MAGKS

Aachen Siegen Gießen Marburg Kassel Göttingen

Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 18-2023

Max Albert

Rationalität, Erkenntnis und Entscheidung
Bayesianismus und kritischer Rationalismus im Vergleich

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Max Albert

Rationalität, Erkenntnis und Entscheidung

Bayesianismus und kritischer Rationalismus im Vergleich

Zweite korr. Fassung (Februar 2024). *Erscheint 2024 in*: Jens Eisfeld (Hg.), *Rationalität im 21. Jahrhundert*, Weilerswist: Velbrück, S. 29–66.
Korrespondenz: Prof. Dr. Max Albert, Verhaltens- und Institutionenökonomik, Justus-Liebig-Universität Gießen, Licher Str. 66, 35394 Gießen. Email: max.albert@wirtschaft.uni-giessen.de.

1. Die Entscheidung zwischen konkurrierenden Rationalitätskonzeptionen

Die Frage, was unter Rationalität zu verstehen ist und in welchem Umfang Menschen rational sind oder sein können, ist eine der wichtigsten Fragen der Philosophie, der Psychologie und der Sozialwissenschaften. Im Folgenden werde ich zwei Rationalitätskonzeptionen einander gegenüberstellen, den Bayesianismus und den kritischen Rationalismus, und der Frage nachgehen, welche dieser Konzeptionen vorzuziehen ist. Diese Frage kann natürlich nur beantwortet werden, wenn klar ist, worauf sich Rationalität bezieht und was sie leisten soll.

Bayesianismus und kritischer Rationalismus zeigen einige Gemeinsamkeiten. Nach beiden Rationalitätskonzeptionen ist die Entscheidung einer Person für eine von mehreren gegebenen Optionen genau dann rational, (a) wenn die Person überzeugt ist, dass keine der nicht gewählten Optionen besser geeignet ist, um ihre Ziele zu erreichen, und (b) wenn auch diese Überzeugung rational ist. Damit werden zwei verschiedene Aspekte der Rationalität angesprochen. Die Entscheidungsfindung bei gegebenen Überzeugungen (a) fällt in den Bereich der praktischen Rationalität, während die Bildung von Überzeugungen oder Urteilsbildung (b) Gegenstand der theoretischen Rationalität ist.

Beide Rationalitätskonzeptionen gehen im Bereich der Urteilsbildung von Ungewissheit oder radikaler Unsicherheit aus. Der Bayesianismus unterstellt Ungewissheit bezüglich aller Aussagen, die über logisch-mathematische Zusammenhänge und die Beschreibung der eigenen Sinneserfahrung hinausgehen. Der kritische Rationalismus unterstellt Ungewissheit bezüglich aller Aussagen.

Beide Konzeptionen ziehen aus dieser Ungewissheit dieselbe Konsequenz: die Urteilsbildung ist eine Frage der Entscheidung; zu den Optionen, zwischen denen man sich entscheiden muss, gehören auch die Überzeugungen. Die theoretische Rationalität wird damit zu einem Spezialfall der praktischen Rationalität; beide Rationalitätskonzeptionen sind Entscheidungstheorien unter Ungewissheit.

Bei Entscheidungstheorien wird zwischen positiven (oder deskriptiven) und normativen (oder präskriptiven) Theorien unterschieden. Aus einer Rationalitätskonzeption erhält man eine positive – psychologische, ökonomische oder sozialwissenschaftliche – Theorie, wenn man unterstellt, dass menschliche

Urteilsbildung und Entscheidungsfindung rational im Sinne dieser Konzeption sind. Die Entscheidung für oder gegen eine solche Theorie erfolgt auf der Grundlage der üblichen wissenschaftlichen Methoden.

Als positive Theorie wurde der Bayesianismus lange Zeit in den Sozialwissenschaften im Rahmen des sogenannten „Rational-Choice"-(RC-)Ansatzes unter der Bezeichnung „subjective expected utility theory" (SEU-Theorie) vertreten. Inzwischen allerdings ist der RC-Ansatz selbst in der Ökonomie, wo er ursprünglich entwickelt wurde, auf dem Rückzug. Zahlreiche psychologische und ökonomische Experimente zeigen, dass menschliche Entscheidungsfindung und Urteilsbildung durch den RC-Ansatz und insbesondere durch die SEU-Theorie nicht erklärt werden können (s. z.B. Kahneman/Tversky 2000). In seiner Rolle als positive Entscheidungstheorie muss der Bayesianismus als widerlegt gelten.

Als normative Entscheidungstheorie und allgemeine Rationalitätskonzeption stößt der Bayesianismus allerdings weiterhin auf breite Zustimmung, insbesondere auch in der Philosophie, wo er die vorherrschende Erkenntnis- und Wissenschaftstheorie liefert. In diesem Bereich konkurriert er mit der Rationalitätskonzeption des kritischen Rationalismus.

Für die Entscheidung zwischen konkurrierenden normativen Rationalitätskonzeptionen gibt es keine allgemein akzeptierten Kriterien. Das liegt unter anderem daran, dass häufig unklar bleibt, welcher Anspruch für eine solche Konzeption erhoben wird.

Eine normative Entscheidungstheorie sagt zunächst einmal, dass Entscheidungen nach bestimmten Regeln getroffen werden sollen. Bei Sollaussagen oder Imperativen wird traditionell zwischen kategorischen und hypothetischen Imperativen unterschieden. Kategorische Imperative wie „Du sollst nicht töten" verlangen eine Befolgung des Imperativs ohne Rücksicht auf die Interessen des Adressaten. Hypothetische Imperative dagegen haben den Charakter einer Empfehlung; ihre Befolgung ist angeblich im eigenen Interesse des Adressaten.

Weder für den kritischen Rationalismus noch für den Bayesianismus kommt eine Interpretation als kategorischer Imperativ in Frage. Beide Konzeptionen gehen davon aus, dass Rationalität nicht Selbstzweck ist, sondern dass es im Interesse eines Entscheiders ist, rational zu entscheiden. Anders ausgedrückt: beide Rationalitätskonzeptionen verstehen sich als Versionen einer instrumentellen Rationalität.[1]

Das bedeutet aber nicht, dass sich die Vertreter dieser Konzeptionen auf den Standpunkt zurückziehen können, die jeweilige Konzeption sei eben als hypothetischer Imperativ gemeint. Die vage Behauptung,

[1] Zur Interpretation normativer Rationalitätsvorstellung speziell in der Wissenschaft siehe Gadenne (2005). Für den Bayesianismus siehe Harsanyi (1977: 16), Sen (2001: 68) und, aus der Perspektive der kognitiven Psychologie, Stanovitch (2011). Für den kritischen Rationalismus siehe Gadenne (2006) und Musgrave (2006: 309–310).

dass die Befolgung gewisser Regeln im Interesse der Regeladressaten liegt, ist keine ausreichende Grundlage, um sich zwischen konkurrierenden Regeln zu entscheiden.

In einem ersten Schritt muss man die Ziele benennen, die mit der Befolgung der Regeln erreicht werden sollen. Die Behauptung, dass durch die Befolgung bestimmter Regeln bestimmte Ziele erreicht werden können, ist eine technologische Aussage im Sinne von Hans Albert (1968).

Technologische Aussagen sind Aussagen darüber, welche Ziele erreichbar oder unerreichbar sind, wie man bestimmte Ziele erreichen kann oder welche Ziele in einem Konflikt stehen. Diese Aussagen sind offensichtlich nicht normativ, sondern es handelt sich um positive Aussagen, also Tatsachenbehauptungen, die wahr oder falsch sind. Allerdings wird man sich für diese positiven Aussagen meist nur interessieren, wenn man die betreffenden Ziele verfolgt.

Beispielsweise unterstellen der Bayesianismus wie der kritische Rationalismus im Bereich der Urteilsbildung Ungewissheit: sichere Wahrheiten sind allgemein oder in bestimmten Bereichen nicht erreichbar. Das ist eine technologische Aussage, die wahr oder falsch ist.

Entsprechend sollte man von den Vertretern einer instrumentellen Rationalitätskonzeption verlangen, dass sie einen technologischen Anspruch formulieren: Wozu sind die Regeln der Rationalität geeignet? Was kann man erreichen, wenn man sich an diese Regeln hält? Wenn die Befolgung der Regeln nicht sicherstellt, dass die vorgegebenen Ziele erreicht oder soweit wie möglich erreicht werden, haben die Regeln den Charakter einer Heuristik. Der mit ihnen verbundene technologische Anspruch ist dann unter Umständen nur, dass die Regeln unter allen bekannten Regeln am besten geeignet sind, um das Ziel zu erreichen.

Regeln mit heuristischem Charakter sind nichts Ungewöhnliches. Regeln für das Verhalten im Straßenverkehr beispielsweise können bestenfalls für sich in Anspruch nehmen, dass keine besseren Regeln zur Vermeidung von Unfällen bekannt sind, wenn man in vertretbarer Zeit den gewünschten Ort erreichen will.

Angewandt auf die beiden Rationalitätskonzeptionen ergibt sich, dass diese Rationalitätskonzeptionen um der Klarheit willen in Form von Regeln dargestellt werden sollten, die mit einem technologischen Anspruch verbunden sind. Auf der Grundlage einer solchen Darstellung lassen sich konkurrierende Rationalitätskonzeptionen kritisch diskutieren: man kann untersuchen, ob der jeweilige technologische Anspruch akzeptiert werden kann.

An diesem Punkt, so könnte man argumentieren, kommen wieder die üblichen wissenschaftlichen Methoden zum Zuge. Schließlich ist ein technologischer Anspruch eine positive Aussage, die man wissenschaftlich prüfen kann. Aber hier gibt es ein Problem. Man muss berücksichtigen, dass beide Rationalitätskonzeptionen für sich in Anspruch nehmen, auch die wissenschaftliche Urteilsbildung zu regeln.

Soweit es um den Status des Bayesianismus als positiver Entscheidungstheorie geht, kann man sich möglicherweise auf einen wissenschaftlichen Konsens zwischen Bayesianern und kritischen Rationalismus berufen, die trotz unterschiedlicher Auffassungen darüber, wie die wissenschaftliche Urteilsbildung erfolgen sollte, zu demselben Ergebnis kommen – nämlich, dass der Bayesianismus als positive Theorie menschliche Entscheidungsfindung und Urteilsbildung nicht erklären kann.

Aber wenn es um die mit der jeweiligen Rationalitätskonzeption verbundenen technologischen Ansprüche geht, sind jedenfalls die kritischen Rationalisten explizit der Auffassung, dass eine Rationalitätskonzeption nur haltbar ist, wenn sie auf Grundlage ihrer selbst akzeptiert werden kann. Genauer heißt das: Wenn ihr eigener technologischer Anspruch auf der Grundlage ihrer eigenen Regeln der Urteilsbildung akzeptiert werden kann. Das ergibt sich daraus, dass eine Rationalitätsauffassung, die nicht nach ihren eigenen Maßstäben akzeptiert werden kann, entweder unvollständig oder widersprüchlich ist.[2]

Die technologischen Ansprüche des Bayesianismus wie des kritischen Rationalismus beziehen sich, historisch betrachtet, auf das sogenannte Induktionsproblem, das in Kapitel 2 vorgestellt wird. Kapitel 3 erläutert die bayesianische Lösung dieses Problems und den sich daraus ergebenden technologischen Anspruch des Bayesianismus, um dann zu zeigen, dass dieser Anspruch durch den Bayesianismus nicht eingelöst werden kann. Kapitel 4 führt die analoge Diskussion für den kritischen Rationalismus, allerdings mit einem positiven Ergebnis. Kapitel 5 enthält einige abschließende Bemerkungen.

2. Das Induktionsproblem

Der Bayesianismus kann genauso wie der kritische Rationalismus als ein Versuch angesehen werden, das Induktionsproblem zu lösen. Dieses Problem wurde von David Hume aufgeworfen, aber Humes eigene Formulierung soll hier keine Rolle spielen. Aus heutiger Sicht können wir dieses Problem wie folgt beschreiben.[3]

Ausgangspunkt des Induktionsproblems ist der klassische Empirismus (im Folgenden kurz Empirismus), also die Auffassung, dass die einzige Quelle der Erkenntnis die Erfahrung ist und dass wir deswegen rationalerweise nur Aussagen akzeptieren können, die wir aus der Erfahrung – oder genauer: aus Aussagen, die unsere Erfahrung beschreiben – erschließen. Für den Empirismus gibt es nur zwei Arten von Aussagen, die ohne weiteres akzeptiert werden können: Aussagen über logische und mathematische Zusammenhänge, die unabhängig von der Erfahrung wahr oder falsch sind, und Aussagen, die unsere Sinneserfahrung beschreiben. Schon Aussagen wie „Hier ist Wasser“ sind problematisch, da sie unterstellen, dass es Wasser gibt, also ein Material mit bestimmten Eigenschaften. Der Empirist akzeptiert

[2] Für den kritischen Rationalismus siehe dazu Musgrave (1999: 330–331). Für den Bayesianismus ist die Lage nicht so eindeutig.
[3] Zum Induktionsproblem siehe Humphreys (1987), Gillies (1988) und Musgrave (1993: 151–166, 2011).

zunächst nur Aussagen wie beispielsweise „Mir kommt es so vor, als sei hier Wasser", die reine Sinneseindrücke wiedergeben sollen, ohne etwas über die dahinterstehenden Tatsachen vorauszusetzen.

Die Frage ist, ob man als Empirist rationalerweise irgendwelche Aussagen akzeptieren kann, die über die Sinneserfahrung hinausgehen. Die Erschließung von über die Erfahrung hinausgehenden Aussagen aus dieser Erfahrung wird als Induktion bezeichnet.

Induktive Argumente können sich nicht auf die deduktive Logik stützen. Gemäß der deduktiven Logik folgt eine Konklusion aus einer Reihe von Prämissen genau dann, wenn es logisch unmöglich ist, dass die Prämissen wahr sind und die Konklusion falsch ist. Die deduktive Logik sichert damit die Wahrheitsübertragung von den Prämissen auf die Konklusion. Weil bei einem gültigen deduktiven Argument – also einem Argument, dass den Regeln der deduktiven Logik folgt – die Konklusion wahr sein muss, wenn die Prämissen wahr sind, ist es unstrittig, dass man angesichts eines solchen Arguments rationalerweise die Konklusion akzeptieren muss, wenn man die Prämissen akzeptiert.

Im Rahmen des Induktionsproblems beschreiben aber die Prämissen nur die bisherige Erfahrung – beispielsweise bisherige Beobachtungen –, während die Konklusion über diese Erfahrung hinausgeht – indem sie beispielsweise neue Beobachtungen vorhersagt. In einem solchen Fall ist es logisch immer möglich, dass die Konklusion falsch ist, selbst wenn die Prämissen wahr sind. Um das einfache Beispiel des nächsten Abschnitts zu benutzen: Auch wenn man bisher immer beobachtet hat, dass Smaragde grün sind, ist es logisch möglich, dass der nächste Smaragd, den man findet, blau ist.

Die Frage ist dann, welche deduktiv ungültigen Argumente akzeptabel sind. Man wird möglicherweise bereit sein zu akzeptieren, dass alle Smaragde grün sind, weil alle bisher gefundenen Smaragde grün sind. Aber vermutlich wird man nicht bereit sein zu akzeptieren, dass die Relativitätstheorie wahr ist, weil alle bisher gefunden Smaragde grün sind. Nach der deduktiven Logik sind beide Argumente gleich schlecht, aber möglicherweise ist aus Sicht irgendeiner induktiven Logik eines der beiden Argumente akzeptabel und das andere nicht.

Für Hume oder zumindest für viele seiner Interpreten ergab sich das Induktionsproblem aus der Kombination von Empirismus und Deduktivismus:

1. Nach dem Empirismus können Aussagen, die über die Erfahrung hinausgehen, nur akzeptiert werden, wenn sie aus der Erfahrung erschlossen werden können.
2. Nach dem Deduktivismus sind nur deduktiv gültige Argumente akzeptabel.

Kombiniert man beide Auffassungen, kommt man zu dem Ergebnis, dass überhaupt keine Aussagen, die über die Erfahrung hinausgehen, akzeptiert werden können. Zur Herleitung solcher

Aussagen aus der Erfahrung würde man ein induktives Argument benötigen – mit wahren Prämissen, die die bisherige Erfahrung beschreiben, und einer Konklusion, die darüber hinausgeht. Die Konklusion dieses Argument könnte zwar wahr, aber genauso gut auch falsch sein. Die Ergebnisse der Induktion sind daher aus Sicht des Deduktivismus genauso sicher oder unsicher wie jede andere über die Erfahrung hinausgehende Vermutung, die sich nicht auf die bisherige Erfahrung stützt.

Aus dieser Perspektive ist keine Vermutung oder Überzeugung, die über die bisherige Erfahrung hinausgeht, vernünftiger oder unvernünftiger als irgendeine andere Vermutung oder Überzeugung. Man kann glauben, was man will, egal, welche Erfahrungen man bisher gemacht hat. Keine Überzeugung ist irrational oder – was auf dasselbe hinausläuft – alle Überzeugungen sind gleichermaßen rational.

Diese Auffassung bezeichnet man auch als radikalen Skeptizismus. Bertrand Russell hat klar formuliert, was radikaler Skeptizismus bedeutet: es gibt keinen intellektuellen Unterschied zwischen Vernunft und Wahnsinn (Russell 1945: 673). Das zentrale Problem jeder Rationalitätskonzeption ist die Vermeidung des radikalen Skeptizismus. Das kann nur gelingen, indem man den Empirismus oder den Deduktivismus ablehnt.

Der Bayesianismus akzeptiert den Empirismus, gibt den Deduktivismus auf und benutzt als induktive Logik die Wahrscheinlichkeitstheorie. Er gesteht zwar zu, dass immer Ungewissheit darüber herrscht, welche über die Erfahrung hinausgehenden Aussagen wahr sind, erlaubt es aber trotzdem, im Lichte der Erfahrung manche Aussagen für wahrscheinlicher zu halten als andere.

3. Bayesianische Rationalität

3.1 Der technologische Anspruch des Bayesianismus

Wir betrachten den bayesianischen Prozess der Urteilsbildung, das bayesianische Lernen, zunächst an einem einfachen Beispiel, dem sogenannten „Blün-Problem“. Das Blün-Problem zeigt auch, welchen technologischen Anspruch der Bayesianismus erheben könnte.[4]

[4] Diese einfache Variante des Induktionsproblems geht auf Goodman (1983) zurück. „Blün“, ein Mischwort aus „blau“ und „grün“, ist die übliche Übersetzung für Goodmans „grue“. Siehe Gillies (1988: 203) für eine knappe Erläuterung von Goodmans sprachphilosophische Intention, die im Kontext eines nur mit Sinneseindrücken arbeitenden Bayesianismus nicht relevant ist. Die bayesianische Analyse des Blün-Problem habe ich in ihren Grundzügen aus einem unveröffentlichten Text von Colin Howson übernommen, der als Grundlage für ein Seminar über den Bayesianismus im Rahmen des Europäischen Forums Alpbach 1997 diente (s. auch Albert 2011a).

Wir betrachten die Hypothese „Alle Smaragde sind grün“ (H_0). Auch wenn diese Hypothese auf alle bisher gefunden Smaragde zutrifft, folgt aus diesen Beobachtungen nicht, dass auch in Zukunft gefundene Smaragde grün sind. Es könnte sein, dass bereits der nächste Smaragd, den wir finden, und alle weiteren Smaragde blau sind.

Aus bayesianischer Sicht können wir unter bestimmten Annahmen trotzdem argumentieren, dass die Hypothese, dass alle Smaragde grün sind, mit jeder Bestätigung immer wahrscheinlicher wird, bis die Wahrscheinlichkeit so hoch ist, dass man sich praktisch darauf verlassen wird, dass die Hypothese wahr ist.

Die Hypothese H_0 sagt voraus, dass wir jedes Mal, wenn wir einen neuen Smaragd finden, feststellen werden, dass er grün ist. Für das Ereignis, dass wir einen grünen Smaragd finden, also für eine Bestätigung von H_0, schreiben wir kurz eine Null. Die Hypothese H_0 sagt also voraus, dass wir immer Nullen beobachten werden.

Für den bayesianischen Lernprozess benötigen wir nun Alternativen zu H_0. Wir betrachten als Alternativen die Hypothesen H_k, $k = 1, 2, \dots, \infty$. Hypothese H_1 sagt, dass der nächste Smaragd, den wir finden, und alle Smaragde, die wir danach finden, blau sind. Für jeden Fund eines solchen blauen Smaragd schreiben wir eine 1. Hypothese H_1 sagt also voraus, dass wir in Zukunft nur Einsen beobachten werden. Hypothese H_2 sagt voraus, dass wir als nächstes eine Null beobachten, aber dann nur noch Einsen. Allgemein sagt Hypothese H_k mit $k > 0$, dass die nächsten $k - 1$ Beobachtungen Nullen sind, aber das wir aber beginnend mit der k-ten Beobachtung nur noch Einsen sehen.

Für die Beobachtungen schreiben wir $X_1, X_2, \dots$, wobei $X_t \in \{0, 1\}$. Damit ist $X_3 = 0$ beispielsweise das Ereignis, dass die dritte beobachtete Ziffer eine Null ist, dass also der dritte der in Zukunft gefundenen Smaragde blau ist.

Jede Hypothese sagt eine bestimmte Folge von Nullen und Einsen, eine binäre Folge, voraus. Wir wollen annehmen, dass wir oder unsere Nachfolger unendlich viele Smaragde finden – auf anderen Planeten oder wie und wo auch immer. Die Folgen, die wir betrachten, sind damit unendlich lang. Damit gibt es auch abzählbar unendlich viele Alternativhypothesen zur Ausgangshypothese.

Sollte die Ausgangshypothese falsch und eine der Alternativhypothesen wahr sein, lernt man offensichtlich eines Tages die Wahrheit, wenn man nämlich eine 1 beobachtet. Dazu benötigt man keinen bayesianischen Lernprozess. Der problematische Fall ist der Fall, in dem die Ausgangshypothese wahr ist. Dann wird man immer nur Nullen beobachten, kann aber nie ausschließen, dass ab dem nächsten Mal nur noch Einsen kommen. Im Folgenden unterstellen wir, dass dieser problematische Fall vorliegt: Smaragde sind tatsächlich immer grün, aber wir wissen das nicht.

Die bayesianische Rationalität erfordert, dass alle Überzeugungen über zukünftige Ereignisse aus einer subjektiven Wahrscheinlichkeitsverteilung, genannt Aprioriverteilung, auf der Menge der Hypothesen hergeleitet werden. Diese Wahrscheinlichkeiten sind persönliche Grade des Glaubens. Jede Aprioriverteilungen ist so rational und irrational wie jede andere; die Rationalität liegt allein darin, dass man eine solche Aprioriverteilung wählt und davon in bestimmter Weise Gebrauch macht.

Wir beschreiben die Aprioriverteilung durch eine Funktion P_0, die jeder Hypothese H_k eine Wahrscheinlichkeit $P_0(H_k)$ zuordnet, wobei die Summe aller dieser Wahrscheinlichkeiten 1 sein muss. Außerdem wollen wir unterstellen, dass alle Hypothesenwahrscheinlichkeiten größer als Null sind. Da es unendlich viele Hypothesen gibt, müssen die meisten Wahrscheinlichkeiten allerdings sehr klein sein. Beispielsweise könnten wir $P_0(H_0) = p$ mit $0 < p < 1$ und $P_0(H_k) = (1-p)/2^k$ wählen; dann gilt $\sum_{k=0}^{\infty} P_0(H_k) = 1$.

Unter der Annahme, dass H_0 wahr ist, beobachten wir immer nur Nullen. Das bayesianische Lernen erfordert bei jeder Beobachtung eine Anpassung der Hypothesenwahrscheinlichkeiten. Wir bezeichnen die Wahrscheinlichkeiten nach Beobachtung von X_t mit P_t. Die neue Wahrscheinlichkeit einer Hypothese H_k ist gleich der alten bedingten Wahrscheinlichkeit von H_k gegeben die Beobachtung: $P_t(H_k) = P_{t-1}(H_k|X_t)$. Die bedingten Wahrscheinlichkeiten werden durch das Bayes'sche Theorem bestimmt, das für deterministische Hypothesen eine besonders einfache Form annimmt. Die Wahrscheinlichkeiten aller Hypothesen, die X_t falsch vorhergesagt haben, fallen auf 0; diese Hypothesen sind durch Beobachtung widerlegt. Die Wahrscheinlichkeiten aller nicht widerlegten Hypothesen werden durch die Summe dieser Wahrscheinlichkeiten dividiert, also durch Multiplikation mit demselben Faktor so angehoben, dass sie sich wieder zu 1 addieren.

Da sich die Wahrscheinlichkeiten der Hypothesen zu 1 addieren müssen, gilt mit $P_0(H_0) > 0$ auch, dass es nur endlich viele Hypothesen gibt, deren Aprioriwahrscheinlichkeit größer ist als die Aprioriwahrscheinlichkeit der Ausgangshypothese H_0. Damit gilt für jede Aprioriverteilung, dass H_0 nach endlich vielen Beobachtungen die wahrscheinlichste Hypothese ist. Da außerdem mit jeder neuen Beobachtung eine weitere Alternativhypothese falsifiziert wird, konvergiert die Wahrscheinlichkeit von H_0 gegen 1. Sobald die Wahrscheinlichkeit von H_0 größer ist als 0,5, können wir nach bayesianischer Auffassung sagen, dass es vernünftiger ist zu glauben, dass H_0 wahr ist, als zu glauben, dass H_0 falsch ist.

Natürlich hängt es von der Aprioriverteilung ab, wie lange es dauert, bis die wahre Hypothese eine hohe Wahrscheinlichkeit hat. Für unser Beispiel oben ergibt sich mit einer Aprioriwahrscheinlichkeit von $p = 0{,}0001$ für H_0 als wahrer Hypothese, dass man nur siebzehn Beobachtungen benötigt, damit H_0 eine Wahrscheinlichkeit von mehr als 0,99 hat. Es ist aber leicht, Aprioriverteilungen anzugeben, bei denen man einen solchen Grad der subjektiven Sicherheit erst nach einer beliebig hohen Zahl von Beobachtungen erreicht.[5] Weil sich keine Obergrenze für die Zahl der Beobachtungen angeben lässt, die man benötigt, damit die wahre Hypothese eine hohe Wahrscheinlichkeit hat, betrachtet man bei der Diskussion des Bayesianismus unendlich lange Beobachtungsfolgen.

Zwar bleiben immer Alternativhypothesen mit positiver Wahrscheinlichkeit übrig, aber nach endlich vielen Beobachtungen ist die Summe der Wahrscheinlichkeiten der noch nicht falsifizierten Alternativhypothesen so klein, dass diese Hypothesen für die Vorhersage praktisch keine Rolle mehr spielen. Die Wahrscheinlichkeit $P_t(X_{t+1} = 0)$ dafür, dass auch der nächste gefundene Smaragd grün ist, ist nie kleiner als $P_t(H_0)$ und konvergiert damit ebenfalls gegen den korrekten Wert von 1.

Dieses Beispiel hat eine Reihe von Eigenschaften, die traditionell für den bayesianischen Lernprozess in Anspruch genommen werden. Die Eigenschaften gelten in unserem Beispiel unabhängig von der genauen Form der Aprioriverteilung und unabhängig davon, welche der betrachteten Hypothesen wahr ist, solange die Wahrheit eine positive Wahrscheinlichkeit hat.

1. *Konvergenz gegen die Wahrheit:* Die Wahrscheinlichkeit für die wahre Hypothese konvergiert gegen 1.
2. *Verbesserung der Vorhersagen:* Die Wahrscheinlichkeiten für Vorhersagen konvergieren gegen die korrekten Vorhersagewahrscheinlichkeiten.
3. *Konvergenz der Überzeugungen:* Die Hypothesenwahrscheinlichkeiten und die Vorhersagewahrscheinlichkeiten von Bayesianern, die mit unterschiedlichen Aprioriwahrscheinlichkeiten beginnen, konvergieren gegeneinander.

Wenn der Bayesianismus überhaupt einen technologischen Anspruch im Bereich der Urteilsbildung erheben kann, dann wird dieser Anspruch durch diese Eigenschaften beschrieben. Würden diese Eigenschaften auch auf den Lernprozess im allgemeinen Induktionsproblem und nicht nur im Blün-Problem zutreffen, könnte man argumentieren, dass der Bayesianismus tatsächlich den radikalen Skeptizismus überwindet und damit das Induktionsproblem löst.

Nebenbei bemerkt würde ein solches Ergebnis es auch möglich machen, den Bayesianismus auf seiner eigenen Grundlage zu akzeptieren. Der oben formulierte technologische Anspruch wurde nämlich mathematisch bewiesen. Da sich im Bayesianismus die Ungewissheit nicht auf mathematische Aussagen

[5] Beispielsweise könnten wir $P_0(H_0) = p$ und $P_0(H_k) = \frac{1-p}{100000} \times 1{,}00001^{-k}$ setzen; auch dann gilt wieder $\sum_{k=0}^{\infty} P_0(H_k) = 1$. Mit $p = 0{,}0001$ benötigen wir über 1,15 Millionen Beobachtungen, damit die Wahrscheinlichkeit von H_0 mindestens 0,99 beträgt, obwohl H_0 von Anfang an die wahrscheinlichste Hypothese ist.

und Beweise erstreckt, könnte ein solcher mathematisch bewiesener technologischer Anspruch ohne weiteres akzeptiert werden.

3.2 Die bayesianische Analyse des Induktionsproblems

3.2.1 Die Problemstellung

Im Vergleich mit dem Induktionsproblem ist das Blün-Problem zu einfach, denn die Alternativhypothesen unterstellen alle, dass wir in der Reihe der gefundenen Smaragde nur einmal einen Wechsel von grün zu blau beobachten. Wenn man aber überhaupt in Betracht zieht, dass ein solcher Wechsel auftritt, dann könnte das ja auch immer wieder geschehen, etwa „grün-blau-grün-grün-blau-grün …“ oder 010010…, wenn man die Beobachtungen binär kodiert. Entsprechend sollte man Alternativhypothesen zulassen, die beliebige binäre Folgen vorhersagen. Bevor man Beobachtungen macht, kann keine dieser binären Folgen ausgeschlossen werden.

Dieser Schritt führt zu dem allgemeinen Vorhersageproblem, wie es heute meistens betrachtet wird. Dieses Problem ist so allgemein, dass es alle denkbaren Erfahrungen umfassen kann, also weit über die Frage hinausgeht, ob eine einzelne Hypothese sich auch in Zukunft als richtig erweist oder nicht. Es ist identisch mit dem Induktionsproblem.

Alle Erfahrungen können auf Sinneseindrücke zurückgeführt werden. Unsere Sinnesorgane können in einem gegebenen Zeitraum nur endlich viele Zustände voneinander unterscheiden. Daher können wir die komplette Erfahrung im digitalen Format betrachten. Die Erfahrungen, die eine Person im Laufe ihres Lebens machen kann – einschließlich der Beobachtung jeglichen Verhaltens anderer Personen, ob verbal oder nicht-verbal –, ergeben digital kodiert immer eine binäre Folge.

Für eine realistische Version des Problems des bayesianischen Lernens könnten wir eine endliche obere Grenze für die Länge der binären Folgen annehmen. Wir haben aber gesehen, dass der technologische Anspruch des Bayesianismus nur auf Konvergenzergebnissen für unendlich lange Beobachtungsfolgen beruht. Die Betrachtung unendlich langer Folgen führt zu mathematischen Problemen, die ich im Folgenden nicht diskutieren möchte. Ich beschränke mich daher zunächst auf die Betrachtung von recht kurzen Folgen. Die grundlegenden Einsichten, die man an diesem Beispiel gewinnt, gelten jedoch auch für den Fall unendlich langer Beobachtungsfolgen.

Betrachten wir also die Menge aller binären Folgen der Länge 4. Wir beschreiben die möglichen Beobachtungen wieder mit Hilfe der Variablen $X_t \in \{0, 1\}$, $t = 1, \dots, 4$. Wir identifizieren jede der sechzehn möglichen Folgen mit einer anderen Hypothese $H_k, k = 0,1, \dots ,15$. Jede Hypothese sagt eine bestimmte Folge von Beobachtungen voraus, eine Folge, die von keiner anderen Hypothese vorausgesagt wird.

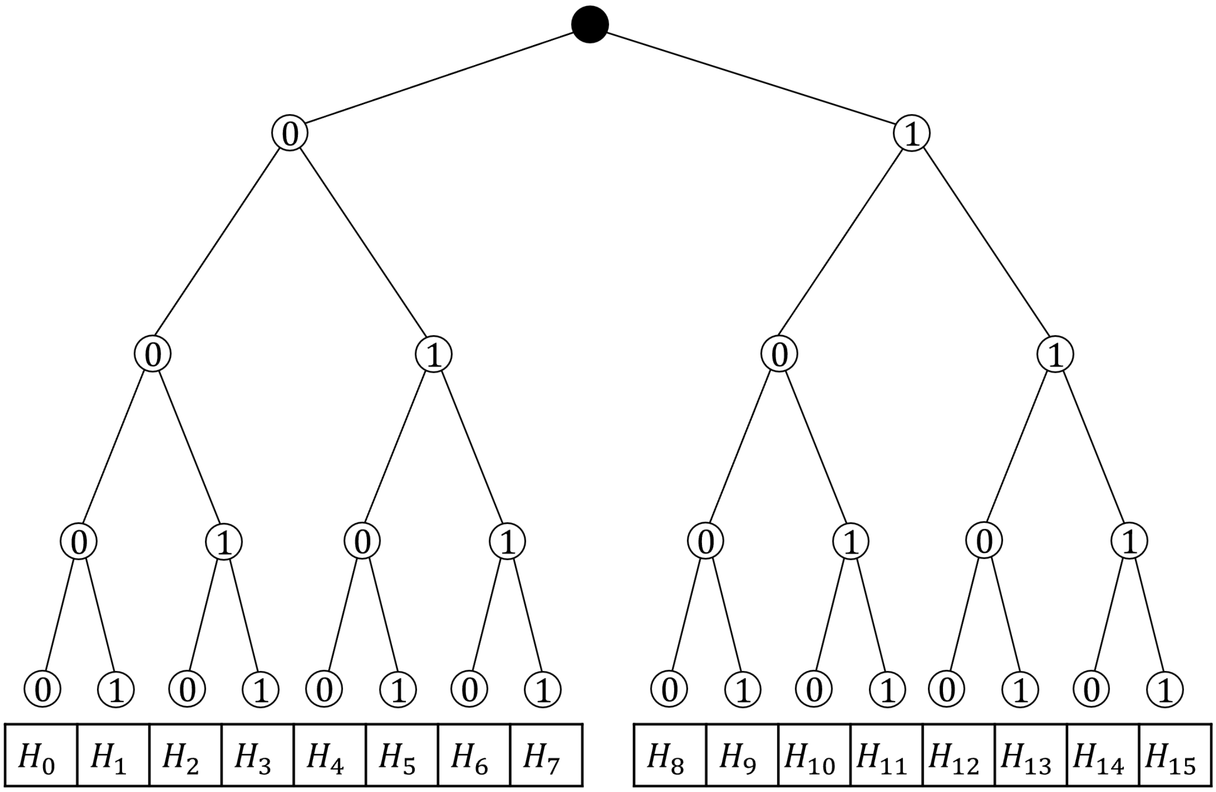


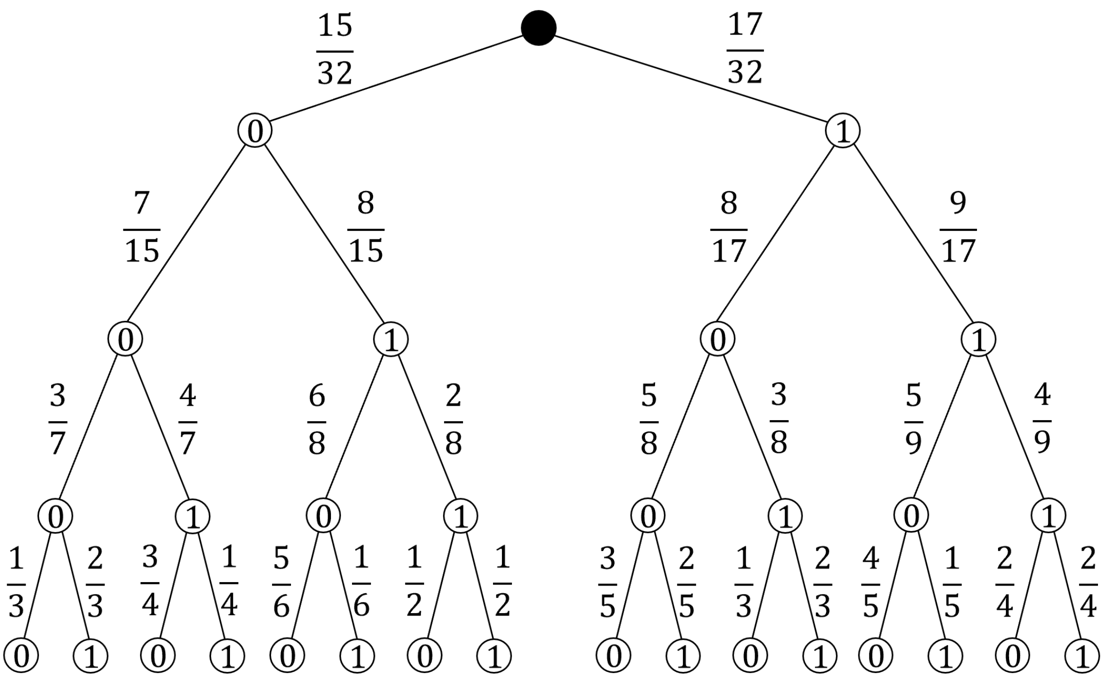
Abbildung 1: Darstellung aller binären Folgen der Länge 4 als binärer Baum. Am Ende jeder der sechzehn Folgen ist die Hypothese $H_k, k = 0, 1, \ldots, 15$ angegeben, die diese Folge voraussagt.

Jede Folge entspricht genau einem vollständigen Pfad durch einen binären Baum, also einem Pfad vom Anfangsknoten bis zu einem der Endknoten (s. Abb. 1). Jeder Knoten nach dem Anfangsknoten steht für die Beobachtung einer Ziffer, 0 oder 1. Von jedem Knoten führen zwei Äste zur nächsten Beobachtung – der linke Ast zu einer Beobachtung von 0, der rechte Ast zu einer Beobachtung von 1.

In dieser Darstellung bestehen die Erfahrungen, die eine Person macht, darin, dass sie Knoten für Knoten den Pfad durch den Baum beobachtet. Aussagen über die Zukunft beziehen sich also immer auf die Frage, welche der noch erreichbaren Knoten in Zukunft erreicht werden.

3.2.2 Wahrscheinlichkeitsrechnung für endliche binäre Bäume

Wir rekapitulieren kurz die Grundzüge der Wahrscheinlichkeitsrechnung für endliche binäre Bäume. Wir betrachten irgendeine Aprioriverteilung auf der Menge der Hypothesen, die den Hypothesen H_k, $k = 0, 1, \ldots, 15$, und damit den entsprechenden vollständigen Pfaden, die Wahrscheinlichkeiten $P_0(H_k)$ mit $\sum_{k=0}^{15} P_0(H_k) = 1$ zuordnet (s. Abb. 2).



	H_0	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}	H_{11}	H_{12}	H_{13}	H_{14}	H_{15}
P_0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{3}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{4}{32}$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$

Abbildung 2: Wahrscheinlichkeitsverteilung P_0 auf der Menge aller Hypothesen bzw. vollständigen Pfade. Aus den Wahrscheinlichkeiten der vollständigen Pfade ergeben sich die an den entsprechenden Ästen angegebenen Wahrscheinlichkeiten für den Übergang von einem Knoten zu seinen beiden unmittelbaren Nachfolgern, die sich jeweils zu 1 addieren. Man kann die Wahrscheinlichkeit jedes vollständigen Pfades wieder aus den Übergangswahrscheinlichkeiten herleiten, indem man alle Übergangswahrscheinlichkeiten entlang des betreffenden Pfades miteinander multipliziert.

Die für die nachfolgenden Argumente wichtigste Eigenschaft einer solchen Wahrscheinlichkeitsverteilung ist, dass sie die bedingten Wahrscheinlichkeiten für den Übergang von einem Knoten zum anderen festlegt.

Wir unterstellen zunächst (wie in Abb. 2), dass alle Hypothesenwahrscheinlichkeiten größer als 0 sind. Betrachten wir einen beliebigen Knoten K. Die Wahrscheinlichkeit, dass Knoten K erreicht wird, ist gleich der Summe der Wahrscheinlichkeiten aller vollständigen Pfade durch K. Wenn K kein Endknoten und K' einer seiner (nicht notwendigerweise: unmittelbaren) Nachfolger ist, dann ist die Wahrscheinlichkeit $P_0(K \wedge K')$ dafür, dass K und K' beide erreicht werden, natürlich gleich der Wahrscheinlichkeit, dass K' erreicht wird. Damit ergibt sich für die bedingte Wahrscheinlichkeit dafür, dass K' erreicht wird, gegeben, dass K erreicht wurde, $P_0(K'|K) = P_0(K \wedge K')/P_0(K) = P_0(K')/P_0(K)$.

Die bedingte Wahrscheinlichkeit $P_0(K'|K)$ kann für Vorhersagen verwendet werden. Sie gibt an, wie wahrscheinlich es ist, dass K' erreicht wird, gegeben, dass K erreicht wird oder bereits erreicht wurde.

Ist K' ein unmittelbarer Nachfolger von K, bezeichnen wir die Vorhersagewahrscheinlichkeit $P_0(K'|K)$ als Übergangswahrscheinlichkeit von K nach K' und schreiben sie an den Ast, der die beiden Knoten verbindet (s. Abb. 2). Die Übergangswahrscheinlichkeiten an den beiden Ästen, die von demselben Knoten ausgehen, addieren sich immer zu 1. Ist K' ein nicht-unmittelbarer Nachfolger von K, erhält man die Vorhersagewahrscheinlichkeit $P_0(K'|K)$ auch als Produkt der Übergangswahrscheinlichkeiten entlang des Pfades von K nach K'. Es lassen sich also alle Vorhersagewahrscheinlichkeiten aus den Übergangswahrscheinlichkeiten berechnen.

Eine Funktion, die allen Ästen eines Baumes Übergangswahrscheinlichkeiten zuordnet, bezeichnen wir als Vorhersagefunktion. Die Wahrscheinlichkeitsverteilung auf der Menge der Hypothesen legt also eine Vorhersagefunktion fest, wenn alle Hypothesenwahrscheinlichkeiten positiv sind. Man kann aber auch die Hypothesenwahrscheinlichkeiten aus der Vorhersagefunktion zurückgewinnen. Wir müssen nur alle Übergangswahrscheinlichkeiten entlang eines vollständigen Pfades miteinander multiplizieren, um die Wahrscheinlichkeit dieses Pfades und damit der Hypothese, die diesen Pfad vorhersagt, zu erhalten.

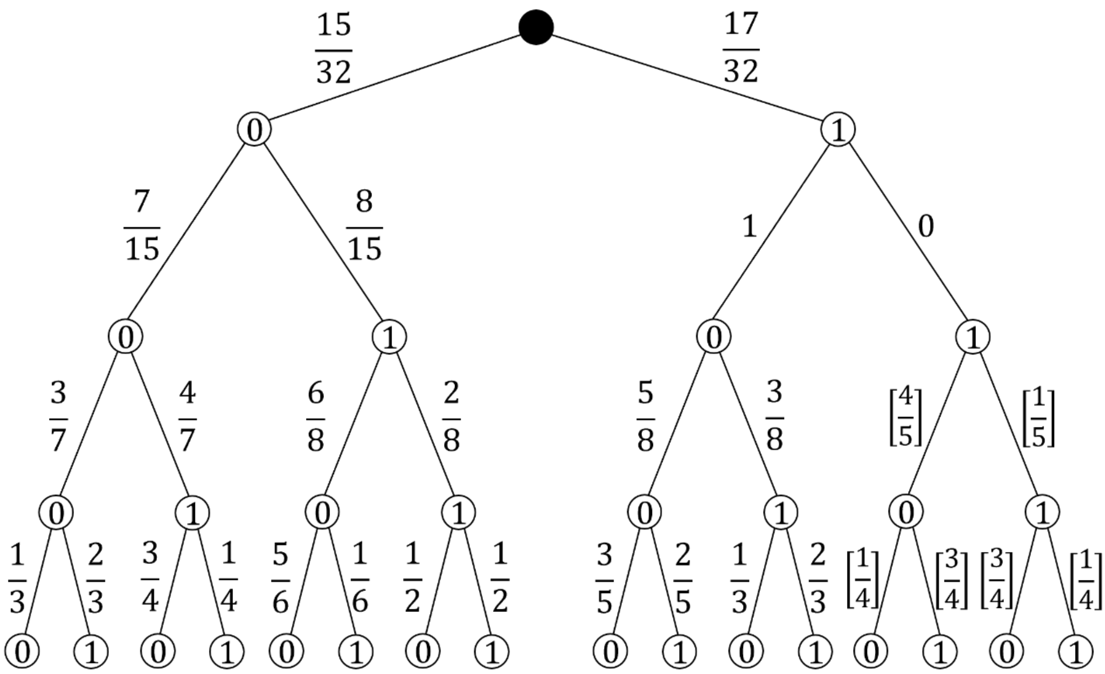
Statt eine Aprioriverteilung auf der Menge der Hypothesen zu wählen und daraus eine Vorhersagefunktion herzuleiten, kann man also genauso gut eine Vorhersagefunktion wählen und daraus Hypothesenwahrscheinlichkeiten herleiten. Beide Aufgaben erfordern die gleiche Zahl von Entscheidungen. Wir benötigen fünfzehn Hypothesenwahrscheinlichkeiten; die sechzehnte Wahrscheinlichkeit ergibt sich aus der Bedingung, dass sich die Hypothesenwahrscheinlichkeiten zu 1 addieren müssen. Alternativ kann man für die fünfzehn Knoten, die keine Endknoten sind, jeweils eine Übergangswahrscheinlichkeit für einen Ast wählen; die Übergangswahrscheinlichkeit für den anderen Ast ergibt sich daraus, dass sich beide Wahrscheinlichkeiten zu 1 addieren müssen.

Unter der bisher gemachten Annahme, dass alle Hypothesenwahrscheinlichkeiten größer als 0 sind, sind auch alle Übergangswahrscheinlichkeiten größer als 0.[6] Um Wahrscheinlichkeiten von Null zu berücksichtigen, müssen wir eine Ergänzung anbringen. Dazu betrachten wir beispielhaft eine Aprioriverteilung mit $P(H_k) > 0$ für $k = 0, \dots, 11$ und $P(H_k) = 0$ für $k = 12, \dots, 15$ (s. Abb. 3). Bei dieser Verteilung ist die Wahrscheinlichkeit, zu Beginn zweimal hintereinander eine Eins zu beobachten ($X_1 = X_2 = 1$), gleich 0.

Auch wenn man irgendeinem Ereignis die subjektive Wahrscheinlichkeit 0 zugeordnet hat, kann dieses Ereignis natürlich eintreten. Wird in unserem Beispiel $X_1 = X_2 = 1$ beobachtet, müssen für die danach allein relevanten Hypothesen H_{12} bis H_{15}, deren Wahrscheinlichkeiten ursprünglich gleich 0 gesetzt

[6] Die binäre Kodierung der Erfahrung ist nur dann mit der Annahme durchgängig positiver Übergangswahrscheinlichkeit vereinbar, wenn die Zahl $m > 1$ der möglichen Erfahrungen zu einem Zeitpunkt gleich 2^n ist. Ist das nicht der Fall und will man trotzdem mit durchgängig positiven Übergangswahrscheinlichkeiten argumentieren, wählt man im einfachsten Fall einen Baum mit m Ästen an jedem Knoten. Alle Ergebnisse gelten auch für solche Bäume (s. Fn. 8 unten).

wurden, neue Wahrscheinlichkeiten $P'(H_k)$ mit $\sum_{k=12}^{15} P'(H\_k) = 1$ gewählt werden. Das gilt ganz allgemein: wenn die ursprüngliche Aprioriverteilung irgendwelchen Ereignissen die Wahrscheinlichkeit 0 zuweist, müssen für den Fall, dass solche Ereignisse eintreten, ergänzende Verteilungen gewählt werden.



	H_0	H_1	H_2	H_3	H_4	H_5	H_6	H_7
P	$\frac{8}{256}$	$\frac{16}{256}$	$\frac{24}{256}$	$\frac{8}{256}$	$\frac{40}{256}$	$\frac{8}{256}$	$\frac{8}{256}$	$\frac{8}{256}$
P'	0	0	0	0	0	0	0	0

	H_8	H_9	H_{10}	H_{11}	H_{12}	H_{13}	H_{14}	H_{15}
P	$\frac{51}{256}$	$\frac{34}{256}$	$\frac{17}{256}$	$\frac{34}{256}$	0	0	0	0
P'	0	0	0	0	$\frac{4}{20}$	$\frac{12}{20}$	$\frac{3}{20}$	$\frac{1}{20}$

Abbildung 3: Alternative Wahrscheinlichkeitsverteilung P. Die Wahrscheinlichkeiten der Hypothesen H_{12} bis H_{15}, die $X_1 = X_2 = 1$ vorhersagen, bzw. die Wahrscheinlichkeiten der entsprechenden vollständigen Pfade sind 0. Entsprechend ist die Wahrscheinlichkeit dafür, nach einer ersten 1 gleich wieder eine 1 zu beobachten, gleich 0, und alle nachfolgenden Übergangswahrscheinlichkeiten sind zunächst nicht definiert. Eine ergänzende Wahrscheinlichkeitsverteilung P' für die H H_{12} bis H_{15} liefert die fehlenden, in Klammern angegebenen Übergangswahrscheinlichkeiten.

Die bayesianische Rationalität wird häufig so aufgefasst, dass die ergänzenden Wahrscheinlichkeitsverteilungen bereits zusammen mit der Aprioriverteilung festgelegt werden, so dass also ein vollständiges – d.h. alle Eventualitäten berücksichtigendes – System von Aprioriverteilungen gewählt wird.[7] Unter dieser Annahme sind alle Vorhersagewahrscheinlichkeiten definiert.

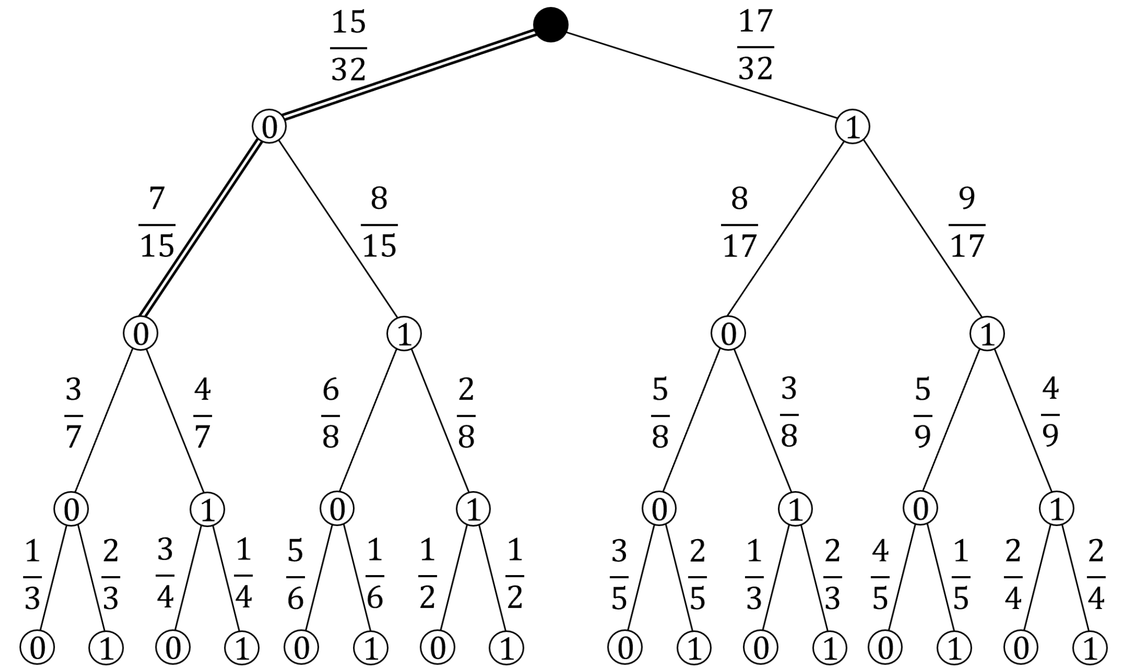
---

[7] Siehe dazu Kiefer/Nyarko (1995) oder Myerson (1991: 168–172) über sequentielle Rationalität in der Spieltheorie.

Mit anderen Worten: Ein vollständiges System von Aprioriverteilungen legt für den gesamten Baum eine Vorhersagefunktion fest, aus der sich wieder dasselbe vollständige System von Aprioriverteilungen herleiten lässt.

3.2.3 Der bayesianische Lernprozess

Das bayesianische Lernen beginnt wieder mit der Wahl einer Aprioriverteilung P_0 auf der Menge der Hypothesen. Wir unterstellen zunächst, dass alle Hypothesenwahrscheinlichkeiten positiv sind.



	H_0	H_1	H_2	H_3	H_4	H_5	H_6	H_7
P_0	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{3}{32}$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{1}{32}$	$\frac{1}{32}$	$\frac{1}{32}$
P_1	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$	$\frac{1}{15}$	$\frac{5}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
P_2	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	0	0	0	0

	H_8	H_9	H_{10}	H_{11}	H_{12}	H_{13}	H_{14}	H_{15}
P_0	$\frac{3}{32}$	$\frac{2}{32}$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{4}{32}$	$\frac{1}{32}$	$\frac{2}{32}$	$\frac{2}{32}$
P_1	0	0	0	0	0	0	0	0
P_2	0	0	0	0	0	0	0	0

Abbildung 4: Die Tabelle gibt drei unterschiedliche Wahrscheinlichkeitsverteilungen P_0, P_1 und P_2 für alle vollständigen Pfade an. Dabei ist P_0 eine Aprioriverteilung, aus der sich die Verteilungen P_1 und P_2 nacheinander durch die Beobachtungen $X_1 = 0$ und $X_2 = 0$ (s. die hervorgehobenen Äste) ergeben. Die jeweilige Wahrscheinlichkeit für die Beobachtung $X_3 = 0$ ergibt sich durch Addition der grau unterlegten Wahrscheinlichkeiten in der betreffenden Zeile. Entsprechend erhält man $P_0(X_3 = 0) = 19/32 \approx 0{,}594$, $P_1(X_3 = 0) = 6/15 = 0{,}600$ und $P_2(X_3 = 0) = 3/8 = 0{,}375$.

Die Aprioriverteilung wird mit jeder Beobachtung X_t angepasst. Nach Beobachtung von X_t mit $t = 1, 2, \ldots$ ist die neue Wahrscheinlichkeit einer Hypothese H_k gleich der alten bedingten Wahrscheinlichkeit von H_k gegeben die Beobachtung: $P_t(H_k) = P_{t-1}(H_k|X_t)$. Wie zuvor ergibt sich ein einfacher Anpassungsprozess. Die Wahrscheinlichkeiten aller Hypothesen, die X_t falsch vorhergesagt haben, fallen auf 0. Die Wahrscheinlichkeiten der nicht widerlegten Hypothesen werden durch die Summe dieser Wahrscheinlichkeiten dividiert, so dass sie sich wieder zu 1 addieren.

Die Anpassung der Hypothesenwahrscheinlichkeit verändert die Wahrscheinlichkeit für zukünftige Ereignisse. Für das Ereignis $X_3 = 0$ ergibt sich beispielsweise mit den Beobachtungen $X_1 = 0$ und $X_2 = 0$ jeweils eine neue Wahrscheinlichkeit (s. Abb. 4). Mit der Beobachtung von X_3 steht dann fest, ob das Ereignis eingetreten ist oder nicht; es gilt für alle $t \geq 3$ entweder $P_t(X_3 = 0) = 0$ oder $P_t(X_3 = 0) = 1$.

Man könnte nun meinen, dass im Zuge des bayesianischen Lernens die Vorhersagefunktion modifiziert wird. Das ist aber nur insofern der Fall, als dass unter der Verteilung P_t für $t > 0$ die Übergangswahrscheinlichkeiten zwischen Knoten, die nicht mehr erreicht werden können, auch nicht mehr definiert sind. Das ist jedoch offensichtlich irrelevant. Dagegen bleiben die Übergangswahrscheinlichkeiten zwischen Knoten, die noch erreicht werden können, unverändert.

Der Grund für das letzte Resultat ist einfach. Wir betrachten einen Knoten K' und seinen unmittelbaren Vorgänger K. Die Übergangswahrscheinlichkeit von K nach K' ist nur relevant, solange K' noch erreicht werden kann, aber noch nicht erreicht wurde. In diesen Fällen gilt $P_t(K'|K) = P_t(K')/P_t(K) = P_0(K'|K)$. Das ergibt sich daraus, dass die Wahrscheinlichkeit eines Knotens gleich der Summe der Wahrscheinlichkeiten aller Hypothesen ist, die das Erreichen dieses Knotens vorhersagen. Mit jedem Lernschritt werden diese Wahrscheinlichkeiten mit demselben positiven Faktor multipliziert, falls nach der neuen Beobachtung K' noch erreicht werden kann. Bei der Berechnung der bedingten Wahrscheinlichkeit $P_t(K'|K)$ kürzt sich dieser gemeinsame Faktor heraus.

Die Vorhersagefunktion eines Bayesianers ist also konstant und wird durch die Aprioriverteilung festgelegt. Da aber auch umgekehrt jede Vorhersagefunktion eine Aprioriverteilung festlegt, aus der sich wieder diese Vorhersagefunktion ergibt, können wir das bayesianische Lernen auch nur mit Hilfe einer Vorhersagefunktion beschreiben: Der Bayesianer wählt einfache eine feste Vorhersagefunktion. Die Wahrscheinlichkeiten für zukünftige Ereignisse, die in dieser Vorhersagefunktion festgeschrieben sind, stimmen zu jedem Zeitpunkt mit den Wahrscheinlichkeiten überein, die sich aus dem bayesianischen Lernen auf der Grundlage der durch die Vorhersagefunktion bestimmten Aprioriverteilung ergeben.

Der bayesianische Lernprozess liefert tatsächlich nur eine unnötig komplizierte Beschreibung der bayesianischen Rationalität. Die bayesianische Rationalität erfordert die Wahl irgendeiner festen Vorhersagefunktion – nicht mehr und nicht weniger. Die durch diese Vorhersagefunktion festgelegten bedingten Wahrscheinlichkeiten liefern die hypothetischen Vermutungen über die Zukunft für alle denkbaren Beobachtungen, also an jedem Knoten des binären Baumes. Können irgendwelche Knoten nicht mehr erreicht werden, hat das keine Auswirkungen auf die Vorhersagefunktion, aber die hypothetischen Vermutungen, die sich an diesen jetzt unerreichbaren Knoten ergeben hätten, sind natürlich nicht mehr relevant.

Die Beziehungen zwischen Aprioriverteilungen, Vorhersagefunktion und bayesianischem Lernen gelten genauso für unendliche lange Folgen und entsprechend unendliche binäre Bäume.[8]

3.3 Das „Anything-goes"-Theorem

Wir betrachten den allgemeinen Fall des Induktionsproblems und unendlich lange Beobachtungsfolgen. Die bayesianische Rationalität fordert, für den entsprechenden binären Baum eine beliebige Vorhersagefunktion zu wählen, aus der sich dann die Vorhersagen ergeben. Diese Vorhersagefunktion unterliegt keinerlei Beschränkungen.

Daraus folgt, dass keine Vermutungen oder Überzeugung über die Zukunft als bayesianisch-irrational ausgeschlossen werden, gleichgültig, welche und wie viele Beobachtungen bisher vorliegen. Das ist das „Anything-goes"-Theorem (AG-Theorem): Der Bayesianismus ist eine probabilistische Variante des radikalen Skeptizismus, also genau der Auffassung, zu deren Überwindung er angetreten war.[9]

Das AG-Theorem ist keine Überraschung. So schreibt Joyce (2004: 149):

> The most persistent objection to Bayesianism is that it engenders an untenable subjectivism in which all manners of preposterous beliefs and ludicrous inferences are immunized from criticism. If the constraints on rational opinion begin and end with the laws of probability and the conditioning rules, then Bayesianism allows a person to draw almost any conclusion on the basis of almost any evidence as long as she starts out with suitable prior beliefs. … In Bayesian epistemology, it appears, just about anything goes.

Soweit es um Vermutungen oder Überzeugungen über zukünftige Ereignisse geht, sind die Einschränkung – „*almost* any conclusion", „*almost* any evidence" und „*just about* anything goes" – tatsächlich hinfällig. Abgesehen davon, dass der Bayesianismus es nicht erlaubt, einem Widerspruch eine positive Wahrscheinlichkeit zuzuordnen, gilt „anything goes" ohne jede Einschränkung.

Joyce (2004: 149–151) diskutiert die Konvergenzidee und verwirft sie, weil er die entsprechenden mathematischen Theoreme für nicht anwendbar hält. Das scheint mir allerdings nicht richtig zu sein.

---

[8] Siehe Albert (2017). Alle Zusammenhänge lassen sich problemlos auf nicht-binäre Bäume verallgemeinern, solange es an jedem Knoten nur endlich viele Äste gibt. Die Zahl der Äste pro Knoten sowie die Länge der vollständigen Pfade durch den Baum kann unterschiedlich sein. Die Verallgemeinerung ergibt sich daraus, dass man jeden Knoten mit mehr als zwei Ästen durch einen endlichen binären Baum ersetzen kann; die Wahrscheinlichkeiten überschüssiger Pfade müssen auf 0 gesetzt werden. Vgl. auch Fn. 6 oben.

[9] Siehe Albert (1996). Zur Relevanz des AG-Theorems siehe auch Howson (2000: 206-208) und Gillies (2006: 83–84). Nyarko (1997: 176), von dem ich die Präsentation mit Hilfe einer Vorhersagefunktion übernommen habe, erwähnt ähnliche mathematische Resultate für bayesianisches Lernen im Kontext der Spieltheorie.

Das von Joyce diskutierte mathematische Theorem besagt folgendes: Wenn jede unendliche Beobachtungsfolge für jede Hypothese H_k zeigt, ob H_k wahr oder falsch ist, konvergiert die Wahrscheinlichkeit der wahren Hypothese gegen 1, wenn die Zahl der Beobachtungen gegen unendlich geht. Die Voraussetzung dieses Theorems ist für die hier betrachtete allgemeine Form des Induktionsproblems trivialerweise erfüllt, weil wir die Hypothesen mit den unendlichen Folgen identifizieren oder weil, in anderen Worten, keine Hypothese über beobachtbare Ereignisse hinausgeht. Damit werden auch die Wahrscheinlichkeiten, die unterschiedliche bayesianische Beobachter, die dieselben Beobachtungen machen, den Hypothesen zuordnen, gegeneinander konvergieren: für alle Beobachter gilt, dass die Wahrscheinlichkeit der wahren Hypothese gegen 1 und die jeder falschen Hypothese gegen 0 konvergiert.

Aus Sicht von Joyce ist dieses Resultat nicht anwendbar, möglicherweise deswegen, weil er konkurrierende Hypothesen unterstellt, die über die beobachtbaren Ereignisse hinausgehen. Aber für die bayesianische Betrachtung des Induktionsproblems ist dieser Einwand nicht relevant. Das Konvergenzresultat bezüglich der Hypothesenwahrscheinlichkeiten ist auf das Induktionsproblem anwendbar. Es ist nur völlig irrelevant, da die Konvergenz keine Konsequenzen für die Wahrscheinlichkeit zukünftiger Ereignisse hat.

Die Konvergenz der Hypothesenwahrscheinlichkeiten ergibt sich daraus, dass mit jeder Beobachtung unendlich viele Hypothesen widerlegt werden. Sei $\mathcal{H}_t$ die Menge der Hypothesen, die zum Zeitpunkt $t = 0, 1, 2, \dots, \infty$ noch nicht widerlegt sind. Dann gilt $\mathcal{H}_{t+1} \subset \mathcal{H}_t$ – die Wahrscheinlichkeitsmasse konzentriert sich auf immer kleinere Teilmengen von Hypothesen, wie das auch in Abb. 4 für den endlichen Fall erkennbar ist.

Allerdings ist im unendlichen Fall die Menge der verbleibenden Hypothesen immer noch überabzählbar unendlich groß und ausreichend, um alle denkbaren zukünftigen Beobachtungen vorherzusagen. An jedem Knoten beginnt ein unendlicher binärer Teilbaum, der die möglichen zukünftigen Entwicklungen beschreibt. Dieser Teilbaum ist nie kleiner ist als der binäre Baum ganz am Anfang, vor jeder Beobachtung. Die Vorhersagewahrscheinlichkeiten für den Teilbaum können immer noch beliebig sein; es gibt keinen Mechanismus, der dafür sorgt, dass diese Vorhersagewahrscheinlichkeiten ab irgendeiner Zahl von Beobachtungen eine höhere Wahrscheinlichkeit dafür ansetzen, auf dem wahren Pfad durch den Baum zu bleiben, als ihn zu verlassen. Konvergenz zur Wahrheit heißt nur, dass man immer mehr über die Vergangenheit weiß.

Entsprechend können auch die Vorhersagefunktionen zweier Bayesianer auf jedem Pfad durch den Baum in alle Ewigkeit völlig unterschiedlich sein. Betrachten wir eine beliebige Vorhersagefunktion, die an Knoten K der Beobachtung von Null die Wahrscheinlichkeit $\mu(K)$ zuweist, und eine alternative Vorhersagefunktion, die an Knoten der Beobachtung von Null die Wahrscheinlichkeit $\nu(K)$ zuweist. Offensichtlich können sich die beiden Funktionen an jedem Knoten des unendlichen Baumes beliebig unterscheiden, weil die Wahrscheinlichkeiten an jedem Knoten frei gewählt werden können.

Für Wahrscheinlichkeiten, die die Zukunft betreffen, gibt es also keine Konvergenzergebnisse, die für jede mögliche Aprioriverteilung gelten.

3.4 Ein zweiter Blick auf das Blün-Problem

Es ist lehrreich, das Blün-Problem anhand der Vorhersagefunktion zu analysieren. Der entsprechende binäre Baum enthält viele Nullen, da das Problem unterstellt, dass nach einer Eins nie wieder eine Null beobachtet wird. Unter der Annahme, dass die Hypothese H_0 (immer nur Nullen) wahr ist, sind fast alle Übergangswahrscheinlichkeiten irrelevant. Relevant und frei wählbar sind nur die Übergangswahrscheinlichkeiten α_t nach einer Folge von t Nullen zur nächsten Null (Abb. 5).

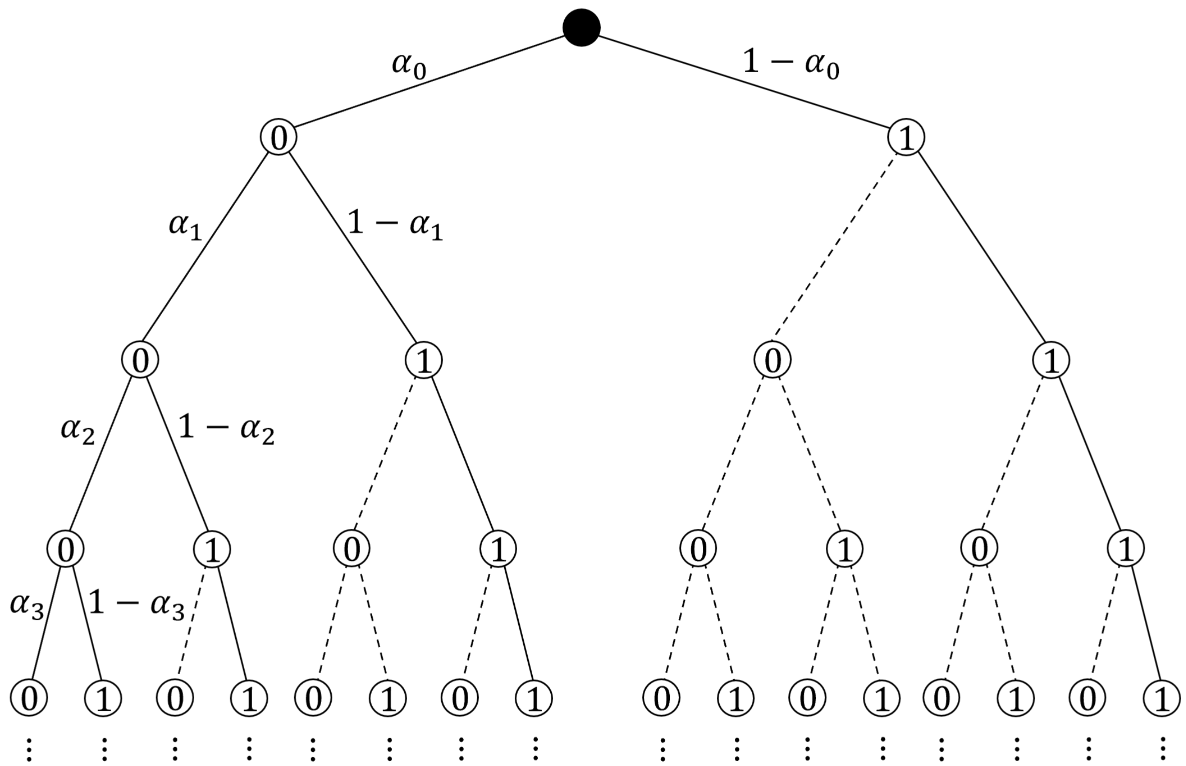


Abbildung 5: Die Vorhersagefunktion für das Blün-Problem. Der binäre Baum besteht aus allen möglichen unendlich langen binären Folgen (angedeutet durch die drei Punkte am Ende der eingezeichneten Pfade). Die Hypothese H_0 ist wahr; sie sagt den äußersten linken Pfad (eine unendliche Folge von Nullen) voraus. Die Alternativhypothesen sagen voraus, dass zu einem bestimmten Zeitpunkt nach lauter Nullen nur noch Einsen beobachtet werden. Die gestrichelt eingezeichneten Äste können daher nicht erreicht werden. Die nicht angegebenen Übergangswahrscheinlichkeiten sind an den durchgezeichneten Ästen gleich 1, an den gestrichelten Ästen 0 oder irrelevant.

Betrachten wir zunächst den Fall $\alpha_t = \alpha$ für $t = 0, 1, ...$ und $0 < \alpha < 1$. Die Wahrscheinlichkeit, dass der nächste gefunden Smaragd auch grün ist, wird also konstant gesetzt. Die Wahrscheinlichkeit der wahren Hypothese H_0 ist $\lim_{t\to\infty} \alpha^t$ =0 und bleibt auch 0. Die Wahrscheinlichkeit aller Hypothesen H_t, $t = 2, 3, ...$, die einen Übergang von Nullen zu Einsen nach $t-1$ Beobachtungen vorhersagen, ist dagegen gleich $1-\alpha$ und damit positiv. Die Wahrscheinlichkeit α kann beliebig nahe an 1 liegen, so dass man für alle praktischen Zwecke, beispielweise für eine Wette, und für eine beliebig hohe endliche Zahl N davon ausgeht, dass die nächsten N Smaragde grün sind.

Betrachten wir nun einen Fall, in dem $\alpha_{t+1} > \alpha_t$ für alle t. In diesem Fall wird es immer wahrscheinlicher, dass der nächste Smaragd grün ist. Sollte α_t mit einer zunehmenden Zahl von Beobachtungen schnell genug gegen 1 konvergieren, könnte $P_t(H_0) > 0$ für alle t gelten, aber das ist unerheblich. Entscheidend ist, wann die Wahrscheinlichkeit α_t groß genug wird, um sich für praktische Zwecke darauf zu verlassen, dass die nächsten N Smaragd grün sind. Mit der Wahl der Übergangswahrscheinlichkeiten α_t legt man willkürlich fest, ob und, wenn ja, wann das der Fall sein wird. Eine solche Festlegung kann man natürlich auch vornehmen, ohne sich mit der Wahrscheinlichkeitsrechnung zu belasten. Es ist nicht erkennbar, welcher Ertrag dem formalen Aufwand gegenübersteht, den der Bayesianismus erfordert.

3.5 Einige mögliche Einwände

Im Folgenden will ich auf einige naheliegende Einwände gegen die obige Kritik des Bayesianismus eingehen.

Der einfachste Einwand ist, dass der Bayesianismus definiert, was Rationalität ist. Das trifft natürlich auf den RC-Ansatz und insbesondere weite Bereiche der ökonomischen Theorie zu. Sobald man aber berücksichtigt, dass es auch aus bayesianischer Sicht im eigenen Interesse sein soll, bei der Urteilsbildung auf einen bayesianischen Lernprozess zu setzen, ist offensichtlich, dass es nicht genügt, Rationalität bayesianisch zu definieren. Man muss zeigen, dass es tatsächlich im eigenen Interesse liegt, bayesianisch vorzugehen. Das ist aber wohl kaum möglich, wenn diese Vorgehensweise darauf hinausläuft, dass man glauben kann, was man will, egal, welche Erfahrungen man bisher gemacht hat.[10]

Ein etwas besserer Einwand verbirgt sich hinter dem Schlagwort „garbage in, garbage out“. Wir können den Bayesianismus vor dem Hintergrund der digitalisierten Darstellung des Induktionsproblems als einen Ansatz für maschinelles Lernen oder künstliche Intelligenz betrachten. In diesem Rahmen werden die sogenannten „No-free-lunch“-Theoreme diskutiert (Forster 1999). Sie besagen, grob formuliert, dass gute Lernergebnisse nicht umsonst zu haben sind; jedes Programm für maschinelles Lernen enthält implizit gewisse Annahmen, die beschränken, was das Programm aus den Daten lernen kann. Aus den Daten alleine kann man nichts herausholen; das Programm muss festlegen, wie diese Daten verarbeitet

[10] Diese Bemerkung trifft insbesondere das sogenannte „Dutch-book“-Argument (s. z.B. Earman 1992: 38-44, Hájek 2009), das bei oberflächlicher Betrachtung einen technologischen Anspruch des Bayesianismus begründet, nämlich, dass die Verwendung subjektiver Wahrscheinlichkeitsverteilungen bei Entscheidungen vor sicheren Verlusten schützt. Siehe Albert (2017: 319–323) für eine detaillierte Widerlegung dieses Anspruchs. Alle nicht-technologischen Aspekte des „Dutch-book“-Arguments sind im vorliegenden Kontext irrelevant.

werden, und davon hängt ab, ob es lernt, gute Vorhersagen zu machen. Wenn das Programm Müll produziert, also unsinnige Ergebnisse liefert, bedeutet das, dass man Müll hineingesteckt hat. Ein solches Ergebnis kann niemanden überraschen, der mit dem Induktionsproblem vertraut ist.

Das „Anything-goes"-Theorem zeigt, dass der Bayesianismus ein Versuch ist, einen „free lunch" zu ergattern, also gute Lernergebnisse zu erhalten, ohne die Kosten dafür zu tragen. Im bayesianischen Kontext stecken alle Annahmen, die darüber entscheiden, ob das Lernergebnis gut oder schlecht ist, in den Aprioriwahrscheinlichkeiten. Dadurch, dass der Bayesianismus beliebige Aprioriverteilungen zulässt, können auch die Lernergebnisse völlig beliebig sein.

Um erfolgreich aus der Erfahrung zu lernen, muss man das Risiko eingehen, den Lernprozess in irgendeiner Weise zu beschränken. Im Rahmen des Bayesianismus heißt das: man muss die Menge der zulässigen Vorhersagefunktionen oder, was dasselbe ist, der zulässigen Aprioriwahrscheinlichkeiten für die Hypothesen beschränken. Solche Beschränkungen werden unter dem Stichwort „objektiver Bayesianismus" diskutiert.

Der objektive Bayesianismus hat eine lange Tradition, gilt aber vielen Bayesianern als gescheitert.[11] Es würde den Rahmen des vorliegenden Aufsatzes sprengen, diese Debatte zu schildern. Ich will nur einen prominenten Vorschlag, wie man zu solchen Beschränkungen kommen könnte, diskutieren.

Das zentrale Problem eines objektiven Bayesianismus ist, dass die Beschränkungen, die er einführt, nicht willkürlich sein dürfen. Dabei reicht es nicht zu argumentieren, die Beschränkungen seien plausibel. Schließlich kann man den Bayesianismus nur gegen konkurrierende Rationalitätskonzeptionen vertreten, wenn man seinen technologischen Anspruch verteidigen kann.

Eine solche Verteidigung ist schwierig. Beispielweise wurde häufig verlangt, die Aprioriverteilung müsse eine Gleichverteilung sein. Das klingt erst einmal vernünftig: wenn alle Vermutungen aus der Erfahrung erschlossen werden müssen, ist es plausibel, dass man anfangs alle Hypothesen gleichbehandelt. Allerdings führt die Gleichverteilungsannahme im binären Baum dazu, dass alle Übergangswahrscheinlichkeiten in alle Ewigkeit gleich $1/2$ und Beobachtungen irrelevant sind, weil zu jedem Zeitpunkt alle möglichen zukünftigen Erfahrungen gleichwahrscheinlich sind.[12]

Man könnte versuchen, eine Beschränkung der Aprioriverteilung aus Annahmen herzuleiten, die allgemein akzeptiert sind. Ein prominenter Vorschlag in diese Richtung firmiert unter dem Schlagwort

[11] Siehe dazu Gillies (1988), Earman (1992: 14-17, 138–141), Albert (2003: 113–114), Joyce (2004: 151–152), Howson/Urbach (2006: 265–297) und Albert (2017: 318 Fn. 2, 324 Fn. 8).

[12] Der Wert $1/2$ ergibt sich aus der Kodierung der Erfahrung im Binärformat. Wenn eine solche Kodierung mit der Gleichverteilungsannahme unvereinbar ist, wählt man eine andere Kodierung (s. Fn. 6 oben). Mit einer passenden Kodierung folgt aus der Gleichverteilungsannahme, dass an jedem Knoten des Baumes alle m Äste gleichwahrscheinlich sind.

„Gleichförmigkeit der Natur" (s. z.B. Humphreys 1987): man setzt voraus, dass alle Erfahrungen auf der Grundlage von Naturgesetzen erklärbar sind.

Solche Beschränkungen könnten als sogenannte „induktive Prinzipien" in die Induktionslogik eingebaut werden.[13] Damit wäre im Falle des Bayesianismus die Hoffnung verbunden, dass die Berücksichtigung dieser Einschränkung hinreichend viele Hypothesen ausschließt, so dass das bayesianische Lernen konvergieren muss.

Es lässt sich jedoch leicht zeigen, dass der Vorschlag nicht weiterhilft. Hier ist ein einfaches dynamisches System, das beliebige Vorhersagen generieren kann:[14]

$$X_t = 2Z_t \text{ div } 1$$
$$Z_{t+1} = 4Z_t \text{ mod } 1$$
$$Z_0 = \theta \in (0,1)$$

Dieses System – eine einfache Modifikation der chaotischen „baker map dynamics" –, beschreibt die Entwicklung der Beobachtungen X_t, also von Nullen und Einsen, durch zwei Gesetze, die man als kausal auffassen kann: die Variable Z_t beschreibt die Ursache sowohl von Z_{t+1} wie der beobachtbaren Variable X_t. Der Parameter θ ist der Anfangszustand, der auch die persönlichen Umstände des Beobachters umfassen kann. Für jede unendliche Folge von Nullen und Einsen gibt es unendlich viele Werte für θ, mit denen das System genau diese Folge produziert.

Jede Vorhersagefunktion für den kompletten binären Baum kann aus diesem dynamischen System hergeleitet werden, indem man eine passende Aprioriverteilung für den Parameter θ auf dessen Definitionsbereich, dem offenen Intervall $(0,1)$, wählt (Albert 2001). Das zeigt, dass die Annahme der „Gleichförmigkeit der Natur" keinerlei Beschränkungen liefert: es gilt immer noch „anything goes".

Als weiteren Einwand könnte man vorbringen, dass das „Anything-goes"-Theorem einen passiven Beobachter unterstellt, also nur Urteilsbildung, nicht aber irgendwelche sonstigen Entscheidungen und insbesondere keine Handlungen berücksichtigt. Unter diese Handlungen fällt insbesondere auch die Durchführung von Experimenten.

Aber auch dieser Einwand trägt nicht weit. Jede Handlung oder Folge von Handlungen hat unbekannte Konsequenzen für die weitere Erfahrung und führt damit zu einem eigenen unendlichen binären Baum. Wenn man beispielsweise unterstellt, dass die Erfahrungen, die man macht, durch das oben angegebene

[13] Siehe Musgrave (2011) für eine allgemeine Kritik an induktiven Prinzipien.
[14] „div" steht für die ganzzahlige Division, „mod" für den Rest dieser Division. Dann ist X_t gleich der Ziffer an der n-ten Nachkommastelle der Binärdarstellung von θ, wobei $n = 1 + 2t$. Siehe Albert (2001: 349-353) für eine ausführliche Darstellung.

dynamische System bestimmt werden, könnte man Handlungen dadurch berücksichtigen, dass man annimmt, dass jede Handlung in irgendeiner Weise den Parameter θ verändert. Damit führt jede Handlung sozusagen zu einem Neustart des Systems mit einem möglicherweise anderen Anfangszustand. Offensichtlich kann man dadurch nichts gewinnen, weil man immer noch keinerlei zukünftige Entwicklungen ausschließen kann.

Ein weiterer Einwand geht auf Savage (1954) zurück (s. auch Binmore 2009: 117). Man gesteht zu, dass der Bayesianismus das Induktionsproblem nicht lösen kann, dass das aber auch keiner anderen Rationalitätskonzeption gelingt. Über ein unlösbares Problem muss man dann nicht weiter diskutieren. Stattdessen behauptet man, dass der bayesianische Lernprozess nur für die Lösung kleiner, übersichtlicher Probleme (sogenannte „small worlds") geeignet ist. Bei solchen Problemen setzt man voraus, dass man bereits genügend gelernt hat – auf welchem Weg auch immer –, um sehr viele Hypothesen ausschließen zu können. Unter dieser Voraussetzung gelten häufig die Konvergenzeigenschaften; der technologische Anspruch des Bayesianismus lässt sich aufrechterhalten.

Die beste Antwort auf diesen Einwand ist der Vorschlag einer Rationalitätskonzeption, die das Induktionsproblem löst. Die angebliche Unlösbarkeit des Induktionsproblems ergibt sich nämlich nur, wenn man versucht, den Empirismus beizubehalten und den Deduktivismus durch einen Induktivismus zu ersetzen. Der kritische Rationalismus geht den anderen Weg: Beibehaltung des Deduktivismus und Ablehnung des Empirismus.

Auch im Kontext des kritischen Rationalismus könnte man unter Umständen bayesianische Verfahren verteidigen. Der Bayesianismus wäre damit aber auf einen subsidiären technischen Ansatz zur Lösung bestimmter Probleme reduziert. In diesem Kontext konkurriert er dann mit anderen Ansätzen, beispielsweise der klassischen Statistik oder nicht-bayesianischen Formen des maschinellen Lernens. Es wäre dann in jedem Anwendungsfall zu zeigen, dass die bayesianische Verfahren bessere Eigenschaften haben als die bekannten Alternativen.

4. Kritische Rationalität

4.1 Das Münchhausen-Trilemma

Der kritische Rationalismus behandelt im Kern die Frage, wie man zu wahren Überzeugungen gelangen kann. Ausgangspunkt für die Beantwortung dieser Frage ist die Einsicht, dass keine Aussage, egal welcher Art, sicher begründet werden kann. Eine sichere Begründung wäre ein unanfechtbarer Beweis, der jede Möglichkeit eines Irrtums definitiv ausschließt. Ein solcher Beweis ist aber unmöglich: er scheitert am Münchhausen-Trilemma (Hans Albert 1968).

Man kann das Münchhausen-Trilemma zunächst für deduktive Beweise formulieren. Ein deduktiver Beweis für A führt mindestens eine Prämisse B an, aus der A folgt. Die Aussage A ist aber nur bewiesen, wenn auch B bewiesen wird, denn der Beweis sichert nur die Wahrheitsübertragung von der Prämisse zur Konklusion, nicht aber die Wahrheit der Konklusion. Für einen unanfechtbaren deduktiven Beweis von A muss man also B beweisen und dazu eine Prämisse B‘ anführen, aus der B folgt. Dann müsste man B‘ beweisen, was eine weitere Prämisse B‘‘ erfordert, und so fort.

Ein solcher Begründungsregress kann nur drei Verläufe nehmen. Entweder ist der Regress unendlich. Oder man kommt zu einem logischen Zirkel, bei dem man beispielsweise B wieder als Begründung für B‘‘ anführt (was, wenn man keinen logischen Fehler gemacht hat, bedeutet, dass die Aussagen B, B‘ und B‘‘ logisch äquivalent sind). Oder man bricht das Verfahren an irgendeinem Punkt mit einer unbegründeten Prämisse ab. In allen drei Fällen ergibt sich offensichtlich kein unanfechtbarer Beweis für die Aussage A.

Das Trilemma tritt aber nicht nur bei deduktiven Beweisen auf, sondern bei allen Arten von Begründungen, insbesondere auch dann, wenn man sich nicht nur auf Argumente beschränkt, sondern auf irgendwelche anderen Verfahren der Begründung zurückgreift: Begründung durch Beobachtung, durch Intuition oder was auch immer. Jede Möglichkeit, einen Zweifel zu äußern, erfordert einen weiteren Begründungschritt, der zeigt, dass an der betreffenden Stelle kein Irrtum vorliegt. Zu jedem Begründungsschritt kann man zwei Fragen stellen: Wurde der Begründungsschritt korrekt durchgeführt? Und leistet das betreffende Begründungsverfahren überhaupt, was es leisten soll? Beide Fragen erfordern eine Antwort und lösen damit jeweils einen eigenen Begründungsregress aus.

Im Falle der Deduktion könnte man unterstellen, dass ein deduktives Verfahren laut Definition eines gültigen Arguments leistet, was es soll, nämlich die Wahrheitsübertragung von den Prämissen auf die Konklusion. Das Trilemma zeigt, dass man trotzdem nicht zu einem unanfechtbaren Beweis kommen kann. Aber darüber hinaus kann man bei jedem einzelnen Schritt die Frage stellen, ob tatsächlich ein gültiges Argument vorliegt, und damit einen zusätzlichen Begründungsregress auslösen.

Deduktive Begründungen haben einen weiteren Nachteil. Bei einer deduktiven Begründung kann die Prämisse nur wahr sein, wenn die Konklusion wahr ist. Daher ist unter dem Aspekt der Sicherheit nichts gewonnen, wenn man eine Prämisse anführt, aus der die zu begründende Aussage deduktiv folgt: die Prämisse ist mindestens so unsicher wie die Konklusion.

4.2 Der technologischer Anspruch

Wie der Bayesianismus kann auch der kritische Rationalismus als eine Entscheidungstheorie unter Ungewissheit aufgefasst werden, die allerdings nicht von subjektiven Wahrscheinlichkeiten Gebrauch macht. Stattdessen erhalten Aussagen, wenn das möglich ist, vorläufige, das heißt jederzeit revidierbare Wahrheitswerte – wahr oder falsch – zugewiesen; wenn eine Zuweisung nicht möglich ist, bleibt ihr Wahrheitswert unbestimmt.

Da der kritische Rationalismus eine Entscheidungstheorie ist, liegt es nahe, ihn explizit in Form von Entscheidungsregeln zu formulieren. Es genügt aber nicht, irgendwelche plausibel klingenden Regeln zu formulieren. Es muss sich um Regeln handeln, die dem Ziel der Erkenntnis dienen. Der kritische Rationalismus muss einen technologischen Anspruch erheben, der sich vielleicht wie folgt formulieren lässt.

TA: Die beste Methode, um zu wahren und informativen Aussagen über die Welt und zu wahren Aussagen über logische oder mathematische Zusammenhänge zu gelangen, besteht darin, die Entscheidungsregeln des kritischen Rationalismus anzuwenden.

Die Entscheidungsregeln des kritischen Rationalismus adressieren eine Person, die, etwas verkürzt gesagt, nach Wahrheit sucht. Dieses Ziel wird angesichts der langen Liste von Theorien, die früher akzeptiert waren, inzwischen aber als falsch gelten, gelegentlich als zu ehrgeizig abgelehnt. Möglicherweise können wir bestenfalls Theorien finden, die in irgendeinem Sinne „wahrheitsähnlich" sind – eine Idee, die auch Popper vertreten hat. Deswegen, so die Argumentation, sollten wir die Suche nach Wahrheit aufgeben und gleich nur nach wahrheitsähnlichen Theorien suchen.[15]

Allerdings ist unklar, wie sich eine Suche nach wahrheitsähnlichen Theorien von der Suche nach wahren Theorien unterscheiden würde und warum man ein Ziel aufgeben sollte, das man, wenn auch vielleicht nur zufällig, erreichen könnte. Aber der entscheidende Punkt ist, dass TA nicht behauptet, dass man mit Hilfe des kritischen Rationalismus jemals wahre Theorien, beispielsweise in der Physik, findet. Vielleicht werden alle großen Theorien, die wir formulieren, falsch sein. Das schließt aber nicht aus, dass viele Folgerungen, die wir aus diesen Theorien ableiten, wahr sind. Das beste historische Beispiel ist die klassische Mechanik, die wir heute für falsch halten, die aber so viele wahre Folgerungen liefert, dass es sehr schwer war, sie empirisch zu widerlegen.

Der Weg zu wahren Aussagen kann ohne weiteres über die irrtümliche Akzeptanz falscher, aber wahrheitsähnlicher Theorien führen. Insbesondere könnte es sein, dass alle Folgerungen aus solchen Theorien, die uns interessieren, wahr sind – darunter nicht nur praktisch nützliche Folgerungen, sondern möglicherweise auch die Folgerungen über fundamentale Naturgesetze und darüber, welche Arten von Dingen existieren.

[15] Siehe z.B. Chakravartty (2017: 22–26) zur sogenannten „pessimistischen Induktion" und zur Wahrheitsähnlichkeit.

4.3 Entscheidungsregeln: Akzeptanz und Kritik

Zu den Entscheidungsregeln, auf die TA verweist, gehört mindestens eine Akzeptanzregel, die festlegt, wann Aussagen akzeptiert werden, und Regeln der Kritik, die im Kontext der Akzeptanzregel relevant sind.[16] Die Akzeptanzregel könnte etwa folgendermaßen lauten:

(AR) Akzeptiere eine Aussage genau dann, wenn sie in eine der folgenden Kategorien fällt und keiner anderen Aussage widerspricht, die ebenfalls in eine dieser Kategorien fällt.

a. Eigene Wahrnehmungsurteile und andere intuitive Überzeugungen, die Du bisher nicht erfolgreich kritisiert hast.

b. Von vertrauenswürdigen Personen übernommene Überzeugungen, insbesondere auch Wahrnehmungsurteile, die Du bisher nicht erfolgreich kritisiert hast.

c. Aussagen, die sich bewährt, also einer strengen kritischen Überprüfung standgehalten haben.

d. Aussagen, die deduktiv aus Aussagen der anderen Kategorien folgen.

Aussagen, die gemäß AR akzeptiert werden, gelten im Sinne des kritischen Rationalismus als begründet. Nur die Aussagen der Kategorie d können als durch deduktive Argumente begründet gelten. Die Aussagen der anderen Kategorien werden nicht durch Argumente, ob induktiv oder deduktiv, begründet. Sie gelten als begründet, weil sie aus einer bestimmten Quelle stammen oder einer kritischen Überprüfung standgehalten haben.

Eine ausführliche Erläuterung der Regeln der Kritik und der damit verbundenen Begriffe würde hier zu weit führen. Einige wenige Bemerkungen müssen genügen.

Um eine Aussage erfolgreich zu kritisieren, muss man entweder die Aussage selbst widerlegen oder die Annahme, dass sie aus einer zuverlässigen Quelle stammt. Ist die Aussage widerlegt, wird ihre Negation akzeptiert, das heißt, es wird akzeptiert, dass sie falsch ist. Wird nur gezeigt, dass die Aussage aus einer unzuverlässigen Quelle stammt, wird die Aussage nicht mehr akzeptiert; es bleibt dann offen, ob sie wahr oder falsch ist.

Eine Widerlegung kann empirisch oder logisch sein. Eine empirische Widerlegung erfolgt anhand eines Widerspruchs mit Wahrnehmungsurteilen. Eine logische Widerlegung erfolgt durch Feststellung eines

[16] Dieser Abschnitt stützt sich wesentlich auf Musgrave (1999: 314–350) und versucht, Musgraves Rekonstruktion des kritischen Rationalismus explizit als Entscheidungstheorie zu formulieren (s. auch Albert 2011b). Musgrave formuliert keine Regeln (also Imperative), sondern – wie weiter unten erläutert – Aussagen über die Bedingungen, unter denen es rational ist, eine Aussage zu akzeptieren. Den technologischen Anspruch des kritischen Rationalismus formuliert Musgrave nicht explizit (vgl. aber Fn. 1 oben). Außerdem führt die unpersönliche Formulierung zu Problemen, die sich nicht leicht vermeiden lassen (Musgrave 1999: 347–350).

inneren Widerspruchs, also eine logisch-mathematische Untersuchung. Widerlegungen sind natürlich nie unanfechtbare Beweise der Falschheit, sondern nur vorläufige Widerlegungen, die selbst wieder überprüft und gegebenenfalls widerlegt werden können.

Eine strenge Kritik ist eine sorgfältige Überprüfung, die gemäß dem sogenannten „Hintergrundwissen“ (d.h., den bisher akzeptierten Aussagen) geeignet ist, die Aussage zu widerlegen, wenn sie falsch sein sollte.

Zur Kategorie a: Wahrnehmungsurteile sind Aussagen, die intuitiv durch Wahrnehmung erworbene Überzeugungen wiedergeben. Sie gehen damit über Aussagen, die Sinneseindrücke beschreiben, hinaus. Wenn man vor sich eine Wand sieht, gelangt man intuitiv zu der Überzeugung, dass man vor einer Wand steht. Das Wahrnehmungsurteil lautet also „Dort ist eine Wand“ und nicht, wie im Empirismus, „Mir scheint, als sei dort eine Wand“ (oder eine ähnliche Formulierung, die vorgeblich nur Sinneseindrücke beschreibt). Eigene Wahrnehmungsurteile sind gemäß AR vorläufig zu akzeptieren, es sei denn, man hätte sie erfolgreich kritisiert.

Wahrnehmungsurteile können beispielsweise dadurch kritisiert werden, dass man die Aussage akzeptiert, dass die Wahrnehmung unter den gegebenen Bedingungen immer zu falschen Urteilen oder zumindest nicht immer zu richtigen Urteilen führt (Kategorie c). An diesem Punkt läuft AR darauf hinaus, den eigenen Sinnen zu trauen, solange nichts dagegenspricht.

Es gibt weitere intuitiv erworbene oder angeborene Überzeugungen, die zunächst ohne weitere Überprüfung weitgehend implizit akzeptiert werden, beispielsweise, dass das eigene Gedächtnis zuverlässig ist, dass Dinge nicht deswegen nicht mehr existieren, weil man sie gerade nicht sieht, oder dass manche Personen (etwa die eigenen Eltern) vertrauenswürdig sind. Würde man von diesen Überzeugungen erst Gebrauch machen, wenn man sie sorgfältig geprüft hätte, könnte man nicht überleben.

Insbesondere lässt sich die Zuverlässigkeit des eigenen Gedächtnisses nicht überprüfen, ohne von diesem Gedächtnis Gebrauch zu machen. Was sich überprüfen lässt, ist die Zuverlässigkeit unter bestimmten Umständen oder in bestimmter Hinsicht.

Zur Kategorie b: Die Übernahme von Überzeugungen anderer Personen erfolgt meist aufgrund von Mitteilungen; allerdings muss man die Personen erst einmal als vertrauenswürdig einstufen. Ausgangspunkt sind intuitive Urteile der Kategorie a darüber, wer vertrauenswürdig ist. Diese intuitiven Urteile können aber kritisiert werden, genauso wie die übernommenen Überzeugungen selbst.

Zunächst können Mitteilungen anderer Personen manchmal anhand der eigenen Wahrnehmung überprüft und widerlegt werden: widerspricht eine Mitteilung einer Aussage der Kategorie a, gilt sie als widerlegt. Andere Möglichkeiten, Mitteilungen zu kritisieren, stützen sich auf Aussagen der Kategorie c, die besagen, dass man den Mitteilungen anderer Personen unter bestimmten Bedingungen nicht trauen kann, beispielsweise weil sie von einer Täuschung profitieren können oder nicht kompetent sind, um zu beurteilen, was sie behaupten. In vielen Bereichen wird man daher (als Erwachsener) nur den Mitteilungen unabhängiger Fachleute trauen.

Aussagen der Kategorie b sind nicht nur überlebenswichtig, sondern sie sind die Grundlage einer kognitiven Arbeitsteilung, die personenübergreifende Lernprozesse und Wissenschaft möglich macht.

Wahrnehmungsurteile und andere intuitive Überzeugungen, insbesondere auch über die Vertrauenswürdigkeit anderer Personen, haben einen besonderen Status. Man beginnt mit der impliziten Überzeugung, dass diese Aussagen wahr sind. Auf der Grundlage von empirischer Kritik wird diese Überzeugung nicht verworfen, sondern in abgeschwächter Form beibehalten, indem man zusätzliche Bedingungen dafür einführt, wann man die betreffenden Aussagen akzeptiert und wann nicht.

Diese Vorgehensweise könnte man auch bei Hypothesen der Kategorie c und insbesondere bei wissenschaftlichen Theorien wählen. Das ist aber nicht der Fall. Eine Hypothese der Kategorie c, die eine Überprüfung nicht übersteht, wird nicht in abgeschwächter Form aufrechterhalten, sondern verworfen: es wird akzeptiert, dass sie falsch ist.

Zur Kategorie c: Die wichtigste Form der kritischen Überprüfung ist die empirische Überprüfung von Hypothesen oder Theorien anhand ihrer Vorhersagen. Sollte man die Überprüfung einer Hypothese selbst durchgeführt haben, würde das Urteil, dass die Hypothese bewährt ist, auf Aussagen der Kategorie a beruhen. Meistens aber fällt das Urteil in Kategorie b, und oft in sehr indirekter Weise, als eine Form von Hörensagen: Man lernt von einer kompetenten Person, beispielsweise einem Lehrer, dass andere kompetente Personen – Wissenschaftler – solche Überprüfungen durchgeführt haben und zu dem Schluss gekommen sind, dass sich die Hypothese bewährt hat.

Offensichtlich spielen bei vielen der obigen Überlegungen logisch-mathematische Aussagen, insbesondere auch Aussagen über Folgerungszusammenhänge, eine große Rolle. Solche Aussagen fallen in einfachen Fällen in Kategorie a: Das intuitive Sprachverständnis führt dazu, dass man auch ohne Logikkurs bestimmte logische Zusammenhänge erkennt.

In schwierigeren Fällen fallen Aussagen über Folgerungszusammenhänge in Kategorie c. Sie gelten dann als bewährt, wenn man auch bei sorgfältiger Prüfung keinen Einwand finden konnte. Diese sorgfältige Prüfung kann beispielsweise darin bestehen, dass man die relevanten Aussagen formalisiert und einen logischen oder mathematischen Beweis führt. Auch in diesem Fall kann man sich aber letztlich nur auf Aussagen der Kategorie a und b berufen. Aussagen der Kategorie b sind relevant, wenn man Ergebnisse aus der Logik oder Mathematik aus als zuverlässig eingeschätzten Quellen übernimmt. Am

Ende aber stützt man sich auf Aussagen der Kategorie a, nämlich auf die Aussage, dass man bei einer sorgfältigen Überprüfung der Formalisierung und der Beweisführung keinen Fehler gefunden hat.[17]

Zur Kategorie d: Zu diesen Aussagen ist nur anzumerken, dass sie im Kontext der Kritik eine große Rolle spielen, weil insbesondere Aussagen der Kategorie c anhand ihrer Folgerungen überprüft werden.

Bei der Anwendung der Akzeptanzregel kann offensichtlich das Problem auftreten, dass man einander widersprechende Aussagen findet, von denen dann keine akzeptiert werden kann, auch wenn jede von ihnen in eine der vier Kategorien fällt. Zu den Regeln des kritischen Rationalismus gehört, dass man in diesen Fällen versucht, eine Entscheidung zwischen diesen Aussagen herbeizuführen.

Dabei gelten unter Umständen Vorrangregeln. Wahrnehmungsurteile haben beispielsweise Vorrang gegenüber anderen Aussagen; bei einem Konflikt werden die Wahrnehmungsurteile akzeptiert und die widersprechenden Aussagen verworfen. Aussagen der Kategorie c können allerdings verwendet werden, um Wahrnehmungsurteile zu kritisieren und damit auch frühere Widerlegungen rückgängig zu machen. In der Wissenschaft gibt es zahlreiche Regeln, die festlegen, wie man richtig beobachtet, Daten erhebt oder Experimente durchführt. Diese Regeln dienen dazu, bekannte Fehlerquellen bei der Beobachtung auszuschließen. Beobachtungsaussagen, die ohne Berücksichtigung dieser Regeln gewonnen werden, werden meist nicht akzeptiert – genauso wenig wie Beobachtungsaussagen, die nicht von kompetenten und unabhängigen Fachleuten stammen.

Es ist möglich, dass zwei einander widersprechende Aussagen oder Theorien beide als bewährt angesehen werden. Ein Beispiel sind die allgemeine Relativitätstheorie und die Quantentheorie, die beide sorgfältig geprüft und unwiderlegt sind, die aber angeblich nicht beide wahr sein können. Musgrave schlägt vor, dass man in einem solchen Fall keine der beiden Theorien akzeptieren sollte.

Man könnte meinen, dass diese Klausel dazu führt, dass man auch die Folgerungen aus zwei bewährten, aber einander widersprechenden Theorien nicht einfach akzeptieren kann, selbst wenn diese Folgerungen der jeweils anderen Theorie nicht widersprechen. Wäre das der Fall, müsste man alle solchen Folgerungen jeweils separat überprüfen, bevor man sich darauf stützten kann. Dieses Problem entsteht aber tatsächlich nicht, weil die Folgerungen aus einer bewährten Aussage oder Theorie ebenfalls als bewährt gelten. Soweit diese Folgerungen also keiner anderen bewährten Theorie widersprechen – das könnte etwa auf die Folgerungen der Relativitätstheorie über die Planetenbewegungen gelten –, sind sie gemäß AR zu akzeptieren.

[17] Eine Frage, die ich im Rahmen dieses Aufsatzes ausklammere, ist, unter welchen Bedingungen metaphysische Aussagen als bewährt angesehen werden können. Der kritische Rationalismus erlaubt es grundsätzlich, metaphysische Aussagen zu akzeptieren. Metaphysische Aussagen, die Folgerungen aus bewährten Theorien sind, gelten als bewährt; dazu gehören Aussagen wie „Es gibt Naturgesetze“, die aus jeder bewährten Theorie folgen, die bestimmte Naturgesetze postuliert. Aber bei anderen metaphysischen Aussagen, die nicht Bestandteil wissenschaftlicher Theorien sind, ist es schwieriger, Regeln der Kritik anzugeben.

Die Anwendung der Entscheidungsregeln lässt sich auch in Form gültiger deduktiver Argumente darstellen. Diese Argumente liefern jedoch keine deduktive Begründung für die zu begründenden Aussagen. Das lässt sich leicht an der von Musgrave (1999) gewählten Darstellung des kritischen Rationalismus zeigen.

Gemäß der Definition der kritischen Rationalität ist es genau dann rational, eine Aussage zu akzeptieren, wenn sie die in AR oben genannte Bedingung erfüllt: sie fällt in eine der vier Kategorien a–d und widerspricht keiner anderen Aussage, die in eine dieser Kategorien fällt. Bezeichnen wir diese Bedingung kurz als Bedingung B. Dann können wir Anwendungen von AR wie folgt in Form gültiger deduktiver Argumente darstellen.

> *Prämisse 1:* Es ist genau dann rational, eine Aussage zu akzeptieren, wenn sie Bedingung B erfüllt.
> *Prämisse 2:* Aussage A erfüllt Bedingung B.
> *Konklusion:* Es ist rational, Aussage A zu akzeptieren.

Was in dieser Formulierung also deduktiv begründet wird, ist nicht die Aussage A, sondern die Entscheidung, A zu akzeptieren. Sofern A nicht in die Kategorie d fällt, wird damit der anfangs genannte Nachteil deduktiver Begründungen vermieden. In einer deduktiven Begründung der Aussage A können die Prämissen des Arguments nur wahr sein, wenn A wahr ist, so dass implizit A als Begründung für sich selbst angeführt wird. Eine Begründung dafür, die Aussage A zu akzeptieren – beispielsweise, weil sie aus einer zuverlässigen Quelle stammt –, kann dagegen Gründe anführen, die logisch unabhängig von A sind.

4.4 Die kognitive Verankerung der Regeln

Der kritische Rationalismus beruft sich unter anderem darauf, dass die Erfolge der Wissenschaft im Großen und Ganzen in Übereinstimmung mit seinen Entscheidungsregeln erzielt wurden und dass die Wissenschaft nicht grundsätzlich anders funktioniert als der Alltagsverstand. Ein solches Argument setzt voraus, dass Menschen die Regeln des kritischen Rationalismus grundsätzlich anwenden können, auch wenn sie das nicht unbedingt immer tun.

Es ist unmöglich, die Bildung von Überzeugungen durch eine Menge von Regeln zu beschreiben, die alle bewusst angewendet werden. Die Anwendung solcher Regeln verlangt Fallunterscheidungen, die wieder auf der Grundlage von Regeln erfolgen müssen, deren Anwendung wiederum Fallunterscheidungen verlangt, und so fort. Will man einen infiniten Regelregress vermeiden, muss man unterstellen, dass an irgendeiner Stelle Fallunterscheidungen ohne bewussten Rückgriff auf eine Regel ausgeführt werden.

Es ist daher unvermeidlich, dass man die Bildung von Überzeugungen als einen psychologischen Prozess betrachtet, der teilweise präattentiv, also unterhalb der Aufmerksamkeitsschwelle abläuft. Im Sinne der „Dual-Process"-Modelle der kognitiven Psychologie (s. z.B. Stanovitch/Stanovitch 2010) handelt es sich bei der Akzeptanzregel AR um eine Regel, die im Zuge des bewussten kritischen Denkens (Typ-2-Prozesse) angewendet werden kann, um die Ergebnisse präattentiver Typ-1-Prozesse zu korrigieren.

Die Bildung von Überzeugungen wird in vielen Fällen automatisch in einem Typ-1-Prozess erfolgen. Das gilt insbesondere für Wahrnehmungsurteile, aber auch für die Einordnung von Aussagen in die vier Kategorien: häufig erinnert man sich, ob man etwas selbst wahrgenommen hat oder ob man nur davon gehört hat. Auch die Überzeugung, dass eine neue Überzeugung keinen anderen Überzeugungen widerspricht, bildet sich im Allgemeinen in einem Typ-1-Prozess. Dasselbe gilt in einfachen Fällen für Überzeugungen, die den logischen Zusammenhang zwischen Aussagen betreffen (Kategorie d), denn diese Überzeugungen geben sich aus dem intuitiven Sprachverständnis.

Vereinfachend können wir annehmen, dass eine Person, die sich auf AR stützt, eine große Menge von Überzeugungen hat, die sie ohne weiteres in eine der obigen Kategorien einordnen kann, wenn sie über eine solche Einordnung bewusst nachdenkt. Aber natürlich erfolgt eine solche Einordnung nicht unbedingt irrtumsfrei.

Die Akzeptanzregel AR kann bewusst angewendet werden, wenn die Person einzelne Überzeugungen oder andere, möglicherweise neue Aussagen kritisch untersucht, also die Frage stellt, ob diese Aussagen tatsächlich gemäß AR akzeptiert werden können. Bei einer solchen kritischen Untersuchung kann auch die Einordnung der kritisch betrachteten Aussage in Frage gestellt werden. Wenn die Zuordnung einer Aussage A zu Kategorie X in Frage gestellt wird, bedeutet das, dass man die Meta-Aussage „Aussage A fällt in Kategorie X" betrachtet und überprüft, ob diese Aussage akzeptiert werden kann.

Weiter vereinfachend können wir zunächst unterstellen, dass bei der kritischen Untersuchung einzelner Aussagen alle anderen Überzeugungen der Person als akzeptiertes „Hintergrundwissen" fungieren. Auf der Grundlage des Hintergrundwissens wird dann in einem Suchprozess entschieden, ob die zu prüfende Aussage vorläufig als wahr oder falsch eingeordnet werden kann oder ob derzeit keine Einordnung möglich ist.

Die vorliegende Beschreibung der Akzeptanz- und Kritikregeln ist sicher nicht vollständig. Vor allem aber ist die Form der Regeln noch nicht wirklich befriedigend. Letztlich legen die Regeln fest, wie sich die Zuweisung von Wahrheitswerten (wahr, falsch und unbestimmt) an eine Menge von Aussagen verändert, wenn man eine Menge von akzeptierten Aussagen bereinigt oder an neu hinzugekommene Aussagen anpasst. Idealerweise würde man eine solche Menge von Aussagen formal beschreiben und dann explizit angeben, wie die Zuordnung von Wahrheitswerten erfolgt. Dieser Zuordnungsprozess ist rekursiv, weil Wahrnehmungsurteile zunächst bestimmte Aussagen widerlegen können, andere Wahrnehmungsurteile auf dem Weg über eine Kritik der Zuverlässigkeit der Wahrnehmung diese Widerlegung aber wieder aufheben können.

In der vorliegenden Form bleibt der rekursive Charakter der Regeln implizit. Eine explizite Formulierung in einem formalen Modell würde ein Verständnis der Regeln erleichtern. Selbst wenn ein solches Modell viele Vereinfachungen enthielte, wäre es vermutlich hilfreich, um die Akzeptanz- und Kritikregeln besser zu verstehen und weiterzuentwickeln. Man könnte in einem ersten Schritt beispielsweise versuchen, eine Art Gleichgewicht des rekursiven Prozesses zu beschreiben, in dem weitere Überprüfungen nicht mehr zu Veränderungen der Zuordnung führen. In einem solchen hypothetischen Zustand könnten Veränderungen nur durch neue Wahrnehmungen, neue Mitteilungen oder andere neue Ideen ausgelöst werden.

4.5 Die Begründung des kritischen Rationalismus

Damit kritische Rationalisten den kritischen Rationalismus als Lösung des Induktionsproblems akzeptieren können, müssen sie zeigen, dass der technologische Anspruch des kritischen Rationalismus (TA) begründet ist.[18] Für TA kommt nur Kategorie c in Frage: Der kritische Rationalist muss behaupten, dass TA bewährt ist, also eine strenge Überprüfung überstanden hat, und keiner anderen bewährten Aussage widerspricht.

Es ist unbestritten, dass TA ernsthaft kritisiert wurde. Die Frage ist daher nur, ob die Kritik erfolgreich war. Die kritischen Rationalisten behaupten, dass keiner der Einwände gegen TA sich als stichhaltig erwiesen hat, so dass sich TA bewährt hat, und dass keine Auffassung, die TA widerspricht, sich ebenfalls bewährt hat. Nehmen wir einmal an, dass sei richtig. Dann wäre TA vorläufig als wahr zu akzeptieren.

Mit anderen Worten: Der kritische Rationalismus könnte auf Grundlage seiner selbst akzeptiert werden. Musgrave (1999: 330) beschreibt diese Begründung als zirkulär, aber damit ist kein logischer Zirkel gemeint. Das ergibt sich schon allein daraus, dass eine strenge Überprüfung hätte zeigen können, dass TA nicht akzeptiert werden kann.

Beispielsweise hätte es sein können, dass der kritische Rationalismus an internen Widersprüchen scheitert. Gelegentlich wird beispielsweise die Argumentation des Münchhausen-Trilemmas als in sich widersprüchlich kritisiert: es werde mit dem Trilemma versucht, sicher zu begründen, dass es keine sicheren Begründungen gibt. Wenn man auf der Grundlage dieser Kritik zu dem Schluss käme, dass sichere

[18] Siehe Musgrave (1999: 330–331, 336–337) für die Auffassung, dass die Entscheidung, rational zu sein, selbst rational getroffen werden muss. Zusammen mit einer instrumentellen Interpretation von „rational“ (s. auch Fn. 1 oben) ergibt sich implizit die Position, dass man als kritischer Rationalist einen technologischen Anspruch des kritischen Rationalismus auf kritisch-rationaler Grundlage akzeptieren können muss.

Begründungen doch möglich sein könnten, dann ließe sich auch TA angreifen, denn die Entscheidungsregeln des kritischen Rationalismus lassen eine Suche nach sicheren Begründungen nicht zu und wären damit möglicherweise nicht die besten Regeln für die Erkenntnis.

Diese Argumentation ist aber rein hypothetisch, denn mit dem Münchhausen-Trilemma wird eben nicht versucht, sicher zu begründen, dass es keine sicheren Begründungen gibt. Es wird gezeigt, dass alle bekannten Vorstellungen darüber, wie eine Begründung aussehen könnte, im Widerspruch zu der Auffassung stehen, dass sichere Begründungen möglich sind. Trotz vieler Versuche konnte bisher niemand zeigen, dass der Widerspruch nicht existiert oder dass es eine Form der Begründung gibt, bei der der Widerspruch nicht auftritt. Die Aussage, dass sichere Begründungen unmöglich sind, hat sich bewährt; auch konnte niemand zeigen, dass diese Aussage im Rahmen des kritischen Rationalismus zu einem Widerspruch führen würde. Aus Sicht des kritischen Rationalismus genügt das, um die Aussage vorläufig zu akzeptieren.

Die Argumentation des Münchhausen-Trilemmas liefert also eine Begründung für die Aussage, dass sichere Begründungen unmöglich sind, aber wie alle Begründungen im kritischen Rationalismus wird nicht der Anspruch erhoben, dass es sich um eine sichere Begründung handelt.

Es hätte auch sein können, dass eine Untersuchung der Wissenschaftsgeschichte ergibt, dass die erfolgreichen wissenschaftlichen Theorien in einem Prozess gefunden und akzeptiert wurden, in dem die Regeln des kritischen Rationalismus systematisch verletzt werden. Die Kritiker des kritischen Rationalismus, von Polanyi über Kuhn und Lakatos bis Feyerabend, haben das immer wieder behauptet, aber diese Behauptungen sind nach derzeitigem Stand nicht haltbar (Anderson 1984, Wootton 2015).

Der kritische Rationalismus setzt, wie gesagt, nicht TA als Prämisse voraus, um daraus TA zu folgern. Die Begründung von TA ist logisch unabhängig von TA: Die Aussage TA kann falsch sein, selbst wenn die Behauptung, die zur Begründung angeführt wird – nämlich, dass TA eine strenge Überprüfung überstanden hat – wahr ist. Die kritischen Rationalisten wählen also bei der Begründung des kritischen Rationalismus nicht den logischen Zirkel, sondern den vorläufigen Abbruch an einer Stelle, an der den Kritikern des kritischen Rationalismus die Argumente ausgehen.

Ein letzter Einwand gegen den kritischen Rationalismus, der immer wieder vorgebracht wurde, ist der Vorwurf, dass der kritische Rationalismus, der doch angeblich Induktion ablehnt, selbst nur eine Form der Induktion beschreibt: man schließe aus vergangenen Erfolgen einer Theorie darauf, dass sie wahr oder wahrscheinlich wahr sei oder zumindest in Zukunft weiter erfolgreich sein werde.

Die obige Darstellung des kritischen Rationalismus zeigt, dass dieser Einwand nicht zutrifft. Die moderne wahrscheinlichkeitstheoretische Form des Induktivismus, der Bayesianismus, funktioniert völlig anders als der kritische Rationalismus. Die andere, traditionellere Form des Induktivismus, die ohne Wahrscheinlichkeiten auskommt, nutzt induktive Argumente, zu deren Prämissen Aussagen über die

Erfahrung gehören und deren Konklusion Aussagen sind, die über die Erfahrung hinausgehen, also beispielsweise wissenschaftliche Hypothesen und Theorien oder Aussagen über zukünftige Ereignisse.

Solche induktiven Argumente kommen aber im kritischen Rationalismus nicht vor. Sie werden ersetzt durch Entscheidungsregeln. Diese Entscheidungsregeln lassen sich, wie bereits erläutert, nicht als Argumente darstellen, deren Konklusion die akzeptierten Aussagen sind.

Der kritische Rationalismus wäre nur dann eine Form des Induktivismus, wenn sein technologischer Anspruch induktiv begründet wäre. Induktivisten sind der Meinung, dass kritische Rationalisten spätestens an dieser Stelle auf eine induktive Begründung zurückgreifen oder zurückgreifen müssen. Beides ist nicht der Fall. Sie greifen auf dieselben Entscheidungsregeln zurück, die sie auch sonst anwenden.

5. Schluss

Bayesianismus und kritischer Rationalismus stimmen in einem wichtigen Punkt überein. Beide präsentieren sich als Entscheidungstheorien unter Ungewissheit. Im kritischen Rationalismus werden auf der Grundlage des Münchhausen-Trilemmas alle Aussagen als ungewiss betrachtet. Im Bayesianismus werden meist Beobachtungsaussagen (üblicherweise aufgefasst als Aussagen über Sinneseindrücke) als sicher wahr betrachtet; außerdem wird unterstellt, dass logische und mathematische Zusammenhänge immer richtig erkannt werden. Aber zumindest bezüglich aller sonstigen Aussagen und insbesondere für Aussagen über die Zukunft unterstellen beide Rationalitätskonzeptionen gleichermaßen radikale Ungewissheit.

Der kritische Rationalismus ist mit einem technologischen Anspruch verbunden: Die beste Methode, um zu wahren und informativen Aussagen über die Welt und zu wahren Aussagen über logische oder mathematische Zusammenhänge zu gelangen, besteht darin, die Entscheidungsregeln des kritischen Rationalismus anzuwenden.

Auch für den Bayesianismus werden technologische Ansprüche diskutiert. Allerdings liegen gegen alle diese Ansprüche schwerwiegende Einwände vor. Aus diesem Grund halten viele Bayesianer den Bayesianismus für nicht mehr als einen formalen Rahmen, innerhalb dessen sich rationales Lernen aus der Erfahrung zwangsläufig abspielen müsse. Die einzige Begründung für diese Auffassung scheint mir zu sein, dass Menschen tatsächlich häufig Aussagen als mehr oder weniger wahrscheinlich oder plausibel einschätzen (s. z.B. Joyce 2004: 132–134). Aber daraus kann man nicht schließen, dass solche Einschätzungen für einen rationalen Lernprozess eine tragende Rolle spielen müssen. Es ist genauso gut möglich, dass solche subjektiven Urteile nur dort relevant sind, wo rationale Überlegungen einen entsprechenden Spielraum lassen.

Für hilfreiche Diskussionen und Hinweise danke ich Hartmut Kliemt sowie den Teilnehmern der Tagung „Rationalität im 21. Jahrhundert“ (1. Kolloquia Triesen 2022).

Albert, Hans (1968): *Traktat über kritische Vernunft*, Tübingen: Mohr Siebeck.

Albert, Max (2001): „Bayesian learning and expectations formation: anything goes“, in: *Foundations of Bayesianism*, David Corfield/Jon Williamson (Hg.), Dordrecht: Kluwer**,** 351–372.

Albert, Max (2003): „Bayesian rationality and decision making: A critical review“, *Analyse & Kritik* 25, 101–117.

Albert, Max (2011a): „Rationales Lernen und die Chaotische Uhr. Zur Kritik des Bayesianismus“, in: *Rationalität und Irrationalität in den Wissenschaften*, Ulrich Arnswald/Hans–Peter Schütt (Hg.), Wiesbaden: VS-Verlag**,** 9–28.

Albert, Max (2011b): „Von der vollkommenen zur kritischen Rationalität. Eine Kritik ökonomischer Rationalitätsauffassungen“, in: *Philosophie und Wirtschaftswissenschaft*, Volker Gadenne/Reinhard Neck (Hg.), Tübingen: Mohr Siebeck**,** 9–28.

Albert, Max (2017): „How Bayesian rationality fails and critical rationality works“, *Homo Oeconomicus* 34, 313–341.

Andersson, Gunnar (1988): *Kritik und Wissenschaftsgeschichte. Kuhns, Lakatos' und Feyerabends Kritik des kritischen Rationalismus*, Tübingen: Mohr Siebeck.

Binmore, Ken (2009): *Rational Decisions*, Princeton: Princeton University Press.

Chakravartty, Anjan (2017): „Scientific realism“, in: *Stanford Encyclopedia of Philosophy. Summer 2017 Edition*, Edward N. Zalta (Hg.), https://plato.stanford.edu/archives/sum2017/entries/scientific–realism/.

Earman, John (1992): *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*, Cambridge, Mass.: The MIT Press.

Forster, Malcolm R (1999): „How do simple rules ‘fit to reality’ in a complex world?“, *Minds and Machines* 9, 543–564.

Gadenne, Volker (2005): „Wozu normative Wissenschaftstheorie? Zur Notwendigkeit und Rechtfertigung von Rationalitätsprinzipien in der Wissenschaftstheorie“, in: *Deskriptive oder normative Wissenschaftstheorie?*, Bernward Gesang (Hg.), Frankfurt: Ontos Verlag**,** 31–47.

Gadenne, Volker (2006): „Methodological rules, rationality, and truth“, in: *Rationality and Reality*, Colin Cheyne/John Worrall (Hg.), Dordrecht: Springer**,** 97–107.

Gillies, Donald (1988): „Induction and probability“, in: *An Encyclopedia of Philosophy*, George H.R. Parkinson (Hg.), London: Routledge**,** 179–204.

Gillies, Donald (2006): *Philosophical Theories of Probability*, London: Routledge.

Goodman, Nelson (1983): *Fact, Fiction and Forecast*, Cambridge, Mass.: Harvard University Press.

Hájek, Alan (2009): „Dutch Book arguments“, in: *The Handbook of Rational and Social Choice*, Paul Anand/Prasanta K. Pattanaik/Clemens Puppe (Hg.), Oxford: Oxford University Press**,** 173–195.

Harsanyi, John C. (1977): *Rational Behaviour and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.

Howson, Colin (2000): *Hume's Problem. Induction and the Justification of Belief*, Oxford: Oxford University Press.

Howson, Colin/Peter Urbach (2006): *Scientific Reasoning. The Bayesian Approach*, Chicago: Open Court.

Humphreys, Paul W. (1987): „Induction“, in: *The New Palgrave: A Dictionary of Economics,* Bd. 3, John Eatwell/Murray Milgate/Peter Newman (Hg.), London: Macmillan**,** 116–120.

Joyce, James M. (2004): „Bayesianism“, in: *The Oxford Handbook of Rationality*, Alfred R. Mele/Piers Rawling (Hg.), Oxford: Oxford University Press**,** 132–155.

Kahneman, Daniel/Amos Tversky (Hg.) (2000): *Choices, Values, and Frames*. New York: Cambridge University Press

Kiefer, Nicholas M. /Yaw Nyarko (1995): „Savage–Bayesian models of economics“, in: *Learning and Rationality in Economics*, Kirman Alan/Salmon Mark (Hg.), Oxford: Blackwell**,** 40–62.

Musgrave, Alan (1993): *Common Sense, Science and Scepticism: A Historical Introduction to the Theory of Knowledge*, Cambridge: Cambridge University Press.

Musgrave, Alan (1999): *Essays on Realism and Rationalism*, Amsterdam: Rodopi.

Musgrave, Alan (2006): „Responses“, in: *Rationality and Reality*, Cheyne Colin/Worrall John (Hg.), Dordrecht: Springer**,** 293–333.

Musgrave, Alan (2011): „Popper and hypothetico–deductivism“, in: *Inductive Logic*, Dov M Gabbay/Stephan Hartmann/John Woods (Hg.), Amsterdam: North–Holland**,** 205–234.

Myerson, Roger B. (1991): *Game Theory. Analysis of Conflict*, Cambridge, Mass.: Harvard University Press.

Nyarko, Yaw (1997): „Savage–Bayesian agents play a repeated game“, in: *The Dynamics of Norms*, Cristina Bicchieri/Richard C. Jeffrey/Brian Skyrms (Hg.), Cambridge: Cambridge University Press**,** 175–197.

Russell, Bertrand (1945): *A History of Western Philosophy*, New York: Simon & Schuster.

Savage, Leonard J. (1954): *The Foundations of Statistics*, New York: Wiley.

Stanovitch, Keith E. (2011): „Normative models in psychology are here to stay“, *Behavioral and Brain Sciences* 34, 268–269.

Stanovitch, Keith E. /Paula J. Stanovitch (2010): „A framework for critical thinking, rational thinking and intelligence“, in: *Educational Psychology. Perspectives on Learning, Teaching and Human Development*, David D. Preiss/Robert J. Sternberg (Hg.), New York: Springer.

Wootton, David (2015): *The Invention of Science. A New History of the Scientific Revolution*, London: Allan Lane.