscheinlichkeitsrechnung sind bisher nicht festgestellt worden. Weder die MARBEsche Lehre vom statistischen Ausgleich noch die (ihr entgegengesetzte) Knäuelungstheorie STERZINGERS, noch das angebliche „Gesetz der Serie" können einen Widerspruch begründen.

3. In den häufig vorliegenden Fällen, in denen eine statistische Aufnahme aus der Abzählung einer Reihe einfacher Alternativen entstanden ist, leistet die LEXISsche Dispersionstheorie mit ihren verschiedenen späteren Zusätzen die Zurückführung auf typische Kollektivs.

4. Benutzt man den Wahrscheinlichkeitsbegriff und Formeln der Wahrscheinlichkeitsrechnung, ohne sich über die auftretenden Kollektivs und ihre gegenseitigen Beziehungen klar zu sein, so gerät man leicht in die Irre.

Rein beschreibende Statistik.

Ich bemerkte schon in der Einleitung zu diesem Vortrag, daß man den Umfang dessen, was man als Statistik bezeichnen will, verschieden abgrenzen kann. Man wird vielleicht geneigt sein, das, wovon wir bisher gesprochen haben, schlechthin „mathematische Statistik" zu nennen. Da will ich doch nicht unerwähnt lassen, daß man vielfach auch mathematische Hilfsmittel in der Statistik verwendet, ohne sich die Aufgabe zu stellen, die uns hier beschäftigt hat. Man begnügt sich oft damit, sei es als erster Schritt der Untersuchung, sei es unter Verzicht auf jede weitere Forschung, zunächst *ohne Bezugnahme auf den Wahrscheinlichkeitsbegriff* die vorgelegte statistische Aufnahme mathematisch zu erfassen.

Diesem Zwecke dient derjenige Zweig der mathematischen Statistik, den man vielleicht als „*beschreibende*" Statistik bezeichnen kann. Dabei ist „Beschreibung" im engeren Wortsinn als bloße Wiedergabe des unmittelbar Vorgelegten mit gegebenen Mitteln zu verstehen, unter Verzicht auf logische Einordnung in ein systematisches Ganzes. Die Aufgabe der beschreibenden Statistik geht somit dahin, eine vorgegebene statistische Reihe durch typische Merkmale möglichst genau zu kennzeichnen. Die einfachsten solchen Merkmale haben wir bereits kennengelernt: es sind der Mittelwert (Durchschnittswert) der Beobachtungen und die Streuung.

Natürlich reichen Mittelwert und Streuung nicht immer aus, wenn man beliebige statistische Aufnahmen miteinander vergleichen will. Hier setzen die verschiedenen Verfahren der beschreibenden Statistik ein, Verfahren, die in der Hauptsache in der Definition bestimmter statistischer Funktionen bestehen. Eine große Menge solcher „Maßzahlen" sind im Laufe der Zeit in die Statistik eingeführt worden, ich nenne nur den Median- oder Zentralwert, verschiedene „mittlere" Abweichungen, Quartile, Dezile u. a., ohne daß damit wesentlich mehr erreicht worden wäre, als schon der gewöhnliche Mittelwert und die Streuung liefert. Eine für spezielle Fragen sehr nützliche statistische Funktion bildet das „Disparitätsmaß" von GINI. In anderer Richtung hat Karl PEARSON, der Begründer einer großen Schule der Statistik in England, die Hilfsmittel der Beschreibung zu vermehren versucht, indem er gewisse typische Verteilungen in Formeln faßte, auf die nun möglichst alle vorkommenden Fälle zurückgeführt werden sollten. Das vom mathematischen Standpunkt vollkommenste und am weitesten reichende Verfahren ist dasjenige, das im Rahmen der sogenannten Kollektivmaßlehre durch H. BRUNS eingeführt wurde. Auf einer Idee des Astronomen BESSEL fußend zeigte BRUNS, wie man eine beliebige statistische Reihe mit jeder gewünschten Genauigkeit durch eine unbegrenzte Folge von „Maßzahlen" charakterisieren kann. Die ersten in dieser systematisch entwickelten Folge sind die schon früher genannten, Mittelwert und Streuung, dann folgen ein Maß der „Schiefe", ein Maß des „Exzesses" usf. Eine wertvolle Ergänzung dieses Gedankenganges und Übertragung auf andere Fälle verdankt man auch dem schwedischen Astronomen CHARLIER.

Auf diese Dinge näher einzugehen ist hier um so weniger Veranlassung, als es sich ja gerade um den Zweig statistischer Überlegungen handelt, der mit der Wahrscheinlichkeitsrechnung unmittelbar nichts zu tun hat. Es ist nur zur Gesamtorientierung wichtig zu wissen, daß die verschiedenen Hilfsmittel und Verfahren der beschreibenden Statistik, die Aufsuchung statistischer Maßzahlen, die PEARSONschen Verteilungstypen, die BRUNSschen und CHARLIERschen Entwicklungen, als Vorstufe oder als Vorarbeit zur eigentlichen theoretischen Durchdringung des Stoffes aufzufassen sind.

Grundlagen der Fehlertheorie.

Ich will jetzt noch ein paar Worte hinzufügen über denjenigen Zweig wahrscheinlichkeitstheoretisch-statistischer Untersuchungen, der wohl die verbreitetste und unumstrittenste Anwendung gefunden hat und der auch zu den im letzten Vortrag zu behandelnden Problemen der theoretischen Physik überleitet, die sogenannte Fehlertheorie.

Bei fast allen Arten von Beobachtungen, die sich auf die Messung oder Größenbestimmung sinnlich wahrnehmbarer Dinge beziehen, treten unvermeidlich Schwankungen auf, als deren Ursache wir sogenannte „Beobachtungsfehler“ ansehen. Wenn wir mit aller uns überhaupt erreichbaren Genauigkeit die Entfernung zweier auf der Erdoberfläche fixierter Punkte mehrmals ermitteln, so geschieht es, daß wir von einem zum andernmal sich ändernde Werte erhalten. Wir stellen uns nun vor, daß jene Entfernung einen für alle Zeiten, oder wenigstens innerhalb der Zeitgrenzen, die für unsere Messungen in Frage kommen, unveränderlichen, ganz bestimmten „wahren“ Wert besitzt und daß demgemäß unsere Meßergebnisse, höchstens bis auf eines, unrichtig sind, daß sie „Fehler“ aufweisen. An Quellen, aus denen Fehler entspringen können, ist kein Mangel. Haben wir einen Meßstab benutzt und, wie sich gehört, den Einfluß der Temperatur auf die Dehnung seiner Länge berücksichtigt, so kann noch immer eine unmittelbare Wirkung der Sonnenbestrahlung, eine noch nicht genügend erkannte Materialveränderung, eine aus irgendwelchen Gründen entstandene unmerklich kleine Krümmung des Stabes im Spiele sein. Dazu kommen die Ungenauigkeiten der Ablesung im Mikroskop — man weiß, daß zwei Beobachter kaum je die gleichen Teilstriche ablesen —, die Störungen durch die Luftbewegung, durch Erschütterungen, elastische Nachgiebigkeiten, durch Mängel des optischen Materials usf.

Ob wir nun daran festhalten wollen oder nicht, daß es einen genauen „wahren“ Wert gibt, das ist eine erkenntnistheoretische Frage, die wir hier nicht zu entscheiden brauchen. Der Tatbestand ist der, daß die Beobachtungszahlen selbst, wenn wir sie für eine genügend lang fortgeführte Versuchsreihe aufzeichnen, die Eigenschaften eines Kollektivs besitzen. Das ist nun freilich nicht immer unmittelbar nachzuweisen in dem Sinn, daß man wirklich prüft, wie sich die Häufigkeit eines bestimmten Resultates verhält, einerseits bei Vermehrung der Beobachtungen schlechthin, anderseits gegenüber dieser oder jener Stellenauswahl. Aber es geht hier so, wie fast immer in den Anwendungsgebieten mathematischer Theorien: Kann man nicht von vornherein das Zutreffen von Voraussetzungen der Theorie einsehen, so lassen sich dafür Folgerungen, die aus der Theorie gewonnen werden, an der Wirklichkeit überprüfen und bestätigen.

Der große Mathematiker Karl Friedrich Gauss war einer der ersten, der die Anwendbarkeit der Wahrscheinlichkeitslehre auf die Untersuchung der Beobachtungsfehler erkannt hat und darauf einen Lehrgang gründete, der heute unter dem Namen „Ausgleichsrechnung“ oder „Methode der kleinsten Quadrate“ vor allem in der Geodäsie und Astronomie, dann aber auch fast in jeder Naturwissenschaft, die mit Beobachtungen zu tun hat, reichliche Anwendung findet. Die tiefere Grundlage der Ausgleichsrechnung liegt in einem Satze, der schon von Laplace herrührt, wenn er auch in seiner eigentlichen Bedeutung erst viel später erkannt wurde. Der Satz ist wesentlich mathematischer Natur und ich kann ihn hier nicht in seinem vollen Umfang oder mit allen Einzelheiten, die nur den Fachmann angehen, wiedergeben. Aber was er im großen ganzen besagt, ist doch zu wichtig, um im Rahmen unserer Vorträge ganz übergangen zu werden, und es läßt sich, wie ich denke, auch ohne Eingehen auf den mathematischen Formalismus einigermaßen verständlich machen.

Die Fehler, oder besser gesagt, die gegenseitigen Abweichungen einer Reihe von Beobachtungen derselben Größe haben, wie schon angedeutet, vielfältige Ursachen; man wird es plausibel finden, wenn ich sage, daß gerade in dieser Vielfältigkeit der in jedem einzelnen Fall wirksamen Fehlerquellen die Erklärung dafür gesucht werden muß, daß überhaupt eine allgemeine Theorie der Beobachtungsfehler möglich ist, daß eine gewisse Einheitlichkeit in dem Verhalten der resultierenden Beobachtungs-Abweichungen besteht. Auf dem Umstand, daß der Beobachtungsfehler eine Summe vieler einzelner Bestandteile ist, beruht denn auch der allgemeine Satz, von dem die Rede sein soll.

Das Galtonsche Brett.

Eine ausgezeichnete Veranschaulichung dessen, was durch das Zusammenwirken zahlreicher einzelner, unabhängiger Fehlerursachen herbeigeführt wird, gewinnen wir durch ein Experiment, das ich Ihnen hier vorführen kann und das unter dem Namen des Galtonschen Brettes allgemein bekannt ist. Auf einer ebenen, schwach geneigten Unterlage sind in 40 waagrechten Reihen Drahtstifte befestigt, alle parallel zueinander und in gleichmäßigen Abständen von 8 mm. Zwischen je zwei waagrechten Reihen solcher Stifte ist ebenfalls eine Entfernung von ungefähr der gleichen Größe eingehalten. Wichtig ist dabei, daß die Stifte in je zwei aufeinanderfolgenden Reihen um 4 mm versetzt sind, so daß alle Stifte z. B. der zehnten Reihe genau über bzw. unter der Mitte je zweier Stifte der elften bzw. der neunten Reihe liegen. Unwesentlich ist, daß die ersten Reihen nicht voll besetzt sind, sondern an den äußeren Rändern des Brettes oben Stifte fehlen.

Nun lasse ich ganz oben, in der Mitte des Brettes eine Stahlkugel von 8 mm Durchmesser so herabrollen, daß sie gerade auf den mittelsten Stift der obersten Reihe trifft und sich nun entscheiden muß, ob sie rechts oder links von ihm weiterrollen will. In jedem Fall trifft sie sofort einen Stift der zweiten Reihe und muß sich wieder einen Weg rechts oder links wählen. Diese Entscheidung und manche folgende ist rascher getroffen, als ich von ihr sprechen kann. Schon hat die Kugel den Weg durch alle 40 Reihen zurückgelegt, wobei sie jedesmal an einen Stift anschlug und dann nach einem unmerklich kurzen Zögern rechts oder links einen Ausweg fand. Wie oft sich die Kugel in ihrem Lauf für rechts, wie oft für links entschieden hat, darüber kann man hinterher eine bestimmte Aussage machen. Denn die Ruhelage, die sie jetzt unten, um vier Stiftabstände nach rechts aus der Mitte verschoben, einnimmt, zeigt, daß die Zahl der Entscheidungen für rechts um acht größer gewesen sein muß, als die für links. Der waagrechte Abstand des Kugelmittelpunktes von der Mittellinie des Brettes nimmt nämlich bei jedem Schritt, d. h. beim jedesmaligen Durchlaufen einer Reihe um die halbe Größe der Stiftteilung zu oder ab, je nachdem die eine oder andere Wahl getroffen wurde. Da im ganzen 40 Reihen vorhanden sind, müssen also 16 Entscheidungen für links und 24 für rechts gefallen sein.

Der ganze Vorgang ist ein Modell für die vierzigmalige Wiederholung einer Alternative mit den Merkmalen $+\,{}^1/_2$ („rechts“) und $-\,{}^1/_2$ („links“), wobei als Resultat der Wiederholung oder als Merkmal des durch vierzigfache Verbindung entstandenen Kollektivs nur die Summe der Einzelmerkmale erscheint. Die Summe der 24 Summanden $+\,{}^1/_2$ und der 16 Summanden $-\,{}^1/_2$ ist eben die Zahl 4, die die Lage der Kugel am Ende angibt.

Die Glockenkurve.

Denken wir uns nun, wir hätten irgendeine physikalische Beobachtung oder Messung vorgenommen, bei der 40 voneinander unabhängige Fehlerquellen wirksam sind, und jede einzelne von ihnen gebe Veranlassung zu einem positiven oder negativen Fehler, dessen Betrag eine halbe Einheit irgendeines Maßsystems ausmacht. Der resultierende Fehler einer Beobachtung entspricht dann genau dem, was in der Endlage unserer Stahlkugel am unteren Rande des GALTONschen Brettes zum Ausdruck kommt. So viel Stiftabstände die Kugel nach rechts oder links aus der Mitte liegt, um so viele Einheiten ist das Beobachtungsergebnis zu groß bzw. zu klein. Wollen wir nun sehen, was eintreten wird, wenn wir nicht nur eine Beobachtung machen, sondern ihrer eine ganze Anzahl, so müssen wir nur ebensoviele Kugeln hintereinander das Brett hinunterrollen lassen. Ich habe hier 400 ganz gleiche, exakt gearbeitete Stahlkugeln und Sie sehen, wie sie sich, eine nach der andern, den Weg zwischen den Stiften suchen. Es dauert nur wenige Minuten, und alle Kugeln haben ihre Zickzackbahn durchlaufen und ruhen in schönen Säulen am unteren Rand des Brettes, jede in der Abteilung, in die sie nach der letzten, in der vierzigsten Reihe vorgenommenen Entscheidung gefallen ist. Mit einem einzigen Blick ist das Gesamtergebnis der vierhundertmaligen Wiederholung des Einzelversuches (der selbst aus 40 unabhängigen Versuchen besteht) zu überschauen: Man sieht, daß der größte Teil der Kugeln in den mittelsten Fächern liegt, dort, wo die Gesamtabweichung aus der Mitte nicht mehr als ein bis zwei Teilungseinheiten beträgt, und daß die Kugelzahl nach beiden Seiten hin ungefähr symmetrisch abfällt. Schon aus weiter Entfernung ist die charakteristische Gestalt der „Glockenkurve“ erkennbar, die die Verteilung der Kugeln auf die verschiedenen Abstände, zugleich die Verteilung innerhalb des durch vierzigfache Verbindung entstandenen Kollektivs darstellt.

LAPLACE war der erste, der für den vorliegenden Fall der Verbindung einer großen Zahl von untereinander gleichen Alternativen die Verteilung richtig berechnete und so theoretisch die Gestalt der Glockenkurve fand, die man heute auch oft die GAUSSsche Kurve nennt. Es ist im Grunde genommen die gleiche Aufgabe wie die von BERNOULLI behandelte, die wir vorhin in der Form kennengelernt haben, daß nach der Anzahl der Kopfwürfe innerhalb einer großen Zahl von Würfen mit einer Münze gefragt wird. Während aber BERNOULLI in dem Gesetz der großen Zahlen (wie es dann von POISSON formuliert wurde) nur eine besondere Eigenschaft der resultierenden Verteilung hervorhob, gelang es LAPLACE, die Verteilung vollständig zu berechnen. Sie wird mathematisch dargestellt durch die sogenannte „e hoch minus x^2-Funktion“, d. h. die Ordinaten der Glockenkurven nehmen von der Mitte nach außen derart ab, daß ihre (negativ genommenen) Logarithmen sich wie die Quadrate der Entfernung von der Mitte verhalten. Dabei ist zu beachten, daß es nicht nur eine Glockenkurve gibt, sondern unendlich viele, die sich durch ihre größere oder geringere Schlankheit unterscheiden. Für den Fall des GALTONschen Brettes läßt sich auch das Maß der Schlankheit oder, wie man zu sagen pflegt, das Präzisionsmaß — das wesentlich von der Zahl der Nägelreihen abhängt — genau berechnen. Die Rechnung benutzt lediglich die Regeln der Addition und der Multiplikation von Wahrscheinlichkeiten, die wir für die Mischung und Verbindung von Kollektivs als gültig erkannt haben.

Das LAPLACEsche Gesetz.

Der allgemeine Satz, auf dem die GAUSSsche Fehlertheorie beruht, ist nun der, daß die Glockenkurve, die bei den Kugeln des GALTONschen Brettes so sichtbar in Erscheinung tritt, immer dann die resultierende Verteilung wiedergibt, wenn das Kollektiv, um das es sich handelt, durch Verbindung und Summenbildung aus sehr vielen Ausgangskollektivs entstanden ist. Es ist nicht notwendig, daß die einzelnen ursprünglichen

Kollektivs einfache Alternativen sind, auch nicht, daß sie alle die gleichen Merkmalwerte und die gleiche Verteilung besitzen. Wenn wir nur das Endkollektiv dadurch bilden, daß wir eine sehr große Zahl von Ausgangskollektivs „verbinden" und dann so „mischen", daß als endgültiges Merkmal die Summe der ursprünglichen Merkmale übrig bleibt, so ist die Endverteilung stets genau von der Beschaffenheit der Glockenkurve.

Man sieht, wie dieser Satz in die Theorie der Beobachtungsfehler eingreift. Macht man nur die Annahme, daß bei jeder Beobachtung sehr zahlreiche Fehlerquellen zusammenwirken, so folgt daraus schon, daß die Wahrscheinlichkeit einer bestimmten Größe des Gesamtfehlers durch die Ordinate einer entsprechenden Glockenkurve wiedergegeben wird. In diesem Sinne gibt es ein bestimmtes „Fehlergesetz", dem die zufallsartigen Beobachtungsfehler folgen und das den Ausgangspunkt der Gaussschen Fehlertheorie und Ausgleichungsrechnung bildet. Alle Formeln, Rechnungsverfahren usw., die man unter dem Namen „Methode der kleinsten Quadrate" zusammenfaßt, sind Folgerungen aus der speziellen Form des Fehlergesetzes, sie fließen also in letzter Linie aus der Annahme, daß die Beobachtungsfehler sehr viele verschiedene Einzelursachen besitzen. Natürlich sind da noch viele, auch grundsätzliche Fragen offen, die ich hier nicht näher behandeln kann, z. B. die Frage, wie das „Präzisionsmaß", das für jede Beobachtungsreihe einen bestimmten Wert haben muß, zu bestimmen ist, und anderes.

Die Anwendungsgebiete der Fehlertheorie.

Gauss hat seine Theorie vor allem auf geodätische und astronomische Messungen angewandt und in diesen Gebieten bilden die auf sie gestützten Rechnungsverfahren, die übrigens vor Gauss schon großenteils Legendre bekannt waren, ein dem Praktiker unentbehrliches Hilfsmittel. Die Theorie bewährt sich aber weit darüber hinaus auch in vielen Fällen, wo von eigentlichen „Fehlern" einer Beobachtung nicht gesprochen werden kann, sondern nur von Schwankungen oder von gegenseitigen Abweichungen der Ergebnisse.

Mißt man z. B. an einer sehr großen Zahl erwachsener Personen die Körperlänge, so erhält man Werte, die, wenn sie etwa auf halbe Zentimeter abgerundet werden, kaum noch Meßfehler im eigentlichen Sinne des Wortes enthalten. Die Abweichungen zwischen den Messungsergebnissen untereinander können aber als zufallsartige Schwankungen angesehen werden, deren Zustandekommen man einer Reihe verschiedenartiger Ursachen zuschreiben wird. Es zeigt sich nun, wenn man die relative Häufigkeit, mit der die einzelnen Längenwerte auftreten, näher untersucht, daß sie ebenfalls dem Gesetz der Glockenkurve folgen. Das ist nicht überraschend, da wir ja gesehen haben, daß diese Form der Verteilung immer resultieren muß, wenn das betrachtete Merkmal als Summe vieler unabhängig voneinander zufallsartig variierender Einzelwerte entsteht. Übrigens gestatten die neueren Ergebnisse, von denen im letzten Vortrag die Rede war, den Satz dahin auszudehnen, daß es sich nicht nur um Summe oder Durchschnitt, sondern um beliebige statistische Funktionen handeln darf. Das Gausssche Gesetz findet so in weiten Bereichen biologischer und ähnlicher Untersuchungen Anwendung, ohne daß es sich da um Beobachtungsfehler im ursprünglichen Sinne handeln würde.

Anderseits darf man den Geltungsbereich der Fehlertheorie nicht, wie das oft geschehen ist, übertreiben. Nicht alle Schwankungen, die es irgendwo gibt, folgen dem Gesetz der Glockenkurve, und nichts wäre verkehrter, als anzunehmen, daß nur dort zufallsartige Abweichungen von einem Mittelwert vorliegen, wo die Verteilung diesem Gesetz entspricht. Die schon erwähnte englische Schule der „Biometriker“, die von Karl Pearson begründet wurde, legt mit Recht großen Nachdruck darauf, daß das Gausssche Gesetz nicht aller Weisheit letzter Schluß in der Statistik ist. Und wenn man auch manchmal in den Untersuchungen, die aus dem Pearsonschen Kreise stammen, eine tiefere wahrscheinlichkeitstheoretische Begründung vermißt, so darf man keinesfalls verkennen, wie sehr sie durch ihre freiere Auffassung vom Wesen einer statistischen Verteilung dem Fortschritt der biologischen Statistik gedient haben.

Mit diesen Hinweisen möchte ich meine Ausführungen über die Anwendung der Wahrscheinlichkeitstheorie in der Statistik beschließen. Auf gewisse Fragen, die zu den Grundlagen der Fehlertheorie in Beziehung stehen, komme ich noch in dem letzten Vortrag zurück, der dem interessanten Gebiet der statistischen Physik und den damit zusammenhängenden Problemen der Kausalität gewidmet sein wird.

Sechster Vortrag.

Probleme der physikalischen Statistik.

Nun habe ich nur noch von einem Anwendungsbereich der Wahrscheinlichkeitsrechnung ausführlicher zu sprechen, der etwa seit einem halben Jahrhundert erschlossen ist, in unserer Zeit stetig steigende Bedeutung erlangt und gerade in grundsätzlicher Hinsicht das größte Interesse in Anspruch nehmen muß. Ich meine damit die Rolle, die der Wahrscheinlichkeitsbegriff heute in der theoretischen Physik spielt, eine Rolle, die geschaffen wurde durch den genialen Gedanken Boltzmanns, einem der wichtigsten Sätze innerhalb des physikalischen Lehrgebäudes, dem sogenannten zweiten Hauptsatz der Wärmetheorie, die Form einer Wahrscheinlichkeitsaussage zu geben. Ich brauche, wenn ich Ihnen die Verhältnisse, um die es sich hier handelt, näherbringen soll, nichts an besonderen physikalischen Kenntnissen vorauszusetzen. Es ist sogar für den ersten Augenblick besser, gar nicht den stofflichen Inhalt des Satzes, auch wenn man ihn voll versteht, ins Auge zu fassen, sondern mehr die logischen Formen, in denen er ausgesprochen wird, zu beachten.

Der zweite Hauptsatz der Thermodynamik.

Die klassische Thermodynamik, die von Robert Meyer, von Joule und vor allem von Carnot begründet wurde, hatte zu der Erkenntnis geführt, daß bei thermischen Vorgängen neben dem, was man Energie nennt, noch eine zweite Variable, eine Zustandsgröße, die sogenannte Entropie, in Frage kommt, und daß für sie der Satz gilt, der eben als zweiter Hauptsatz neben den ersten, den Satz von der Konstanz der Energie, tritt: Bei jedem beobachtbaren Vorgang wird die Entropie vermehrt. Seit den berühmten Arbeiten Boltzmanns aber, die bis zum Jahre 1866 zurückreichen, heißt es: Es ist mit außerordentlich großer Wahrscheinlichkeit zu erwarten, daß der Entropiewert zunimmt; eine irgendwie erhebliche Abnahme der Entropie besitzt außerordentlich geringe Wahrscheinlichkeit.

Übersetzen wir diesen Satz in eine Sprache, die das Wort „Wahrscheinlichkeit“ durch seine Definition ersetzt, so erhalten wir die Aussage: Wird der gleiche physikalische Vorgang unbeschränkt oft wiederholt, so wird sich mit sehr großer relativer

Häufigkeit der Fall ereignen, daß die Entropie wächst, und nur äußerst selten der entgegengesetzte. Freilich bedarf hier noch mancher Punkt näherer Erklärung, vor allem, was man unter „gleichen“ Vorgängen zu verstehen hat, wie die Größe der Wahrscheinlichkeit mit der Größe der erwarteten Zu- oder Abnahme der Entropie verknüpft ist usf.

Kein Zweifel kann aber darüber bestehen, daß der Sinn der Aussage der ist, etwas über die Häufigkeit des einen oder anderen Versuchsausganges zu behaupten. Die Gleichsetzung von Wahrscheinlichkeit und relativer Häufigkeit bei unbeschränkter Wiederholung des Versuches wird hier wohl kaum einem Einwand begegnen. Der Physiker Smoluchowski, dem man viele wertvolle Beiträge zur physikalischen Statistik verdankt, sagt einmal in diesem Zusammenhange: „Mathematische Wahrscheinlichkeit, das ist die relative Häufigkeit des Eintrittes bestimmter auffallender Ereignisse.“ Dies ist, wenn auch etwas ungenau und mit einer belanglosen Einschränkung versehen, unsere Häufigkeitsdefinition der Wahrscheinlichkeit, die wir der klassischen Gleichmöglichkeitsdefinition gegenübergestellt haben. Unter den Physikern besteht kaum eine Meinungsverschiedenheit darüber, daß Wahrscheinlichkeit als relative Häufigkeit zu deuten sei, wenn man auch gewöhnlich jedes nähere Eingehen auf diese Frage unterläßt.

Determinismus und Wahrscheinlichkeit.

Was ist nun das Neue, was erscheint oder erschien als das Revolutionierende an der BOLTZMANNschen Auffassung? Die Frage ist leicht zu beantworten, wenn man sich überlegt, welche Vorstellung wir uns seit Jahrhunderten von sogenannten Naturgesetzen gebildet haben. „Alle Körper fallen mit gleicher Beschleunigung zu Boden“, sagte GALILEI. „Jeder Lichtstrahl wird beim Eintritt in ein dichteres Medium zum Lot hin gebrochen“, lautet ein Teil des SNELLIUSschen Brechungsgesetzes. „Zwei in gleicher Weise elektrisch geladene Kügelchen stoßen einander ab“, ist eine Teilaussage des COULOMBschen Anziehungsgesetzes. In allen diesen Fällen wird das Eintreten einer bestimmten Erscheinung unter genau bekannten Voraussetzungen mit vollster Sicherheit vorausgesagt. Es heißt nicht „die meisten Körper fallen“ oder „die Körper fallen fast immer“ zur Erde, die Aussage ist eine determinierte, die in keiner Weise (so scheint es wenigstens) mit dem Wahrscheinlichkeitsbegriff in Verbindung gebracht werden kann. „Nach ewigen, ehernen, großen Gesetzen müssen wir alle unseres Daseins Kreise vollenden", in diesen Worten drückt GOETHE die Überzeugung von der eindeutigen Bestimmtheit des Naturgeschehens aus.

Das erste Beispiel einer deterministischen Erklärung eines umfassenden Tatsachengebietes gab die NEWTONsche Mechanik, wie sie 1687 in den „Philosophiae naturalis principia mathematica" niedergelegt wurde. Fast zweihundert Jahre lang konnte man sich eine Naturerklärung nicht in anderer Form vorstellen als in der der NEWTONschen Prinzipien — in diesem Punkte stimmt auch die Auffassung GOETHES mit der seines bitter bekämpften Gegners NEWTON überein. Die extremste Formulierung fand der NEWTONsche Determinismus in der LAPLACEschen Idee des „mathematischen Dämons", der imstande sein soll, allein vermöge seiner unbeschränkten Fähigkeit zur mathematischen Deduktion den Ablauf aller Vorgänge in der Welt vorauszusagen, sobald ihm alle den augenblicklichen Zustand charakterisierenden Größen bekannt wären. Die Auffassung, daß etwas Derartiges zumindest für die ganze unbelebte Natur, vielleicht auch noch für die Pflanzen- und Tierwelt möglich sein müsse — eventuell mit dem Zusatz, daß auch die weiter zurückliegende Vergangenheit mit von Einfluß sein kann — ist allmählich zur festen Überzeugung aller Naturforscher geworden. Nur dem menschlichen Handeln blieb, im Einklang mit der religiösen Lehre vom freien Willen, nach der Meinung der Mehrzahl, eine gewisse Wahlfreiheit vorbehalten. So fährt auch GOETHE in dem eben angeführten Gedichte fort: „Nur allein der Mensch vermag das Unmögliche: er unterscheidet, wählet und richtet." Von der Ausnahmestellung des Menschen sind wir, zumindest gefühlsmäßig, durchaus überzeugt, und dies beeinflußt auch unsere Ansicht über die Möglichkeit „zufallsartigen" Geschehens in entscheidender Weise.

Es erscheint uns keineswegs verwunderlich, daß die Zahl der Todesfälle oder der Selbstmorde in einem bestimmten Lande sich nicht exakt für die nächsten Jahre vorausberechnen läßt, etwa in der Art, wie man den Zeitpunkt jeder Mondesfinsternis genau vorhersagen kann. Denn in jenen beiden Erscheinungen tritt mittelbar oder unmittelbar der menschliche freie Wille ins Spiel, etwa hinsichtlich der Lebensführung, der Wahl des Aufenthaltsortes, des Eingehens von Gefahren usf. Auch wenn wir in allen unseren Entscheidungen durch äußere Umstände mitbestimmt werden, halten wir doch diese Bestimmung nicht für eine zwingend eindeutige. Im Gegenteil haben die älteren Autoren, wie der englische Historiker BUCKLE, über nichts so sehr gestaunt wie über die starke Gleichförmigkeit der statistischen Zahlen und die daraus entspringende Möglichkeit einer Voraussage in nicht allzu weiten Schranken.

Auch bei den Glücksspielen, deren Ablauf doch den Vorgängen in der unbelebten Natur viel näher steht, ist das Dazwischentreten einer freien Willenshandlung erkennbar. Wir heben den Würfelbecher mit der Hand, führen eine Schüttelbewegung aus usf.; das Lotterielos wird von dem Waisenknaben oder dem Verwaltungsbeamten aus dem Glücksrad gezogen und es bleibt bis zum letzten Augenblick in jeglichem Sinne ungewiß, welches Röllchen seine Hand erfassen wird. Daß derartige Vorgänge oder Erscheinungen „zufallsartig" verlaufen, d. h. von den Gesetzen der Wahrscheinlichkeitstheorie beherrscht werden, ist eine Erkenntnis, an die wir uns schon seit langem mehr oder weniger gewöhnt haben.

Aber die eigentlichen mechanischen oder physikalischen Prozesse, in die keine Menschenhand eingreift, die Vorgänge in der unbelebten Natur, die weisen wohl einen grundsätzlich anderen Charakter auf? Ein mechanischer Webstuhl vollführt seine verwickelten Bewegungen mit unfehlbarer Regelmäßigkeit, und was er leistet, das Erzeugnis seiner Webkunst, ist Stück für Stück unveränderlich das gleiche? Wenn wir im Sinne der sogenannten kinetischen Gastheorie annehmen, daß das, was man ein Gas nennt, nichts anderes ist als ein ungeheurer Haufen von im einzelnen unsichtbar kleinen, gegeneinander bewegten Atomen, so scheint es uns (oder schien uns noch vor kurzem) denknotwendig zu sein, daß diese Bewegungen gesetzmäßig, in vollkommen eindeutiger Weise vorausbestimmt, vor sich gehen. Unser intellektuelles Gewissen sträubt sich dagegen, anzunehmen, daß Zufall oder „Zufallsgesetze" derartige Vorgänge beherrschen können.

Zufallsmechanismen.

Aber dieser Widerstand, dieses Sträuben des Intellekts, erweist sich, wie das so oft der Fall ist, bei näherer Betrachtung als ein Vorurteil, das sich in keiner Weise aufrechterhalten läßt, wenn man der Sache auf den Grund geht. Worin besteht denn der Unterschied zwischen dem „rein mechanischen" System der in einem abgeschlossenen Gefäß gegeneinander bewegten Atome oder Molekel und dem Mechanismus eines Glücksspiels? Nehmen wir als Beispiel ein bekanntes, seit Jahrhunderten erprobtes Verfahren, „reine" Zufallsergebnisse zu erzielen, die Lotterieziehung. Der wesentliche Teil des Spieles besteht in folgendem. Es werden die Nummern von 1 bis, sagen wir, 100000 auf einzelne, untereinander gleiche Zettel gedruckt, die Zettel möglichst gleichförmig zu Röllchen geschlossen, diese in einen großen Behälter geschüttet, der ständig in drehender oder andersartiger Bewegung erhalten wird; schließlich greift eine Hand in den Behälter und zieht willkürlich ein Röllchen heraus. Das, worauf in der Hauptsache geachtet werden muß, ist, daß bei der Bewegung des Behälters die Röllchen in unübersichtlicher Weise durcheinandergerüttelt werden; wie das einzelne Röllchen am Schlusse ergriffen und aus dem Behälter entfernt wird, erscheint ohne Bedeutung, wenn nur dem Ziehenden jede Einsichtnahme in etwaige Unterschiede der Röllchen entzogen ist. Aber gerade diese Bedingung wird besonders vollständig erfüllt, wenn überhaupt kein intelligentes Wesen das Herausnehmen besorgt, sondern irgendeine mechanische Einrichtung getroffen ist, die bewirkt, daß nach genügender Durchrüttelung des Behälters ein einzelnes Röllchen, etwa durch eine trichterförmige Öffnung, herausfällt. In der Tat ist schon oft, zur Vermeidung von Unterschleif, also zur vollständigen Sicherstellung des „Zufalls", vorgeschlagen worden, den gsamten Ziehungsvorgang in dieser oder anderer Weise zu „mechanisieren".

Daß dies möglich ist, unterliegt keinem Zweifel. Eine derartige, durch einen Elektromotor getriebene Ziehungsmaschine würde dies leisten: an einer Stelle des Apparates wird ein langes Papierband eingeführt, das automatisch bedruckt, zerschnitten und gerollt, in Form von hunderttausend gleichen Röllchen in einen Behälter fällt; dieser wird hinreichend stark gerüttelt und gibt schließlich ein Röllchen durch eine vorgesehene Öffnung als den Träger des Gewinnes (bzw. des ersten Gewinnes) nach außen ab. Damit haben wir eine, bis auf den motorischen Antrieb ganz in sich geschlossene, mechanische Einrichtung vor uns, die, sofern man sie mehrmals in Tätigkeit setzt, keineswegs stets das gleiche Ergebnis zeitigt, sondern, wie wir völlig überzeugt sind, durchaus nach Zufallsgesetzen, d. h. nach Art eines Kollektivs, bald die eine, bald die andere Losnummer als die gewinnende ausweisen wird.

Man kann sich auch einen noch einfacheren automatisierten Zufallsmechanismus hergestellt denken, indem man das schon erwähnte Bajazzospiel (bei dem eine zwischen Nägeln, wie beim GALTONschen Brett, herabfallende Kugel aufgefangen werden soll) durch einen Motor ergänzt, der das Hinaufheben der herabgefallenen Kugel (das jetzt der Spieler durch Drehen eines Knopfes leistet) mechanisch bewirkt; der Auffangbecher müßte natürlich an irgendeiner Stelle feststehen, damit — nach dem früher Ausgeführten — überhaupt ein Glücksspiel und kein Geschicklichkeitsspiel vorliegt.

Die zufallsartigen Schwankungen.

Hat man nun einmal erkannt, daß ein automatischer Mechanismus zufallsartig schwankende Resultate ergeben kann, so liegt kein Grund mehr vor, die analoge Annahme für die Gasmolekel abzulehnen. Aber man kann in der Überlegung noch einen Schritt weitergehen. Wir haben vorhin davon gesprochen, daß eine Maschine von der Art eines mechanischen Webstuhls nach der gängigen Auffassung vollkommen zwangsläufig ein Webstück genau gleich dem andern herstellt. Ist denn diese Ansicht überhaupt richtig? Wenn wir mit etwas schärferen Mitteln als dem freien Auge zwei Erzeugnisse desselben Webstuhles betrachten, so finden wir zweifellos — auch wenn wir vorerst der Einfachheit halber von Fehlern in der Beschaffenheit der Kette und der Einschußfäden absehen — kleinere und größere Abweichungen in der Maschenweite, in den Winkeln, die die Fäden miteinander bilden usf. Ja, wir haben an einer Reihe von gewebten Stücken gleich eine mehrfache Mannigfaltigkeit von Schwankungen vor uns, indem wir einmal die Maschenunterschiede, die nebeneinander an einem Stück bestehen, ein andermal die Unterschiede betrachten können, die an den entsprechenden Stellen der verschiedenen Stücke auftreten. Und solche Abweichungen, solche Fehler gibt es bei jedem, auch bei dem exaktesten maschinellen Arbeitsvorgang.

Als wir von den Anwendungen der Statistik in der Industrie sprachen, erwähnte ich, daß man sich bei der Überprüfung der für Kugellager bestimmten Stahlkugeln mit Vorteil der Gaussschen Fehlertheorie bedient. Hier wird also auf das Resultat eines der vollkommensten mechanischen Arbeitsprozesse, den die moderne Technik kennt, eine Theorie angewandt, die sich auf dem rationellen Kollektivbegriff aufbaut, obgleich jenes Resultat, nämlich die ganz automatisch hergestellte Stahlkugel, nach den Grundgesetzen der Mechanik völlig determiniert sein müßte. Gerade das, was die deterministische Auffassung der Naturvorgänge als das Allereindeutigste ansehen muß, das Werk einer exakten Maschine, erweist sich bei näherem Zusehen als etwas zufallsartig Schwankendes. Gibt es dann überhaupt noch einen Wiederholungsvorgang, der wirklich genau durch die Vorhersage der Mechanik oder der analogen physikalischen Theorie gedeckt wird? Es wäre müßig, darüber zu streiten, ob z. B. bei einem mathematischen Pendel ein Ausschlag genau gleich dem andern verläuft. Haben wir eine Anordnung gefunden, bei der sich Abweichungen nicht feststellen lassen, so ist es nur vernünftig, anzunehmen, daß genauere Meßmethoden uns doch noch welche zeigen werden und — nach allen bisherigen Erfahrungen — daß diese Abweichungen die Eigenschaften eines Kollektivs besitzen, vielleicht auch den Annahmen der Fehlertheorie gehorchen dürften.

Kleine Ursachen, große Wirkungen.

Ein Unterschied freilich zwischen dem Verhalten der Ziehungsmaschine und den eben erwähnten „Ungenauigkeiten“ in der Arbeit beliebiger Werkmaschinen tritt auffallend zutage. Man kann im letzten Falle das wesentliche Ergebnis des mechanischen Prozesses voraussagen (bzw. es ist die Maschine im Hinblick auf dieses wesentliche Ergebnis gebaut) und nur kleine Abweichungen der Einzelresultate untereinander unterliegen den Zufallsgesetzen. Bei der automatischen Ziehungsvorrichtung aber steht es so, daß alles, worauf es ankommt, durch den Zufall entschieden wird. In dieser Formulierung liegt noch etwas Subjektives, nur durch den Verwendungszweck, durch die Absicht, der die Maschine dienen soll, Bedingtes. Von einer Maschine, die dazu bestimmt ist, kleine Kistchen zu nageln und bei der die einzelnen Nägel aus einem Behälter in der gleichen Weise wie bei der vorhin entworfenen Ziehungsmaschine in zufallsartiger Folge zur Wirkung kommen, sagen wir wohl nicht, daß sie wesentlich Zufallsergebnisse zeitigt, obwohl hier die mechanischen Verhältnisse nicht andere sind als bei der Ziehungsvorrichtung.

Man kann, vom rein mechanischen Standpunkt aus, ohne auf den Zweck, der mit einer Einrichtung verfolgt wird, Rücksicht zu nehmen, mag es sich also um die gleichgültige Auswahl der Kistennägel oder die bedeutungsvolle der Lotterieröllchen handeln, eine charakteristische Eigenschaft „eigentlicher" Zufallsmaschinen feststellen: Bei ihnen weist der Bewegungs- oder Arbeitsverlauf von einem Mal zum andern Unterschiede auf, die nicht den Typus kleiner Abweichungen oder Ungenauigkeiten besitzen, sondern ein bestimmtes Maß nicht unterschreiten. So ist es bei zwei Lotterieziehungen das Mindeste, daß die als „erste Treffer" erscheinenden Losnummern sich um eine Einheit unterscheiden, wenn sie nicht gerade zusammenfallen. Dagegen kann der Durchmesser einer maschinell erzeugten Stahlkugel jeden noch so kleinen Fehler gegenüber dem beabsichtigten Normalmaß aufweisen, und Fehler oberhalb einer gewissen geringen Größe kommen praktisch überhaupt nicht vor. Bedenkt man nun, daß der Bewegungsvorgang in jedem Mechanismus zu Beginn wenigstens annähernd determiniert verläuft — bei der Ziehungsmaschine z. B. bis zum Einrollen der bedruckten Zettel — und während dieser Teilperiode höchstens kleine Schwankungen aufweisen kann; so gelangt man dazu, ein ausschlaggebendes Merkmal der „Zufallsmaschinen" von folgender Art herauszuheben: Bei ihnen entwickeln sich aus anfänglich nur kleinen, vielleicht verschwindend kleinen Ungenauigkeiten im Verlauf des Prozesses große, endliche Unterschiede im mechanischen Ergebnis.

Das Schlagwort, mit dem wir, dem Physiker M. v. Smoluchowski folgend, diese Verhältnisse am besten kennzeichnen können, lautet: Kleine Ursachen, große Wirkungen. Ich will nicht so weit gehen, wie SMOLUCHOWSKI und teilweise vor ihm schon POINCARÉ es getan hat, und ein solches Mißverhältnis zwischen Wirkung und Ursache geradezu als die entscheidende Eigenschaft jeder Massenerscheinung, auf die die Wahrscheinlichkeitsrechnung anwendbar sein soll, ansehen. Aber für die „Zufallsmechanismen", d. h. für solche Einrichtungen, die menschlichen Eingriffen entzogen, automatisch verlaufende Wiederholungsvorgänge vom Typus eines Kollektivs mit endlich verschiedenen (arithmetischen) Merkmalen aufweisen, für diese trifft das Kennzeichen zweifellos zu. Die gerollten Lose fallen in der Reihenfolge, in der sie gedruckt wurden, in den Behälter und zeigen in diesem, solange er nicht bewegt wird, noch eine bestimmte Anordnung, die sich bei Wiederholung des ganzen Prozesses ungefähr reproduziert. Wird jetzt der Behälter zwangsläufig in Bewegung versetzt, so bewirken die kleinen, schon vorhandenen Unterschiede der Lage, daß die „Anfangsbedingungen" des einsetzenden Bewegungsvorganges von einem Mal zum andern nicht ganz die gleichen sind. Aus den kleinen Abweichungen der Anfangslage entwickeln sich nun im langen Verlauf der Rüttelbewegung so erhebliche Unterschiede in der späteren Konfiguration, daß das erste schließlich herausfallende Los jedesmal ein ganz anderes wird.

Noch viel durchsichtiger ist der Vorgang beim GALTONschen Brett oder, was für uns dasselbe ist, beim Bajazzospiel. Trifft die Kugel auf den ersten Stift auf, so entscheidet eine ganz unmerklich geringe Abweichung nach rechts oder links darüber, ob sie um eine halbe Teilung nach der einen oder der andern Seite verschoben wird. Beim zweiten Stift, der getroffen wird, wiederholt sich diese Sachlage, wobei man durchaus nicht annehmen muß, daß nur Schwankungen der ursprünglichen Anfangsbedingungen weiterwirken. Es ist vielmehr plausibel, daß fortdauernd neue Einflüsse von außen, Luftbewegungen, Erschütterungen usf. sich geltend machen, die alle das Gemeinsame haben, daß sie einzeln sehr „kleine Ursachen" darstellen, aber in ihrer Gesamtheit über ein „Entweder-Oder" entscheiden, über das Weiterrollen der Kugel rechts oder links von der Aufstoßstelle.

Kinetische Gastheorie.

Es ist nicht schwer, die Überlegungen, die wir an die „Zufallsmechanismen" geknüpft haben, in der Richtung zu ergänzen, daß sie uns Einblick in das Wesen der Vorgänge gewähren, die man in der physikalischen Statistik zu untersuchen pflegt. Als das älteste Beispiel eines solchen, das in gewissem Sinn für fast alle anderen Fälle charakteristisch ist, betrachten wir das Bild, das die sogenannte kinetische Gastheorie von einem „idealen“ Gase entwirft. Es wird hier bekanntlich angenommen, daß die Gasmasse nicht stetig den Raum erfüllt, wie es die rohe Beobachtung vermuten läßt, sondern daß der „gaserfüllte“ Raum mit enormer Geschwindigkeit von sehr vielen, sehr kleinen Atomen oder Molekülen durchschwirrt wird, die in kurzen Zeitabständen aneinanderstoßen und so unaufhörlich aus ihrer sonst geradlinig verlaufenden Bahn abgelenkt werden.

Betrachtet man als Merkmal eines solchen Molekels seine augenblickliche Geschwindigkeit, oder genauer die drei Geschwindigkeitskomponenten in den Richtungen eines Koordinatenkreuzes, so erweist es sich als zweckmäßig, d. h. als mit den Beobachtungen gut übereinstimmend, die Gesamtheit der Molekel als Kollektiv anzusehen und die Regeln der Wahrscheinlichkeitsrechnung auf sie anzuwenden. Dabei ist man, wenn man Übereinstimmung mit der Erfahrung erzwingen will, nicht mehr frei in der Festsetzung der Abmessungen, der Anzahl der in einer Volumeinheit befindlichen Moleküle sowie der Ausgangsverteilungen. Es ist für uns wichtig, zunächst einige Zahlen kennenzulernen.

Nach den üblichen Annahmen, zu denen, wie gesagt, die Anpassung an die Erscheinungen zwingt, enthält unter durchschnittlichen Verhältnissen ein Kubikmillimeter nicht weniger als 30000 Billionen Gasmolekel; ihre Größe ist derart, daß drei Millionen nebeneinander gelegt die Länge von 1 mm ergeben; die durchschnittliche Geschwindigkeit beträgt einige hundert Meter in der Sekunde; nach einem Weg von etwa einem Zehntausendstel Millimeter wird das Molekel durch Zusammenstoß abgelenkt; ungefähr fünf Milliarden solcher Zusammenstöße eines bestimmten Moleküles mit anderen sind auf eine Sekunde zu rechnen. Man sieht, daß hier die Anwendung der Ausdrücke „sehr viel“, „sehr klein“ usw. im Sinne des üblichen Sprachgebrauches schon gerechtfertigt ist. Auch wird es verständlich, daß die Regeln der Wahrscheinlichkeitsrechnung, die eigentlich für unendlich große Wiederholungszahl gelten, sich gut bewähren, da man es hier mit so enormen Elementenzahlen zu tun hat wie sonst nie.

Wie steht es aber hier mit dem Schlagwort „kleine Ursachen, große Wirkungen“? In welcher Weise tritt hier dieses entscheidende Kennzeichen des Zufallsmechanismus auf? Nun, um das einzusehen, müssen wir nur noch eine Zahlengröße aus den früheren Angaben berechnen. Der Durchmesser des Molekels beträgt, wie ich sagte, ein Dreimillionstel Millimeter, die „freie Weglänge“, d. i. die durchschnittlich ohne Zusammenstoß durchlaufene Strecke, etwa ein Zehntausendstel, demnach 300mal so viel. Die geometrischen Verhältnisse sind also die gleichen, wie wenn Kugeln von 1 cm Durchmesser nach durchschnittlichen Weglängen von 3 m aneinanderprallen. Es ist dann einleuchtend, daß eine verschwindend kleine Abweichung in der Flugrichtung einer Kugel die Wirkung des Zusammenpralles entscheidend beeinflußt. Bewegt sich z. B. eine 1-cm-Kugel zentrisch auf eine ebenso große zu, die wir uns im Abstand von 3 m augenblicklich festgehalten denken wollen, so wird sie durch den elastischen Aufprall geradewegs zurückgeworfen, kehrt also an ihren Ausgangspunkt zurück. Wenn aber die Bewegungsrichtung nur um etwas weniger als neun Bogenminuten (d. i. 0,0004 des Kreisumfanges) von der geraden Verbindungslinie der Kugelmittelpunkte abweicht, so folgt aus den Stoßgesetzen, daß der Rückprall unter 45° geschieht, so daß die bewegte Kugel auf dem Rückweg in 3 m Entfernung an ihrem Ausgangspunkt vorbeiläuft. Bei 16 Bogenminuten Abweichung würde die Kugel schon im rechten Winkel gegen ihre Ankunftsrichtung abgelenkt und wenn die Bahnrichtung um 23 Bogenminuten, also um weniger als einen halben Grad, von der ursprünglich angenommenen abweicht, würde gar kein Stoß mehr erfolgen, sondern die Kugel ungestört weiter fliegen.

Die eben ausgesprochenen Sätze folgen allerdings nur aus der ältesten, noch von Boltzmann vertretenen, seither aber fallen gelassenen Annahmen, daß die Molekel vollkommene Kugelgestalt haben und sich beim Stoße wie rein elastische Körper verhalten. Aber sie sind doch sehr geeignet, uns einen Einblick zu gewähren in die große Empfindlichkeit des Bewegungsablaufes gegenüber kleinen Ungenauigkeiten. Bedenkt man, daß nach neueren, allerdings auch schon großenteils überholten, Vorstellungen jedes einzelne Gasmolekel eine größere oder geringere Zahl überaus rasch gegeneinander bewegter Elektronen umfaßt, so wird sich der Eindruck nur verstärken, daß das Resultat eines Zusammenstoßes in außerordentlichem Maße von kleinen und kleinsten Ursachen abhängt. Denn wenn die Elektronen den Atomkern mit Geschwindigkeiten umkreisen, die gegenüber den Fluggeschwindigkeiten des ganzen Molekels ungeheuer groß sind, so muß die geringste Variation in der Größe oder Richtung der Fluggeschwindigkeit die relative Lage der Atomteile im Augenblick des Zusammenstoßes von Grund auf verändern. Kurz, man kann hier mit noch weit mehr Recht als bei den meisten Glücksspielmechanismen von „kleinen Ursachen und großen Wirkungen" sprechen.

Der sich selbst überlassene Haufen von bewegten Gasmolekeln erscheint in dieser Weise nicht grundsätzlich verschieden von der automatischen Ziehungsmaschine oder dem GALTONschen Brett (Bajazzospiel) mit selbsttätiger Hebevorrichtung für die herabgefallenen Kugeln. Daß es eine auf dem Wahrscheinlichkeitsbegriffe aufgebaute kinetische Gastheorie gibt, die mit der Erfahrung gut übereinstimmende Resultate liefert, ist nicht um ein Haar wunderbarer als die Anwendbarkeit der Wahrscheinlichkeitsrechnung auf irgendwelche Glücksspielprobleme. Im Gegenteil, die besonders gute Übereinstimmung wird plausibler zufolge der beiden schon erwähnten Umstände: der außerordentlich großen Zahl von Elementen, mit denen man es in der Gastheorie zu tun hat, und des überaus stark ausgeprägten Mißverhältnisses zwischen den die Verschiedenheiten des Stoßeffektes herbeiführenden Ursachen und ihren Wirkungen. Wir sind jedenfalls nicht berechtigt, einseitig in dem einen Fall, dem der Gastheorie, einen Widerspruch gegen das Kausalitätsgesetz oder auch nur eine besondere Belastung unseres Bedürfnisses nach kausaler Erklärung der physischen Erscheinungen zu erblicken.

Größenordnung der „Unwahrscheinlichkeit".

Vielleicht werden Sie aber den Einwand machen, daß der Annahme einer statistischen Fassung für ein bestimmtes physikalisches Grundgesetz nicht nur intellektuelle Bedenken, sondern auch rein praktische gegenüberstehen. Wie gewinnt man denn jemals die Überzeugung von der Richtigkeit einer derartigen Aussage? Ist es nicht so, daß die physikalische Beobachtung entscheidend dazu drängt zu erklären, die Entropie nehme immer zu bezw. sie nehme unter den und den Umständen sicher zu? Der nicht näher Orientierte wird vermuten, daß es sozusagen eine Tatsachenfrage ist, ob man mit einem sicheren Zunehmen der Entropie zu rechnen habe oder nicht, und daß im letzteren Fall sich auf Grund der Erfahrung zumindest Bedingungen werden angeben lassen, unter denen die Zunahme zu erwarten ist. Um die hier vorliegenden Verhältnisse aufzuklären, muß ich noch ein wenig näher auf den Entropiebegriff eingehen und dabei noch einmal auf die früher schon mitgeteilten „großen Zahlen“ der Gasmolekel usw. zurückkommen.

Nach den Vorstellungen der kinetischen Gastheorie befinden sich, wie schon vorhin angegeben, in einem von Luft erfüllten Würfel, der 1 cm Seitenlänge besitzt, 30 Trillionen lebhaft durcheinander schwirrende Molekel, die in jedem Augenblick den Raum von 1 ccm mehr oder weniger gleichförmig erfüllen. Die Entropie ist nun — ebenfalls im Rahmen der kinetischen Theorie betrachtet — im wesentlichen ein Maß für die Gleichförmigkeit dieser Verteilung. Auf jeden einzelnen der 1000 kleinsten Würfelchen von der Größe eines Kubikmillimeters, in die der ganze betrachtete Würfel zerfällt, kommen im Durchschnitt 30000 Millionen Molekel. Man schreibt nun nach Boltzmann der Zustandsgröße, die Entropie heißt, einen höheren Wert zu, wenn die tatsächliche Verteilung der Molekel in höherem Maße der gleichförmigen Verteilung, die gerade ein Tausendstel der Gesamtzahl in jeden Kubikmillimeter verweist, entspricht, und einen geringeren, wenn eine auffällige Ungleichheit vorhanden ist, z. B. wenn ein kleines Würfelchen von einem Kubikmillimeter nur die Hälfte der auf ihn durchschnittlich entfallenden Molekelzahl enthält. Dabei ist natürlich noch auf einige Nebenumstände zu achten. So ist vorausgesetzt, daß innerhalb des ganzen Raumes, den wir ins Auge fassen, also in unserem Beispiel innerhalb des Kubikzentimeters, keinerlei merkliche physikalische Ungleichmäßigkeiten bestehen, keine Druck- oder Temperaturdifferenzen zwischen einzelnen Punkten, wie sie in einem größeren Raum, natürlich mannigfaltig festgestellt werden könnten. Ferner ist zu bemerken, daß zur genauen Abgrenzung des Entropiebegriffes bzw. zur Bestimmung der Größe der Entropie in einem konkreten Falle nicht nur die örtliche Verteilung der

Molekel, sondern auch ihr Geschwindigkeitszustand heranzuziehen wäre, was aber an den grundsätzlichen Zusammenhängen nichts ändert.

Sieht man von den Geschwindigkeiten ab, so kann man sagen, alle Erfahrung weise darauf hin, daß eine solche Abnormität, wie die Verminderung der Molekelzahl in einem Kubikmillimeter auf die Hälfte (bei entsprechender, geringer Vermehrung in den übrigen 999 Würfelchen) niemals eintritt. Die statistische Theorie schließt nun allerdings die Möglichkeit dieser Erscheinung nicht vollständig aus, sondern erklärt sie nur für äußerst unwahrscheinlich. Man muß aber das Maß der Unwahrscheinlichkeit kennen, das die Theorie angibt. Es ist mir leider ganz unmöglich, die Zahl hier anzuschreiben, die den von der Theorie gelieferten Wahrscheinlichkeitsbruch zum Ausdruck bringt. Die Zahl ist nämlich so beschaffen, daß hinter dem Komma erst einmal hundert Millionen Nullen stehen, bevor die erste von Null verschiedene Ziffer kommt!

Eine derart kleine Größe entzieht sich jeder Vorstellung, wir wollen daher unser Beispiel etwas milder gestalten. Fragen wir nur nach der Wahrscheinlichkeit dafür, daß die Zahl der Molekel in dem äußersten Winkel des betrachteten Kubikzentimeters, sagen wir etwa um wenigstens ein Zehntausendstel nach oben oder unten von der auf den einzelnen Kubikmillimeter entfallenden Durchschnittszahl 30000 Millionen abweicht, daß sie also um höchstens drei Millionen zu groß oder zu klein ist. Dieser Wahrscheinlichkeitsbruch läßt sich auch noch nicht in gewöhnlicher Form anschreiben, er hat noch mehr als 60 Nullen hinter dem Komma, aber man kann sich schon ganz gut einen Begriff von dem Ausmaß dieser Wahrscheinlichkeit machen. Spielt man nämlich zehnmal hintereinander bei einer Lotterie mit, bei der jedesmal eine Million Lose ausgegeben werden, so besitzt man ungefähr die angegebene Wahrscheinlichkeit dafür, alle zehnmal den Haupttreffer zu gewinnen. Jeder normal Denkende wird finden, daß zwischen diesem Grad von „Unwahrscheinlichkeit“ und der vollkommenen Unmöglichkeit kaum noch ein praktisch feststellbarer Unterschied besteht.

So sieht man also, daß durch die Erfahrung, durch die Beobachtung, eine Entscheidung für oder gegen die statistische Auffassung nicht gegeben wird. Die Behauptungen, zu denen die statistische Theorie führt, unterscheiden sich praktisch nicht von deterministischen.

Kritik der Gastheorie.

Ich habe hier einige einfache Rechnungen, die an die sogenannte kinetische Gastheorie anknüpfen, nur deshalb ein wenig ausgeführt, weil diese Theorie das älteste und sicher bekannteste Beispiel einer Anwendung des Wahrscheinlichkeitsbegriffes in der Physik ist. Aber das Beispiel ist aus mehrfachen Gründen nicht ganz zureichend. Einer dieser Gründe ist schon erwähnt worden: Man glaubt heute nicht mehr an die einfachen, kugelförmigen Moleküle von Clausius, Maxwell und Boltzmann, auch die vor etwa zwei Jahrzehnten entstandene Vorstellung von den Atomen als kleinen Planetensystemen, die auf Rutherford und Niels Bohr zurückgeht, ist schon in mancher Hinsicht umstritten und hat einer viel allgemeineren, von Ernst Mach vorausgeahnten Auffassung, wonach man den Atomen weder Ort noch Zeit im gewöhnlichen Sinn zuschreiben darf, weichen müsssen. Aber für uns ist noch wesentlicher der Umstand, daß man im Bereiche der alten Gastheorie eigentlich niemals zu einer restlosen Klärung der begrifflichen Grundlagen, gerade in Hinsicht auf die statistischen Fragen gelangt ist. Boltzmann selbst stand noch mit einem Fuß in dem Anschauungskreis der deterministischen Mechanik und er versuchte die statistische Auffassung des zweiten Hauptsatzes zu vereinigen mit einer Beschreibung der Stoßvorgänge zwischen den Gasmolekeln, die ganz in den Gedankengängen der Newtonschen Mechanik verlief. Es war der Standpunkt Boltzmanns, der von der Mehrzahl seiner Zeitgenossen geteilt wurde, daß einerseits die Erscheinungen „im Großen", d. h. die, die an einer ausgedehnten Gasmasse beobachtet werden, statistischen Gesetzmäßigkeiten genügen, während die Vorgänge „im Kleinen", d. h. die Bewegungen der einzelnen Moleküle den Gesetzen der klassischen Mechanik folgen. Mit Recht wandte Ernst Mach dagegen ein, daß aus den mechanischen Gesetzen niemals ein Verhalten, wie es der zweite Hauptsatz der Thermodynamik fordert, gefolgert werden könne.

Die Machsche Kritik an der Boltzmannschen Ableitung des zweiten Hauptsatzes ist so wichtig und sie ist leider so oft in ganz falschem Sinn dargestellt worden, daß ich nicht unterlassen kann, darauf etwas näher einzugehen. MACH findet, daß ein „Analogon der Entropievermehrung in einem rein elastischen System aus absolut elastischen Atomen nicht existiert"; daß ein Verhalten der Atome gemäß dem zweiten Hauptsatz aus den mechanischen Gesetzen sich nicht nachweisen lasse. Er sagt: „Wie könnte auch ein absolut konservatives System elastischer Atome durch die geschicktesten mathematischen Betrachtungen, die ihm doch nichts anhaben können, dazu gebracht werden, sich wie ein nach einem Endzustand strebendes System zu verhalten?" Diese Einwände, über die man sich vor zwei bis drei Jahrzehnten, als die kinetische Gastheorie in hohem Ansehen bei den Physikern stand, meinte hinwegsetzen zu können, treffen genau den Kern der Sache. In der Tat ist es niemals gelungen und auch in keiner Weise möglich, aus den Gesetzen der Mechanik einen Bewegungsverlauf abzuleiten, der dem entspricht, was der zweite Hauptsatz fordert: einer Bevorzugung derjenigen Zustände, die als „ungeordnetere" gelten können, wie die gleichförmige Verteilung der Molekel über die tausend einzelnen Millimeterwürfel des früheren Beispieles.

Was ich bisher ausgeführt habe, nöchte ich jetzt wie folgt zusammenfassen. Erstens die BOLTZMANNsche Wendung, an Stelle determinierter Naturgesetze in bestimmten Fällen statistische treten zu lassen, steht in vollem Einklang mit unserer Auffassung der Glücksspiele und der auf dieser Auffassung aufgebauten Häufigkeitstheorie der Wahrscheinlichkeit. Zweitens eine experimentelle Entscheidung für oder gegen die statistische Fassung des Entropiegesetzes ist nicht möglich angesichts der enorm geringen Wahrscheinlichkeit, die sie dem von der deterministischen Theorie ausgeschlossenen Fall zuweist. Endlich drittens die statistische Auffassung im großen ist nicht vereinbar mit Determinismus im kleinen, man kann statistische Aussagen nicht aus den Differentialgleichungen der klassischen Physik herleiten.

Die neueste Entwicklung der Atomphysik, auf die ich am Schlusse dieses Vortrages zu sprechen komme, hat diesem Standpunkt, den ich seit langem vertrete, namentlich in dem an dritter Stelle genannten Punkte, durchaus recht gegeben. Jetzt wollen wir noch einige konkrete Aufgaben der statistischen Physik etwas näher verfolgen, bevor wir auf die allgemeinen Fragen zurückgreifen.

BROWNsche Bewegung.

Vor etwa hundert Jahren entdeckte der englische Botaniker BROWN in gewissen organischen Flüssigkeiten unter dem Mikroskop sichtbare kleine Körperchen, die eine merkwürdige, ruhelose Zickzackbewegung vollführten. Inzwischen hat man gefunden, daß sich diese „BROWNsche Bewegung" unter den mannigfachsten Umständen beobachten läßt und eine den Gesetzen der Wahrscheinlichkeitsrechnung folgende Massenerscheinung bildet. Wir können, da es uns nur auf das Grundsätzliche ankommt, die Vorstellung dahin vereinfachen, daß wir annehmen, die Bewegung der Partikelchen erfolge ausschließlich innerhalb einer waagrechten Ebene. Wir sehen also von Auf- und Niedersteigen ab oder betrachten lediglich die Projektion der Zickzackbewegung auf die Horizontalebene.

Denken wir uns etwa den Boden eines Gefäßes mit einem regelmäßigen Netz von rechtwinklig zueinander stehenden Geraden überzogen, so daß er in eine große Anzahl, sagen wir N kleine Quadrate zerfällt, so können wir ein Kollektiv folgendermaßen festlegen: Element ist die Beobachtung eines BROWNschen Teilchens, Merkmal die Nummer des Quadrates, in dem es sich befindet. Man erzielt gute Ergebnisse, wenn man annimmt, daß jede der hier in Frage kommenden Wahrscheinlichkeiten den gleichen Wert, also $1 : N$ besitzt. Die Annahme läßt sich in der Weise prüfen, daß man aus dem angegebenen Ausgangskollektiv ein anderes ableitet, dessen Element die Beobachtung einer Anzahl von, sagen wir n Partikelchen ist, mit den n Nummern ihrer Aufenthaltsquadrate als Merkmal und dann daraus durch Mischung eine einfache Alternative herstellt z. B. mit dem Merkmal: In dem Quadrat Nr. 25 befinden sich drei Partikel oder nicht drei Partikel. Macht man eine Reihe von mikrophotographischen Aufnahmen der ganzen Emulsion, wobei das Liniennetz auf der photographischen Platte angebracht ist, so kann man zählen, mit welcher relativen Häufigkeit das eben bezeichnete Merkmal auftritt. Die Übereinstimmung mit der Rechnung erweist sich dabei als sehr befriedigend. In einer Versuchsreihe von SVEDBERG, die noch näher besprochen werden wird, betrug die aus der mittleren Teilchenzahl 1,54 berechnete Wahrscheinlichkeit für den Aufenthalt von genau 3 Teilchen in einem

Quadrat 0,130. Tatsächlich beobachtet wurde bei 518 Zählungen eine relative Häufigkeit der Anzahl 3 gleich 0,133, nämlich 69mal auf 518. Um die Voraussetzungen der Rechnung genau zu erfüllen, müßte man allerdings vor jeder Aufnahme die Teilchen gewaltsam durcheinandermengen, damit die einzelnen Beobachtungen „unabhängig“ werden — so wie man es beim Kartenmischen oder bei der Lotterieziehung macht. Daß die Ergebnisse der Zählung auch ohne ein solches Mischen, ohne Unterbrechung des natürlichen Bewegungsablaufes, mit der rechnungsmäßigen Erwartung gut übereinstimmen, das ist eine sehr bemerkenswerte Tatsache, über die ich jetzt noch Näheres sagen muß.

Der zeitliche Ablauf.

Denn das eigentliche und stärkste Interesse in einer Frage der physikalischen Statistik richtet sich nicht auf das räumliche Nebeneinander der Erscheinungen, sondern auf das Nacheinander, auf den zeitlichen Ablauf der Ereignisse. Auch in der Gastheorie interessierte uns vor allem die zeitliche Veränderung der Entropie. Man betrachtet bei der BROWNschen Bewegung unter dem Mikroskop vorzugsweise ein bestimmtes quadratisches Feld des Netzes und zählt die Partikelchen, die von Sekunde zu Sekunde oder in sonst geeigneten Intervallen in ihm zu finden sind. Wie lassen sich die Regeln der Wahrscheinlichkeitsrechnung auf diese Beobachtungsreihe anwenden? Bilden etwa die aufeinanderfolgenden Partikelzahlen die Ergebnisreihe der Elemente eines Kollektivs, das sich aus dem früher genannten Ausgangskollektiv ableiten läßt? Lassen sie sich überhaupt als ein Kollektiv auffassen?

Eine einfache Überlegung zeigt schon, daß das kaum der Fall sein kann. Denken wir etwa an das Auftreten einer bestimmten Teilchenzahl, z. B. 7, in einem der Quadrate. Wenn wir eine genügend lange Zeitfolge untersuchen, wird sich gewiß eine relative Häufigkeit der Beobachtungszahl 7 herausstellen, von der man annehmen kann, daß sie einem Grenzwert zustrebt. Berücksichtigen wir aber nicht nur jedesmal den augenblicklichen Zustand, sondern auch das, was im unmittelbar vorausgegangenen Zeitelement im gleichen Feld geschehen ist, so wird sich zweifellos finden, daß die Zahl 7 auf die Zahl 6 oder 8 viel öfter folgt als etwa auf die Zahl 1. Das zweite Axiom, die Regellosigkeit der Merkmalreihe, trifft also auf die zeitliche Folge der Partikelzahlen in einem bestimmten, ins Auge gefaßten Felde gar nicht zu. Dies rührt natürlich, physikalisch gesehen, daher, daß die einzelnen Teilchen innerhalb des Zeitintervalles von einer Beobachtung zur nächsten doch nur eine beschränkte Bewegung machen, so daß kleine Veränderungen der Teilchenzahlen eher zu erwarten sind als größere. Vom Standpunkt der Wahrscheinlichkeitstheorie aus müssen wir sagen, daß die zeitlich aufeinanderfolgenden Teilchenzahlen zwar kein Kollektiv bilden, aber in irgendeiner Weise mit dem Begriff des Kollektivs in Zusammenhang gebracht werden können. Wir haben schon im vorhergehenden Vortrag, im Rahmen der allgemeinen Statistik von solchen Fällen der Zurückführung auf Kollektivs gesprochen. Es ist nicht schwer, einzusehen, wie man im vorliegenden Fall vorgehen muß; ich will es zunächst an dem einfachsten Beispiel eines Glücksspieles, an dem Kopf-und-Adlerspiel, erläutern.

Wahrscheinlichkeits-Nachwirkung.

Denken wir uns die Ergebnisse einer Alternative durch die Zahlen 0 und 1 bezeichnet, so ergibt eine Beobachtungsfolge eine aus Nullen und Einsern bestehende Zahlenreihe, die alle Eigenschaften eines Kollektivs aufweist. Insbesondere kommt die Zahl 1 hinter einer 0 ebenso häufig vor wie hinter einer 1. Bilden wir aber jetzt die Summe aus je zwei aufeinanderfolgenden Ergebniszahlen, und zwar derart, daß wir das erste und zweite, dann das zweite und dritte, dann das dritte und vierte Ergebnis usf. addieren, so entsteht aus der Beobachtungsfolge

$$1\,0\,1\,1\,1\,0\,1\,0\,0\,1\,1\,0\,0\,0\,1\,0\,1\,..$$

die Summenfolge

$$1\,1\,2\,2\,1\,1\,1\,0\,1\,2\,1\,0\,0\,1\,1\,1\,..,$$

die natürlich Nullen, Einser und Zweier enthält. Es wird niemanden überraschen, daß diese Folge ebenso Grenzwerte der relativen Häufigkeiten aufweist, wie die ursprüngliche. Aber man erkennt doch hier anderseits, daß die Regellosigkeit nicht mehr besteht: denn unmöglich kann die Zahl 2 auf eine 0 oder die Zahl 0 auf eine 2 folgen. Eine 2 an fünfter Stelle der Summenfolge besagt nämlich, daß das fünfte und sechste Ergebnis der ursprünglichen Reihe 1 war, eine 0 an sechster Stelle der Summenfolge sagt, daß der sechste und siebente Versuch 0 ergeben hat, was offenbar unvereinbar ist: die Zahl, die in der Ausgangsreihe an sechster Stelle steht, kann nicht gleichzeitig 0 und 1 sein. Wir haben hier das Schulbeispiel einer Zahlenreihe vor uns, die aus einem Kollektiv, aus der Beobachtung einer zufallsartigen Erscheinungsreihe gewonnen wurde, die auch noch zum Teil die Eigenschaften eines Kollektivs besitzt, aber nicht mehr die charakteristische Eigenschaft der (vollen) Regellosigkeit aufweist. Die mathematische Analyse wird nun folgendermaßen dieses Vorkommnisses Herr.

Man betrachtet vorerst ein aus dem ursprünglichen Kollektiv abgeleitetes, dessen Element die Zusammenfassung von je drei ursprünglichen Elementen ist, und zwar ohne Übergreifungen, also des ersten bis dritten, des vierten bis sechsten usf. Als zweidimensionales Merkmal des Elementes wählen wir die beiden Summen der Beobachtungsergebnisse, nämlich die des jeweils ersten und zweiten, sowie die des zweiten und dritten innerhalb der Dreiergruppe. Durch Anwendung der geläufigen Regeln erhält man die Wahrscheinlichkeit dafür, daß in der betrachteten Elementenfolge die Kombination 0-1 oder 1-0 usw. auftritt; die Kombinationen 0-2 und 2-0 sind nach dem früher Gesagten unmöglich. Ein zweites Kollektiv, das zu den analogen Rechnungen führt, erhält man bei Zusammenfassung der ursprünglichen Elemente mit den Nummern 2 bis 4, 5 bis 7, 8 bis 10 usf. (an Stelle von 1 bis 3, 4 bis 6 usf.), ein drittes, wenn man die Elemente 3 bis 5, 6 bis 8 usf. jeweils vereinigt. Diese drei Kollektivs haben die gleiche Verteilung, d. h. in jedem von ihnen ist die Wahrscheinlichkeit einer bestimmten Kombination von zwei Teilsummen, etwa 2-1, die gleiche. Dieselbe Größe ist daher auch der Grenzwert, dem die relative Häufigkeit dieser Kombination in der Gesamtfolge aller Summen von je zwei Nachbarelementen des ursprünglichen Kollektivs zustrebt. Der so für die Häufigkeit der Aufeinanderfolge 2-1 in der Summenreihe gefundene Grenzwert wird im allgemeinen ein anderer sein als der für die Kombination 1-1 oder 0-1 sich ergebende, und das hindert uns, im Sinne unserer früheren Definition von einer „Wahrscheinlichkeit des Auftretens der 1“ schlechthin zu sprechen.

M. v. Smoluchowski, der diese Erscheinung an der Brownschen Bewegung zum erstenmal eingehend untersucht hat, hat einen sehr anschaulichen Namen dafür eingeführt, der freilich — wie das oft mit solchen suggestiven Bezeichnungen der Fall ist — auch zu sehr falschen Vorstellungen Anlaß geben kann. Man spricht hier von einer „Wahrscheinlichkeits-Nachwirkung“ und meint damit, daß trotz des zufallsartigen Charakters des ganzen Vorganges doch eine Beeinflussung der einzelnen Beobachtung durch die vorangegangene vorliegt. Für uns ist es wichtig festzustellen, daß die Erscheinung der Wahrscheinlichkeits-Nachwirkung im Rahmen unserer Theorie der Kollektivs, die der Bedingung der Regellosigkeit genügen, volle Erledigung findet.

Die Verweilzeit und ihre Voraussage.

Kehren wir nun zu unserem Problem, der Brownschen Bewegung zurück. Die Analogie mit dem eben vorgebrachten Beispiel wird dadurch hergestellt, daß wir die Koordinaten eines Teilchens zu irgendeinem Zeitpunkt als die Summe aus den Anfangskoordinaten und den sprungartig zu denkenden Änderungen der Koordinatenwerte während der einzelnen Beobachtungsintervalle auffassen. Betrachten wir zwei aufeinanderfolgende Zeitpunkte, so ist eine Koordinate des Teilchens im zweiten gleich der Summe aus der Koordinate, die im früheren Zeitpunkt galt, und der im Zeitintervall erfolgten Änderung. Wir müssen also außer den Wahrscheinlichkeiten für die Anwesenheit eines Teilchens an einer bestimmten Stelle im Anfangszeitpunkt als neues Ausgangskollektiv, das für die Betrachtung des zeitlichen Ablaufs der Brownschen Bewegung wesentlich ist, das folgende einführen: die Beobachtung einer Koordinatenänderung, eines Sprunges, mit der Sprunggröße, genauer der Koordinatendifferenz in der einen und der anderen Richtung als Merkmal. Wir haben demnach als gegeben anzusehen die Wahrscheinlichkeiten dafür, daß die Koordinaten des Aufenthaltsortes eines Teilchens in je einem Zeitintervall Änderungen von beliebiger Größe erfahren. Diese Änderungs- oder Übergangs-Wahrscheinlichkeiten, die unser zweites Ausgangskollektiv bilden, sind für die Behandlung der Brownschen Bewegung und in ähnlicher Weise für alle Aufgaben der physikalischen Statistik von ausschlaggebender Bedeutung.

Nun bilden wir aus dem neuen und dem schon früher eingeführten Ausgangskollektiv ein abgeleitetes mit folgendem Element: Beobachtet wird die Gesamtheit von n Partikeln durch k Zeitpunkte hindurch. Merkmal ist zunächst die Gesamtheit der zweimal $n \times k$ Koordinaten aller Partikel, dann nach entsprechender Mischung die relative Anzahl $x : k$ der Zeitpunkte, in denen beispielsweise drei Partikel sich in einem bestimmten Felde befinden (bestimmte Koordinatenwerte aufweisen). Mit anderen Worten heißt dies: Wir können die Wahrscheinlichkeit dafür berechnen, daß sich z. B. in dem hundertsten Teil der beobachteten Zeitmomente drei Partikelchen in dem herausgehobenen, im Mikroskop, wie man sagt, „optisch isolierten" Raumteil befinden. Das Ergebnis dieser Rechnung ist ein äußerst charakteristisches und soll gleich kurz angegeben werden.

Es zeigt sich, wenn man — den Tatsachen entsprechend — die Teilchenzahl n und die Felderzahl N sehr groß annimmt, auch einen genügend ausgedehnten Zeitraum, also großes k, voraussetzt, daß unter allen möglichen (d. h. zwischen 0 und 1 liegenden) Werten für die relative „Verweilzeit" $x : k$ immer ein bestimmter Wert erdrückend große, kaum von 1 zu unterscheidende Wahrscheinlichkeit besitzt. Ich hatte vorhin angedeutet, wie man, von dem ersten Ausgangskollektiv herkommend, eine Wahrscheinlichkeit dafür berechnen kann, daß ein beliebiges der N kleinen Quadrate, wir sprachen dort von Nr. 25, in irgendeinem Augenblick gerade drei Partikel beherbergt. Diese Wahrscheinlichkeit ist natürlich ein echter Bruch, beispielsweise in dem Fall der SVEDBERGschen Versuchsreihe, die schon erwähnt wurde, ergab sich für die Wahrscheinlichkeit der Anwesenheit von drei Partikeln $w = 0{,}130$. Nun, das Ergebnis der weiteren Rechnung, die auch das zweite Ausgangskollektiv (die Sprungwahrscheinlichkeiten) berücksichtigt, geht dahin, daß bei Betrachtung einer genügend langen Folge von Zeitpunkten annähernd der Wert $x : k = 0{,}130$ mit außerordentlich großer Wahrscheinlichkeit zu erwarten ist. Um dies Resultat kurz und präzise aussprechen zu können, ist es gut, einige Bezeichnungen einzuführen.

Die im voraus gefundene Wahrscheinlichkeitsgröße $w = 0{,}130$ war aus der einfachen Annahme, daß die Wahrscheinlichkeit für ein Teilchen, sich in einem bestimmten der N Felder aufzuhalten, gerade gleich $1 : N$ ist, im wesentlichen durch Überlegungen, die der Kombinatorik angehören, gefunden worden. Wir wollen sie daher, um einen kurzen Ausdruck zu haben, als die „kombinatorische Wahrscheinlichkeit“ der Teilchenzahl 3 bezeichnen. Für den echten Bruch $x : k$, Quotient aus der Zahl x der Zeitmomente, in denen drei Teilchen beobachtet werden, durch die Gesamtzahl k der Beobachtungszeiten, haben wir schon den Ausdruck relative „Verweilzeit“ benutzt. Gemeint ist natürlich die Zeit des „Verweilens“ der Zahl 3 in dem betrachteten Feld, nicht etwa die des Verweilens bestimmt individualisierter Partikel. Wir können nun kurz das Ergebnis der Rechnung so zusammenfassen: Es ist nahezu mit Sicherheit zu erwarten, daß die relative Verweilzeit der Zahl 3 in einem Felde ungefähr gleich ist der kombinatorischen Wahrscheinlichkeit der Teilchenzahl 3; für stark davon abweichende Werte der Verweilzeit besteht nur ganz verschwindend geringe Wahrscheinlichkeit.

Dieser Satz weist die charakteristische Form aller Aussagen, die in der physikalischen Statistik gemacht werden können, auf. Es wird niemals eine determinierte Behauptung über den zeitlichen Ablauf künftigen Geschehens aufgestellt, wie dies in den sogenannten kausal aufgebauten Teilen der Naturwissenschaft, z. B. der klassischen Mechanik der Fall ist, sondern es wird ein bestimmter, zahlenmäßig beschriebener Verlauf als mit erdrückender Wahrscheinlichkeit zu erwartend hingestellt. Dabei ist die Größenordnung der entgegenstehenden Unwahrscheinlichkeit ganz außerordentlich gering. Dies ist alles schon früher, bei Besprechung der Gastheorie gesagt worden. Was aber jetzt neu hinzugekommen ist, ist dieses: Wir haben gezeigt, wie man mit den Hilfsmitteln der Wahrscheinlichkeitsrechnung, ohne irgendwie die Mechanik od. dgl. hinzuziehen, zu konkreten Aussagen gelangt darüber, was in einer Erscheinungsfolge geschieht, obwohl diese Folge für sich genommen kein Kollektiv bildet, nämlich nicht die volle Regellosigkeit besitzt. Dies ist ganz allgemein die Struktur aller Aussagen der statistischen Physik.

Zwei kurze Bemerkungen noch hierzu. In die Rechnung, deren Ergebnis wir eben besprachen, gehen die Annahmen über die Verteilung innerhalb des zweiten unserer Ausgangskollektivs ein. Man muß für die Rechnung wissen, wie groß die Wahrscheinlichkeiten größerer und kleinerer Sprünge eines Partikels sind. Aber es zeigt sich: das qualitative Endergebnis wird, wenn nur die Zahlen n, N, k genügend groß sind, von der zahlenmäßigen Verteilung dieser Wahrscheinlichkeiten so gut wie ganz unabhängig. — Das zweite ist, daß auch die Nachwirkungserscheinungen in der gleichen Form als Resultat der Rechnung erscheinen. Man findet nämlich: Es ist mit außerordentlich großer Wahrscheinlichkeit zu erwarten, daß eine bestimmte Aufeinanderfolge von Teilchenzahlen, z. B. 3 auf 2 folgend, mit einer relativen (zeitlichen) Häufigkeit erscheint, die der Größe nach annähernd gleich ist einer in entsprechender Weise kombinatorisch berechneten Wahrscheinlichkeit und diese ist verschieden, von der Wahrscheinlichkeit der Aufeinanderfolge 3 auf 1 oder 3 auf 5 usw. Man sieht, daß in dieser Weise vollkommen das Dilemma gelöst ist, daß die zeitliche Aufeinanderfolge der Teilchenzahlen zwar kein Kollektiv bildet (keine volle Regellosigkeit aufweist), sich aber dennoch mit den Gedankenmitteln der Wahrscheinlichkeitsrechnung beherrschen, d. h. rationell behandeln läßt.

Entropiesatz und Markoffsche Kette.

An dieser Stelle greife ich wieder auf den Grundgedanken der Boltzmannschen Gastheorie zurück, die wir jetzt besser verstehen können. Es war da die Rede von der zeitlichen Änderung der Entropie, also einer Zustandsgröße des Gases. Man kann nicht annehmen, auch wenn man sich keineswegs auf den deterministischen Standpunkt stellt, daß die zeitlich aufeinanderfolgenden Werte einer physikalischen Variablen unmittelbar die Eigenschaften eines Kollektivs besitzen, d. h. völlig regellos verlaufen wie die Resultate eines Würfelspieles. Die Anwendung des Wahrscheinlichkeitsbegriffes auf zeitliche Abläufe ist im allgemeinen nur in dem Sinne möglich, wie wir es eben hinsichtlich der Teilchenzahlen bei der Brownschen Bewegung kennengelernt haben. Deshalb ist diese Untersuchung für uns von so ausschlaggebender Bedeutung.

Im allgemeinen Fall muß man den Nachwirkungsbegriff durch einen etwas weiteren ersetzen, den der Verkettung von Ereignissen oder, wie man häufig sagt, durch den Begriff der MARKOFFschen Kette. Das Wesentliche ist dabei, daß als ein gegebenes Kollektiv ein solches angesehen wird, dessen Elemente aus Übergängen oder Sprüngen von einem Zustand in einen andern bestehen. In die Daten der Aufgabe geht also vor allem ein System von Übergangs-Wahrscheinlichkeiten ein. Daß es meist nicht auf die speziellen numerischen Werte sondern nur auf den generellen Typus dieser Wahrscheinlichkeiten ankommt, ist für den Ansatz und das logische Verständnis der Theorie nebensächlich.

Bei der BROWNschen Bewegung haben wir die Teilchenzahl in einem optisch isolierten Gebiet als Zustandsgröße betrachtet. Für ein beliebiges Gasvolumen tritt an deren Stelle die gesamte Aufteilung der Moleküle auf die verschiedenen Teilgebiete des Geschwindigkeitsraumes (allgemeiner des sogenannten Phasenraumes). Jede Aufteilung besitzt eine bestimmte „kombinatorische" Wahrscheinlichkeit, d. h. eine nach den elementaren Regeln der Addition und Multiplikation von Wahrscheinlichkeiten zu berechnende Wahrscheinlichkeitsgröße. Die vollständige Untersuchung, die sich auf die Betrachtung der Übergangswahrscheinlichkeiten stützt, lehrt nun wieder: Es ist mit außerordentlich großer Wahrscheinlichkeit zu erwarten, daß in der natürlichen Aufeinanderfolge der Zustände (die nicht ganz regellos verläuft, also kein Kollektiv bildet) gleichwohl die verschiedenen möglichen Aufteilungen annähernd mit der relativen Häufigkeit erscheinen, die durch die kombinatorisch errechnete Wahrscheinlichkeit zum Ausdruck gebracht wird. Statt von „kombinatorischer" spricht man hier auch manchmal von „thermodynamischer" Wahrscheinlichkeit.

Die (kombinatorische) Rechnung ergibt, daß gewisse Aufteilungen der Moleküle allen anderen gegenüber eine übergroße, thermodynamische Wahrscheinlichkeit besitzen. Daraus folgt dann, unter Zuhilfenahme der eben erwähnten Überlegung, daß das Gasgemenge, sich selbst überlassen, fast immer in einem Zustand solcher bevorzugter Aufteilung gefunden werden wird, und weiter: Wenn das System sich einmal erheblich von dem bevorzugten Zustand entfernt hat, so wird fast immer der nächstfolgende Augenblick einen Zustand aufweisen, der dem bevorzugten näher liegt. Man nennt diese Aufteilungen, die fast die ganze Wahrscheinlichkeit (und mithin, wie wir sehen, fast den gesamten Zeitablauf) für sich in Anspruch nehmen, MAXWELL-BOLTZMANNsche Aufteilung, weil sie vor BOLTZMANN schon von dem englischen Physiker MAXWELL studiert worden sind.

Was nun BOLTZMANN entscheidend Neues hinzugefügt hat, war, daß er in der Größe der thermodynamischen Wahrscheinlichkeit, genauer gesagt in ihrem Logarithmus, ein Maß der Entropie erkannt hat. Daß auf einen an sich unwahrscheinlicheren Zustand fast immer ein wahrscheinlicherer folgt, ist die neue Form des Entropiegesetzes: Fast immer vergrößert sich die Entropie in einem sich selbst überlassenen System. Man sieht, daß hier das Wörtchen „fast", dieses Kennzeichen des Indeterminismus, nicht zu unterdrücken ist.

Diese Andeutungen müssen hier genügen. Ein näheres Eingehen auf die Fragen der Gastheorie verbietet sich mit Rücksicht auf die vielen thermodynamischen Begriffe, die dazu eingeführt werden müßten, und die damit verbundenen mathematischen Schwierigkeiten. Ich kehre wieder zu dem einfacheren und anschaulicheren Beispiel der BROWNschen Bewegung zurück, um Ihnen an Hand konkreter Versuchszahlen die Ergebnisse der Theorie noch näherzubringen.

Versuchsreihe von SVEDBERG.

Eine Nachprüfung der Theorie der BROWNschen Bewegung durch Versuche ist nur in einem durch die Natur der Sache beschränkten Maße möglich. Man kann natürlich nicht so lange experimentieren, d. h. so viele Versuchsreihen vornehmen, daß erwartet werden kann, auch einmal ein Ergebnis von jener horrenden Unwahrscheinlichkeit zu erhalten, wie sie vorhin an dem gastheoretischen Beispiel angedeutet wurde. Vielmehr ist anzunehmen, daß etwa eine einmalige, nicht zu kurze Reihe von Zeitbeobachtungen mit dem übereinstimmen wird, was nach der Rechnung mit erdrückend großer Wahrscheinlichkeit eintreten muß. Praktisch bedeutet dies, daß die zeitliche Häufigkeit, mit der eine beliebige Teilchenzahl beobachtet wird, zu vergleichen ist mit der kombinatorischen Wahrscheinlichkeit, die für diese Zahl berechnet wurde. Darin liegt, nebenbei bemerkt, auch die Aufklärung der früher erwähnten Tatsache, daß die Versuche ohne Durcheinanderrütteln der ganzen Emulsion schon zu befriedigender Übereinstimmung zwischen der Beobachtung und den kombinatorischen Wahrscheinlichkeiten führen. Das folgende Beispiel einer von dem schwedischen Physiker SVEDBERG durchgeführten Untersuchung der BROWNschen Bewegung gibt über das Maß der Übereinstimmung Auskunft.

SVEDBERG beobachtete in einer bestimmten Goldlösung die Partikelchen, die sich in Intervallen von je $^1/_{30}$ Minute in einem optisch isolierten Raumteil aufhielten. Er fand bei insgesamt 518 Zählungen

die Teilchenzahl		0	112mal,	relat. Häufigkeit		0,216
,,	,,	1	168 ,,	,,	,,	0,325
,,	,,	2	130 ,,	,,	,,	0,251
,,	,,	3	69 ,,	,,	,,	0,133
,,	,,	4	32 ,,	,,	,,	0,062
,,	,,	5	5 ,,	,,	,,	0,010
,,	,,	6	1 ,,	,,	,,	0,002
,,	,,	7	1 ,,	,,	,,	0,002

Die mittlere Teilchenzahl, der Durchschnittswert, beträgt 1,54. Rechnet man daraus nach den Regeln der Wahrscheinlichkeitsrechnung die Wahrscheinlichkeiten w für das Auftreten von 0, von 1, 2 .... bis 7 Teilchen, so erhält man die acht Größen: 0,212, 0,328, 0,253, 0,130, 0,050, 0,016, 0,004, 0,001, die, wie man sieht, recht gut mit der letzten Spalte der eben angeführten Zahlentafel übereinstimmen. Nach der Theorie war mit sehr nahe an 1 liegender Wahrscheinlichkeit zu erwarten, daß beispielsweise die Teilchenzahl 3 mit einer nahe an 0,130 liegenden relativen Häufigkeit erscheint. Die Beobachtung ergab für die eine durchgeführte Versuchsreihe die Häufigkeit 0,133.

Von Interesse ist es nun, in dem vorliegenden Falle auch die „Nachwirkungserscheinung“ zu untersuchen. Wie eben mitgeteilt, ist 168mal die Teilchenzahl 1 beobachtet worden. In vier Fällen blieb die anschließende Beobachtung aus, so daß wir mit nur 164 Fällen einer Eins als erster Zahl eines Paares zu rechnen haben. Unter diesen 164 Fällen erschien 40mal eine Null, 55mal eine Eins, 40mal eine Zwei, 17mal eine Drei usw. als zweite Zahl. Die relativen Häufigkeiten der Null, Eins, Zwei und Drei betrugen daher $40 : 164 = 0{,}246$, bzw. 0,336, 0,246, 0,104, sobald man nur diejenigen Beobachtungen zählt, die auf eine vorangegangene Eins folgen. Berechnet man in analoger Weise die Häufigkeiten der Zahlen 0 bis 3, indem man nur die auf eine Drei folgenden berücksichtigt, so ergeben sich die Werte 0,087, 0,334, 0,319, 0,189. Man sieht, daß die Null hier viel seltener, die Drei viel öfter auftritt als in den Paaren, die mit einer Eins beginnen. Die „Regellosigkeit", d. h. die Unempfindlichkeit der relativen Häufigkeit gegenüber einer Stellenauswahl ist nicht vorhanden: die Reihe der fortgesetzt beobachteten Teilchenzahlen bildet kein Kollektiv.

Aber es ist gleichwohl möglich, wie ich es eben vorhin ausgeführt habe, die auf den Kollektivbegriff aufgebauten Regeln der Wahrscheinlichkeitsrechnung anzuwenden. Man kann für die angegebenen Paare von Teilchenzahlen diejenigen Häufigkeiten berechnen, die mit überaus großer Wahrscheinlichkeit zu erwarten sind, und man erhält — ich gehe auf die Art der Berechnung nicht näher ein — für das Paar 1-0 die Wahrscheinlichkeit 0,246, für das Paar 3-0 die Wahrscheinlichkeit 0,116 gegenüber den Beobachtungen, die 0,246 und 0,087 ergaben. Die Wahrscheinlichkeit der Null schlechthin war vorher zu 0,212 gefunden worden. Man muß die Übereinstimmung zwischen Rechnung und Beobachtung im Hinblick auf den nicht sehr großen Umfang der Versuchsreihe als eine sehr gute bezeichnen.

Radioaktive Strahlung.

Eine andere physikalische Erscheinung, die ähnlich der Brownschen Bewegung ein fruchtbares Feld der Anwendung wahrscheinlichkeitstheoretischer Überlegung darstellt, finden wir in der radioaktiven Strahlung. Auch hier handelt es sich nicht wie bei den Atomen und Molekülen der Gastheorie um hypothetische, im Wechsel der Zeiten und Schulen schwer umstrittene Vorstellungen, sondern um Dinge, die in verhältnismäßig einfachen experimentellen Anordnungen beobachtet werden können. Der „Zerfall" eines Metalls der Radiumgruppe geht in der Weise vor sich, daß der Körper sogenannte α-Teilchen, Elementarbestandteile des Atoms mit positiver elektrischer Ladung, aussendet, die, auf einen aus geeigneter Substanz bestehenden Schirm auffallend, jedesmal ein Aufleuchten, eine Szintillation hervorrufen. Die Zeitabstände zwischen je zwei aufeinanderfolgenden Szintillationen weisen nun einen zufallsartigen Charakter auf. Man gelangt zu einer sehr einfachen Darstellung der Verhältnisse, indem man annimmt, daß für eine sehr große Zahl von Einzelteilen des Körpers, die man ruhig als „Atome“ bezeichnen kann, in jedem Augenblick oder, genauer gesagt, in außerordentlich kurzen Zeitabständen, die Alternative: Zerfall oder Nichtzerfall besteht.

Die analoge Glücksspielaufgabe wäre die: Man spielt in kurzen regelmäßigen Zeitabständen einen Würfel aus und beobachtet die Zahl der Zeitintervalle zwischen zwei aufeinanderfolgenden Sechser-Würfen. Aus der einfachen Alternative 6 oder Nicht-6 als Ausgangskollektiv leitet man zunächst eine k-fache „Verbindung“ ab, deren Element die Zusammenfassung von k Würfen und deren Merkmal die Zusammenfassung der k Ergebnisse (6 oder Nicht-6) ist. Daraus bildet man durch Mischung ein Kollektiv mit denselben Elementen und dem zweidimensionalen Merkmal: $z =$ Anzahl der Würfe, die dem ersten Sechs-Wurf vorangehen, und $x =$ Anzahl der Würfe vom ersten bis zum zweiten Sechs-Ergebnis (so daß $x - 1$ Nicht-6-Würfe und x Zeitintervalle dazwischenliegen). Den Fall, daß weniger als 2 Sechser fallen, können wir außer acht lassen, wenn wir k sehr groß voraussetzen.

Die Wahrscheinlichkeit für das Eintreten eines bestimmten z und bestimmten x rechnet sich nach bekannten Regeln, wenn p die Wahrscheinlichkeit des Sechs-Wurfes, $q = 1 - p$ die des Nicht-6-Wurfes bezeichnet, gleich dem Produkt $q^z\, p \,.\, q^{x-1}\, p =$ $= p^2\, q^{x-1} \,.\, q^z$. Mischt man noch einmal, indem man alle Fälle mit irgendeinem Wert von z zusammenwirft, so erhält man die Wahrscheinlichkeit dafür, daß zwischen dem ersten und zweiten Sechs-Ergebnis gerade x Zeitelemente liegen, gleich der Summe $p^2\, q^{x-1}\, (1 + q + q^2 + q^3 + \ldots .)$. Die Summe ist eigentlich nur fortzusetzen bis zu der höchstmöglichen Zahl von z, d. i. $k - x$. Wenn wir aber k sehr groß wählen, dürfen wir für den Ausdruck in der Klammer die unendliche (geometrische) Reihe $1 + q +$ $+ q^2 + q^3 + \ldots .$ setzen, deren Wert, wie man weiß, $1 : (1 - q) =$ $= 1 : p$ ist. Demnach wird die Wahrscheinlichkeit des Zeitabstandes x zwischen dem ersten und zweiten Sechs-Wurf $p\, q^{x-1}$, also für die Zeitabstände 1, 2, 3.... gleich $p, p\, q,\ p\, q^2 \ldots .$ usf.

Nach den früheren Erklärungen können wir daraus den Erwartungswert des Zeitabschnittes rechnen, indem wir jede dieser

Wahrscheinlichkeiten mit dem zugehörigen x-Wert multiplizieren und die Produkte addieren. Man erhält so: $1 . p + 2 . pq + 3 . pq^2 + \ldots = p\,(1 + 2\,q + 3\,q^2 \ldots)$, wobei die Summe in der Klammer, wie die Algebra lehrt, den Wert $1 : (1 - q)^2 = 1 : p^2$ besitzt. Somit ist der Erwartungswert des Zeitabstandes zwischen dem ersten und zweiten Sechs-Wurf gleich $1 : p$, also etwa bei einem „richtigen" Würfel mit $p = {}^1/_6$ gleich 6. Dieses Resultat ist gewiß nicht überraschend und, wer der oben dargelegten Rechnung nicht folgen will oder kann, wird es auch so hinnehmen. Ich wollte nur die Gelegenheit benutzen, um zu zeigen, wie unsere rationelle Methode des Operierens mit den Kollektivs diese Frage (wie jede andere) zwangsläufig zu beantworten erlaubt.

Wollten wir den Abstand zwischen dem zweiten und dritten oder zwischen dem dritten und vierten Sechs-Wurf rechnen, so bekämen wir immer die gleiche Zahl $1 : p$. Wir können daher sagen: Der im Mittel zu erwartende Zeitabstand zwischen irgendeinem Sechs-Wurf und dem nächstfolgenden beträgt $1 : p$ (im Spezialfall des richtigen Würfels 6) solcher Zeitintervalle, die zwischen einem Ausspielen des Würfels und dem nächsten liegen.

Die Voraussage der Zeitabstände.

Bei der Übertragung dieser Überlegungen auf die radioaktive Strahlung tritt vor allem die Schwierigkeit auf, daß man von vornherein weder die Größen p und q noch die Länge des Zeitelementes zwischen zwei aufeinanderfolgenden „Spielen" kennt. Auch muß man annehmen, daß dieses Zeitintervall so außerordentlich klein ist, daß die Genauigkeit der Zeitmessung nicht ausreicht, um so geringe Zeitunterschiede zu beobachten. Gleichwohl läßt sich die Theorie in folgender Weise durch das Experiment überprüfen.

Wir rechnen zunächst die Wahrscheinlichkeit dafür aus, daß zwischen zwei aufeinanderfolgenden Szintillationen **höchstens** drei Zeitelemente (also entweder eins oder zwei oder drei) liegen. Nach der Additionsregel ist dies die Summe $p + pq + pq^2 = p\,(1 + q + q^2) = p\,\frac{1 - q^3}{1 - q} = 1 - q^3$. Die Wahrscheinlichkeit des Gegenteils, also dafür, daß mehr als drei Zeitelemente von einer Szintillation zur nächsten verstreichen, ist daher q^3 oder $(1 - p)^3$. Nehmen wir statt der Zahl 3 eine beliebige Zahl n, so erhalten wir in $(1 - p)^n$ die Wahrscheinlichkeit dafür, daß das Intervall zwischen zwei benachbarten Szintillationen mehr als n Zeitelemente umfaßt. Nun lehrt die Analysis, daß wenn p eine sehr kleine, n eine sehr große Zahl ist, für $(1 - p)^n$ genau genug gesetzt werden darf $(1 : e)^{np}$, wobei e eine bestimmte Zahl, die sogenannte Basis des natürlichen Logarithmus 2,718 (reziproker Wert $1 : e = 0{,}3679$) bedeutet. Es ist also die (np)-te Potenz von 0,3679 die Wahrscheinlichkeit dafür, daß Zeitabstände von mehr als n Zeitelementen zwischen zwei Ereignissen verstreichen. Dieser Wert läßt sich mit der Beobachtung wie folgt vergleichen.

Man dividiert, nach hinreichend langer Beobachtungszeit die (in Sekunden gemessene) Zeitdauer zwischen der ersten und letzten Szintillation durch die Anzahl der Zeitintervalle (nämlich durch die um 1 verminderte Anzahl der beobachteten Oszillationen) und erhält damit die mittlere Länge eines Intervalles, sagen wir gleich 5 Sekunden. Wenn die Theorie richtig ist und die Beobachtungszeit groß genug war, so muß dieser Zeitraum von 5 Sekunden annähernd mit dem Erwartungswert des Zeitabstandes übereinstimmen, der, wie vorhin festgestellt, $1 : p$ Zeitelemente umfaßt. Ein Zeitelement beträgt, somit $5\,p$ Sekunden, n Zeitelemente betragen $5\,n\,p$ Sekunden, und die Intervalle, die mehr als n Zeitelemente umfassen, sind jene, die mehr als $5\,n\,p$ Sekunden dauern. Wählt man z. B. $n\,p = 2$, so gehören dazu die Intervalle von mehr als 10 Sekunden Länge, zu $n\,p = 3$ gehören die von mehr als 15 Sekunden usf. Man braucht also nur die Anzahl der wirklich beobachteten Intervalle von mehr als 5, mehr als 10, mehr als 15 .... Sekunden Dauer zu zählen, die Zahlen durch die Gesamtzahl der überhaupt beobachteten Intervalle zu dividieren und die so erhaltenen relativen Häufigkeiten mit den früher errechneten Wahrscheinlichkeiten zu vergleichen, nämlich der ersten, zweiten, dritten .... Potenz von 0,3679.

Allgemein kann man sagen: Wenn a die aus den Beobachtungen ermittelte mittlere Länge des Zeitintervalls (früher 5 Sekunden) ist, so verlangt die Theorie, daß Intervalle, die das n-fache der Länge a überschreiten, mit einer relativen Häufigkeit gleich der n-ten Potenz von 0,3679 auftreten.

Versuchsreihe von Marsden und Barratt.

Eine Versuchsreihe, über die E. Marsden und T. Barratt berichten, umfaßte 7563 Intervalle zwischen aufeinanderfolgenden Szintillationen und dauerte insgesamt 14595 Sekunden. Die mittlere Länge des Intervalles betrug daher $a = 1{,}930$. In der untenstehenden Tabelle sind zunächst die beobachteten Häufigkeiten der Intervalle von mehr als 1, von mehr als 2...., von mehr als 9 Sekunden Dauer angegeben und daneben die durch 7563 dividierten, also relativen, Häufigkeiten. Zum Vergleich sind hinzugefügt die berechneten Wahrscheinlichkeiten. Da $1 : a = 0{,}518$, so hat man zuerst die 0,518te Potenz von 0,3679 zu bilden, das gibt 0,596, sodann die zweite Potenz, also das Quadrat, der ersten Zahl usf. bis zu ihrer neunten Potenz. Die so berechneten Potenzen sind in der letzten Spalte der Tabelle eingetragen.

Intervalle von mehr als	Anzahl	Relative Häufigkeit	Berechnete Wahrscheinlichkeit
0 Sekunden	7563	1,000	1,000
1 ,,	4457	0,590	0,596
2 ,,	2694	0,356	0,355
3 ,,	1579	0,209	0,211
4 ,,	921	0,122	0,126
5 ,,	532	0,070	0,075
6 ,,	326	0,043	0,045
7 ,,	196	0,026	0,027
8 ,,	110	0,015	0,016
9 ,,	68	0,009	0,009

Man sieht, daß eine sehr gute Übereinstimmung zwischen den beiden letzten Spalten der Tabelle, also zwischen den beobachteten und den nach der Theorie berechneten relativen Häufigkeiten besteht. Es soll nicht verschwiegen werden, daß die Untersuchung noch nach manchen Richtungen weitergeführt werden muß, wenn man die Verhältnisse in allen Einzelheiten übersehen will, und daß manche Frage auf diesem Gebiet heute noch ungeklärt ist. Aber für uns kommt nur das Grundsätzliche in Betracht, nur die allgemeine F o r m der wahrscheinlichkeitstheoretisch abgeleiteten Aussagen über den Ablauf physikalischer Vorgänge. In dieser Richtung ist zur Präzisierung des bisher Gesagten noch folgendes zu bemerken.

Unsere Theorie führt natürlich nicht zu der Folgerung, daß, wenn 7563 Szintillationsintervalle in einem Zeitraum von 14595 Sekunden beobachtet werden, genau 0,211 von ihnen (also $0{,}211 \,.\, 7563 = 1596$) eine Länge von mehr als 3 Sekunden haben müssen. Wir wissen nur, daß die Wahrscheinlichkeit der angegebenen Intervallgröße 0,211 beträgt und wenn wir dies mit dem vorhin ausführlich dargelegten ersten Gesetz der großen Zahl in Verbindung bringen, so folgt daraus: Es ist mit sehr nahe an 1 liegender Wahrscheinlichkeit zu erwarten, daß annähernd $211^0/_{00}$ der 7563 Intervalle eine Länge von mehr als drei Sekunden besitzen. Oder noch genauer: Wiederholt man sehr oft die Serie von 7563 Beobachtungen, so müssen in einer sehr stark überwiegenden Mehrzahl der Fälle annähernd 1596 Intervalle von mehr als 3 Sekunden Länge auftreten. Diese Aussage hat genau die gleiche Form wie die, die wir in der Theorie der Brownschen Bewegung kennengelernt haben. Dort hieß es: Wenn wir die (lange) Beobachtungsreihe für ein Feld hinreichend oft wiederholen, wird in der überwiegenden Mehrheit der Fälle die relative Häufigkeit des Auftretens einer bestimmten Teilchenzahl annähernd gleich der vorausberechneten sein. Das Wesentliche ist: Nicht, was bei einer Versuchsreihe geschieht, wird durch die Theorie vorausgesagt, sondern es wird eine Behauptung darüber aufgestellt, was bei sehr vielfältiger Wiederholung des Versuches in der überwiegenden Mehrheit der Fälle geschehen wird.

Neuere Entwicklung der Gastheorie.

Ich komme jetzt noch einmal auf Fragen der kinetischen Gastheorie zurück, von denen dieser Vortrag den Ausgang genommen hat. Es wurde schon gesagt, daß die ältesten Ansätze, in deren Rahmen Boltzmann die statistische Deutung des Entropiesatzes schuf, mehrfache Abänderungen erfahren haben. Diese zielen in erster Linie dahin, eine Erklärung der sogenannten Gasentartung, d. h. der Abweichungen im Verhalten der Gase bei sehr niederen Temperaturen gegenüber dem normalen zu geben. Vom statistischen Standpunkte lassen sich die verschiedenen, hier zur Geltung gekommenen Theorien kurz kennzeichnen: sie bestehen darin, daß die Ausgangswahrscheinlichkeiten, die, wie wir wissen, als gegebene Größen in jeden Ansatz der Wahrscheinlichkeitsrechnung eingehen, variiert werden. Man versucht, solche Annahmen für die Ausgangswahrscheinlichkeiten zu finden, daß bei mittleren und höheren Temperaturen sich wesentlich Übereinstimmung mit der älteren CLAUSIUS-BOLTZMANNschen Theorie ergibt, bei Annäherung an den Nullpunkt der Temperaturskala aber die der Beobachtung entsprechende Verminderung der spezifischen Wärme herauskommt.

Üblicherweise werden die Ansätze für die Ausgangsverteilungen in die Form gekleidet, daß gewisse Bereiche der Merkmalwerte oder des Merkmalraumes als „gleichwahrscheinlich" angenommen werden. Dadurch, daß man in den verschiedenen Theorien jeweils andere Bereiche, also andere Merkmalgruppen als untereinander gleichwahrscheinlich voraussetzt, variiert man eben die Ausgangsverteilung. Daher sind hier auch gegen die Verwendung des „Gleichwahrscheinlichkeits"-Begriffes nicht die Einwände zu erheben, die wir vom Standpunkt der Grundlagen der Wahrscheinlichkeitsrechnung machen mußten. Man kann schließlich jede Verteilung als „Gleichverteilung" ansehen, wenn man die Teilgebiete entsprechend definiert. Nur darauf kommt es an, daß man sich bewußt ist, Annahmen, Hypothesen zu machen und nicht meint, die Rechnungsgrundlage sei „a priori" gegeben.

Ansätze für die Gasentartung. Elektronentheorie der Metalle.

In der älteren Gastheorie nahm man an, daß für das einzelne Molekel alle überhaupt möglichen Geschwindigkeitswerte in folgendem Sinne gleichwahrscheinlich seien: Denkt man sich in einem rechtwinkligen Koordinatensystem die drei Geschwindigkeitskomponenten in den Achsrichtungen aufgetragen, so sollen in dem so gebildeten „Geschwindigkeitsraum" (der für unser Ausgangskollektiv der „Merkmalraum" ist) gleich großen Volumenteilen gleich große Wahrscheinlichkeiten entsprechen. Jeden dieser Volumteile wollen wir als einen „Platz" für ein Molekel bezeichnen. Geht man zu einem Kollektiv über, dessen Element die Betrachtung von mehreren, sagen wir von n Molekeln ist, so wird jetzt jede mögliche Komplexion, d. h. jede individuelle „Aufstellung" der n Molekel auf den m Plätzen des Geschwindigkeitsraumes gleichwahrscheinlich. Haben wir z. B. nur zwei

Molekel A und B und drei verschiedene Geschwindigkeitsgebiete a, b, c, so gibt es 9 (im allgemeinen m^n) verschiedene individuelle Aufstellungen: Man kann zuerst für A auf dreifache Weise einen Platz suchen, nämlich entweder a oder b oder c, dann jedesmal für B noch einen der drei Plätze a, b oder c wählen. Jeder dieser 9 Aufstellungen kommt dann nach der Annahme der klassischen Theorie die Wahrscheinlichkeit $^1/_9$ zu.

Eine neuere, von dem Inder BOSE begründete, von EINSTEIN weiter ausgeführte Theorie betrachtet nun nicht die eben genannten 9 Aufstellungen als gleichwahrscheinlich, sondern geht von dem Begriff der „Aufteilung" aus. Eine Aufteilung ist dann gegeben, wenn man weiß, wie viele Molekel sich auf jedem Platz (des Geschwindigkeitsraumes) befinden, ohne daß es darauf ankommt, welche individuelle Molekel es sind. In dem Beispiel von 2 Molekeln auf 3 Plätzen gibt es nur 6 verschiedene Aufteilungen, nämlich: Es können entweder beide Molekel zusammen in a, b oder c sich befinden, oder je eines in a und in b, bzw. in b und in c, bzw. in a und in c. Nach der BOSE-EINSTEINschen Theorie besitzt jeder dieser Fälle die gleiche Wahrscheinlichkeit, also $^1/_6$. Nach der klassischen Theorie käme den drei erstgenannten Möglichkeiten je die Wahrscheinlichkeit $^1/_9$, den anderen drei aber je $^2/_9$ zu; denn diese Aufteilungen umfassen je zwei verschiedene individuelle Aufstellungen: Es kann bei der ersten von ihnen A in a und B in b oder umgekehrt A in b und B in a sein usw.

Schließlich hat der Italiener FERMI eine statistische Gastheorie entwickelt, in der angenommen wird, daß nur solche Aufteilungen, bei denen auf jedem Platz höchstens ein Molekel sich befindet, möglich sind und daß ihnen gleiche Wahrscheinlichkeiten zukommen. In unserem Beispiel käme also den drei Aufteilungen: je ein Molekel in a und in b, bzw. in b und in c, bzw. in a und in c je die Wahrscheinlichkeit $^1/_3$ zu. Die Prüfung dieser und der anderen Annahmen geschieht in der Weise, daß man nach dem BOLTZMANNschen Satz in der Wahrscheinlichkeit eines Zustandes des Gases ein Maß für seine Entropie sieht und nun verlangt, daß die Abhängigkeit der Entropie von der Temperatur und der Gesamtmasse (die in den Größen m und n stecken) sich den Beobachtungen gemäß ergibt. Diese Art der Statistik hat sich ganz besonders in der Elektronentheorie der Metalle bewährt, d. h. in der rechnerischen Verfolgung der Auffassung, daß die Elektrizitätsleitung in Metallen durch Vermittlung von Elektronen geschieht, die, einem stark verdünnten Gase ähnlich, zwischen den Atomen des Metalles hindurchschwirren.

Es kann nicht unsere Aufgabe sein, diese Fragen hier weiter zu verfolgen oder gar zu den einzelnen Theorien kritisch Stellung zu nehmen. Die experimentelle Forschung ist in vollem Gange, um das Verhalten von Gemengen geringer Dichte weiter aufzuklären und allmählich wird sich erst zeigen, in welchem Maße eine Theorie, die mit so einfachen Ausgangsverteilungen arbeitet, die Beobachtungen richtig wiederzugeben vermag.

Für die Fragen, die uns im Hinblick auf die prinzipiellen Grundlagen der Wahrscheinlichkeitsrechnung interessieren, ist hier nur festzustellen, daß die Gastheorie mit allen ihren modernen Wendungen sich zwanglos dem allgemeinen Rahmen einfügt, den wir für alle Aufgaben der Wahrscheinlichkeitstheorie als maßgebend erkannt haben: Aus Ausgangsverteilungen, die als gegeben vorausgesetzt werden, rechnet man die Verteilungen in abgeleiteten Kollektivs; lassen sich, wie das in der Gastheorie der Fall ist, nicht unmittelbar die Ausgangswerte durch Beobachtungen prüfen, so muß man die errechneten Größen mit den Ergebnissen des Experiments vergleichen.

Quantentheorie.

Einen ähnlichen Gegenstand wahrscheinlichkeitstheoretischer Untersuchung wie die kinetische Gastheorie bildet die seit Beginn dieses Jahrhunderts zur Geltung gekommene Quantentheorie. Sie findet ihren Ausgangspunkt in der weittragenden Entdeckung von Max Planck, die im Dezember 1899 publiziert wurde. Man besaß seit langem die Vorstellung, die namentlich in den chemischen Erscheinungen ihre Stütze fand, daß die Materie aus letzten, nicht weiter teilbaren Elementarteilen, den Atomen und Molekülen, bestehe. Daneben hatte sich auch allmählich die Annahme einer atomistischen Struktur der Elektrizität als unabweisbar herausgestellt, die Annahme nämlich, daß alle elektrischen Ladungen ganzzahlige Vielfache einer Ladungseinheit, der des Elektrons, seien. Trotzdem war es ein kühner Schritt, als Planck dazu überging, eine Atomistik der Strahlungsenergie einzuführen, aus der sich dann die allgemeine atomistische Auffassung des Lichtes entwickelte.

Um die Temperaturabhängigkeit der schwarzen Strahlung (der reinen Wärmestrahlung) in Übereinstimmung mit dem vorhandenen Beobachtungsmaterial deuten zu können, nahm PLANCK an, daß es als Träger der thermischen und optischen Schwingungen submikroskopisch kleine „Oszillatoren" gäbe, die Schwingungsenergie nur in ganzen Vielfachen einer konstanten Energiemenge aufnehmen können. Diese Einheit, das Energiequant, ist von der Schwingungszahl abhängig, ihr direkt proportional. Nennt man, wie üblich, ν die Frequenz, d. h. die Zahl der Schwingungen pro Sekunde, so ist die Größe der für einen Oszillator in Betracht kommenden Energieeinheit das Produkt $h\nu$, worin h die universelle „PLANCKsche Konstante" oder das Wirkungsquantum heißt. Im gebräuchlichen physikalischen Maßsystem ist $h = 6{,}55 \cdot 10^{-27}\,\mathrm{gcm^2/sec^2}$, so daß der Oszillator eines gelben Lichtstrahles mit einer Frequenz von ungefähr $\nu = 5 \cdot 10^{14}$ ein Energiequant von $h\nu = 3{,}3 \cdot 10^{-12}$ Erg besitzt (1 Erg ist die Arbeit, die erforderlich ist, um ein Milligrammgewicht 1 cm hoch zu heben).

Der Vorgang einer monochromatischen Strahlung, d. h. einer Strahlung, die nur eine Schwingungsfrequenz aufweist, ist nun so zu denken, daß eine große Menge n von Oszillatoren vereinigt ist, deren jeder entweder 1 oder 2 oder 3 usw. Energiequanten h enthält. Man nimmt an, daß die Oszillatoren ein Kollektiv mit der Qantenzahl als Merkmal bilden. Die Verteilung in diesem Ausgangskollektiv, also die Wahrscheinlichkeiten dafür, daß ein Oszillator 1, 2, 3 oder mehr Quanten besitzt, muß natürlich angenommen werden. Man versucht es mit der Annahme der Gleichverteilung über alle möglichen Werte von Quantenzahlen und findet damit gute Resultate. In Rechnung gestellt wird z. B. das folgende abgeleitete Kollektiv: Element ist eine Gesamtheit von n Oszillatoren mit einer bestimmten Gesamtenergie E; das Merkmal besteht in der „Aufteilung" der verschiedenen möglichen Quantenzahlen, d. h. in der Angabe, wie viele Oszillatoren ein Quant, wie viele zwei Qanten usf. besitzen. Unter den verschiedenen möglichen Aufteilungen gibt es eine wahrscheinlichste und bei genügend großem n kann man es, im Sinne der Gesetze der großen Zahlen als „fast sicher" annehmen, die Energie in einem beliebig herausgegriffenen Augenblick in diesem Aufteilungszustand oder einem nahe benachbarten anzutreffen (Analogie zur MAXWELL-BOLTZMANN-Aufteilung in der Gastheorie). Diese Rechnungen führen, wenn man die aus der Thermodynamik und der kinetischen Gastheorie bekannten Beziehungen zwischen Temperatur, Energie usf. hinzunimmt, zu dem vorerwähnten, mit den Beobachtungen übereinstimmenden Abhängigkeitsgesetz zwischen Strahlungsenergie und Temperatur.

Alle prinzipiellen, uns vom Standpunkt der Wahrscheinlichkeitstheorie interessierenden Überlegungen verlaufen bei dieser Problemgruppe analog wie im Falle der BOLTZMANNschen Gastheorie. Ich kann also auf das früher Gesagte, namentlich in Hinsicht auf die Bedeutung der Übergangswahrscheinlichkeiten für den zeitlichen Ablauf der Erscheinungen, verweisen.

Erwähnt sei noch daß der PLANCKsche Ansatz im Jahre 1905 durch ALBERT EINSTEIN eine nähere Begründung und weitere Ausdehnung erfahren hat. Mit EINSTEIN spricht man heute von Lichtquanten, deren Energieeinheit $h\nu$ ist und faßt das Licht schlechthin als aus diesen Lichtatomen oder Photonen bestehend auf. Allein die auf diesen Ideen aufgebaute statistische Theorie des Lichtes tritt aus dem Rahmen der bisher besprochenen, an die Gaskinetik anschließenden Statistik heraus. Ich muß noch in der Richtung auf allgemeinere Fragen ein wenig ausholen, bevor ich auf die moderne Quantenstatistik zu sprechen komme.

Statistik und Kausalgesetz.

Bei Besprechung der BROWNschen Bewegung und der radioaktiven Strahlung habe ich bereits ausdrücklich das gekennzeichnet, was im Rahmen der physikalischen Naturerklärung den Unterschied zwischen einer statistischen und einer rein kausalen Theorie ausmacht. Während die kausale Theorie den Anspruch erhebt, auf Grund der Ausgangswerte der Rechnung den Ablauf einer Erscheinung präzise voraussagen zu können, stellt die statistische Theorie lediglich eine Behauptung darüber auf, was bei oft genug wiederholtem Versuch in der überwiegenden Mehrheit der Fälle geschehen wird. Dieser Übergang zur indeterministischen Auffassung gegenüber bestimmten Problemen der Physik wurde, wie ich eingangs erwähnte, im wesentlichen durch BOLTZMANNS Wendung des klassischen Entropiegesetzes um 1870 begründet. Er bedeutete einen

Bruch mit Ansichten und Denkgewohnheiten, die sich durch mehr als zwei Jahrhunderte physikalischer Forschung in den Kreisen der Wissenschaft und allmählich auch in der Laienwelt eingebürgert hatten. Aber wir leben in einer Zeit äußerst schneller Entwicklung der physikalischen Theorien: Heute bezeichnet man schon die BOLTZMANNsche Gastheorie und alles, was in ähnlichem Sinne verläuft, wie z. B. die eben erörterten Theorien der BROWNschen Bewegung, der radioaktiven Strahlung oder der PLANCKschen Oszillatoren als „klassische Statistik" und stellt dieser ganz neuartige Überlegungen gegenüber, die unter dem Namen „Quantenstatistik" zusammengefaßt werden und die sich noch viel weiter von den altgewohnten Anschauungen der Physik früherer Zeiten entfernen. Bevor ich auf diesen letzten Punkt meiner Ausführungen eingehe, muß ich noch einige Bemerkungen einschalten über das Verhältnis der klassischen Statistik zum sogenannten Kausalgesetz.

Es ist nun freilich hier nicht möglich und wäre auch durchaus nicht angebracht, alle Fragen zu erörtern, die sich gebräuchlicherweise an das Wort Kausalgesetz anknüpfen. Ich kann dies um so eher unterlassen, als wir kürzlich von PHILIPP FRANK eine eingehende, ausgezeichnete Untersuchung dieses Fragenkomplexes erhalten haben. Was uns hier angeht und kurz besprochen werden soll, ist allein der in der Naturwissenschaft der Gegenwart lebende Begriff einer „kausalen Erklärung" von Naturvorgängen. Wir können von der älteren Geschichte seiner Entstehung aus vorwissenschaftlichen Gedankengängen absehen und seinen Ursprung dorthin verlegen, wo der Beginn der heute als „klassisch" zu bezeichnenden Naturwissenschaft liegt: in das schon einmal von mir genannte grundlegende Werk von NEWTON, die „Philosophiae naturalis principia mathematica". Hier war, in der Mechanik der festen, punktförmigen Körper, das erste Modell einer derartigen Beschreibung beobachtbarer Vorgänge gegeben, wie man sie als eine „kausale" Erklärung oder schlechthin als eine „Erklärung" bezeichnet. Auf einige wenige ganz knappe Grundgesetze oder Axiome war eine außerordentliche Fülle und Mannigfaltigkeit von Erscheinungen zurückgeführt, in dem Sinne, daß alle diese Erscheinungen als logische Folgerungen aus jenen Axiomen hervorgingen. Welches war das entscheidende Hilfsmittel, das diesen Aufbau der Theorie ermöglichte?

Das Schema der „kausalen" Erklärung.

Es ist kein Zufall, daß NEWTON zugleich einer der Erfinder der Differential- und Integralrechnung war. Untrennbar verbunden mit der Grundidee der NEWTONschen Mechanik, der klassischen Physik überhaupt, ist die Anwendung der Infinitesimalrechnung zur Durchführung der Deduktionen, die von den Axiomen bis zur Beantwortung konkreter Einzelfragen führen. Das Schema der Überlegungen ist regelmäßig dieses: Die Axiome oder Grundgesetze liefern eine Differentialgleichung, d. h. eine Beziehung zwischen den verschiedenen Werten der physikalischen Variablen innerhalb eines örtlich und zeitlich eng begrenzten Gebietes. Man kann sich vorstellen, daß — etwa in der Mechanik — eine Beziehung zwischen den Geschwindigkeiten und Lagen des Körpers in zwei unmittelbar aufeinanderfolgenden Zeitpunkten gegeben ist. Unter „Integration" versteht man dann den mathematischen Vorgang, der aus dieser Beziehung eine solche zwischen den Lagen und Geschwindigkeiten in weiter auseinanderliegenden Zeiten ableitet.

Das Ergebnis der Integration kann aber immer nur dahin gehen, daß man instand gesetzt wird, aus der Kenntnis des Anfangszustandes, d. i. der Lage und Geschwindigkeit zu einer Zeit, auf den Endzustand, d. i. die Lage und Geschwindigkeit zu einer späteren Zeit, zu schließen. Schon hierin liegt eine starke Einschränkung in der Befriedigung unseres Kausalitätsbedürfnisses. Wir erfahren nicht, warum die Erde sich heute mit einer Geschwindigkeit von 29,6 km/sek in einer Entfernung von etwa 23400 Erdradien von der Sonne in einer bestimmten Richtung bewegt; es sei denn, wir lassen es als Beantwortung, als kausale Erklärung gelten: weil sie vor einem Monat die und die Geschwindigkeit, sowie die und die Entfernung von der Sonne hatte. In der Tat ist es im wesentlichen nur eine Frage der Gewohnheit, wenn wir in den Ergebnissen, zu denen die NEWTONsche Mechanik führt, jeweils eine kausale Begründung der Naturvorgänge sehen. Der Laie, der zum erstenmal hört, daß die gesamte exakte Mechanik nicht mehr leistet, als aus den anfänglichen Geschwindigkeiten die Geschwindigkeiten späterer Zeit vorauszusagen, wird dies viel eher eine Beschreibung denn eine Erklärung nennen wollen. Nur die Fachleute waren entrüstet, als ihnen durch KIRCHHOFF im Jahre 1874 die von Ernst MACH schon früher vertretene Formulierung bekannt wurde, unsere Mechanik könne lediglich eine (möglichst einfache) Beschreibung der Bewegungsvorgänge liefern.

Noch deutlicher werden die Verhältnisse, wenn wir ein wenig auf die historische Entwicklung eingehen. Wir sehen bei KEPLER, welche Fragen ursprünglich im Zusammenhang mit der Bewegung der Himmelskörper aufgeworfen wurden. Man kennt aus dem Schulunterricht die durch mühsame Einzelbeobachtungen aufgefundenen „drei KEPLERschen Gesetze", die durch die NEWTONsche Mechanik, wie man zu sagen pflegt, eine einheitliche Erklärung erhalten haben. Aber man vergißt oder übersieht dabei, daß KEPLER noch viele andere Gesetze aufgestellt hat, die sich in keinerlei Beziehung zur NEWTONschen Mechanik bringen lassen. So meinte er, daß die Halbachsen der Bahnen der fünf großen Planeten sich zueinander verhalten müßten, wie die Seiten der fünf platonischen Körper (Tetraeder, Würfel, Oktaeder usf.), die einer und derselben Kugel einbeschrieben sind. Es kommt hier nicht darauf an, ob diese Behauptung numerisch richtig ist oder nicht. Entscheidend ist, daß die NEWTONsche Mechanik in keiner Weise zu der Frage nach der Größe der Planetenbahnen Stellung nimmt oder nehmen kann. Es geht eben die Anfangslage des Planeten, also seine Entfernung von der Sonne zu irgendeinem Zeitpunkt, als eine gegebene, bzw. als gegeben vorausgesetzte Größe in die mechanischen Gleichungen ein. Nur ein starker intellektueller Verzicht gegenüber dem, was zur Zeit KEPLERS als „Planetenproblem" galt, ermöglicht uns heute die Auffassung, die in der NEWTONschen Mechanik eine Lösung dieses Problems sieht. Daran hat auch alle Weiterbildung der NEWTONschen Ansätze einschließlich der in unserer Zeit so berühmt gewordenen Relativitätstheorie nichts geändert.

Die Schranken der NEWTONschen Mechanik.

Man wird nun dagegen den Einwand erheben, daß die Tragweite der Mechanik nicht nur bis zur Darstellung der heutigen Gestirnbewegungen reicht, sondern darüber hinaus auch Schlußfolgerungen über die Entstehung des Sonnen- und Planetensystems gestattet, sich also doch mit der eben erwähnten KEPLERschen Fragestellung berührt. In der Tat könnten wir annehmen, daß die Welt, etwa im Sinne der LAPLACEschen Theorie, ursprünglich aus einer zusammengeballten, flüssigen Masse bestand, aus der sich dann einzelne Teile als selbständige Himmelskörper loslösten. Diesen ganzen Vorgang könnten wir uns selbstverständlich den Gesetzen der NEWTONschen Mechanik unterworfen denken. Aus den Anfangszuständen müßten dann auch die Endzustände folgen, also diejenigen Daten, die den Ausgangspunkt für die in unseren Zeitläuften stabil gebliebenen Himmelsbewegungen liefern. Aber hier greift eine andere Schwierigkeit ein und ich muß jetzt noch auf eine zweite, viel wesentlichere Schranke der klassischen Mechanik eingehen, die ich vorhin nur der Einfachheit halber vorläufig außer acht gelassen habe.

In den Differentialgleichungen der Mechanik und den aus ihnen abgeleiteten Integralen treten nicht nur die Koordinaten und Geschwindigkeiten des Körpers auf, sondern noch weitere Größen, die man „Kräfte" nennt, und die bekannt, zumindest in der Art ihrer Abhängigkeit von den erstgenannten Veränderlichen bekannt sein müssen, wenn man aus den Gleichungen Schlüsse der früher geschilderten Art ziehen will. Wir müßten also, um mit den NEWTONschen Ansätzen die Entstehung der Himmelskörper beschreiben zu können, erst die Kraftgesetze kennen, durch die sie bedingt war. Wenn wir sagen, daß durch die NEWTONsche Mechanik die Bewegungen der Himmelskörper erklärt oder in einfacher Weise beschrieben werden, so liegt dies daran, daß in alle Gleichungen, die zur Darstellung dieser Bewegungen aufgestellt werden, nur eine einzige Art von Kräften, man kann sagen, nur ein einziges Kraftgesetz eingeht. Sämtliche astronomische Vorgänge (mit einer einzigen kleinen Ausnahme) werden vollkommen befriedigend wiedergegeben, wenn man mit NEWTON annimmt, daß zwischen je zwei Körpern in der Richtung ihrer Verbindungslinie eine dem Quadrat des Abstandes umgekehrt proportionale Anziehungskraft wirkt. Es war ein ungeheurer Schritt zur Vereinfachung unserer Naturerkenntnis, als NEWTON nachwies, daß das Fallen eines Apfels vom Baum und die Drehung des Mondes um die Erde durch Gleichungen beschrieben werden, in die nur eine und dieselbe Annahme über das Vorhandensein einer äußeren „Kraftwirkung" eingeht. Unter diesen Umständen waren wir gerne geneigt, diese Darstellung als eine „kausale Erklärung" anzusehen, als Zurückführung der Erscheinungen auf eine einzige

„Ursache“, eben die genannte Gravitationskraft. Allmählich zeigte sich, daß man fast alle Feinheiten der astronomischen Bewegungen herausbekam, wenn man die Annahme der allgemeinen Gravitation in alle Konsequenzen durchführte. Aber darüber hinaus, d. h. über das, was vor Beginn der jetzt beobachtbaren Bewegungen, vor endgültiger Bildung des heutigen Sonnensystems stattfand, darüber lehrt das Anziehungsgesetz nichts.

Allerdings wenn wir von NEWTONscher Mechanik sprechen, meinen wir ein umfassenderes Lehrgebäude als das der astronomischen Theorie allein. Die Bewegungen aller Körper auf der Erde werden durch die Differentialgleichungen, die NEWTON angegeben hat, richtig dargestellt, wenn wir nur die richtigen Kraftgrößen oder Kraftgesetze in die Gleichungen einführen. Beispielsweise bewährt sich die klassische Mechanik ohne Einschränkung, wenn wir die Bewegungsvorgänge etwa der Teile einer Dampfmaschine oder einer Lokomotive untersuchen wollen. Hier müssen wir eben nur wissen, welche Kräfte im Dampfzylinder infolge der darin stattfindenden thermischen Vorgänge entwickelt werden.

Betrachten wir nun aber den mechanischen Vorgang am GALTONschen Brett, an dem eine große Zahl von gleichmäßigen Stahlkugeln zwischen den Nägelreihen herabfällt! Es erscheint zunächst selbstverständlich, daß die NEWTONschen Ansätze auch hier in Geltung bleiben. Allein was sollen wir als wirksame Kräfte ansehen und in die Gleichungen einführen? Nehmen wir an, daß nur die Schwerkraft einwirkt, so bekommen wir die feineren Züge des Vorganges, das, worauf es uns allein ankommt, nicht heraus. So oft eine Kugel auf einen der Nägel auftrifft, tritt ein Zustand ein, in dem ein außerordentlich geringer Krafteinfluß darüber entscheidet, ob die Kugel rechts oder links vom Nagel weiterrollt. Dieser Einfluß kann vielleicht von der Luftbewegung herkommen, die infolge der unvermeidlichen Druck- und Temperaturschwankungen im Zimmer besteht. Man müßte also den Verlauf der Luftbewegungen kennen, um aus der NEWTONschen Mechanik auf den Ablauf der Bewegung am GALTONschen Brett schließen zu können. Das „Kausalitätsbedürfnis“ wird aber noch keineswegs befriedigt, wenn wir eine Annahme über die Luftbewegung machen, aus der die beobachteten Tatsachen über die Bewegung der Kugeln sich ableiten lassen. Wir werden unbedingt weiterfragen, woher diese Luftbewegung kommt, also auch auf sie die klassische Mechanik anwenden und fragen, auf welche Kraftgesetze sich die Bewegung der Luft zurückführen läßt.

Nur wenn wir in letzter Linie aus einer einfachen Annahme alles herleiten können, werden wir das Gefühl haben, für die ins Auge gefaßten Vorgänge eine kausale Erklärung gefunden zu haben.

Die Einfachheit ein Kriterium der Kausalität.

Es zeigt sich demnach bei näherer Betrachtung, daß nicht das Bestehen des Schemas der klassischen Mechanik (oder der klassischen Physik überhaupt), also eines Differentialgleichungsansatzes, der bei gegebenen Anfangsbedingungen zu integrieren ist, eine kausale Erklärung der Naturvorgänge verbürgt, sondern daß noch ein gewisses Kriterium der Einfachheit der in die Gleichungen eingehenden Annahmen hinzukommt. Ein sehr lehrreiches Beispiel hierzu finden wir schon in der Theorie der Planetenbewegung selbst.

Hier hatte sich ergeben, daß eine bestimmte Einzelheit, nämlich die allmähliche Drehung der Ellipse, in der die Merkurbahn verläuft, nicht unter der einfachen Voraussetzung der Gravitation zwischen den beobachtbaren Massen erklärt werden konnte. Nun war es ja gewiß möglich, Kräfte anzunehmen, die diese Abweichung bewirkten, ohne daß die mechanischen Gleichungen verletzt würden; ja man konnte die Kräfte sogar als Gravitationskräfte ansetzen, wenn man sich vorstellte, daß etwa staubartig verteilte und daher unsichtbare Massen in der Nähe der Sonne vorhanden seien. Allein das „Kausalitätsbedürfnis" beruhigte sich bei diesen Hypothesen nicht. Als dann Einstein zeigte, daß man durch eine veränderte Art, Zeiten und Geschwindigkeiten zu messen, die Perihelbewegung des Merkur herausrechnen könne, ohne jede zusätzliche Annahme über die wirksamen Kräfte, da neigte sich die übereinstimmende Ansicht aller vernünftigen Physiker dieser Auffassung zu, obwohl sie in anderer Richtung starke intellektuelle Opfer forderte. Für unser Problem der kausalen Naturerklärung können wir hieraus nur lernen: Die Ansätze der klassischen Physik, die alle Vorgänge formal als eindeutig determiniert erscheinen lassen, befriedigen nur dann unser Kausalitätsbedürfnis, können nur dann als eine kausale Naturerklärung aufgefaßt werden, wenn es gelingt, mit genügend einfachen Annahmen in den Gleichungsansätzen auszukommen.

Diese Voraussetzung der Einfachheit ist nun eben bei denjenigen physikalischen Vorgängen, auf die wir die Wahrscheinlichkeitsrechnung anwenden, nicht erfüllt und nicht erfüllbar. Es ist eine vollkommene Illusion, zu meinen, daß etwa die Bewegung der Kugeln am GALTONschen Brett durch die Differentialgleichungen der Mechanik „kausal" erklärt werden können, weil diese der äußeren Form nach den Bewegungsablauf eindeutig bestimmen. Glaubt man etwa durch Verfeinerung der Meßmethoden Aufschlüsse gewinnen zu können, so gerät man in noch unsichereres Fahrwasser. Suchen wir nämlich die Einzelvorgänge in der Luft oder an der Oberfläche der festen Körper unter dem Mikroskop näher zu verfolgen, so finden wir nur eines: daß überall Vorgänge vom Typus der BROWNschen Bewegung auftreten. Statt also auf etwas Einfacheres (im Sinne der klassischen Physik) gelangt man auf immer kompliziertere und unübersichtlichere Vorgänge — vom Standpunkt der deterministischen Auffassung aus gesprochen. Man weiß heute, daß alle unsere Präzisionsmessungen, auch etwa die simple Längenmessung, eine unübersteigliche Genauigkeitsgrenze in dem Vorhandensein der BROWNschen Bewegung finden. Andere, prinzipiell bedeutsamere Schwierigkeiten genauer Messungen werden noch zu besprechen sein. Jedenfalls muß man auf diesem Wege zur Erkenntnis kommen, daß die gewohnten, anschaulichen ja suggestiven Vorstellungen der klassischen Physik nicht unbegrenzt anwendungsfähig sind.

Verzicht auf die Kausalitätsvorstellung.

Nun erscheint hier, in der Wahrscheinlichkeitsrechnung, eine Theorie auf dem Plan, die für ganz bestimmte Vorgänge — die von uns als Massenerscheinungen gekennzeichneten — mit Hilfe bestimmter, logisch einwandfreier Überlegungen aus einfachen Voraussetzungen (über die Ausgangswahrscheinlichkeiten) zu bestimmten Schlüssen über den Ablauf der Vorgänge gelangt. Die Aussagen dieser statistischen Theorie unterscheiden sich schon der Form nach von denen der deterministischen — nur über das, was in der Mehrzahl der Fälle bei hinreichender Wiederholung der Versuche geschieht, wird etwas behauptet — so daß sie nicht in Widerspruch, ja nicht einmal in Konkurrenz mit jenen treten können. Die empirische Brauchbarkeit der Theorie, die Übereinstimmung ihrer Schlußfolgerungen mit den Beobachtungsergebnissen, ist durch zahllose Versuche auf allen Gebieten erwiesen. Das, was der allgemeinen Annahme der statistischen Theorie heute noch im Wege steht oder bis vor kurzem im Wege gestanden hat, ist ein gewisser intellektueller Widerstand, der mehr psychologisch als logisch zu erklären ist.

Man darf nicht unterschätzen, welche Macht die jahrhundertelange Gewöhnung an eine voll deterministische Naturauffassung auf unseren Intellekt ausübt. Wir sind seit einigen Jahrzehnten — wie ich an anderer Stelle einmal auszuführen Gelegenheit hatte — in einen Zeitraum eingetreten, der durch eine besondere Blüte der spekulativen Naturwissenschaften, durch eine früher ungeahnte Erweiterung und Umwandlung unseres naturwissenschaftlichen Weltbildes gekennzeichnet wird. Das erste große Ereignis dieser entwicklungsgeschichtlichen Epoche war das Entstehen der Einsteinschen Relativitätstheorie, die in weiten Kreisen über die der Physiker hinaus Aufsehen erregt hat. Aber alles, was uns hier an Verzicht auf althergebrachte, durch Jahrhunderte sozusagen geheiligte Vorstellungen und Denkgewohnheiten zugemutet wird, die Aufgabe oder Modifikation des üblichen Zeit- und Raumbegriffes, erscheint fast geringfügig gegenüber den grundstürzenden Neuerungen, zu denen man von den Anfängen der kinetischen Gastheorie her über Wahrscheinlichkeitstheorie und Statistik, in der Quantentheorie, in der Wellenmechanik gelangt. Es stellt sich als unabweisbar heraus, noch eine liebgewordene Position aufzugeben, die aus dem praktischen Leben, dem vorwissenschaftlichen Denken stammt und von eilfertigen Philosophen in die unantastbare Höhe ewiger Denkkategorien erhoben wurde: den naiven Kausalitätsbegriff.

Das Kausalgesetz.

Wenn man die reichlich unbestimmten und wechselvollen Fassungen betrachtet, die das Kausalgesetz im Laufe der Zeit bei den maßgebenden Philosophen gefunden hat, so muß man zu dem Schluß kommen, daß kaum die Gefahr, vielleicht nicht einmal die Möglichkeit besteht, mit ihm in Widerspruch zu geraten.

Wenn es in der ersten Auflage von KANTS Kritik der reinen Vernunft heißt: „Alles, was geschieht (anhebt zu sein), setzt etwas voraus, worauf es nach einer Regel folgt", oder diese Formulierung in der zweiten Auflage durch die folgende ersetzt wird: „Alle Veränderungen geschehen nach dem Gesetze der Verknüpfung von Ursache und Wirkung", so scheint es nicht allzu schwer, beliebige, auch statistische Gesetzmäßigkeiten in eine Form zu kleiden, die sich diesem „Prinzip" anpaßt. Es kommt eben nur darauf an, was man im gegebenen Fall als eine Veränderung oder als ein Geschehen bezeichnet, was man unter Ursache und Wirkung versteht.

Als GALILEI auf dem Wege der Beobachtung das Trägheitsgesetz fand, widersprach es gewiß der damaligen Vorstellung von Kausalität, daß dauernde Ortsänderung ohne fortwirkende Ursache vor sich gehen soll. Nachdem aber Generationen von Physikern den Trägheitssatz und alle seine Folgerungen als eine brauchbare Grundlage für die systematische Beschreibung der Bewegungserscheinungen erkannt hatten, fügte sich auch die philosophische Betrachtungsweise. Heute steht in der Beurteilung sämtlicher Philosophen das Trägheitsgesetz nicht nur im Einklang mit dem Kausalitätsprinzip, sondern es ist nach SCHOPENHAUER und vielen anderen eine unausweichliche, denknotwendige Folge desselben: nur die Änderung des Geschwindigkeitszustandes, nicht die des Ortes bedarf einer Verursachung. Hätten die Beobachtungen der Physiker ergeben, daß erst die dritten Ableitungen der Koordinaten nach der Zeit durch unabhängig feststellbare Umstände, die dann eben die Kräfte hießen, bestimmt werden, so würde ohne Zweifel der Satz, daß ein sich selbst überlassener Körper, ohne Einwirkung äußerer Ursachen, eine parabolische Bahn durchläuft, als Erfüllung oder gar als notwendige Konsequenz des Kausalprinzips angesehen werden.

So lassen sich auch für statistische Aussagen Formen finden, die dem Gesetz von Ursache und Wirkung entsprechen. Wenn bei lang dauerndem Spiel mit zwei Würfeln die Doppelsechs durchschnittlich alle 36mal erscheint, so hat dieses Geschehen die „Ursache", daß jede Seite der beiden Würfel durchschnittlich gleich oft fällt, und jede merkliche Abweichung von der Häufigkeit $^1/_{36}$ hat zur „Ursache", daß der eine oder andere Würfel ein „falscher" ist. Oder wenn die Kugeln, die am GALTONschen Brett herabfallen, eine Glockenkurve zeigen, so ist die „Ursache" hierfür die, daß eine einzelne Kugel, die man oft genug fallen läßt, gleich häufig rechts und links an einem Nagel vorbeigeht. Gewiß kann man dem entgegenhalten, daß für das Einzelergebnis beim Würfelspiel oder beim GALTONschen Brett die Ursache fehlt, bzw. hier nicht angegeben wird, oder daß Abweichungen innerhalb kurzer Serien nicht auf Ursachen zurückgeführt werden; aber auch beim Trägheitssatz hat man ja schließlich auf die von der naiven Auffassung geforderte Ursache der Ortsveränderung verzichtet und begnügt sich mit einer Ursache für die Änderung der Geschwindigkeit. So scheint mir der folgende Ausblick am Platze zu sein: Wenn erst einmal die Physik oder ganz allgemein die auf fortschreitende Beobachtungen gegründete Naturwissenschaft die Schlußweisen und Gedankengänge der Statistik sich voll zu eigen gemacht und als unentbehrliche Hilfsmittel anerkannt hat, dann wird nach einiger Zeit niemand mehr die Empfindung haben, daß hier irgendeine Denknotwendigkeit unerfüllt, irgendein philosophisches Bedürfnis unbefriedigt geblieben ist. Mit einem Wort: Das Kausalprinzip ist wandelbar und wird sich dem unterordnen, was die Physik verlangt.

Die neue Quantenstatistik.

Ist sonach keineswegs zu befürchten, daß aus sogenannten philosophischen Gründen, besser gesagt aus Gründen der Anpassung an ältere Denkgewohnheiten, ein bleibendes Hindernis für die Aufnahme der statistischen Theorie entsteht, so kann anderseits nicht verschwiegen werden, daß die Fortbildung der physikalischen Theorie zu Schwierigkeiten geführt hat, die zumindest ein Teil der heutigen Physiker für nicht überwunden hält. Ich will mit einer kurzen Besprechung dieser Fragen, die in der von DE BROGLIE, SCHRÖDINGER, HEISENBERG und BORN geschaffenen neuen Quantenstatistik auftreten, meinen Vortrag beschließen.

Erinnern wir uns zunächst daran, was am Ende des letzten Vortrages zum Gegenstand der Fehlertheorie gesagt worden war. Jede physikalische Messung ist ein wiederholbarer Vorgang, der das Element eines Kollektivs bildet. Ein Kollektiv ist vollständig gegeben, wenn man seine Verteilung kennt, d. i.

die Gesamtheit der Wahrscheinlichkeiten, die den einzelnen Wertbereichen des Merkmales, hier also des Messungsergebnisses zukommt. Um eine Verteilung zu charakterisieren, kann man. wie wir ebenfalls schon wissen, zwei Größen benutzen, die durch die Verteilung bestimmt werden, ihren Mittelwert und ihre Streuung. Den Mittelwert erhält man, indem man jeden möglichen Wert des Merkmales mit seiner Wahrscheinlichkeit multipliziert und diese Produkte addiert, für den normalen Würfel also $\frac{1}{6}(1+2+3+\ldots+6)=\frac{21}{6}=3{,}5$. Die Streuung wird gebildet, indem man zunächst von jedem der möglichen Werte den Mittelwert abzieht, die Differenz ins Quadrat erhebt, sie dann mit der Wahrscheinlichkeit des betreffenden Wertes multipliziert und schließlich diese Produkte aufsummiert. Im Würfelspiel $\frac{1}{6}(1-3{,}5)^2+\frac{1}{6}(2-3{,}5)^2+\ldots+\frac{1}{6}(6-3{,}5)^2 =$
$= \frac{1}{6}(6{,}25+2{,}25+0{,}25+0{,}25+2{,}25+6{,}25)=\frac{1}{6}.17{,}5=2{,}92$
Natürlich werden zwar durch eine Verteilung Mittelwert und Streuung eindeutig bestimmt, umgekehrt aber gehören noch sehr viel verschiedene Verteilungen zu gegebenen Beträgen von Mittelwert und Streuung.

Handelt es sich um ein Kollektiv, das aus den Messungen einer physikalischen Größe besteht, so nennt man den Mittelwert der Verteilung, in Anlehnung an ältere, aus der klassischen Physik stammende Vorstellungen, den wahren Wert der zu messenden Größe. In der Streuung sieht man üblicherweise ein Maß für die Ungenauigkeit oder Unschärfe der Messung; wie weit das berechtigt ist, werden wir noch sehen. Die Streuung ist ihrer Definition nach nur dann gleich Null, wenn die ganze Wahrscheinlichkeit 1 auf einen einzigen Wert entfällt, der dann natürlich auch der Mittelwert ist. Hat man daher einen Fall mit sehr kleiner Streuung, so kann man annehmen, daß fast die ganze Wahrscheinlichkeit sich auf Werte konzentriert, die nahe an einem festen Werte liegen, daß also fast sicher nur sehr geringe Abweichungen vom „wahren Wert“ als Messungsergebnisse auftreten.

Die Frage, um die es sich nun dreht, ist die: Kann man für jede unbekannte Größe ein Meßverfahren angeben, dessen Streuung beliebig klein, in letzter Linie also gleich Null ist? Die klassische Physik nahm etwas derartiges als selbstverständlich an, und dieser Auffassung entsprang auch die Vorstellung vom „wahren Wert" der Messung. In der ersten Periode der statistischen Theorien, in der man sich mit den Gasmolekülen, BROWNschen Partikeln, Alphateilchen usw. beschäftigte, nahm man an, daß es in dieser Hinsicht einen prinzipiellen Unterschied zwischen zweierlei Arten von Größen gibt, den „makroskopischen" und „mikroskopischen", entsprechend zwei prinzipiell verschiedenen Teilen der Physik, der „Makrophysik" und „Mikrophysik". Zu den ersteren zählte man die Variablen, die nur an einer großen Gesamtheit von Molekülen usw. definiert sind, wie z. B. die Länge eines Stabes von festem Material. Derartige Längen, meinte man, seien niemals vollkommen scharf meßbar, eben weil sie nur an statistischen Gesamtheiten auftreten. Dagegen galt es als stillschweigend vorausgesetzt, daß für alles, was ein einzelnes Elementarteilchen betrifft, z. B. die Koordinaten und Geschwindigkeitsgrößen eines Moleküls, völlig exakte Meßverfahren existieren, d. h. solche, bei denen es keine von Null verschiedene Streuung gibt.

Gibt es exakte Messungen?

Die neue Quantenstatistik aber lehrt, daß auch für die Elementarteilchen, und für diese ganz besonders, Meßverfahren prinzipiell ausgeschlossen sind, mit denen man gleichzeitig alle ein Teilchen betreffenden Größen *genau* bestimmen könnte. Es besteht vielmehr eine von der Theorie gelieferte, von vornherein bekannte Beziehung zwischen den Streuungen verschiedener zusammengehöriger Messungen, und diese Beziehung läßt erkennen, daß immer, wenn eine Streuung sehr klein wird, eine bestimmte andere, die mit ihr zusammenhängt, sehr groß werden muß. Diese Verhältnisse wollen wir etwas näher studieren.

Woher kommt man eigentlich zu der Annahme, daß es Meßverfahren gibt, mit denen sich eine Größe beliebig genau bestimmen läßt? Wenn wir etwa mit dem Metermaß die Länge einer Tischkante nachmessen und dabei die Ablesung auf volle Zentimeter beschränken, so gehört nicht viel Sorgfalt dazu, um ein *eindeutiges* Resultat zu erzielen, d. h. um zu erreichen, daß eine beliebig oft vorgenommene Wiederholung der Messung immer wieder dieselbe ganze Zahl von sagen wir 97 cm liefert. Wir haben hier ein Kollektiv vor uns, das praktisch die Streuung Null besitzt. Aber dies bezeichnen wir doch nicht als eine genaue oder exakte Messung, weil das Verschwinden der Streuung nur erreicht wurde durch die Wahl einer genügend groben Maßeinheit. Das Resultat besagt nur, daß die fragliche Länge in ein Gebiet fällt, das ungefähr zwischen 96,5 und 97,5 cm liegt. Darüber etwa, daß diese Grenzen genau innegehalten seien, können wir auch nichts behaupten.

Es gibt nun natürlich viel genauere Verfahren der Längenbestimmung. Wenn der Geodät eine sogenannte Basislänge festlegt, so liest er mit Hilfe einer geeigneten, sehr kostspieligen Meßvorrichtung im Mikroskop Tausendstel von Millimetern ab. Wiederholte aufeinanderfolgende Messungen geben verschiedene Werte der (ganzen) Zahl, die abgelesen wird. Man hat es hier mit einem richtigen Kollektiv zu tun, dessen Streuung sicherlich nicht Null ist. Gleichwohl sagt man, dieses Meßverfahren sei ein genaueres, weil sein Ergebnis nach der üblichen Auffassung sich in der Form aussprechen läßt, die gesuchte Größe falle in ein Gebiet von vielleicht 0,1 mm Ausdehnung od. dgl. Würde man bei diesem Meßverfahren sich von vornherein auf Größen von ein Zehntelmillimeter beschränken, so würde man vielleicht auch hier die Streuung Null erhalten.

Man könnte sich etwa vorstellen, daß durch weitere Verfeinerung der technischen Hilfsmittel Verfahren gewonnen werden, die auch ein Tausendstelmillimeter oder noch weniger ohne Streuung abzulesen gestatten. Aber dies hätte kaum viel Sinn. Man kann nicht annehmen, daß die Erdkruste derart starr sei, daß zwei Punkte in 200 m Entfernung mit der Zeit nicht Schwankungen erleiden, die sie um plus minus ein Tausendstelmillimeter nähern oder entfernen. Hier ist von Fragen der Mikrophysik noch gar nicht die Rede, es liegt ein durchaus makroskopischer Fall vor, in dem auch nach den älteren Anschauungen eine genaue Messung gar nicht möglich sein soll.

Wie steht es nun, wenn das Meßobjekt der Mikrophysik angehört?

Ort und Geschwindigkeit eines Materieteilchens.

Der Physiker W. Heisenberg, einer der Begründer der Quantenmechanik, hat zum erstenmal in konkreter Form dargestellt, wohin es führt, wenn man versucht, die Meßgenauigkeit für die mechanischen Variablen, Ort und Geschwindigkeit oder Ort und Impuls eines Materieelementes beliebig weit zu steigern.

Wollen wir zunächst das Teilchen möglichst genau lokalisieren, d. h. seinen Ort feststellen, so müssen wir es im Gesichtsfeld eines Mikroskops beleuchten, um sein Zusammenfallen mit einem bestimmten Teilstrich des Maßstabes zu konstatieren. Die Schärfe, mit der ein Mikroskop Entfernungen zu unterscheiden gestattet, hängt von der Wellenlänge des verwendeten Lichtes ab. Die kleinste Entfernung, die man wahrnimmt, ist der Wellenlänge direkt proportional. Will man also möglichst genau beobachten, so muß man ein Licht von geringer Wellenlänge, d. h. großer Schwingungszahl verwenden.

Nun besteht aber, nach der heutigen Auffassung von der Natur des Lichtes, das Beleuchtetwerden eines Körpers darin, daß eine große Menge von bewegten Lichtquanten auf ihn auffällt, was, im ganzen genommen, einen statistischen Vorgang von der Art der BROWNschen Bewegung oder der Molekülbewegung in einem Gas darstellt. Die Energie der einzelnen Lichtquanten ist der Schwingungszahl des Lichtes proportional. Durch die Stoßenergie der Lichtquanten wird jetzt der Bewegungszustand des betrachteten Materieteilchens beeinflußt, naturgemäß um so stärker, je größer die Energie, also die Schwingungszahl des Lichtes ist (sogenannter COMPTON-Effekt). Will man also die Geschwindigkeit des Teilchens möglichst richtig bestimmen, so muß man dafür sorgen, daß die Schwingungszahl des verwendeten Lichtes möglichst gering ist. Man sieht demnach: Steigerung der Meßgenauigkeit für den Ort des Materieelementes bedeutet Verringerung der Meßgenauigkeit für die Geschwindigkeit und umgekehrt. Die Aufgabe, beide Arten von mechanischen Bestimmungsstücken, Koordinaten und Impulse, gleichzeitig an einem Teilchen genau zu messen, ist prinzipiell unlösbar.

Man hat es oft so dargestellt, als wäre es das Wesentliche an dem hier geschilderten Tatbestand, daß der Meßvorgang den Zustand des Meßobjekts beeinflußt und dadurch der Genauigkeit Schranken gesetzt werden. Aber daran liegt es nicht. Auch in der Makrophysik, wenn man z. B. den hydrodynamischen Druck an einer Stelle einer Strömung mißt, wird durch den Meßapparat selbst der Druck etwas verändert, aber man weiß hier, welche Korrekturen anzubringen sind. Das Entscheidende ist, daß das Auftreffen der Lichtquanten auf das Materieteilchen als ein regelloser, nur statistisch erfaßbarer Vorgang aufgefaßt wird, für den es eine deterministische Theorie im Sinne der NEWTONschen Mechanik nicht gibt.

Das wichtigste Ergebnis der HEISENBERGschen Betrachtung läßt sich dahin aussprechen: Messungsergebnisse sind immer Kollektivs. Bei makroskopischen Messungen, wenn wir etwa die Länge eines Lineals zu bestimmen suchen, ist das Meßobjekt selbst eine statistische Gesamtheit, nämlich ein Haufen bewegter Moleküle, und daher hat der Begriff „absolut exakte Länge" überhaupt keinen faßbaren Sinn. In der Mikrophysik, d. h. wenn wir als Meßobjekt ein einzelnes Teilchen der statistischen Gesamtheit nehmen, wird durch das Meßmittel, die Gesamtheit der Lichtquanten, das Regellose, das Indeterminierte hineingebracht. Wir entgehen auf keinem Fall dem Indeterminismus, wenn wir auf die konkreten Tatbestände, von denen die physikalische Theorie handelt, zurückgreifen.

HEISENBERGS Unschärferelation.

Die heutige Quantenmechanik bildet nach der jetzt allgemein angenommenen Auffassung eine rein statistische Theorie. Ihre Axiome sind Differentialgleichungen, die die Wahrscheinlichkeitsverteilungen für Koordinaten und Impulse der Materieteilchen zu verschiedenen Zeiten miteinander verknüpfen. Man bemüht sich noch vielfach um eine Deutung dieser Gleichungen und eine „Ableitung" aus Begriffsbildungen der klassischen Mechanik, mit denen sie sicher durch vielfache formale Beziehungen verknüpft sind. Aber hier wird es vielleicht so gehen, wie es einmal mit den MAXWELLschen Gleichungen der Elektrodynamik gegangen ist. Nachdem man jahrzehntelang versucht hat, sie mechanisch zu „deuten", durch Einführung verborgener Massen und komplizierter mechanischer Systeme, hat man sich schließlich daran gewöhnt, sie ohne solche „Deutung" als Elementargesetze hinzunehmen. Allerdings liegt es in der Quantenmechanik schwieriger, weil hier die einzelnen Ansätze jedesmal Beziehungen zu speziellen mechanischen Systemen erkennen lassen.

Unter den Konsequenzen, zu denen die Quantenmechanik führt, hat besonders die folgende die Aufmerksamkeit in Anspruch genommen. Zwischen den beiden Verteilungen, die einerseits der Koordinate, anderseits der Impuls- (oder Geschwindigkeits-) Größe eines Teilchens zugehören, besteht eine bestimmte Beziehung, die unter anderm darin zum Ausdruck kommt, daß das Produkt der beiden Streuungen einen festen Wert hat, der von den sonstigen speziellen Daten des Problems unabhängig ist. Der Größenordnung nach ist dies Produkt gleich dem Quadrat des früher erwähnten PLANCKschen Wirkungsquantums, das in den üblichen physikalischen Maßeinheiten, Zentimeter, Sekunde, Grammasse, etwa 6.10^{-27} beträgt. In dem neuen Satz über das Produkt der Streuungen, der als die HEISENBERGsche Unschärferelation bezeichnet wird, spricht sich der Tatbestand allgemein aus, den wir vorhin an dem Beispiel einer Beobachtung im Mikroskop dargelegt haben: Je genauer die Messung der Koordinate, um so ungenauer ist die der Geschwindigkeit.

Natürlich ist der Satz von dem Produkt der Streuungen ein rein theoretisches Ergebnis und er ist demgemäß im mathematisch präzisen Sinn zu verstehen. D. h. es wird dabei angenommen, daß bei jeder einzelnen der individuellen Beobachtungen, deren Gesamtheit das Kollektiv bildet, die betreffende Zahl völlig genau abgelesen wird. Wenn wir also eine Meßeinrichtung derart gestalten, daß die aufeinanderfolgenden Teilstriche der Längenskala Abstände von einem Billionstelmillimeter (10^{-13} cm) aufweisen und anderseits die Impulsgrößen auch nur auf eine Einheit von 10^{-13} gcm/sec abgelesen werden, so steht vom Standpunkt der Theorie dem nichts im Wege, daß als Messungsresultat jedesmal die gleichen Zahlen erscheinen, daß also die Messung sowohl des Ortes wie der Geschwindigkeit praktisch ohne Streuung erfolgt. Wir befinden uns dann, bei dieser Beschränkung der Maßeinheit, genau in dem gleichen Fall wie bei der makroskopischen Messung einer Kantenlänge mittels eines Meterstabes, der nur Zentimeterteilung besitzt.

Manche Physiker glauben seit der Auffindung der Unschärferelation, allen Boden unter den Füßen verloren zu haben: Wenn es prinzipiell keine exakten Messungen gibt, was hätten dann unsere exakten physikalischen Theorien noch für Sinn? Ich glaube nicht, daß solche Befürchtungen berechtigt sind. Die Resultate der Wellen- oder Quantenmechanik sind in demselben

Sinne brauchbar wie alle bewährten Sätze der klassischen Makrophysik. Was stört es uns, wenn wir den Eintritt einer Sonnenfinsternis auf Sekunden genau vorhersagen können, daß es nicht möglich und auch nicht sinnvoll ist, auf Billionstelsekunden genau zu prophezeien! Im Grunde genommen liegt hier nur eine gewisse Inkongruenz zwischen rein mathematischen Begriffsbildungen (Präzision ohne Ende) und der physikalischen Wirklichkeit vor, ein Tatbestand, der uns doch nicht fremd ist.

Was ist also in letzter Linie der Sinn der HEISENBERGschen Unschärferelation? Wir müssen in ihr einen großen Schritt zur Vereinheitlichung des physikalischen Weltbildes erblicken. Denn bis dahin dachte man, es gäbe zwei grundsätzlich verschiedene Arten von Beobachtungen in der Natur, solche, die statistischen Charakters seien, und daher nicht zu beliebiger Genauigkeit gesteigert werden können; und andere „rein atomare“ Erscheinungen, die in mathematischem Sinne exakt determiniert sind. Wir sehen jetzt, daß diese Zweiteilung ein unbegründetes Vorurteil war. Gewiß will ich nicht sagen, daß es nun gar keinen Unterschied mehr zwischen den extremsten Bereichen der Physik, etwa der Mechanik der Planetenbewegung und der Theorie des radioaktiven Zerfalls gibt. So einfach ist die Naturbeschreibung nicht, daß sich alles mit einem Schema behandeln ließe. Aber eine gewisse, scheinbare Gegensätzlichkeit wird durch die neue Betrachtungsweise aufgelöst.

Folgerungen für das physikalische Weltbild.

Die Folgerungen, die sich aus all dem Gesagten für das allgemeine naturwissenschaftliche Weltbild ergeben, können hier nur lose angedeutet werden.

Wir finden vor allem keinen Grund, uns in unserem Glauben an die Zweckmäßigkeit deterministischer Theorien in weiten Bereichen der Physik erschüttern zu lassen. Diese Theorien, auf sichere Erfahrungsgrundlagen gestützt, liefern Ergebnisse, die mit Beobachtungen gut übereinstimmen, und gestatten, künftige Erscheinungen vorauszusehen in einem Umfang, der das ganze menschliche Leben von Grund auf umgestaltet hat. Der größte Teil der heutigen Technik, dies Wort im weitesten Sinne genommen, beruht noch auf den Voraussagen der klassischen Mechanik und Physik.

Seit langem war es bekannt oder zumindest jedem Einsichtigen klar, daß die aus den mathematischen Ansätzen der klassischen Theorien gezogenen Folgerungen sich in der Wirklichkeit nicht mit unbegrenzter Genauigkeit im mathematischen Sinn verifizieren lassen. Schon die atomistischen Vorstellungen des Altertums wiesen darauf hin und auch die Auffassung des Lichtes als einer Wellenbewegung mußte den Gedanken an die Existenz solcher Grenzen nahelegen. Den ersten umfassenden Deutungsversuch für die Natur der Genauigkeitsgrenzen der Beobachtungen lieferte die Entwicklung der Atomstatistik (kinetischen Gastheorie) in der zweiten Hälfte des 19. Jahrhunderts. Sie lehrt uns, daß die Voraussagen der klassischen Physik nur in dem Sinne determiniert sind wie Sätze der Wahrscheinlichkeitsrechnung, die sich auf die Gesetze der großen Zahlen stützen: „Es ist bei großem n fast sicher, daß...“. Dabei zeigt das Eingehen auf die ins Spiel kommenden Werte von n (Atom-, Molekülzahlen usf.), wie enorm groß, von der Sicherheit kaum verschieden, in normalen Fällen die Wahrscheinlichkeiten sind, um die es sich handelt. In diesem Stadium der Entwicklung und vielfach noch bis in die letzten Jahre wurde die — mit unserer Auffassung des Wahrscheinlichkeitsbegriffes unvereinbare — Annahme gemacht, daß das atomare Geschehen, die Bewegung der einzelnen Elementarteilchen, durch die Gesetze der deterministischen Mechanik bestimmt wird.

Von dieser Beschränkung, die eine logisch einwandfreie Fassung der Theorie verhinderte, sind wir jetzt, dank den Erkenntnissen der Quantenmechanik, befreit. Es gibt zwar, wie wir jetzt wissen, neben der klassischen Physik der Erscheinungen „im großen“ noch eine Mikrophysik, eben die Quanten- oder Wellenmechanik. Aber deren Differentialgleichungen verknüpfen nur Verteilungsfunktionen, also Wahrscheinlichkeiten, miteinander und sie gestatten daher hinsichtlich der Elementargrößen nur solche Voraussagen für die Zukunft, die die Struktur von Wahrscheinlichkeitsaussagen besitzen. Eine exakte, d. h. von Streuung freie Messung ist auch im atomaren Gebiet nur möglich, wenn man sich in der Wahl der Maßeinheit von vornherein einer bestimmten Beschränkung unterwirft. Daß es sich dabei im wesentlichen um Billionstel Millimeter handelt, ist nur praktisch, nicht prinzipiell von Bedeutung.

Ich habe mich in diesem Vortrag von einer Ausdehnung der Betrachtungen über das anorganische Gebiet hinaus auf das der Biologie ferngehalten und will diese mit Absicht innegehaltenen Grenzen auch jetzt nicht überschreiten. Das soll aber nicht heißen, daß ich eine solche Übertragung für unmöglich oder unzulässig ansehe. Ich glaube nur, daß die Vorgänge, die man biologische nennt, noch weit verwickelter liegen als das, was man heute als Physik (oder Chemie) bezeichnet, und daß noch wesentliche Ergänzungen der physikalischen Theorien erforderlich sein werden, ehe man hier wird einiges Begründete sagen können.

Schlußbetrachtung.

Überblicken wir noch einmal in raschem Fluge den ganzen Weg, den wir in diesen Vorträgen durchlaufen haben! Wir waren ausgegangen von einer Reinigung des sprachüblichen Wortbegriffes „Wahrscheinlichkeit“. In einer bestimmten Klasse beobachtbarer, geläufiger Erscheinungen, wie dem Würfelspiel u. ähnl., fanden wir eine geeignete Grundlage, um Definitionen und Axiome einer exakt-naturwissenschaftlichen Theorie daran anzuknüpfen. Kollektiv, Grenzwert der relativen Häufigkeiten, Regellosigkeit der Anordnung bezeichnen die Ausgangspunkte, die vier Rechnungsoperationen, die an Kollektivs vorgenommen werden können, Auswahl, Mischung, Teilung und Verbindung, stellen das entscheidende Instrument für den Aufbau der Wahrscheinlichkeitstheorie dar.

Daß immer nur Beziehungen zwischen Wahrscheinlichkeiten gesucht und gefunden werden können, daß in diesem Sinne die Wahrscheinlichkeitsrechnung sich nicht von anderen Gebieten der exakten Wissenschaften unterscheidet, ließ uns einen sicheren Standpunkt gegenüber den erkenntnistheoretischen Unzulänglichkeiten älterer Auffassungen, namentlich der sogenannten Gleichmöglichkeitsdefinition, gewinnen. Die neueren Versuche zu einer Modifikation der von mir entwickelten Grundlagen wurden erörtert, ohne daß sich daraus zwingende Gesichtspunkte für die Annahme wesentlich veränderter Auffassung ergaben. Die klassischen Gesetze der großen Zahlen und ihre modernen Erweiterungen erhielten im Rahmen der auf den Häufigkeitsbegriff gestützten Auffassung eine klare Deutung als Aussagen über bestimmte beobachtbare Erscheinungen.

Das erste große Anwendungsgebiet der Wahrscheinlichkeitstheorie, das wir behandelten, bildete die Statistik im gewöhnlichen Sinne dieses Wortes: das Studium der Zahlenreihen, die wir durch Abzählung gleichförmiger Vorgänge im menschlichen Leben bilden. Wir sahen, wie z. B. die umfassende Statistik Marbes über das Geschlechtsverhältnis der Geburten sehr gute Übereinstimmung mit Rechnungsergebnissen der Wahrscheinlichkeitstheorie aufweist. Andere statistische Reihen, wie die Todesfalls- oder Selbstmordstatistik, konnten, ohne daß sie unmittelbar Kollektivs im Sinne der mathematischen Definition darstellen, auf Kollektivs zurückgeführt werden. Hier bieten die Methoden, die auf wahrscheinlichkeitstheoretische Begriffe gestützt sind, z. B. die Lexissche Dispersionstheorie, ein Mittel rationeller, systematischer, zusammenfassender und einteilender Beschreibung der Vorgänge, kurz das, was man im gewöhnlichen Sprachgebrauch eine „Erklärung" nennt. Die Fehlertheorie, d. i. die Statistik der Messungsergebnisse und ähnlicher Schwankungserscheinungen, stellt einen Übergang zu den Fragestellungen der physikalischen Statistik dar.

Diese Problemgruppe nimmt heute unser Interesse in stärkstem Maß in Anspruch, da mit ihr eine Umwälzung des gesamten naturwissenschaftlichen Weltbildes verknüpft ist. Wir sehen, wie zunächst Boltzmann vor fast 70 Jahren den kühnen Schritt unternimmt, eine Wahrscheinlichkeitsaussage an Stelle eines Naturgesetzes zu formulieren. Noch ist die Auffassung zögernd und in sich widerspruchsvoll, denn es wird versucht, das der Wahrscheinlichkeitstheorie entsprechende Verhalten aus Gesetzen der klassischen, deterministischen Mechanik abzuleiten, was nicht möglich ist und den energischen Widerspruch von Mach hervorruft. Wie reine wahrscheinlichkeitstheoretische Betrachtungen in guter Übereinstimmung mit nachweisbaren Beobachtungen die Erklärung bestimmter Erscheinungsreihen liefern, haben wir u. a. an den Beispielen der Brownschen Bewegung und der durch Alphastrahlung hervorgerufenen Szintillationen gesehen. Diese Untersuchungen leiten naturgemäß zu den Fragen nach der Bedeutung des sogenannten Kausalgesetzes, nach dem Verhältnis von Determinismus und Indeterminismus hinüber. Wir mußten erkennen, wie hier die Fortschritte der Physik, in gleicher Weise, wie es seit Jahrhunderten der Fall ist, zum schrittweisen Aufgeben vorgefaßter Meinungen führen, auch solcher, die durch hergebrachte philosophische Systeme dogmatisiert wurden. Die neue Quantenmechanik, und in ihr die HEISENBERGsche Unschärferelation, vollendet gewissermaßen den Aufbau der statistischen Naturauffassung, indem sie zeigt, daß keine Beobachtung, sei es im großen oder im kleinen, völlig „exakt", d. h. ohne Dazwischenkunft statistischer Vorgänge erfolgen kann.

So glaube ich, in diesen Vorträgen dargetan zu haben, was ich im Titel und in der Einleitung zum Ausdruck bringen wollte: daß man, von einem logisch geklärten, auf allgemeine Erfahrungsgrundlagen gestützten Wahrscheinlichkeitsbegriff ausgehend, durch Betrachtungen, die nach dem gewöhnlichen Sprachgebrauch als statistische bezeichnet werden, zur Erkenntnis der Wahrheit in einem weiten Bereich menschlicher Interessen gelangen kann.

Zusammenfassung der sechs Vorträge in fünfzehn Leitsätzen.

1. Die Aussagen der Wahrscheinlichkeitsrechnung lassen sich nicht richtig erfassen, wenn man die Bedeutung des Wortes „Wahrscheinlichkeit" dem allgemeinen Sprachgebrauch entnimmt; sie gelten vielmehr nur für einen bestimmten, künstlich abgegrenzten, rationellen Wahrscheinlichkeitsbegriff.

2. Von Wahrscheinlichkeit kann in rationellem Sinn erst dann die Rede sein, wenn in jedem einzelnen Fall das Kollektiv genau beschrieben ist, innerhalb dessen die Wahrscheinlichkeit betrachtet wird; Kollektiv heißt eine gewissen Forderungen genügende Massenerscheinung, ein Wiederholungsvorgang, allgemein eine Folge von Beobachtungen, die man sich unbegrenzt fortsetzbar denkt.

3. Wahrscheinlichkeit eines Merkmals (Beobachtungsergebnisses) innerhalb eines Kollektivs heißt der Grenzwert, dem die relative Häufigkeit des Auftretens dieses Merkmals in der Beobachtungsfolge bei unbegrenzter Fortsetzung der Versuche sich nähert; dieser Grenzwert bleibt ungeändert, wenn man die Beobachtungsfolge einer beliebigen Stellenauswahl unterwirft: Prinzip der Regellosigkeit oder des ausgeschlossenen Spielsystems.

4. Die Aufgabe der Wahrscheinlichkeitsrechnung im engeren Sinne besteht ausschließlich darin, aus den als gegeben voraus-

gesetzten Wahrscheinlichkeiten innerhalb gewisser „Ausgangskollektivs" die Wahrscheinlichkeiten innerhalb solcher Kollektivs zu rechnen, die aus den ersteren abgeleitet werden; die hier gemeinte Ableitung neuer Kollektivs läßt sich stets auf die, eventuell wiederholte, Durchführung von vier einfachen Grundoperationen zurückführen.

5. Von dem Zutreffen eines Wahrscheinlichkeitswertes in irgendeinem konkreten Fall — es mag sich um die Ausgangswerte einer Rechnung oder ihre Ergebnisse handeln — kann man sich allein nur durch den statistischen Versuch, d. h. die Durchführung einer genügend langen Beobachtungsreihe, überzeugen; es gibt weder eine a-priori-Erkenntnis der Wahrscheinlichkeit noch auch die Möglichkeit, mit Hilfe einer anderen Wissenschaft, etwa der Mechanik, auf Wahrscheinlichkeitswerte zu schließen.

6. Was man in der klassischen Theorie als Definition der Wahrscheinlichkeit ansieht, ist nur der Versuch, den allgemeinen Fall auf den speziellen der Gleichverteilung (bei dem alle Merkmale innerhalb des Kollektivs gleiche Wahrscheinlichkeit besitzen) zurückzuführen; eine solche Zurückführung ist oft unmöglich, wie bei der Sterbenswahrscheinlichkeit, oft führt sie zu Widersprüchen, wie beim BERTRANDschen Problem; keinesfalls enthebt sie der Notwendigkeit, die Wahrscheinlichkeit im Falle der Gleichverteilung erst noch zu definieren. Ohne die Ergänzung durch die Häufigkeitsdefinition kann die Wahrscheinlichkeitsrechnung überhaupt zu keiner Aussage über den Ablauf von Erscheinungen führen.

7. Die sogenannten Gesetze der großen Zahlen enthalten nur dann sinnvolle Aussagen über den Ablauf einer Beobachtungsfolge, wenn man von vornherein von der Häufigkeitsdefinition der Wahrscheinlichkeit ausgeht; sie stellen dann bestimmte, aus der Regellosigkeit folgende Eigenschaften der Anordnung der Beobachtungsergebnisse fest; bei Verwendung der klassischen „Definition" liefern die Gesetze der großen Zahlen lediglich rein arithmetische Eigenschaften gewisser Gruppen von ganzen Zahlen, also Aussagen, die in keinerlei Beziehung zum zeitlichen Ablauf von Vorgängen stehen.

8. Die Aufgabe, die der Wahrscheinlichkeitsrechnung innerhalb der sogenannten mathematischen Statistik zufällt, be-

steht darin, zu untersuchen, ob eine vorgegebene statistische Aufnahme ein Kollektiv bildet, bzw. ob und in welcher Weise sie sich auf Kollektivs von möglichst einfacher Verteilung zurückführen läßt. Eine solche Zurückführung liefert eine zusammenfassende, systematische Beschreibung, also eine „Erklärung" der statistischen Erscheinungen.

9. Keine der bisher bekannt gewordenen Theorien, die einen Widerspruch gegen die Wahrscheinlichkeitsrechnung konstruieren, wie die Theorie vom statistischen Ausgleich, die Knäuelungstheorie, das Gesetz der Serie, haben sich als begründet erwiesen.

10. Die Theorie der Beobachtungsfehler, die an der Grenze zwischen der allgemeinen und der physikalischen Statistik steht, beruht auf der Auffassung, daß jede physikalische Messung das Element eines Kollektivs bildet, dessen Mittelwert „wahrer Wert" der zu messenden Größe heißt. Durch zusätzliche Annahmen über dieses Kollektiv gelangt man zu den verschiedenen Aussagen der Fehlertheorie.

11. Die Sätze einer statistischen Theorie in der Physik unterscheiden sich grundsätzlich von denen jeder deterministischen Theorie: sie stellen immer nur eine Behauptung darüber auf, welchen Verlauf ein Versuch oder eine größere Versuchsreihe bei hinreichend häufiger Wiederholung in der überwiegenden Mehrheit der Fälle nehmen wird; diese Mehrheit kann eine so starke sein, daß praktisch der Unterschied aufgehoben wird.

12. Die zeitlich aufeinanderfolgenden Beobachtungen an einem sich selbst überlassenen physikalischen System bilden unmittelbar kein Kollektiv. Aber sie lassen sich mit den Gedankenmitteln der rationellen Wahrscheinlichkeitstheorie (Nachwirkungs-Erscheinungen, Markoffsche Ketten) in befriedigender Weise erklären.

13. Die Auffassung, daß eine statistische Theorie „im großen" (Makrophysik) mit einer deterministischen Theorie für die Elementarteilchen (Mikrophysik) vereinbar sei, steht im Widerspruch zu der Auffassung des Wahrscheinlichkeitsbegriffes, des Kollektivs usf., wie sie in diesem Buche vertreten wird.

14. Die Quanten- oder Wellenmechanik erscheint heute als eine rein statistische Theorie, deren Grundgleichungen Ver-

knüpfungen von Wahrscheinlichkeitsverteilungen zum Ausdruck bringen. Die aus ihnen folgende Unschärferelation besagt, daß Messungen im atomaren Bereich, ebenso wie die der Makrophysik, Kollektivs bilden, also nur dann streuungsfrei verlaufen können, wenn die Maßeinheit nicht zu fein gewählt ist.

15. Daß eine statistische Theorie eine „nur vorläufige" Art der Naturerklärung gegenüber einer deterministischen, das „Kausalitätsbedürfnis" befriedigenden darstellt, ist ein Vorurteil, das aus der geschichtlichen Entwicklung der Naturwissenschaften verstanden werden kann, das aber mit zunehmender Einsicht verschwinden muß.

Anmerkungen und Zusätze.

Eine knappe, populäre Darstellung der Hauptgedanken des vorliegenden Buches findet man in dem ersten Teil meines Aufsatzes: Marbes „Gleichförmigkeit in der Welt und die Wahrscheinlichkeitsrechnung", Die Naturwissenschaften (Julius Springer, Berlin), 7. Jahrg. 1919, H. 11, 12, 13, S. 168 bis 175, 186 bis 192, 205 bis 209. — Die mathematische Begründung habe ich erstmals gegeben in einer Abhandlung „Grundlagen der Wahrscheinlichkeitsrechnung", Mathematische Zeitschrift (Julius Springer, Berlin), Bd. 5, 1919, S. 52 bis 99. — Eine Übersicht der Theorie in französischer Sprache findet sich in den Annales de l'Institut H. Poincaré Paris, Vol. III, 1932, pag. 137 bis 190 (Wiedergabe von Vorträgen, die im November 1931 im Institut gehalten wurden). — Eine ausführliche lehrbuchmäßige Darstellung erschien 1931 unter dem Titel: Vorlesungen aus dem Gebiete der angew. Mathematik, Bd. I: Wahrscheinlichkeitsrechnung u. ihre Anwendung in der Statistik, Fehlertheorie und in der theoretischen Physik (Franz Deuticke, Wien und Leipzig). — Mit dem Thema des vorliegenden Buches, namentlich dem Gegenstand des letzten Abschnittes, beschäftigen sich auch die folgenden Vorträge des Verfassers: „Über die gegenwärtige Krise der Mechanik", Ztschr. f. angew. Math. u. Mech., Bd. I, 1921, S. 425 bis 431 und Die Naturwissenschaften, Bd. 10, 1922, S. 25 bis 29; „Über kausale und statistische Gesetzmäßigkeit in der Physik", Die Naturwissenschaften, Bd. 18, 1930, S. 145 bis 153 und Erkenntnis, Bd. 1, 1930, S. 189 bis 210; „Über das naturwissenschaftliche Weltbild der Gegenwart", Rede beim Stiftungsfest der Berliner Universität, Berlin 1930 und Die Naturwissenschaften, Bd. 18, 1930, S. 885 bis 893.

Erster Vortrag: Definition der Wahrscheinlichkeit.

Zu S. 2, Zeile 2: Georg Christoph Lichtenberg, Vermischte Schriften, erster Teil, II, 1. (Neue, von dessen Söhnen veranstaltete Originalausgabe, Göttingen 1853, Bd. 1, S. 79.)

Zu S. 3, Zeile 3: Deutsches Wörterbuch von JAKOB und WILHELM GRIMM, Bd. 13, Leipzig 1922, S. 994. Der heutige englische Sprachgebrauch kennt „verylike“ nicht.

Zu S. 3, Zeile 21: R. EISLER, Wörterbuch der philosophischen Begriffe, 3. Aufl., Berlin 1910, S. 1743 (dritter Band).

Zu S. 4, Zeile 19: W. SOMBART, Der proletarische Sozialismus, 10. Aufl., Jena 1924, Bd. 1, S. 4.

Zu S. 5, Zeile 4: J. KANT, Kritik der reinen Vernunft, Methodenlehre, 1. Hauptstück, 1. Abschn., 2. Aufl. 1787, S. 758. Mit der hier dargelegten Lehre von den Definitionen stimmen die folgenden Ausführungen des Buches nicht überein.

Zu S. 6, Zeile 14: Über den Begriff der mechanischen Arbeit und über die Begriffsbildung in den exakten Wissenschaften überhaupt erhält der Leser (auch der nicht-mathematische) die beste Auskunft in: E. MACH, Die Mechanik in ihrer Entwicklung, historisch-kritisch dargestellt (1883), 7. Aufl., Leipzig 1921; noch präziser in desselben Verfassers: Prinzipien der Wärmelehre, historisch-kritisch entwickelt (1896), 4. Aufl., Leipzig 1923, S. 406 bis 432 (Abschnitte: Die Sprache, Der Begriff, Der Substanzbegriff). Dem von MACH vertretenen Standpunkt folgt im wesentlichen die Auffassung des Textes.

Zu S. 7, Zeile 9: Über den Streit um das „wahre Kraftmaß“ vgl. etwa MACH, Mechanik, a. a. O., S. 247 und 288. Der durch LEIBNIZ angeregte Streit dauerte 57 Jahre und wurde erst 1743 durch D'ALEMBERTS „Traité de dynamique“ beigelegt.

Zu S. 10, Zeile 34: GOETHES Aufsatz aus den Propyläen, ersten Bandes, erstes Stück (Werke, Ausg. letzter Hand, 12°, Bd. 38, 1830, S. 143 bis 154), der das Wort Wahrscheinlichkeit im Sinne von Illusion gebraucht, zeigt ein viel feineres Sprachgefühl als die früher angeführten Sätze der Philosophen. Eine auf dem Theater dargestellte Szene muß nicht wahr scheinen, aber einen Schein von Wahrheit vermitteln.

Zu S. 11, Zeile 1: A. A. MARKOFF sagt in seinem Lehrbuch der Wahrscheinlichkeitsrechnung, deutsche Ausgabe von H. LIEBMANN, Leipzig und Berlin 1912, S. 199: „Wir können daher keineswegs mit dem Akademiker BUNJAKOWSKI übereinstimmen (Grundlagen der mathematischen Theorie der Wahrscheinlichkeit, S. 326), daß man eine bekannte Klasse von Erzählungen abtrennen muß, an denen zu zweifeln er für ungehörig hält.“

Zu S. 11, Zeile 5: PIERRE SIMON (später Marquis DE) LAPLACE, 1749 bis 1827, veröffentlichte 1814 den „Essai philosophique des probabilités“, der dann als „Introduction“ der späteren Auflagen seines Hauptwerkes der „Théorie analytique des probabilités“ (1. Aufl. 1812, 2. Aufl. 1814, 3. Aufl. 1820) diente. Der „Essai“, der einen unbeschränkten Determinismus vertritt, ist ein charakteristischer Ausdruck der philosophischen Anschauungen im Frankreich des 18. Jahrhunderts. Eine handliche Neuausgabe erschien in der

Sammlung Les maîtres de la pensée scientifique, Paris 1921. Deutsche Übersetzung herausg. von R. v. MISES in Ostwalds Klassiker der exakten Wissenschaften, Bd. 233, Leipzig 1932.

Zu S. 11, Zeile 14: SIMÉON DENIS POISSON, 1781 bis 1840, veröffentlichte 1837 unter dem Titel „Recherches sur la probabilité des jugements en matière criminelle et en matière civile" ein durchaus mathematisch gehaltenes Lehrbuch der Wahrscheinlichkeitsrechnung, das nur im fünften Kapitel auf das im Titel genannte Problem eingeht, im übrigen eines der für die Entwicklung der mathematischen Theorie wichtigsten Werke darstellt. Deutsche Ausgabe: Lehrbuch der Wahrscheinlichkeitsrechnung und deren wichtigsten Anwendungen, bearbeitet von C. H. SCHNUSE, Braunschweig 1841.

Zu S. 16, Zeile 28: Die mathematische Definition des Grenzwertes kann etwa so gefaßt werden. Wir sagen von einer unbegrenzten Reihe fortlaufend numerierter Zahlen, die alle zwischen 0 und 1 liegen mögen, sie strebe einem Grenzwerte zu, wenn folgendes zutrifft: Wie groß auch k ist, sollen die ersten k Dezimalstellen der Zahlen unserer Reihe von einer bestimmten (von k abhängenden) Nummer ab unverändert bleiben.

Zu S. 19, Zeile 32: Deutsche Sterblichkeitstafeln aus den Erfahrungen von 23 Lebensversicherungsgesellschaften, Berlin 1883. — Kurze Angaben über diese und andere Tafeln findet man bei E. CZUBER, Wahrscheinlichkeitsrechnung, 2. Bd., 3. Aufl., Leipzig und Berlin 1921, S. 140ff.

Zu S. 21, Zeile 35: JOH. V. KRIES, Die Prinzipien der Wahrscheinlichkeitsrechnung, eine logische Untersuchung (1886), 2. Abdruck, Tübingen 1927, S. 130.

Zu S. 22, Zeile 33: LAPLACE, Essai (Anm. 11/5), S. 11 der deutschen Ausgabe.

Zu S. 25, Zeile 34: Das Lehrbuch von POISSON ist oben (Anm. 11/14) genannt worden, ebenso das von LAPLACE (Anm. 11/5). Die übrigen Schriften sind: JOHN VENN, The logic of chance, London and Cambridge 1866. TH. FECHNER, Kollektivmaßlehre, herausg. von G. F. LIPPS, Leipzig 1897. H. BRUNS, Wahrscheinlichkeitsrechnung und Kollektivmaßlehre, Leipzig und Berlin 1906. G. HELM, Die Wahrscheinlichkeitslehre als Theorie der Kollektivbegriffe in Annalen der Naturphilosophie, Bd. 1, 1902, S. 364 bis 381.

Zweiter Vortrag: Elemente der Wahrscheinlichkeitsrechnung.

Zu S. 35, Zeile 8: In dem Referat von E. CZUBER in der Enzyklop. d. mathem. Wissensch., Bd. I (2. Teilband), S. 736, heißt es über die Auffassung der Subjektivisten: „Nach dem ersten (sc. dem Prinzip des mangelnden Grundes) stützt sich die Konstatierung der Gleichmöglichkeit auf absolutes Nichtwissen über die Bedingungen des Daseins oder der Verwirklichung der einzelnen Fälle . . ."

Zu S. 42, Zeile 36: Der mathematisch unterrichtete Leser wird erkennen, daß die Wahrscheinlichkeitsdichte der Differentialquotient der Wahrscheinlichkeit nach den das Merkmal bestimmenden Variablen ist. Bezeichnen wir mit x den Zahlenwert, den etwa eine zu messende physikalische Größe annehmen kann, und mit $w(x)$ die Wahrscheinlichkeitsdichte, so bedeutet $w(x)\,dx$ die Wahrscheinlichkeit dafür, daß das Meßergebnis in das Intervall x bis $x + dx$ fällt. Die Dichte muß dabei der Bedingung genügen, daß $\int_{-\infty}^{\infty} w(x)\,dx = 1$.

Zu S. 50, Zeile 38: Die Mischungsregel für den Fall stetiger Verteilungen lautet folgendermaßen. Es ist die Wahrscheinlichkeit dafür, daß eine zu beobachtende Größe, die die Wahrscheinlichkeitsdichte $w(x)$ besitzt, in den Bereich $x = a$ bis $x = b$ fällt, gleich $\int^{b} w(x)\,dx$.

Zu S. 56, Zeile 7: Die berühmte Abhandlung von TH. BAYES ist erst nach dem Tode des Verfassers von R. PRICE herausgegeben worden unter dem Titel: An essay toward solving a problem in the doctrine of chances, London, Philos. Transactions 53 (1763), S. 376 bis 398 und 54 (1764), S. 298 bis 310. Eine deutsche Übersetzung erschien in Ostwalds Klassikern der exakten Wissenschaften, Nr. 169, Leipzig 1908.

Zu S. 68, Zeile 14: Die Kenntnis dieser Fragestellung verdanken wir dem Briefwechsel zwischen FERMAT und PASCAL, der uns erhalten ist. Man findet darüber (und über andere historische Fragen der älteren Zeit bis zu LAPLACE) Auskunft bei J. TODHUNTER, A history of the mathematical theory of probability. Reprint 1931 New-York, G. E. Stechert.

Dritter Vortrag: Kritik der Grundlagen.

Zu S. 77, Zeile 32, 35: JOHN MAYNARD KEYNES, Treatise on probability, London 1921. Deutsche Übersetzung von F. M. URBAN, u. d. T. Über Wahrscheinlichkeit, Leipzig 1926. Hauptsächlich das Kap. VIII (S. 73 bis 90 der deutschen Ausgabe) wendet sich gegen die Häufigkeitstheorie. — JOHN VENN, s. Anm. 25/32.

Zu S. 79, Zeile 23: HAROLD JEFFREYS, Scientific inference, Cambridge 1931. Hauptsächlich S. 8 bis 35.

Zu S. 82, Zeile 38: KANT, Kritik der reinen Vernunft, S. 40 (1. Ausgabe 1781): „Denn die geometrischen Sätze sind insgesamt apodiktisch, d. i. mit dem Bewußtsein ihrer Notwendigkeit verbunden." Dagegen z. B. B. RIEMANN, Abhandl. d. Kön. Gesellsch. d. Wissensch. Göttingen, 13. Bd., 1867: „. . . es lassen sich mehrere Systeme einfacher Tatsachen angeben, welche zur Bestimmung der Maßverhältnisse des Raumes hinreichen; am wichtigsten ist . . . das von Euklid zugrunde gelegte. Diese Tatsachen sind wie alle Tatsachen nicht notwendig, sondern nur von empirischer Gewißheit . . ."

Zu S. 83, Zeile 8: SCHOPENHAUER, Über die vierfache Wurzel des Satzes vom Grunde (Werke, herausg. v. FRAUENSTÄDT, Bd. I, Leipzig 1873, S. 42): „Aus dem Gesetze der Kausalität ergeben sich zwei wichtige Korollarien, welche eben dadurch ihre Beglaubigung als Erkenntnis a priori, mithin als über allen Zweifel erhaben und keiner Ausnahme fähig, erhalten: nämlich das Gesetz der Trägheit und das der Beharrlichkeit der Substanz."

Zu S. 83, Zeile 22: ED. V. HARTMANN, Philosophie des Unbewußten, 11. Aufl., Leipzig 1904, 1. Bd., S. 36 bis 47. Derselbe, Die Grundlagen des Wahrscheinlichkeitsurteils, in Vierteljahrschr. f. wiss. Philos. u. Soziologie, Bd. 28, 1904, S. 281. — Zur Kritik des Wahrscheinlichkeitsbegriffes bei HARTMANN vgl. auch die erstgenannte Stelle, S. 445 bis 452.

Zu S. 83, Zeile 29: TH. FECHNER, s. Anm. 25/32.

Zu S. 84, Zeile 16: H. POINCARÉ, Calcul des probabilités, 1. éd., Paris 1912, S. 24. Eine Übersicht über die Literatur zur Grundlegung der Wahrscheinlichkeitsrechnung (bis etwa 1916) gibt das Buch von E. CZUBER, Die philosophischen Grundlagen der Wahrscheinlichkeitsrechnung, Leipzig und Berlin 1923.

Zu S. 85, Zeile 7: JACOB BERNOULLI, Ars conjectandi, Basel 1713, z. B. S. 6, Propositio II: „Si a, b vel c expectem, quorum unumquodque pari facilitate mihi obtingere possit, expectatio mea aestimanda est $\frac{a+b+c}{3}$."

Zu S. 85, Zeile 22: A. MEINONG, Über Möglichkeit und Wahrscheinlichkeit, Beiträge zur Gegenstands- und Erkenntnistheorie, Leipzig 1915, 760 Seiten 8⁰!

Zu S. 87, Zeile 9: In der Theorie analytique von LAPLACE (s. Anm. 11/5) ist das 7. Kapitel (S. 402 bis 407 der 1. Ausg.) überschrieben: „De l'influence des inégalités inconnues qui peuvent exister entre des chances que l'on suppose parfaitement égales." Hier heißt es: „... $\frac{1+\alpha}{2}$ soit la possibilité d'amener pile ..." Es wird also statt probabilité das Wort possibilité gesetzt, weil die ursprüngliche Definition von probabilité nicht mehr anwendbar ist; dann aber werden für diese possibilité stillschweigend die Regeln als gültig angesehen, die für die probabilité abgeleitet wurden.

Zu S. 94, Zeile 35: J. M. KEYNES, a. a. O. (deutsche Ausgabe), S. 29ff. Als Hauptvertreter der Subjektivisten nenne ich C. STUMPF, Über den Begriff der mathematischen Wahrscheinlichkeit, Sitz.-Ber. d. Bayr. Akad. d. Wiss., philos.-hist. Kl., 1892, S. 41. Hier findet sich die folgende ausführlichere Darlegung über die Rolle des Nichtwissens: „Gleich möglich sind Fälle, in bezug auf welche wir uns in gleicher Unwissenheit befinden. Und da die Unwissenheit nur dann ihrem Maß nach gleichgesetzt werden kann, wenn wir absolut nichts darüber wissen, welcher von den unterscheidenden Fällen eintreten wird, so können wir noch bestimmter diese Erklärung dafür einsetzen."

Zu S. 96, Zeile 24: JOH. V. KRIES, Die Prinzipien der Wahrscheinlichkeitsrechnung, eine logische Untersuchung (1886), 2. Abdruck, Tübingen 1927. Die zitierte Formulierung, S. 157.

Zu S. 97, Zeile 17: Über das BERTRANDsche Paradoxon berichten die meisten bekannten Lehrbücher der Wahrscheinlichkeitsrechnung, z. B. E. CZUBER, Wahrscheinlichkeitsrechnung, Bd. 1, 3. Aufl., Leipzig und Berlin, S. 116ff. Mein Lehrbuch (s. oben), S. 78 bis 83.

Zu S. 102, Zeile 31: Zu der „älteren Generation", die sich ablehnend verhält, rechne ich vor allem eine Gruppe italienischer Mathematiker, die in den Jahren 1916 bis 1917, also vor Erscheinen meiner Arbeiten, die angebliche „Unzulässigkeit" der Annahme eines Grenzwertes der relativen Häufigkeiten bewiesen haben will. Vgl. F. P. CANTELLI, Annales de l'Institut Henri Poincaré, t. 5, 1935, p. 1—50. Tatsächlich wird hier gezeigt, daß man zu Widersprüchen gelangt, wenn man die Existenz dieses Grenzwertes voraussetzt, gleichzeitig aber dem Wort Wahrscheinlichkeit in den Sätzen der Wahrscheinlichkeitsrechnung eine andere Deutung beilegt als die, diesem Grenzwert gleich zu sein.

S. 103, Zeile 14: FRÉCHET et HALBWACHS, Le calcul des probabilités à la portée de tous, Paris 1924. — J. L. COOLIDGE, An introduction to mathematical probability, Oxford 1925. Deutsche Ausgabe: Einführung in die Wahrscheinlichkeitsrechnung, deutsch von F. M. URBAN, Leipzig 1927.

Zu S. 104, Zeile 10: HANS BLUME, Zur axiomatischen Grundlegung der Wahrscheinlichkeitsrechnung, 1934 (Dissertation Münster). Ferner zwei Aufsätze in Zeitschr. f. Physik, Bd. 92, 1934, S. 232 bis 252 und Bd. 94, 1935, S. 192 bis 203. Die Besprechung von K. KOLMOGOROFF in Zentralblatt für Mathematik, Bd. 10, 1935, S. 172.

Zu S. 104, Zeile 36: THEODOR FECHNER, s. Anm. 25/32.

Zu S. 107, Zeile 4: CARL G. HEMPEL, Erkenntnis (Annalen der Philosophie), Bd. 5, 1935, S. 228 bis 260.

Zu S. 108, Zeile 6ff.: Der hier besprochene Einwand bildet den Hauptgegenstand der Kritik von CANTELLI und anderer Italiener, von der Anm. 102/31 die Rede war.

Zu S. 112, Zeile 6: Eine gut verständliche Einführung in die Ideenbildung der intuitionistischen Mathematik (und ihr Verhältnis zur formalistischen) findet man bei A. FRAENKEL, Einleitung in die Mengenlehre, 3. Aufl., Berlin 1928, insbes. § 14. Hier heißt es S. 239: „Als extremes Beispiel einer intuitionistischen Konstruktion diene etwa die Bestimmung der aufeinander folgenden Ziffern eines Dezimalbruches durch Auswürfeln."

Zu S. 116, Zeile 5ff.: Zu den BERNOULLIschen Folgen vgl. E. BOREL, Traité du calcul des probabilités et de ses applications, t. II, fascic. 1, Applications à l'arithmétique, Paris 1926, hauptsächlich Chap. I. — Von den Arbeiten TORNIERS, der im übrigen einen wechselnden und nicht immer klaren Standpunkt in der Grundlagenfrage vertritt, seien genannt: Wahrscheinlichkeitsrechnung und Zahlentheorie, Journal f. d. reine und angew. Mathematik, Bd. 160, 1929, S. 177

bis 198; Die Axiome der Wahrscheinlichkeitsrechnung, ebda. Bd. 163, 1930, S. 45 bis 64; Grundlagen der Wahrscheinlichkeitsrechnung, Acta mathematica, Bd. 60, 1933, S. 239 bis 280. — Einige Arbeiten von A. H. COPELAND sind: Independent event histories, American Journal of mathematics, vol. 51, 1929, p. 612—618; The theory of probability from the point of view of admissible numbers, Annals of mathematical statistics, 1932, p. 143—156; Admissible numbers in the theory of geometrical probability, American Journal of mathematics, vol. 53, 1931, p. 153—162.

Zu S. 116, Zeile 9: E. KAMKE: Einführung in die Wahrscheinlichkeitstheorie, Leipzig 1932.

Zu S. 117, Zeile 6: COPELAND, s. Anm. 116/5. — H. REICHENBACH, Wahrscheinlichkeitslehre, Leiden 1935. Vgl. besonders § 29. — K. POPPER, Logik der Forschung (Schriften zur wissenschaftl. Weltauffassung, Bd. 9), Wien 1935. Vgl. besonders Nr. 50 bis 60.

Zu S. 118, Zeile 20: Außer den früher genannten Arbeiten COPELANDS vgl. die neu erschienene: Point set theory applied to the random selection of the digits of an admissible number, American Journal of mathematics, vol. 58, 1936, p. 181—192. — In der gleichen Richtung noch wesentlich weiter gehende Resultate bei A. WALD, Sur la notion de collectif dans le calcul des probabilités, Comptes rendus de l'Acad. des sciences, Paris, t. 202, 1936, p. 180—183.

Zu S. 118, Zeile 27—39: E. TORNIER, Mathematische Annalen, Bd. 108, 1933, S. 320. — E. KAMKE, Jahresber. d. deutschen Mathematiker Vereinigung, Bd. 42, 1933, S. 14 bis 27. — K. DÖRGE, s. folgende Anmerkung.

Zu S. 120, Zeile 9: Eine gut lesbare Übersicht der DÖRGEschen Aufstellungen findet man in seinem Aufsatz: Eine Axiomatisierung der VON MISESschen Wahrscheinlichkeitstheorie, Jahresber. d. deutschen Mathematiker-Vereinigung, Bd. 43, 1934, S. 39 bis 47. Die mathematischen Ausführungen sind gegeben in zwei Aufsätzen der Mathematischen Zeitschrift, Bd. 32, 1930, S. 232 bis 258 und Bd. 40, 1935, S. 161 bis 193.

Zu S. 124, Zeile 32: A. KOLMOGOROFF, Grundbegriffe der Wahrscheinlichkeitsrechnung (Ergebnisse der Mathematik und ihrer Grenzgebiete, 2. Bd., Heft 3), Berlin 1933.

Zu S. 126, Zeile 9: A. KOLMOGOROFF, a. a. O., S. 3.

Zu S. 126, Zeile 28: Besprechung von DOETSCH, Jahresber. d. deutschen Mathematiker-Vereinigung, Bd. 45, 1935, S. *153*.

Vierter Vortrag: Die Gesetze der großen Zahlen.

Zu S. 129, Zeile 6: Der wesentliche Inhalt dieses Abschnittes bis etwa S. 141 ist vom Verfasser erstmals veröffentlicht worden in Die Naturwissensch., Jg. 15, 1927, S. 479 bis 502.

Zu S. 129, Zeile 29: Das Werk von POISSON ist oben angeführt worden (Anm. 11/14). — JAC. BERNOULLI, Ars conjectandi, Basel 1713; deutsche Ausgabe von R. HAUSSNER in Ostwalds Klassiker

d. exakten Wissensch., Nr. 107 und 108, Leipzig 1899. Das hier in Frage kommende Gesetz findet sich im 4. Teil, Kap. 5, S. 236 (deutsche Ausgabe, 2. Bd., S. 104). — Das Zitat aus der POISSONschen Einleitung ist aus dem Original S. 7 übersetzt.

Zu S. 131, Zeile 26: Der allgemeine Satz von P. L. TSCHEBYCHEFF ist 1867 in russischer Sprache veröffentlicht, sodann im Journ. de Liouville, sér. II, t. 12, 1867. Die Ableitung des Satzes ist durchaus elementar.

Zu S. 135, Zeile 28: Den deus ex machina führt z. B. H. WEYL ganz unverhüllt ein: Philosophie der Mathem. u. Naturwissensch. (S.-A. aus Handb. d. Philos.), München und Berlin 1927, S. 151.

Zu S. 137, Zeile 24: Über die mathematischen Fragen, die an das „Gegenbeispiel" anknüpfen, vgl. G. PÓLYA und G. SZEGOE, Aufgaben und Lehrsätze aus der Analysis, Bd. I, Berlin 1925, S. 72 und 238. Hier wird auch das einschlägige Verhalten der Logarithmentafel erklärt. Weitere Beispiele und Untersuchungen ähnlicher, hier in Frage kommender Zahlenfolgen vgl. des Verf. Aufsatz in Mathematische Annalen, Bd. 108, 1933, S. 757 bis 772, Über Zahlenfolgen, die ein kollektivähnliches Verhalten zeigen.

Zu S. 140, Zeile 17, 21: G. PÓLYA, Nachr. v. d. Ges. d. Wissensch. Göttingen, math.-phys. Kl. 1921: „Alle Schriftsteller, die die Wahrscheinlichkeit als Grenzwert der Häufigkeit definieren, präsupponieren gewissermaßen ... nicht nur das Bestehen des BERNOULLIschen Gesetzes ..." — H. WEYL, a. a. O., S. 152.

Zu S. 143, Zeile 29: Die Originalarbeit von BAYES ist oben angeführt worden (Anm. 56/7). Das „BAYESsche Theorem" findet sich tatsächlich hierin nicht. Die Bezeichnung rechtfertigt sich nur dadurch, daß es sich um die Lösung eines von BAYES angeregten Problems handelt. Beides, Bezeichnung und Lösung, stammen von LAPLACE.

Zu S. 154, Zeile 10: Die hier referierte Verschärfung des Gesetzes der großen Zahlen wurde zum erstenmal von F. P. CANTELLI 1917 mitgeteilt. Unabhängig von ihm fand sie G. PÓLYA 1921. Der Gedankengang der im Text gegebenen Erklärung stammt von K. KOLMOGOROFF.

Zu S. 157, Zeile 6: Die hier gegebene Erweiterung der beiden Gesetze der großen Zahlen ist zum erstenmal angedeutet worden in des Verfassers „Vorlesungen", S. 192 bis 197. Die Einführung des allgemeinen Begriffes der statistischen Funktionen erfolgte in der Abhandlung „Deux nouveaux théorèmes de limite dans le calcul des probabilités", Revue de la Faculté des Sciences d'Istanbul, Bd. 1, 1935, Heft 1, S. 61 bis 80. Eine ausführliche Ableitung findet sich in dem Aufsatz „Die Gesetze der großen Zahlen für statistische Funktionen", Monatshefte f. Mathem. u. Physik, 43. Bd., 1936, S. 105 bis 128.

Fünfter Vortrag: Anwendungen in der Statistik und Fehlertheorie.

Zu S. 166, Zeile 32: KARL MARBE, Die Gleichförmigkeit in der Welt, Untersuchungen zur Philosophie und positiven Wissenschaft, München 1916; Mathematische Bemerkungen zu meinem Buch: „Die Gleichförmigkeit in der Welt“, ebenda, 1916. Die angeführte Stelle im 1. Band, S. 375.

Zu S. 167, Zeile 18: KARL MARBE, Grundfragen der angewandten Wahrscheinlichkeitsrechnung und theoretischen Statistik, München und Berlin 1934. Eine Besprechung dieses Werkes habe ich in Die Naturwissenschaften, 22. Jahrg., 1934, S. 741 bis 743, veröffentlicht.

Zu S. 168, Zeile 25: Vgl. dazu meinen zu Beginn dieser Anmerkungen genannten Aufsatz in den Naturwissenschaften 1919, sowie für die mathematische Erledigung des Problems meine Arbeit „Zur Theorie der Iterationen“, Ztschr. f. angew. Math. u. Mech., 1, 1921, S. 298 bis 307. Hier wird gezeigt, daß die Wahrscheinlichkeit für das x-malige Auftreten einer langen Iteration unter n Einzelbeobachtungen sich aus der Formel

$$w = \frac{e^{-a} a^x}{x!}$$

rechnet, wobei $a = n \,.\, 2^{-m-1}$ und m die Länge der Iteration bedeutet.

Zu S. 170, Zeile 22: O. STERZINGER, Zur Logik und Naturphilosophie der Wahrscheinlichkeitslehre, Leipzig 1911. PAUL KAMMERER, Das Gesetz der Serie, eine Lehre von den Wiederholungen im Lebens- und im Weltgeschehen, Stuttgart und Berlin 1919.

Zu S. 172, Zeile 22: F. EGGENBERGER und G. PÓLYA, Ztschr. f. angew. Math. u. Mech., Bd. 3, 1923, S. 279 bis 289.

Zu S. 176, Zeile 1: W. LEXIS, Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft, Freiburg i. B. 1877. Eine Darstellung dieser Theorie findet man in den meisten Lehrbüchern der Statistik.

Zu S. 183, Zeile 28: Die Zahlen sind entnommen der Österreichischen Statistik, Bd. 88, 1911, Heft 3, S. 20 und 120.

Zu S. 185, Zeile 4: Der algebraische Satz ist die sog. SCHWARZsche Ungleichheit. Sind $p_1, p_2, \ldots p_z$ die einzelnen Wahrscheinlichkeiten und p ihr Mittelwert, so gilt

$$p_1 (1 - p_1) + p_2 (1 - p_2) + \ldots p_z (1 - p_z) \leqq z \,.\, p\,(1 - p).$$

Zu S. 185, Zeile 23: Die Zahlen aus: Statistisches Jahrbuch für das Deutsche Reich, z. B. Bd. 43, 1923, S. 35.

Zu S. 188, Zeile 1: Zur Theorie der „wesentlichen Schwankungskomponente“ und zu den Zahlen der Selbstmordstatistik vgl. L. v. BORTKIEWICZ, Homogeneität und Stabilität in der Statistik, in Skandinavisk Aktuarietidskrift 1918, Heft 1/2. Das Zusatzglied zur Formel S. 184 lautet, wenn $p_1, p_2, \ldots p_n$ die Einzelwahrscheinlichkeiten und p ihr Mittelwert ist: $\frac{1}{n}\,[(p_1 - p)^2 + (p_2 - p)^2 + \ldots \ldots (p_n - p)^2]$.

Zu S. 191, Zeile 38: AD. QUETELET, Physique sociale ou essai sur le développement des facultés de l'homme, Brüssel-Paris-St. Petersburg 1869. Deutsche Ausgabe von VALENTINE DORN, Jena 1914.

Zu S. 192, Zeile 21: Die grundlegenden Arbeiten von MENDEL (1866 und 1870) sind von E. v. TSCHERMAK u. d. T. Versuche über Pflanzenhybriden, in Ostwalds Klassiker d. exakten Wissensch., H. 130, 4. Aufl., Leipzig 1923, neu herausgegeben worden. Zu den Zahlenbeispielen vgl. E. CZUBER, Statistische Forschungsmethoden, Wien 1921, S. 184.

Zu S. 193, Zeile 27ff.: Über die wichtigsten Fragen der Statistik in der Technik vgl. etwa: G. RÜCKLE und F. LUBBERGER, Der Fernsprechverkehr als Massenerscheinung mit starken Schwankungen, Berlin 1924; EMIL KOHLWEILER, Statistik im Dienste der Technik, München und Berlin 1931.

Zu S. 195, Zeile 10: Die angegebene Überlegung findet sich in dem sonst geistvollen Buche von E. BLEULER, Das autistisch-undisziplinierte Denken in der Medizin und seine Überwindung, Berlin 1921, S. 132 bis 145. Hierzu die Stellungnahme des Mathematikers PÓLYA, S. 145 bis 148.

Zu S. 199, Zeile 10: Über die gebräuchlichen statistischen Maßzahlen erhält man Auskunft in E. CZUBER, Die statistischen Forschungsmethoden, Wien 1921, insbesondere S. 57 bis 108. Der Nichtmathematiker mag auch F. ZIZEK, Die statistischen Mittelwerte, Leipzig 1908, zur Hand nehmen.

Zu S. 199, Zeile 16: Über das Disparitätsmaß von GINI vgl. den Bericht II von L. v. BORTKIEWICZ in XIX[e] Session de l'Institut international de Statistique Tokio 1930, La Haye 1930.

Zu S. 199, Zeile 17: Die PEARSONschen Verteilungstypen werden dargestellt bei W. PALIN ELDERTON, Frequency curves and correlation, London 1906.

Zu S. 199, Zeile 25 und 32: H. BRUNS, Wahrscheinlichkeitsrechnung und Kollektivmaßlehre, Leipzig und Berlin 1906. C. V. L. CHARLIER, Vorlesungen über die Grundzüge der mathematischen Statistik, Lund 1920. Die rationelle Methode der Beschreibung, von der im Text die Rede ist, ist die Entwicklung der Verteilungsfunktion in eine unendliche Reihe, deren Koeffizienten die charakteristischen Zahlen liefern. Die BRUNSsche Entwicklung geht von der GAUSschen Funktion e^{-x^2} aus, die CHARLIERsche von der POISSONschen Funktion $\frac{a^x e^{-x}}{x!}$. Die weiteren Entwicklungsglieder sind die Differential- bzw. Differenzenquotienten der ersten. Die mathematische Theorie der BRUNSschen Reihe ist in meiner Arbeit im Jahresber. d. deutschen Mathematiker-Vereinigung, 21, 1912, S. 9 bis 20, die der CHARLIERschen Reihe von H. POLLACZEK-GEIRINGER in Skandinavisk Aktuarietidskrift 1928, S. 98 bis 111 entwickelt. Vgl. auch den Aufsatz der letztgenannten Verfasserin: Die Statistik seltener Ereignisse in Die Naturwissenschaften, Jg. 16, 1928, S. 815 bis 820.

Zu S. 201, Zeile 12: Die grundlegenden Arbeiten von C. F. GAUSS (1821 und 1826) sind u. d. T. Abhandlungen zur Methode der kleinsten Quadrate von C. F. GAUSS, in deutscher Ausgabe von A. BÖRSCH und P. SIMON, Berlin 1887, erschienen.

Zu S. 201, Zeile 20: Der hier erwähnte LAPLACEsche Satz bildet einen Hauptgegenstand der mathematischen Untersuchungen zur Statistik und Wahrscheinlichkeitsrechnung. Vgl. u. a. meine Abhandlung: Fundamentalsätze der Wahrscheinlichkeitsrechnung, Mathem. Ztschr., Bd. 4, 1919, S. 1 bis 96, und für die neuere Entwicklung das Büchlein von A. KHINTCHINE, Asymptotische Gesetze der Wahrscheinlichkeitsrechnung (Ergebnisse der Mathematik und ihrer Grenzgebiete, 2. Bd., Heft 4), Berlin 1933.

Zu S. 206, Zeile 27: Die Arbeiten der PEARSONschen Schule findet man in der seit 1902 erscheinenden Zeitschrift: Biometrika, a journal for the study of biological problems. Das Hauptwerk von KARL PEARSON ist: Grammar of science, London 1900.

Sechster Vortrag: Probleme der physikalischen Statistik.

Zu S. 207, Zeile 28: Für die BOLTZMANNsche Theorie vgl. L. BOLTZMANN, Vorlesungen über Gastheorie, 2 Teile, 2. Abdr., Leipzig 1910. Die Grundgedanken der Theorie werden in allen Darstellungen der theoretischen Physik wiedergegeben.

Zu S. 208, Zeile 12: M. v. SMOLUCHOWSKI, Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik. Die Naturwissenschaften, Jg. 6, 1921, S. 253 bis 263 (PLANCK-Heft).

Zu S. 209, Zeile 15: Die Idee des LAPLACEschen „Dämon“ findet sich in dem Anm. 11/5 angeführten Essai philosophique.

Zu S. 210, Zeile 6: H. TH. BUCKLE, Geschichte der Civilisation in England. Übersetzt von J. H. RITTER, in 5 Bdn. Berlin o. J., 1. Bd., S. 25. Hier auch Angaben über die ersten Anfänge der Statistik.

Zu S. 214, Zeile 38: M. v. SMOLUCHOWSKI, s. Anm. 208/12. H. POINCARÉ in Science et méthode, Edition définitive, Paris o. J., S. 71. Deutsche Ausgabe von F. und L. LINDEMANN (Wissenschaft und Hypothese, Bd. 17), Leipzig 1914.

Zu S. 219, Zeile 12: Eine leicht verständliche Einführung in die ältere Atomtheorie mit vielen Zahlenangaben gibt das Buch von J. PERRIN, Die Atome, deutsch von A. LOTTERMOSER, 2. Aufl., Dresden und Leipzig 1920.

Zu S. 222, Zeile 1: ERNST MACH, Die Leitgedanken meiner naturwissenschaftlichen Erkenntnislehre und ihre Aufnahme durch die Zeitgenossen, Leipzig 1919, S. 9. Vgl. auch Wärmelehre, 4. Aufl., S. 364.

Zu S. 226, Zeile 39: Die grundlegenden Arbeiten von M. v. SMOLUCHOWSKI finden sich in den Sitzungsberichten der Wiener Akademie, math.-naturw. Kl., Abt. IIa, Bd. 123, 1914, S. 2381 bis 2405, und Bd. 124, 1915, S. 339 bis 368.

Zu S. 230, Zeile 20: Die Aufklärung der grundsätzlichen Fragen hinsichtlich des zeitlichen Ablaufes dieser und ähnlicher Erscheinungsreihen habe ich gegeben in der Abhandlung: Ausschaltung der Ergodenhypothese in der physikalischen Statistik, Physikal. Ztschr. 21, 1920, S. 225 bis 232 und 256 bis 262.

Zu S. 231, Zeile 1: Über die Rolle der MARKOFFschen Ketten in der physikalischen Statistik vgl. meine „Vorlesungen", § 16.

Zu S. 232, Zeile 20: Die Versuche von SVEDBERG sind beschrieben in: SVEDBERG, Existenz der Moleküle, Leipzig 1912, S. 148ff. Vgl. auch E. FÜRTH, Schwankungserscheinungen in der Physik, Braunschweig 1920.

Zu S. 234, Zeile 22: Zu diesem Gegenstand vgl. L. v. BORTKIEWICZ, Die radioaktive Strahlung als Gegenstand wahrscheinlichkeitstheoretischer Untersuchungen, Berlin 1913.

Zu S. 238, Zeile 1: E. MARSDEN und T. BARRATT, Proceed. of the Physical Society of London, XXIII, 1911, S. 367 bis 373.

Zu S. 240, Zeile 23: Über die neuere Entwicklung der Gastheorie, die Elektronentheorie der Metalle, wie auch über die weiter noch zu besprechenden Fragen der Quantentheorie usf. findet der Leser nähere Aufschlüsse in ARTHUR HAAS, Materiewellen und Quantenmechanik, 4. und 5. Aufl., Leipzig 1934. Dieses Buch setzt keine allzu hohen mathematischen Kenntnisse voraus.

Zu S. 242, Zeile 27: Die grundlegenden Abhandlungen von M. PLANCK, Theorie des Gesetzes der Energieverteilung im Normalspektrum, erschienen in den Verhandl. d. Deutschen Physikal. Gesellschaft, 2, 1900, S. 237 bis 245 und in den Annalen der Physik, 4, 1901, S. 553 bis 563.

Zu S. 245, Zeile 20: PHILIPP FRANK, Das Kausalgesetz und seine Grenzen (Schriften zur wissensch. Weltauffassung, Bd. 6), Wien 1932.

Zu S. 246, Zeile 1: Zu den folgenden Ausführungen vgl. meinen Vortrag „Über die gegenwärtige Krise der Mechanik", 1921. Hierin wird zum erstenmal die Unvereinbarkeit einer deterministischen Theorie „im Kleinen" mit statistischer Auffassung „im Großen" dargelegt.

Zu S. 252, Zeile 13: Mein Vortrag: Naturwissenschaft und Technik der Gegenwart, Leipzig und Berlin 1922; auch in Ztschr. d. Vereines dtsch. Ing., 64, 1920, S. 687 bis 690 und 717 bis 719.

Zu S. 252, Zeile 34: Dieser Abschnitt ist entnommen meinem Vortrag „Über kausale und statistische Gesetzmäßigkeit etc.", 1930.

Zu S. 254, Zeile 31: Originaldarstellungen: LOUIS DE BROGLIE, Einführung in die Wellenmechanik, übersetzt von PEIERLS, Leipzig 1929. W. HEISENBERG, Die physikalischen Prinzipien der Quantentheorie, Leipzig 1930. E. SCHRÖDINGER, Abhandlungen zur Wellenmechanik, Leipzig 1927. Zusammenfassende Übersicht gibt das Anm. 240/23 genannte Buch von HAAS.

Zu S. 259, Zeile 19: Die Auffassung der HEISENBERGschen Unschärferelation ist heute noch vielfach schwankend. Vgl. dazu M. v. LAUE, Die Naturwissenschaften, Jg. 22, 1934, S. 439 und meine Erwiderung ebenda, S. 822.

Namenverzeichnis.