

Die Relativität der Information

Andreas de Vries*

*FH Südwestfalen University of Applied Sciences
Haldener Straße 182, D-58095 Hagen, e-mail: de-vries@fh-swf.de*

Inhaltsverzeichnis

1	Einleitung	2
2	Definitionen von Information	3
2.1	Information und Wissen	3
2.2	Shannonsche Information	4
2.3	Algorithmische Information	5
2.4	Fisher-Information	8
2.5	Quanteninformation	11
3	Anwendungen der Information	17
3.1	Information und Energie	17
3.2	Entropie Schwarzer Löcher	18
3.3	Information in der Finanzmathematik	24
3.4	Information in der Biologie	26
4	Erste Vereinheitlichung: Rényi	29
4.1	Axiomatischer Zugang	29
4.2	Shannonsche Information	31
4.3	Brukner-Zeilinger-Information	32
5	Die Relativität der Information	33
5.1	Relative Wahrscheinlichkeitsverteilungen	33
5.2	Wo ist die Objektivität der Information?	35

*© Der wesentliche Inhalt dieses Beitrags ist erschienen in R. Muno (Hrsg.): *Jahresschrift der Bochumer Interdisziplinären Gesellschaft eV 2004*, ibidem-Verlag, Stuttgart 2006 (ISBN 978-3-89821-656-2)

Zusammenfassung

In diesem Beitrag wird ein Überblick über gängige mathematische Definitionen von Information und deren Anwendungen in verschiedenen Wissenschaften gegeben. Insbesondere der Zusammenhang von Information und Energie und die Wirkung Schwarzer Löcher auf Information wird dargelegt, sowie die Rolle von Information in der Finanzmathematik und bei der Entwicklung des irdischen Lebens.

Durch Axiomatisierung wird weiters eine erste Vereinheitlichung der Informationsmaße angegeben, die auch die kürzlich eingeführte Brukner-Zeilingersche Information umfasst. Hierbei spielt das „Rekursionspostulat“ für das Informationsmaß eine dem Parallelenaxiom der Euklidischen Geometrie vergleichbare Rolle. Schließend wird die Relativität der Information und ihre Auswirkung auf die physikalisch notwendige Objektivität von Information dargestellt.

1 Einleitung

Was ist Information? In meinen Anfängervorlesungen der Informatik definiere ich salopp: „Information ist eine Antwort auf eine Frage.“ Diese Aussage zielt einerseits auf die Definition von Claude Shannon [43] von 1948 ab, die Information als Entropie sieht und die Kapazität eines Kanals zur Übertragung von Nachrichten quantifizierbar macht, so dass sie sich eignet, die in der Informationstechnologie fundamentalen Größen *Speicherplatz* und *Übertragungsrate* zu messen. Insbesondere muss es bei dieser Definition zwei Kommunikationsteilnehmer geben, den Fragenden und den Antwortenden, also Sender und Empfänger, die Nachrichten austauschen.

Andererseits versucht die Aussage, den gerade für den alltäglichen Sprachgebrauch von Information wichtigen Aspekt der *Bedeutung* zu erfassen. Denn schließlich steckt hinter jeder Frage ein Zweck, ein Kontext, in dem die Antwort eine inhaltliche Bedeutung hat, also den *semantischen* Aspekt umfasst.

In diesem Beitrag wird die These vertreten, dass der Shannonsche Informationsbegriff nicht nur das technische „Bit-Zählen“ von Nachrichten als kodierte Zeichenfolgen umfasst, sondern durchaus den semantischen Aspekt, bereits einschließt — wenn auch auf indirekte Weise über den Bayesschen Wahrscheinlichkeitsbegriff. Dieser Begriff wird jedoch seit jeher wegen seiner vermeintlichen Subjektivität vorwiegend skeptisch betrachtet. In der Tat sollte Subjektivität in Betrachtungen grundlegender Begriffe nicht vorkommen: Was interessierte meinen Kühlschrank Entropie, wenn sie nur ein subjektiver Begriff wäre?

Ein Ausweg aus diesem Dilemma könnte der Begriff der *Relativität der Information* weisen. Erstens ist Relativität nicht Subjektivität, zweitens ist das Prinzip der Relativität in der Geschichte der Physik bereits von Galilei und Einstein erfolgreich angewandt worden, und drittens sind die Zustände eines Systems einer der genauesten

physikalischen Theorien, der Quantenmechanik, sogar *beobachterabhängig*. Warum sollte der Begriff Relativität bei dem offensichtlich schwierigen Begriff Information nicht auch zum Erfolg führen?

Der große Wurf, die letzte Klärung des Begriffs Information wird auch in diesem Beitrag nicht gelingen. Er stellt jedoch die wichtigsten mathematischen Konzepte der Information vor, zeigt einige der Anwendungen in verschiedenen Wissenschaften und versucht, sie über eine geeignete Axiomatisierung zu strukturieren. Leitlinie ist dabei stets Offenheit gegenüber allen mathematisch formulierbaren Ansätzen und Erscheinungsformen. So ergibt sich ein sehr facettenreiches Bild, dennoch lassen sich die meisten der vorgestellten Ansätze als Folgerungen der Rényi-Information betrachten und damit unter ein gemeinsames logisches Dach bringen.

2 Definitionen von Information

2.1 Information und Wissen

Wie mit allen fundamentalen („letzten“) Begriffen [57, §8.2.C], [58] kann eine Definition nur vage ausfallen: *Information* ist der darstellungs-, sender- und empfängerinvariante Gehalt einer Mitteilung, einer Nachricht oder eines Ereignisses [34].

Nach Weizsäcker [57, §5.1] muss zunächst zwischen zwei Arten von Information unterschieden werden: Einerseits gibt es die *faktische* oder *aktueller* Information, „die man bereits weiß“, andererseits die *potentielle* oder *virtuelle* Information, „die man (noch) nicht weiß“. Faktische Information, also *Wissen*, kann ausschließlich von bereits vergangenen Ereignissen stammen. Zeit und Wissen hängen also unauflöslich zusammen. Im Umkehrschluss liefern zukünftige Ereignisse potentielle Information.

Eine exakte Definition gibt die Informationstheorie. Sie befasst sich speziell mit der potentiellen Information, die in drei Arten zerfällt [30, §1]:

1. Die *syntaktische Information* bezieht sich auf die Symbole oder Zeichen, mit denen Nachrichten übertragen werden. Hieraus leitet sich direkt die Einheit der Information ab, insbesondere das Bit.
2. Die *semantische Information* bezieht sich auf die Bedeutung von Nachrichten.
3. Die *pragmatische Information* bezieht sich auf die Wirkung und den Nutzen von Nachrichten.

Lyre [31] spricht suggestiv von der „Dreidimensionalität der Information“ um zu verdeutlichen, dass die drei Informationsaspekte lediglich verschiedene Sichtweisen auf ein und dasselbe Phänomen darstellen.

Die syntaktische Information wurde historisch zuerst 1948 von Shannon [43] für eine Quelle eingeführt, die Zeichen mit einer gegebenen Wahrscheinlichkeitsverteilung aus einem endlichen Zeichenvorrat sendet. Nach Jaynes [23, 24] kann dieser Ansatz auf prinzipiell jede Wahrscheinlichkeitsverteilung verallgemeinert werden.

Grundlage hierfür ist die Idee, dass eine Wahrscheinlichkeitsverteilung Wissen (A-priori-Kennntnis) darstellt, also faktische Information. Auf diese Weise kann die semantische Information mathematisch präzise gefasst werden, denn die Wahrscheinlichkeitsverteilung selbst ist formaler Ausdruck des Wissens über die betreffende Situation.

Wenn beispielsweise — wie in Lehrbüchern der Wahrscheinlichkeitstheorie üblich — für einen Würfel die Gleichverteilung $p_1 = \dots = p_6 = \frac{1}{6}$ angenommen wird, so wird in Wahrheit das Wissen formal ausgedrückt, dass es sich um einen *idealen* Würfel handelt. Die Gleichverteilung kann andererseits formaler Ausdruck für das Nicht-Wissen sein, welche Eigenschaften über den betrachteten Würfel vorausgesetzt werden können: also setzt man rationalerweise gar keine voraus. Weiß man dagegen mehr, beispielsweise dass der empirisch beobachtete Mittelwert eines konkreten Würfels 4,5 beträgt (und nicht 3,5), so darf man keine Gleichverteilung mehr annehmen. Welche Verteilung aber dann? Eine mathematisch eindeutige Antwort gibt das aus der Statistischen Physik bekannte Prinzip der „maximalen Entropie“. Mit ihm lässt sich die Verteilung bei gegebenem Wissen also Optimierungsproblem bei gegebenen Randbedingungen errechnen [24].

Allerdings ist dieser so-genannte Bayessche Ansatz [32], obwohl sogar auf die konzeptionellen Ideen der Begründer der Wahrscheinlichkeitstheorie Laplace und Bernoulli zurückführbar [25], bis heute nicht unumstritten. Von Kritikern wird ihm die Subjektivität angelastet, sie akzeptieren nur *objektiv* definierte Wahrscheinlichkeiten. Am Ende dieses Beitrags wird auf diesen Konflikt der zwei verschiedenen Sichtweisen eingegangen.

2.2 Shannonsche Information

The game Twenty Questions provides a beautiful opportunity to express the most varied sorts of information in the form of successive binary choices between ... alternatives (...). It is played by two people, the first of whom dreams up something that the second player has to guess in twenty questions or less (...). The questions have to be answered either “yes” or “no”; each is a binary choice. For the second player it is advantageous to make the questions as close as possible to a choice between equally probable alternatives. Knowing that the thing is mineral, for example, the questioner would be ill-advised to ask right away whether it is the Hope diamond. Instead, he or she might ask, “Is it natural [as opposed to being manufactured or modified by humans]?” (...) When the probabilities of a yes and no answer are equal, each question will elicit one bit of information (the most that such a question can extract). Twenty bits of information correspond to a choice from among $[2^{20} =] 1,048,576$ equally probable alternatives.

Murray Gell-Mann [15, S. 37]

Shannons bahnbrechende Idee von 1948 war es, die Information H als Funktion einer gegebenen Wahrscheinlichkeitsverteilung p zu betrachten, $H = H(p)$, und zwar als Maß für die Unsicherheit des Eintretens eines bestimmten Ereignisses, bzw. als Maß für die entsprechende Überraschung oder den Neuigkeitswert. Die Shannonsche Information entspricht genau der Entropie eines statistischen Systems [61, §5.17.1].

Der *Informationsgehalt* eines Ereignisses A , das mit der Wahrscheinlichkeit p eintritt, ist

$$I_A = -k \ln p, \quad (1)$$

Hier ist k eine Konstante, die die Einheit des Informationsgehalts bestimmt. Für $k = 1/\ln 2$ z.B. ist sie „Bit“, d.h. $I_A = -\log_2 p$. Ist $p = (p_1, \dots, p_n)$ ein Wahrscheinlichkeitsvektor [19], so ist die *durchschnittliche Information* oder auch die *Entropie* $H(p)$ definiert als

$$H(p) = -k \sum_{i=1}^n p_i \ln p_i. \quad (2)$$

Mit $k = 1/\ln 2$ ist $H(p)$ die Information der Verteilung p in Bit und entspricht daher der minimalen Anzahl von Ja-Nein-Fragen, um das eingetretene Ereignis sicher herauszufinden. Entropie ist also der durchschnittliche Informationsgehalt aller möglichen Ereignisse.

Die Shannonsche Information hängt somit lediglich von einer gegebenen Wahrscheinlichkeitsverteilung p ab und bezieht sich nicht direkt auf den Inhalt oder die Bedeutung der zu Grunde liegenden Ereignisse. Nur die *Wahrscheinlichkeiten*, mit denen die Ereignisse eintreten, sind relevant, nicht die Ereignisse selber. Die Semantik liegt also implizit bereits in der Bestimmung der Wahrscheinlichkeiten.

Gibt es nun zwei Wahrscheinlichkeitsvektoren p und q mit der Eigenschaft, dass $q_i \neq 0$, so ist die so genannte *Kullback-Leibler-Information* oder die *relative Entropie* [32, §2.6] von p bezüglich q definiert durch

$$K(p; q) = \sum_{i=1}^n p_i \ln \frac{p_i}{q_i}. \quad (3)$$

Die Kullback-Leibler-Information ist ein Maß für die Abweichung zweier Verteilungen. Hierbei kann die Wahrscheinlichkeitsverteilung q aufgefasst werden als mathematische Beschreibung des Wissens oder der anfänglichen Vorkenntnis (*Bayesian prior distribution* [24]); auf Grund neuer Erkenntnisse, etwa durch eine Messung oder durch Lernen, wird das Wissen verändert und führt zu einer geänderten Wahrscheinlichkeitsverteilung p . Die Kullback-Leibler-Information ist dann der entsprechende Informationsgewinn, also die Informationsdifferenz, die das Wissen p von dem Vorwissen q unterscheidet.

2.3 Algorithmische Information

Nothing you can know that isn't known.

Nothing you can see that isn't shown.

The Beatles, *All You Need is Love* (1967)

Als Tycho Brahe (1546–1601) vor über 400 Jahren in mühseliger Arbeit die Positionsdaten des Mars in bis dahin nicht erreichter Genauigkeit und mit bloßem Auge (das Fernrohr war noch nicht erfunden!) in seine Tafeln schrieb, hat er „Information

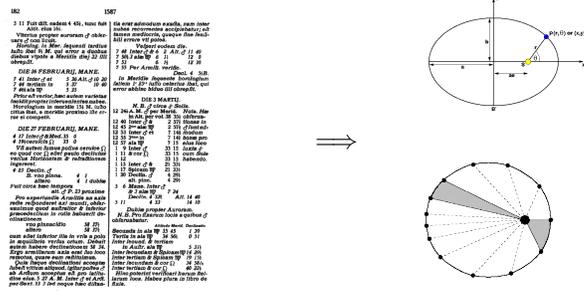


Abbildung 1: Brahes Marsbeobachtungen vom 26.2. – 10.3.1587 (aus *Tychohis Brahe Opera Omnia*) und die ersten zwei Keplerschen Gesetze.

gesammelt“. Sein letzter Assistent und Nachfolger Johannes Kepler (1571–1630) leitete aus diesen Daten die ersten beiden der drei heute nach ihm benannten Gesetze der Planetenbahnen [48, §2.1.2] empirisch ab.¹ Mit diesen Gesetzen konnten Brahes riesige Beobachtungsprotokolle ersetzt werden, die Information wurde durch sie sogar größer, denn man kann nun auch zukünftige Marsdaten *berechnen*. Ein Gesetz *komprimiert* Information.

Das abstrakte Modell eines Computers ist in der Theoretischen Informatik die *universelle Turing-Maschine*, die mit einem Programm p und Daten w gefüttert wird und daraus eine Ausgabe x erzeugt. Hierbei sind p , w und x Worte oder Strings, d.h. Zeichenfolgen aus einem gegebenem Alphabet Σ . In Kurzschreibweise heißt das für eine universelle Turing-Maschine U , dass

$$U(p, w) = x, \quad p, w, x \in \Sigma^*, \tag{4}$$

mit der Menge Σ^* der Strings. Üblicherweise ist das Alphabet gegeben durch $\Sigma = \{0, 1\}$, womit die Menge der Strings

$$\Sigma^* = \{\Lambda, 0, 1, 10, 11, 110, 111, 1110, \dots\}.$$

lautet (Λ bezeichnet den leeren String). Wenn eine Berechnung keine Eingabedaten w benötigt, also wenn $U = U(p, \Lambda)$, so schreiben wir kurz $U(p)$ statt $U(p, \Lambda)$. Der Vollständigkeit halber sollte erwähnt sein, dass die Eingabedaten einer universellen Turing-Maschine aus einer *Präfixmenge* S von *Präfixcodes* w stammen müssen, für die die Kraft-Ungleichung [30, §2.2], [32, §5.2] gilt:

$$\sum_{w \in S} 2^{-|w|} \leq 1. \tag{5}$$

Hier bezeichnet $|w|$ die Länge des Strings w , d.h. die Anzahl seiner Zeichen. Üblicherweise wird $|w|$ in Bits angegeben. Da $2^0 = 1$ besagt die Kraft-Ungleichung insbesondere, dass die Präfixmenge entweder nur aus dem leeren String besteht („keine

¹Eine schöne Darstellung von Keplers Vorgehensweise findet sich in [27, §1], ein Schulbuch, das mich als Jugendlicher so sehr faszinierte, dass ich es abschrieb. Man kann den revolutionären Schritt Keplers gar nicht überbewerten, er war von der damals noch umstrittenen Kopernikanischen Idee überzeugt und wollte sie mit Brahes Daten verifizieren.

Eingabe“) oder Teilmenge aller nichtleeren Strings $\Sigma^* \setminus \{\Lambda\}$ ist („Eingabe zwingend“). Da außerdem $\sum_1^\infty 2^{-n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1$, existieren Präfixmengen, die jeweils nur einen String w der Länge $|w| = i$, für $i = 1, 2, \dots$, besitzen [30, Ex. 2.3].

Die *algorithmische Wahrscheinlichkeit* eines Strings $x \in \Sigma^*$ ist definiert als

$$P(x) = \sum_{U(p)=x} 2^{-|p|}, \quad (6)$$

also als die Wahrscheinlichkeit, dass die universelle Turing-Maschine U , deren Programm p per Zufall aus der Menge aller möglichen Programme gewählt wird, genau die Ausgabe x erzeugt. Für $x, w \in \Sigma^*$ definieren wir entsprechend die algorithmische Wahrscheinlichkeit eines Strings x relativ zur Eingabe w als

$$P(x|w) = \sum_{U(p,w)=x} 2^{-|p|}. \quad (7)$$

Die *algorithmische Entropie* $H(x)$ eines Strings x ist definiert [5] als die kleinste natürliche Zahl, die größer ist als $-\log_2 P(x)$, d.h.

$$H(x) = \lceil -\log_2 P(x) \rceil. \quad (8)$$

Entsprechend ist die *bedingte algorithmische Entropie* $H(x|w)$ definiert durch $H(x|w) = \lceil -\log_2 P(x|w) \rceil$.

2.3.1 Der algorithmische Informationsgehalt AIC

In den 1960er Jahren führten unabhängig voneinander Kolmogorov, Chaitin (als 15-Jähriger) und Solomonoff den Begriff des *Algorithmischen Informationsgehalts* AIC (*algorithmic information content*) ein als die Länge $|x^*|$ eines Minimalprogramms des Strings x . Genauer ist für einen String x ein *Minimalprogramm* x^* von x definiert als

$$x^* \in \{p \in \Sigma^* : U(p) = x, |p| \leq |q| \forall q \in \Sigma^*\}$$

Mit anderen Worten ist x^* eine Lösung des Minimierungsproblems

$$|p| \rightarrow \min \quad \text{unter der Nebenbedingung } U(p) = x.$$

Ein String x heißt *komprimierbar* um k Bits, wenn ein Minimalprogramm x^* mindestens k Bits kürzer ist als x selbst. Betrachten wir einen sehr langen String, z.B. $x = 110110110 \dots 110$ mit $|x| = 3 \cdot 10^6$ Bits. Ein Minimalprogramm x^* von x lautet $x^* =$ „gib 1000000-mal „110“ aus“. In einem Hexacode-Editor sieht dieser Text so aus:

```
0000: 67 69 62 20 31 30 30 30 30 30 2D 6D 61 6C 20
0010: 22 31 31 30 22 20 61 75 73
```

Das ist ein ASCII-Text der Länge 25 Bytes = 200 Bits. In diesem Falle wäre also x um $k = 2999800$ Bits komprimierbar. Ein lauffähiges Java-Programm (als Bytecode), das den String x erzeugt, umfasst 436 Byte = 3488 Bits. Mit dieser konkreten Realisierung eines Programms $p = x^*$ wäre also x um $k = 2996512$ Bits komprimierbar. Der algorithmische Informationsgehalt von x ist also sehr niedrig.

Wie sehen denn eigentlich nichtkomprimierbare Strings aus? Ein Minimalprogramm eines solchen Strings x kann nur aus der Anweisung „gib x aus“ bestehen. Solch ein String hat einen AIC, der seiner Länge entspricht, also $AIC(x) = |x|$. Es gibt keine Regel, kein Gesetz, keinen Algorithmus, wodurch die Beschreibung von x vereinfacht werden könnte. Solch ein String ist ein *zufälliger* String (*random string*). Mit anderen Worten ist der algorithmische Informationsgehalt maximal für zufällige Strings. Daher nennt man nichtkomprimierbare Strings auch *algorithmisch zufällig*.

Chaitin [9] konnte zeigen, dass der AIC nicht berechenbar ist. Er benutzt dabei ein Argument, das den Beweis der Gödelschen Unvollständigkeitssätze verallgemeinert. Zwar wissen wir den AIC für einen zufälligen String, man kann mathematisch sogar zeigen, dass die meisten aller möglichen Strings zufällig sind – allein, wir können bei einem gegebenem String nie sicher sein, ob er tatsächlich zufällig ist. Denn es könnte einen uns unbekanntem Satz oder einen raffinierten Algorithmus geben, durch den dieser String komprimierbar wird. Es *gibt kein allgemein gültiges* Verfahren, mit dem wir alle Theoreme finden könnten, die die Komprimierbarkeit eines Strings vollständig überprüfen.

2.4 Fisher-Information

Gegeben sei eine Wahrscheinlichkeitsdichte $f(x, \vec{\lambda})$ mit einer (diskreten oder stetigen) Zufallsgröße $x \in X \subset \mathbb{R}^n$ und einem Satz von Parametern $\vec{\lambda} = (\lambda_1, \dots, \lambda_p)$. Wird nun eine Stichprobe x entnommen bzw. ein Messergebnis x beobachtet, wobei man jedoch die wahren Parameterwerte $\vec{\lambda}$ nicht kennt, dann seien

$$L(\vec{\lambda}, x) = f(x, \vec{\lambda}), \quad \ell(\vec{\lambda}, x) = \ln L(\vec{\lambda}, x). \quad (9)$$

L heißt eine *Likelihood*² [55, §35] oder *A-posteriori-Wahrscheinlichkeit*, und ℓ ist die *Log-Likelihood*. Für den Fall nur eines Parameters ($p = 1$) und der Differenzierbarkeit der Likelihood bezüglich λ ist der *Beitrag* („score“) V die Ableitung der Log-Likelihood,

$$V(\lambda, x) = \frac{\partial}{\partial \lambda} \ell(\lambda, x) \quad (10)$$

Der Erwartungswert $E[V(\lambda)]$ über die Zufallsgrößen x ist 0, denn $E[V(\lambda)] = \int_X V(\lambda, x) f(x, \lambda) dx = \int_X \frac{\partial f}{\partial \lambda} dx = \frac{\partial}{\partial \lambda} \int_X f(x, \lambda) dx = \frac{\partial}{\partial \lambda} 1 = 0$. Das bedeutet, dass

²Eine deutsche Übersetzung wird dadurch erschwert, dass *likelihood* ebenso wie *probability* einfach Wahrscheinlichkeit heißt. Man muss jedoch deutlich unterscheiden zwischen einer Wahrscheinlichkeitsdichte und einer Likelihood-Funktion, deren Unbekannte λ ja gar nicht vom Zufall abhängen. „Plausibilität“ wäre ein möglicher Begriff. Für eine Übersetzungstabelle wichtiger englischer Begriffe der Statistik siehe [55, S. 353ff]

der Mittelwert der Beiträge nach 0 tendiert, wenn der Umfang der Stichprobe sehr groß wird.

Die *Fisher-Information* oder *Information der Stichprobe bezüglich λ* ist definiert [6, §7] als die Varianz des Beitrags V ,

$$I(\lambda) = E[V^2(\lambda)] = E \left[\left(\frac{\partial}{\partial \lambda} \ell(\lambda, x) \right)^2 \right] \quad (11)$$

Sie ist stets nichtnegativ, und sie verschwindet, wenn die Likelihood nicht von λ abhängt oder λ ein Extremwert von L ist. Ist L zweimal differenzierbar bezüglich λ , so gilt [6, Gl. (7.3.9)]

$$I(\lambda) = -E \left[\left(\frac{\partial^2}{\partial \lambda^2} \ell(\lambda, x) \right) \right] \quad (12)$$

Beispiel 2.1 Nach dem Satz von Bayes (S. 33) ist

$$P(B|A) = \frac{P(A|B)}{P(A)} P(B),$$

und somit sind $P(A|B)$ und $P(A|B)/P(A)$ als Funktionen von B Likelihood-Funktionen. \square

Beispiel 2.2 Gegeben seien zwei asymmetrische Münzen A und B derart, dass A mit der Wahrscheinlichkeit $\frac{1}{3}$ und B mit der Wahrscheinlichkeit $\frac{2}{3}$ Kopf zeigt. Ohne zu wissen, um welche Münze es sich handelt, wirft man sie fünf Mal und erhält zweimal Kopf und dreimal Zahl. Um welche Münze handelt es sich „wahrscheinlich“?

Die Likelihood $L(\lambda, x)$ für die diskreten Werte $\lambda = A$ oder B und $x = \text{Kopf (K)}$ oder Zahl (Z) ist gegeben durch die folgende Tabelle:

(λ, x)	(A, K)	(A, Z)	(B, K)	(B, Z)
$L(\lambda, x)$	$\frac{1}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{1}{3}$

Die Likelihood für fünf Würfe $x = \text{KKZZZ}$ lautet also

$$L(A, x) = \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^3 = \frac{8}{3^5}, \quad L(B, x) = \left(\frac{2}{3}\right)^2 \cdot \left(\frac{1}{3}\right)^3 = \frac{4}{3^5},$$

d.h. der „Likelihood-Quotient“ $L(A, x)/L(B, x) = 2 > 1$. Daher ist es plausibler, dass die Münze A verwendet wurde. \square

Beispiel 2.3 (*Spamfilter*) Spams³ sind Massensendungen (meist Werbesendungen), die per E-Mail den Empfängern unverlangt zugesandt werden. Sie stellen ein zunehmendes persönliches und ökonomisches Problem dar, es gibt Angaben, dass weltweit

³Etymologisch stammt das Wort aus Sketchen der britischen Fernsehreihe *Monty Python's Flying Circus*, in denen ständig bei Restaurantbesuchen der Gesang des Wortes „SPAM“ (*spiced pork and ham*) als Werbung für Pressfleisch in Dosen den jeweiligen Dialog übertönte und bildhaft für etwas Unausweichliches stand. Oft wird die Abkürzung umgedeutet in *send phenomenal amounts of mail*. Leider haben weder Humor noch Verstand der Verursacher von Spams das Niveau Monty Pythons und ihrer in diesem Fall sogar prophetischen Ideen.

über die Hälfte aller versendeten E-Mails Spam sind.⁴ Es gibt eine Reihe von Spamfiltern, die Spams zu erkennen versuchen und sie markieren bzw. in Quarantäne setzen, z.B. das Open-Source-Projekt SpamAssassin⁵. In einer Auswertung aus 2004⁴ hat es eine Genauigkeit (*precision*) von 99,7% und eine Trefferquote (*recall*) von 96,1%. Hierbei ist die Genauigkeit der Anteil der Nachrichten, die zurecht als Spam klassifiziert wurden, und die Trefferquote der Anteil der Spams, die tatsächlich erkannt wurden. Wie wahrscheinlich ist es, dass eine Ihrer E-Mails tatsächlich Spam ist, wenn $0 \leq \lambda \leq 1$ der Anteil aller als Spam klassifizierten E-Mails in Ihrer Mailbox sind?

Sei A das Ereignis, dass die E-Mail ein Spam ist, und B das Ereignis, dass eine E-Mail als Spam klassifiziert wird. Dann ist $P(A|B)$ die Genauigkeit, also $P(A|B) = 0,997$, $P(B|A)$ ist die Trefferquote, also $P(B|A) = 0,961$, und $P(B) = \lambda$. Mit dem Satz von Bayes (Satz 5.1) gilt

$$P(A) = \frac{P(A|B)P(B)}{P(B|A)} = \frac{0,997\lambda}{0,961} \approx 1,0375\lambda.$$

(Damit $P(A)$ eine Wahrscheinlichkeit ist, muss also $\lambda \leq 0,9639$ gelten). Die Likelihood, dass eine an Sie gesendete E-Mail Spam ist, ist also etwas *größer* als der Anteil Ihrer als Spam markierten E-Mails! Mit $\ell = \ln 1,0375 + \ln \lambda$ folgt $\ell' = 1/\lambda$ und $\ell'' = -1/\lambda^2$, d.h. die Fisher-Information ist

$$I(\lambda) = \frac{1}{\lambda^2}.$$

□

Beispiel 2.4 Was ist die Information $I(\lambda)$ einer Stichprobe $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}$ der Größe N , die aus einer Normalverteilung bekannter Varianz σ^2 , aber unbekanntem Mittelwert $\lambda = a$ entnommen ist?

Da die Dichte für jede einzelne Stichprobe $f(\mathbf{x}^{(j)}, \lambda) = \frac{1}{\sqrt{2\pi}} e^{-(\mathbf{x}^{(j)} - \lambda)^2 / 2\sigma^2}$, lautet die Likelihood

$$L(\lambda, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}) = \frac{1}{\sqrt{2\pi}^N} \prod_{j=1}^N \exp\left(-\frac{(\mathbf{x}^{(j)} - \lambda)^2}{2\sigma^2}\right)$$

Mit $\ell(\lambda) = \ln L(\lambda, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)})$ ist damit der Beitrag $V = \ell' = \frac{1}{\sigma^2} \sum_{j=1}^N (\mathbf{x}^{(j)} - \lambda)$ und $\ell'' = -N/\sigma^2$. Gleichung (12) liefert dann

$$I(\lambda) = \frac{N}{\sigma^2}. \quad (13)$$

Je größer also die Stichprobe ist, umso größer ist ihre Information bezüglich dem Mittelwert. □

⁴<http://sam.holden.id.au/writings/spam2/>

⁵<http://spamassassin.apache.org>

Beispiel 2.5 (Mittlere Zerfallszeit) Die Wahrscheinlichkeit dafür, dass ein radioaktiver Kern, der zum Zeitpunkt $t = 0$ existiert, im Zeitintervall $t + dt$ zerfällt, ist

$$f(t) dt = \frac{1}{\tau} e^{-t/\tau} dt$$

wo τ die gesuchte mittlere Lebensdauer des Kerns ist. Für beobachtete Zerfallszeiten t_1, t_2, \dots, t_N ist die Likelihood also

$$L = \frac{1}{\tau} \exp\left(-\frac{1}{\tau} \sum_{j=1}^N t_j\right) = \frac{1}{\tau} e^{-N\bar{t}/\tau}. \quad (14)$$

(\bar{t} ist der Mittelwert der Stichprobe). Ihr Logarithmus ist $\ell(\tau) = \ln L = -N(\bar{t} + \ln \tau)$, d.h. $\ell' = \frac{N}{\tau^2}(\bar{t} + \tau)$ und $\ell'' = -\frac{N}{\tau^3}(2\bar{t} + \tau)$. \square

2.5 Quanteninformation

[Playing the game of Twenty Questions,] I come into the room, I say “Is it edible?” and my friend that I asked thinks, thinks, thinks, and says finally, “No.” Then I asked the next person, “Is it mineral?” Thinks and thinks. “Yes.” And so it goes. But before my twenty questions run out, I must arrive at some word. “Is it ‘cloud’?” And the respondent that I ask this time thinks and thinks and thinks. I couldn’t understand why the delay because supposedly the word already existed in the room, and all he’d have to do is give an answer “yes” or “no.” But instead he thinks. Finally he says “yes” and everybody bursts out laughing. They explained to me that when I came in there was no word in the room, that everybody could answer my questions as he wished with one small proviso – that if I challenged, and he couldn’t supply a word compatible with all the previous answers, he lost and I won. So it was just as difficult for them as it was for me. Thus the word was brought into being by the choice of questions asked and answers given. The game reminds us of how we find out about the electron, its position or its momentum. What we get depends partly on our choice of question asked. However, we don’t have the whole voice in it. “Nature” has the rest of the answer.

John Archibald Wheeler [59]

2.5.1 Qubits

In Analogie zu dem klassischen Bit heißt ein Quantensystem, das zwei Zustände einnehmen kann, ein *Qubit*.⁶ In der Diracschen „bra-ket“-Notation wird ein Qubit als $|0\rangle$ oder $|1\rangle$ dargestellt. In der Physik ist es üblich, Qubits als Spaltenvektoren darzustellen,

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad \text{geometrisch: } \begin{array}{c} \uparrow |1\rangle \\ |0\rangle \rightarrow \end{array} \quad (15)$$

Ein Qubit kann in zwei Zuständen sein, genau wie ein klassisches Bit, ein wesentlicher Unterschied zwischen Bit und Qubit ist jedoch, dass ein Bit nur in einem eindeutigen Zustand sein kann:

⁶Mathematisch gesehen nimmt ein Qubit einen Wert in dem Vektorraum \mathbb{C}^2 an und ist somit ein „Spinor“.

Postulat 1 (Superpositionsprinzip) Ein Qubit $|\psi\rangle$ kann nur in einem Zustand sein, der eine Linearkombination von zwei Basiszuständen $|0\rangle$ and $|1\rangle$ ist,

$$|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} \quad (16)$$

wo $\alpha_i \in \mathbb{C}$ komplexe Konstanten sind, die „Wahrscheinlichkeitsamplituden“, für die $|\alpha_0|^2 + |\alpha_1|^2 = 1$ gilt. Für $\alpha_0, \alpha_1 \neq 0$ heißt der Zustand $|\psi\rangle$ eine „Superposition“ der Basiszustände.

Ein Qubit kann also durch eine einfache Vektorsumme von zwei Basisvektoren $|0\rangle$, $|1\rangle$ illustriert werden (Abb. 2). Beachten Sie, dass diese Illustration die Gefahr einer zu

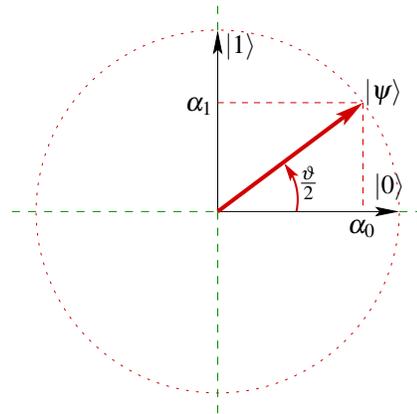


Abbildung 2: Ein Qubit $|\psi\rangle$ als eine Superposition von Basiszuständen $|0\rangle$ and $|1\rangle$, d.h. $|\psi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle$, mit den Amplituden α_0 and α_1 . Ein Qubit entspricht also einem Vektor auf dem gepunkteten Kreis, hier illustriert als *reeller* Unterraum des \mathbb{R}^2 .

starken Vereinfachung in sich birgt: natürlich sind die erlaubten Amplituden *komplex*, d.h. ein Qubit liegt auf der komplexen Einheitskugel⁷ des \mathbb{C}^2 .

Beispiel 2.6 Ein allgemeines Qubit kann durch

$$|\psi\rangle = e^{i\delta} \begin{pmatrix} e^{-i\varphi/2} \cos \frac{\vartheta}{2} \\ e^{i\varphi/2} \sin \frac{\vartheta}{2} \end{pmatrix} \quad (17)$$

mit den drei Winkeln $\delta, \varphi, \vartheta \in [0, 2\pi)$. dargestellt werden.

Insbesondere für den Schnitt $\delta = \varphi = 0$, ergibt Gl. (17) das Qubit

$$|\psi\rangle = \begin{pmatrix} \cos \frac{\vartheta}{2} \\ \sin \frac{\vartheta}{2} \end{pmatrix} \quad (18)$$

⁷ Komplexe Zahlen sind darstellbar als zweidimensionale reelle Vektoren ($\mathbb{C} \cong \mathbb{R}^2$); somit ist die geometrische Figur möglicher Qubits eine *dreidimensionale reelle Hyperkugel* im 4-dimensionalen reellen Zustandsraum $\mathbb{C}^2 \cong \mathbb{R}^4$.

□

Beispiel 2.6 zeigt, dass ein Qubit von drei Parametern abhängt, den Winkeln δ , φ , und ϑ . Die Winkel δ und φ haben keine direkte physikalische Bedeutung, da sie nicht gemessen werden können. Der Winkel ϑ jedoch kann direkt gemessen werden, er ist eine physikalisch beobachtbare Größe.

Postulat 2 (Messung eines Qubits) Wenn ein Qubit $|\psi\rangle$ in (16) gemessen wird, ist das Messergebnis „0“ mit der Wahrscheinlichkeit $|\alpha_0|^2$, und „1“ mit der Wahrscheinlichkeit $|\alpha_1|^2$:

$$P(j) = |\alpha_j|^2 \quad \text{für } j = 0, 1. \quad (19)$$

Nach der Messung ist das Qubit entweder in Zustand $|0\rangle$ oder in Zustand $|1\rangle$:

$$|\psi'\rangle = \begin{cases} |0\rangle & \text{wenn das Messergebnis „0“ war,} \\ |1\rangle & \text{wenn das Messergebnis „1“ war.} \end{cases} \quad (20)$$

Hier bezeichnet $|\psi'\rangle$ den Zustand des Qubits nach der Messung. Oft sagt man, dass durch eine Messung des Zustands $|\psi\rangle$ das System in die „Eigenzustände“ $|0\rangle$ oder $|1\rangle$ „kollabiert“.

Die Wirkung einer Messung, d.h. der Wechsel von Zustand $|\psi\rangle$ nach Zustand $|\psi'\rangle = |0\rangle$ oder $|1\rangle$, ist in Abbildung 3 skizziert. Eine wesentliche Eigenschaft der Quanten-

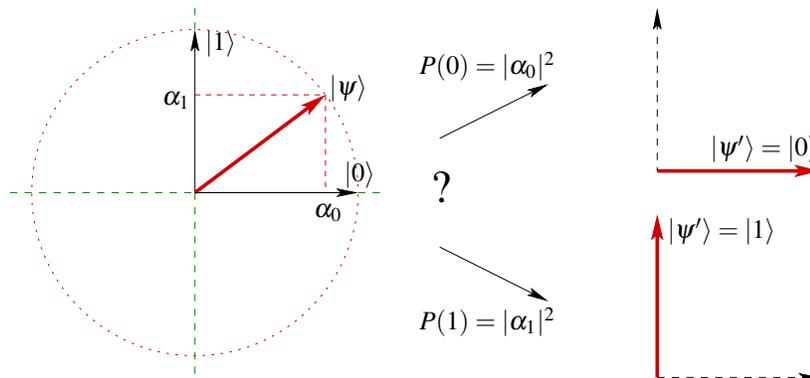


Abbildung 3: Die Wirkung einer Messung eines Qubits: Der Qubit-Zustand $|\psi\rangle$ kollabiert zu $|\psi'\rangle$, und zwar entweder $|0\rangle$ mit Wahrscheinlichkeit $P(0) = |\alpha_0|^2$ oder $|1\rangle$ mit Wahrscheinlichkeit $P(1) = |\alpha_1|^2$.

mechanik ist, dass die Messung eines Quantensystems, hier eines Qubits, *den Zustand selbst verändert*. Das ist ganz im Gegensatz zu unserer Alltagserfahrung, nach der die reine Beobachtung von Objekten deren Zustand nicht ändert. Beispielsweise können Sie ein Bild anschauen, ohne dass Sie dadurch seinen Inhalt oder seine Farben ändern. Nach der Messung ist der Zustand exakt bekannt, ein so- genannter *reiner Zustand*.

Allgemein heißt der Kollaps einer Superposition in einen reinen Zustand *Dekohärenz*. Dekohärenz ist ein weit allgemeineres Phänomen, das immer dann auftritt, wenn ein Quantensystem mit seiner Umgebung wechselwirkt. Eine Messung ist ein spezieller Fall von Dekohärenz.

Das Postulat der Messung impliziert das *Heisenbergsche Unschärfeprinzip* [61, Gl. (133)]. Es besagt, dass man den vollständigen Zustand eines Quantensystems, also beispielsweise Ort und Impuls eines Elementarteilchens, prinzipiell nicht exakt messen kann.

2.5.2 von-Neumann-Entropie

Für einen gegebenen Zustand ψ mit entsprechendem *Dichteoperator* $\rho = |\psi\rangle\langle\psi|$ ist die *von-Neumann-Entropie* $S(\rho)$ definiert als

$$S(\rho) = -\text{tr}(\rho \ln \rho). \quad (21)$$

Es ergibt die durchschnittliche Information eines Quantensystems [18]. Für ein Quantensystem in einem N -dimensionalen Hilbert-Raum reicht die Entropie von $\ln N$ für einen vollständig gemischten Zustand ($\rho^2 \neq \rho$) bis herunter zu 0 für einen reinen Zustand ($\rho^2 = \rho$).

Beispiel 2.7 Betrachte ein Qubit ψ wie in Beispiel 2.6. Dann gilt für den Dichteoperator

$$\begin{aligned} \rho = |\psi\rangle\langle\psi| &= \begin{pmatrix} e^{-i\varphi/2} \cos \frac{\vartheta}{2} \\ e^{i\varphi/2} \sin \frac{\vartheta}{2} \end{pmatrix} \cdot \begin{pmatrix} e^{i\varphi/2} \cos \frac{\vartheta}{2} & e^{-i\varphi/2} \sin \frac{\vartheta}{2} \end{pmatrix} \\ &= \begin{pmatrix} \cos^2 \frac{\vartheta}{2} & e^{-i\varphi} \cos \frac{\vartheta}{2} \sin \frac{\vartheta}{2} \\ e^{i\varphi} \cos \frac{\vartheta}{2} \sin \frac{\vartheta}{2} & \sin^2 \frac{\vartheta}{2} \end{pmatrix}. \end{aligned}$$

(Beachte: $\rho = \rho(\vartheta, \varphi)$.) Mit den Additionstheoremen [62, S. 56] folgt

$$\rho = \frac{1}{2} \begin{pmatrix} 1 + \cos \vartheta & e^{-i\varphi} \sin \vartheta \\ e^{i\varphi} \sin \vartheta & 1 - \cos \vartheta \end{pmatrix}. \quad (22)$$

Da ρ also die Eigenwerte $\lambda_1 = 0$ und $\lambda_2 = 1$ hat, ist

$$\text{tr}(\rho \ln \rho) = -\sum_{i=1}^2 \lambda_i \ln \lambda_i = 0. \quad (23)$$

Andererseits hat der gemischte Zustand $\rho' = \frac{1}{2} \mathbf{1}$ einen doppelten Eigenwert $\frac{1}{2}$, d.h.

$$\text{tr}(\rho \ln \rho) = -\sum_{i=1}^2 \lambda_i \ln \lambda_i = \ln 2. \quad (24)$$

□

Ein *Quantenregister* der Größe n ist ein System von n Qubits. Auf den ersten Blick könnte man vermuten, dass ein Zustand eines Quantenregisters einen unbegrenzten Betrag an Information beinhalten kann, da die Wahrscheinlichkeitsamplituden ja unendlich viele Werte annehmen können und so sich das Quantenregister sich in unendlich vielen Zuständen befinden kann. Das ist zwar richtig, aber wegen der Besonderheiten einer Quantenmessung ist der Betrag an klassischer Information, den wir aus einem Quantenzustand extrahieren können, tatsächlich begrenzt.

Der fundamentale Satz von Holevo besagt, dass in einem Quantenregister der Größe n nur n klassische Bits gespeichert werden können, so dass sie auch wieder als klassische Information abgerufen werden kann.

2.5.3 Fisher-Quanteninformation

Sei (ρ, θ, m) ein „parametrisches Quantenmodell“, für das definitionsgemäß eine Messung in einer Beobachtung eines Messwertes x mit der Wahrscheinlichkeitsdichte

$$p(x; \theta) = \text{tr}[\rho(\theta)m(x)] \quad (25)$$

resultiert. Hierbei ist ρ die so-genannte *Dichtematrix*, $\theta \in \Theta$ ein Parameter im Parameterraum Θ , und $m(x)$ ist die Dichte der Messung. Falls der Zustandsraum \mathcal{H} des Quantensystems endlichdimensional ist, also $\mathcal{H} = \mathbb{C}^d$, so sind ρ und $m(x)$ selbstadjungierte $(d \times d)$ -Matrizen. Zur Vereinfachung nehmen wir den Parameterraum als eindimensional an. Die parameterabhängige Wahrscheinlichkeitsdichte (25) impliziert also eine Log-Likelihood

$$\ell(\theta) = \ln \text{tr}[\rho(\theta)m(x)]. \quad (26)$$

Zur Berechnung der Ableitungen der Log-Likelihood verwendet man die *symmetrische logarithmische Ableitung* (den *symmetrischen Quantenbeitrag*) $\rho_{//\theta}$ von ρ , der implizit definiert ist als die selbstadjungierte Lösung der Gleichung

$$\rho_{/\theta} = \frac{1}{2} (\rho\rho_{//\theta} + \rho_{//\theta}\rho), \quad (27)$$

wo $\rho_{/\theta} = \partial\rho/\partial\theta$ die partielle Ableitung bezeichnet. Die Observable

$$J(\theta) = -\rho_{//\theta/\theta}(\theta) \quad (28)$$

ist die *beobachtbare Quanteninformation*. Damit lautet die *Fisher-Quanteninformation* [1, §6.1]

$$I(\theta) = \text{tr}[\rho(\theta)\rho_{//\theta}(\theta)^2] = \text{tr}[\rho(\theta)J(\theta)]. \quad (29)$$

Die zweite Gleichung ist das Quantenanalogon der klassischen Gleichung $i(\theta) = E[j(\theta)]$ mit $j(\theta) = -\ell''(\theta)$ ist. J ist eine Observable, während j entsprechend eine Zufallsgröße ist.

Die klassische Fisher-Information eines Quantensystems mit Dichteoperator ρ im Zustandsraum \mathcal{X} einer Messung M mit Dichte m ist dann gegeben durch

$$i(\theta; M) = \int_{\mathcal{X}} \frac{1}{p(x; \theta)} (\operatorname{Re} \operatorname{tr}\{\rho(\theta) \rho_{//\theta}(\theta) m(x)\})^2 \mu(dx). \quad (30)$$

Weder I noch J hängen von der Messung ab, im Gegensatz zur klassischen Fisher-Information. Die klassische Fisher-Information einer Messung ist maximal, wenn für μ -fast alle Messergebnisse $x \in \{x: p(x; \theta) > 0\}$ gilt

$$\sqrt{m(x)} \rho_{//\theta}(\theta) \sqrt{\rho(\theta)} = r \sqrt{m(x)} \sqrt{\rho(\theta)} \quad (31)$$

mit einer Konstanten $r \in \mathbb{R}$. In diesem Fall nimmt die klassische Fisher-Information den Wert der Fisherschen Quanteninformation an. Für jedes θ gibt es solche Messungen, die diese Obergrenze annehmen. Insbesondere kann man die Messung M so wählen, dass jedes $m(x)$ eine Projektion auf einen Eigenraum des Quantenbeitrags (*quantum score*) $\rho_{//\theta}(\theta)$ ist.

Beispiel 2.8 Betrachten wir ein Qubit ψ (Beispiel 2.6) in einem reinen Zustand, also mit Dichtematrix $\rho = |\psi\rangle\langle\psi|$, wo $|\psi\rangle = |\psi(\vartheta, \varphi)\rangle$. Als selbstadjungierte Matrix kann dann ρ dann dargestellt werden durch $\rho = (\mathbf{1} + \mathbf{u}\vec{\sigma})/2$, wo $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ die drei Pauli-Matrizen sind und $\mathbf{u} = \mathbf{u}(\vartheta, \varphi) \in S^2$ ein Punkt auf der Bloch-Sphäre [37, §1.2] mit Polarkoordinaten (ϑ, φ) , d.h.

$$\mathbf{u} = \begin{pmatrix} \cos \varphi \sin \vartheta \\ \sin \varphi \sin \vartheta \\ \cos \vartheta \end{pmatrix}, \quad \frac{\partial \mathbf{u}}{\partial \varphi} = \sin \vartheta \begin{pmatrix} -\sin \varphi \\ \cos \varphi \\ 0 \end{pmatrix}. \quad (32)$$

Das ergibt exakt Gleichung (22). Nehmen wir an, dass der Breitengrad ϑ bekannt ist, $0 < \vartheta < \pi$. Dann ist der Längengrad φ der unbekannte Parameter, d.h. $\rho = \rho(\varphi)$. Da alle Zustände $\rho(\varphi)$ rein sind, hat man mit (32)

$$\rho_{//\varphi}(\varphi) = 2 \rho_{/\varphi}(\varphi) = 2 \sin \vartheta \begin{pmatrix} \cos \varphi \\ \sin \varphi \\ 0 \end{pmatrix} \cdot \vec{\sigma}.$$

und damit mit Gleichung (28)

$$J(\varphi) = 2 \sin \vartheta \begin{pmatrix} 0 & e^{-i\varphi} \\ e^{i\varphi} & 0 \end{pmatrix} \quad (33)$$

Mit (22) ergibt sich $\rho J = \begin{pmatrix} (1 - \frac{\sin \vartheta}{\cos \vartheta}) e^{i\varphi} & (1 + \frac{\cos \vartheta}{\sin \vartheta}) e^{-i\varphi} \end{pmatrix}$. Mit (29) erhalten wir so die Fishersche Quanteninformation

$$I(\varphi) = \sin^2 \vartheta. \quad (34)$$

Für ein Qubit $|\psi(\vartheta, \varphi)\rangle$ ist also die Fishersche Quanteninformation bei gegebenem ϑ unabhängig von dem Parameter φ . Sie wird um so kleiner, je näher sich der Bloch-Vektor \mathbf{u} an einem der Pole befindet. Am Nordpol ergibt $\vartheta = 0$ ergibt sich sicher $|\psi\rangle = |0\rangle$, am Südpol $\vartheta = \pi$ entsprechend $|1\rangle$. Die Fisher-Quanteninformation ist an den Polen 0 — kein Wunder, man kennt ja mit ϑ bereits den vollständigen Zustand des Qubits! \square

3 Anwendungen der Information

3.1 Information und Energie

Mit Information kann man Energie gewinnen, wie uns die folgenden Überlegungen zeigen. Der *Maxwellsche Dämon* ist ein Gedankenexperiment des schottischen Physikers James Clerk Maxwell aus dem Jahre 1867 [29, §1.2.1], in dem zwei gasgefüllte, thermisch isolierte Kammern gleicher Temperatur durch eine kleine Luke miteinander verbunden sind, welche (nahezu) ohne Arbeit geöffnet oder verschlossen werden kann. Die Luke ist so klein, dass nur ein einziges Gasteilchen hindurchfliegen kann. Ein

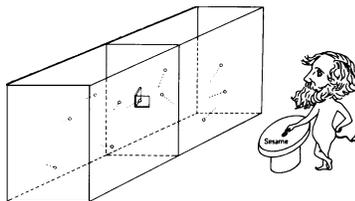


Abbildung 4: Der Maxwellsche Dämon steuert eine Luke zwischen zwei Kammern mit anfänglich gleicher Temperatur. Abb. aus [35].

„Dämon“ öffnet oder schließt die Luke für auf sie zurasende Teilchen, um abhängig von ihrer Geschwindigkeit alle überdurchschnittlich schnellen Teilchen in die rechte Kammer und alle überdurchschnittlich langsamen in die linke zu lassen. Die linke Kammer kühlt sich also ab und die rechte erwärmt sich, während am Ende beide dieselbe Anzahl Teilchen beinhalten. Die Temperaturdifferenz kann später zur Verrichtung von Arbeit verwendet werden, die damit netto gewonnen wurde, solange das Öffnen und Schließen der Luke weniger Arbeit erfordert als die Überschussenergie eines schnellen Teilchens. Der Aufbau verrichtet also Arbeit durch Abkühlung eines Wärmereservoirs, das sich im thermischen Gleichgewicht mit der Umgebung befindet, und stellt damit ein Perpetuum Mobile zweiter Art dar, dessen Existenz der zweite Hauptsatz verbietet. Wie lässt sich das Paradox lösen, also der Dämon vertreiben?

Bei dem Versuch der Vertreibung des Maxwellschen Dämons brachte Szilard 1929 erstmals eine Messung durch ein intelligentes Wesen in Verbindung mit Entropie [47, S. 845]: „*Es ist nun naheliegend, anzunehmen, daß die Vornahme einer Messung prinzipiell mit einer ganz bestimmten mittleren Entropieerzeugung verbunden ist . . .*“ Obwohl diese Annahme falsch ist, stellt er das interessante und für die weitere Ent-

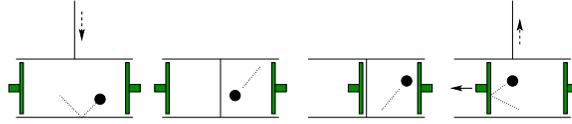


Abbildung 5: Szilards Maschine, ein Gedankenexperiment [47] zur Gewinnung der Energie $\Delta E = kT \ln 2$ aus einem Bit Information.

wicklung bedeutsame Gedankenexperiment in Abbildung 5 vor, aus dem man korrekt folgert, dass ein Bit Information eine Energie vom Betrag

$$\Delta E = kT \ln 2 \quad (35)$$

ergibt, denn die Entropieabnahme S durch die Halbierung des Volumens von V auf $V/2$ bei gegebener Temperatur T ist mit $P = kT/V$ („isotherme Expansion“ [8, Gl. (5.29)])

$$S = \frac{1}{T} \int_V^{V/2} P dV = k \int_V^{V/2} \frac{dV}{V} = -k \ln 2,$$

d.h. $\Delta E = -TS = kT \ln 2$. Nach Landauer [28] muss jedoch eine „logisch reversible“ Operation, also eine Operation, bei der man vom Ergebnis eindeutig auf die Eingabe schließen kann, nicht notwendig Entropie erzeugen, denn jede logisch reversible Operation kann durch einen physikalisch reversiblen Prozess realisiert werden:

$$\boxed{\text{logisch reversibel} = \text{thermodynamisch reversibel.}} \quad (36)$$

Eine Messung ist aber im Wesentlichen ein Kopieren vorhandener Information und daher umkehrbar.

Doch wenn die Messung nicht prinzipiell Entropie kostet, was ist es dann? Das *Löschen* von Information als eine „logisch irreversible“ Operation erzeugt Entropie, und zwar stets mindestens so viel wie durch den Informationsgewinn der Szilard-Maschine erniedrigt wird: Nach dem *Landauerschen Prinzip* [37, §3.2.5] erhöht das Löschen eines Bits an Information die Entropie der Umgebung um mindestens $kT \ln 2$.

Da der Dämon keinen unendlichen Speicherplatz besitzt, muss er zwangsläufig Information, die er durch seine Messungen gewonnen hat, irgendwann löschen. Das erhöht die Entropie.

3.2 Entropie Schwarzer Löcher

Schwarze Löcher sind spezielle Lösungen der Einsteinschen Feldgleichungen der Allgemeinen Relativität. Ein *Schwarzes Loch* ist eine Raumregion, die von einer Art Membran, dem *Ereignishorizont*, umgeben ist, durch die zwar Materie und Energie von außen ins Innere gelangen kann, nicht aber umgekehrt. Nach einem berühmten Eindeutigkeitssatz im Rahmen der Allgemeinen Relativität, dem „Keine-Haare-Theorem“ [20, §2.2], ist jedes Schwarze Loch allein durch drei Parameter spezifiziert, seine Masse $M > 0$, seinen Drehimpuls a und seine elektrische Ladung Q , es hat keine

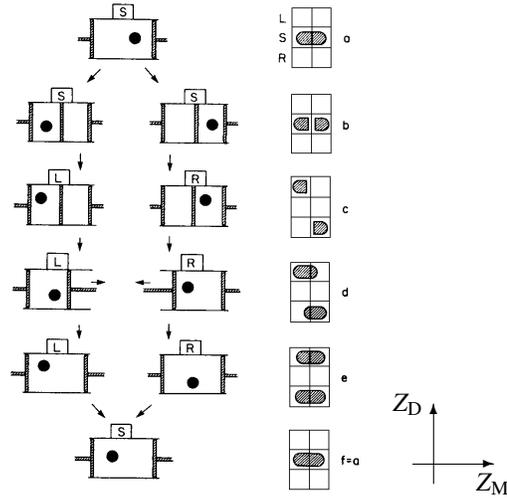


Abbildung 6: Bennetts Ein-Teilchen-Version des Maxwell'schen Dämons. Der Zustandsraum in der rechten Spalte setzt sich zusammen aus dem Geisteszustand Z_D (dem „Wissen“ des Dämons), und dem Zustand Z_M des Moleküls. Der Geisteszustand kann nur „S“ (Standard) für Unwissenheit, „L“ nach der Messung „Molekül in linker Kammer“ und „R“ nach der Messung „Molekül in rechter Kammer“ sein. In Schritt d wird die Information in Energie umgesetzt, das unvermeidliche Löschen geschieht von Schritt e nach f. Abb. aus [4].

weiteren physikalischen Eigenschaften. Üblicherweise werden diese Größen in geometrisierten Einheiten angegeben, also in einer Längeneinheit [53, 54]. Ein solches allgemeines Schwarzes Loch wird *Kerr-Newmansches Schwarzes Loch* [10, 51] genannt. Es heißt *extremal*, wenn $M^2 = a^2 + Q^2$, d.h. wenn zu gegebenem Drehimpuls a und gegebener Ladung Q die Masse M minimal ist.

Eine *Reissner-Nordström-Raumzeit* ist speziell gegeben für $a = 0$, eine *Schwarzschild-Raumzeit* für $a = Q = 0$ [53].

Die Oberfläche A eines Kerr-Newmanschen Schwarzen Lochs [49, §A2] ist gegeben durch

$$A = 4\pi(2Mr_+ - Q^2), \quad (37)$$

wo $r_+ = M + \sqrt{M^2 - a^2 - Q^2}$ den Radius des Ereignishorizonts bezeichnet, den *Schwarzschild-Radius*. Bekenstein [3] definierte heuristisch die Entropie eines Schwarzen Lochs als

$$S_{\text{BH}} = \frac{1}{4} k N_B \quad (38)$$

mit der Boltzmann-Konstanten k und der *Bekenstein-Zahl*

$$N_B = \frac{A}{l_{\text{P}}^2}, \quad \text{wo } l_{\text{P}}^2 = \frac{\hbar G}{c^3}. \quad (39)$$

Die Konstante l_{P}^2 heißt *Planck-Fläche*. Geometrisch ist N_B also die Anzahl der Planck-Flächen, die in die Oberfläche A des Schwarzen Lochs passen. Die Entropie eines Schwarzen Lochs ist die Menge an Information, die sich hinter seinem Ereignishorizont befindet.

Der folgende Satz hat als Konsequenz, dass ein rotierendes Schwarzes Loch sich nicht im thermodynamischen Gleichgewicht befindet, seine Rotationsenergie also stets an die Umgebung abgegeben wird.

Satz 3.1 *Die Entropie eines Schwarzen Lochs der Masse M und der elektrischen Ladung Q ist maximal dann und nur dann, wenn es nicht rotiert, d.h. wenn $a = 0$.*

Beweis. Betrachtet man die Bekenstein-Zahl als eine Funktion von a auf, also $N_B = N_B(a)$, so ergibt sich mit (37) und wegen $\frac{\partial r_+}{\partial a} = -a/\sqrt{M^2 - a^2 - Q^2}$,

$$\frac{\partial N_B}{\partial a} = -\frac{8\pi c^3}{\hbar G} \cdot \frac{aM}{\sqrt{M^2 - a^2 - Q^2}}. \quad (40)$$

Da für ein Schwarzes Loch stets $M^2 \geq a^2 + Q^2$, erhalten wir $\partial N_B/\partial a < 0$ für $a > 0$, und $\partial N_B/\partial a > 0$ für $a < 0$. Q.E.D.

Physikalisch wird die Dissipation der Rotationsenergie eines Schwarzen Lochs von der *Superradianz* verursacht, durch die einfallende Wellen mit ganzzahligem Spin (z.B. Skalare, Photonen und Gravitonen) mit genügend kleinen Frequenzen verstärkt werden [10, 49, 50, 56]. Ein allgemeines Schwarzes Loch im stabilen thermodynamischen Gleichgewicht ist also stets statisch ($a = 0$) und (i.A.) elektrisch geladen, also eine Reissner-Nordström-Raumzeit.

So elegant diese Überlegungen auch sind, so sind sie semiklassisch und lediglich in Analogie zur Thermodynamik begründet. Sie sind zwar deswegen nicht unbedingt falsch, aber für eine strenge Herleitung der Entropie müsste man die Quantenzustände eines Schwarzen Lochs zählen können. Das gelang erst mit der Superstring-Theorie.

3.2.1 String-Theorie und Entropie Schwarzer Löcher

Die String-Theorie, oder genauer Superstring-Theorie und ihre allgemeine Formulierung, die M-Theorie, ist eine vielversprechende Quantentheorie gravitativer Wechselwirkungen. Sie verwirft radikal das vor 2500 Jahren von Demokrit eingeführte Konzept idealisierter Punktteilchen und sieht stattdessen jedes Teilchen als eine schwingende (offene oder geschlossene) Saite, einen *String* mit Längen von der Größenordnung der Planck-Länge $l_p = 1.616 \cdot 10^{-35}$ m. Nach der M-Theorie wird diese Verallgemeinerung der Teilchendarstellung noch weiter getrieben zu p -dimensionalen p -Branen, Untermanifaltigkeiten einer 11-dimensionalen Mannigfaltigkeit $X = M_4 \times K_7$, wo M_4 eine vierdimensionale Mannigfaltigkeit ist, die die Raumzeit darstellt, und K_7 eine siebendimensionale kompakte Mannigfaltigkeit [26, §11],

$$\text{Punkt (0-dim)} \quad \rightarrow \quad \text{String (1-dim)} \quad \rightarrow \quad p\text{-Bran (} p\text{-dim)}.$$

Als eine Quantentheorie der Gravitation müsste die Superstring-Theorie offene Probleme der Physik Schwarzer Löcher lösen können. Tatsächlich konnte die Bekenstein-

Hawkingsche Entropie-Formel (37) mit statistischen Mitteln Mitte der 1990er Jahre hergeleitet werden [26, §13.7]. Dies war ein außerordentlicher Durchbruch für die String-Theorie.

Durch Anwendung von Methoden der M-Theorie konnten Strominger and Vafa [45] die Quantenzustände eines extremalen Reissner-Nordströmschen Schwarzen Lochs ($M = |Q|$) zählen. Sie verwendeten “BPS gesättigte Zustände”:⁸ Das sind supersymmetrische Zustände gegebener Ladung, die einer Minimalbedingung, z.B. der kleinstmöglichen Masse, genügen und dadurch eindeutig sind. Zudem ist dieser Zustand exakt berechenbar, also ohne die sonst in der Stringtheorie oft notwendigen störungstheoretischen Rechnungen und hängen zusammen mit der Existenz von p -Bran-Zuständen, vgl. [17, S. 350], [26, §12.1].

Gehen wir etwas mehr ins Detail [21, 33], [63, §16.7]. Wir betrachten die Typ IIB Superstring-Theorie,⁹ kompaktifiziert auf dem fünfdimensionalen Torus T^5 . Die Niedrigenergie-Theorie ist die maximal supersymmetrische Supergravitationstheorie in fünf Dimensionen. Wir betrachten ein Schwarzes Loch mit den folgenden drei ganzzahligen Ladungen.

- Q_5 : Anzahl der um T^5 gewickelten D-Fünfbrannen,
- Q_1 : Anzahl der um S^1 gewickelten D-Einbrannen,
- N : Windungszahl in Richtung i des D-Strings
($i = 9, P = N/R$ Impuls, mit Stringradius R)

Beginnen wir mit der zehndimensionalen Lösung [33, Gl. (2)]

$$ds^2 = \frac{1}{\sqrt{f_\alpha f_\gamma}} \left[-dt^2 + dx_9^2 + \frac{r_+^2}{r^2} (\cosh \sigma dt + \sinh \sigma dx_9)^2 \right] + f_\alpha [dx_5^2 + \dots + dx_8^2] + \sqrt{f_\alpha f_\gamma} \left[\frac{r^2 dr}{r^2 - r_+^2} + r^2 d\Omega_3^2 \right] \quad (41)$$

mit

$$f_x = 1 + \frac{r_+^2 \sinh^2 x}{r^2}. \quad (42)$$

Diese Lösung ist mit den vier unabhängigen Größen α , γ , σ und r_+ parametrisiert. Zwei weitere Parameter R und V erscheinen durch die Bedingungen der Ladungsquantisierung,

$$R = R_9, \quad V = R_5 R_6 R_7 R_8, \quad (43)$$

wo R_i den Radius der i -ten Dimension der fünf kompakten Richtungen bezeichnet. Das

⁸BPS steht für E. Bogomol’nyi, Manoj Prasad und Charles Sommerfield.

⁹d.i. eine zehndimensionale Theorie geschlossener Strings

ergibt die drei Ladungen

$$\begin{aligned} Q_1 &= \frac{Vr_+^2}{2g} \sinh 2\alpha, \\ Q_5 &= \frac{r_+^2}{2g} \sinh 2\gamma, \\ N &= \frac{R^2 V r_+^2}{2g^2} \sinh 2\sigma, \end{aligned} \quad (44)$$

wo g die Stringkopplung ist, für die gilt [21, (2.9)], [63, §13.4]

$$g^2 = \frac{G^{(10)}}{8\pi^6} = \frac{4RVG^{(5)}}{\pi}. \quad (45)$$

Weitere Lösungen Schwarzer Löcher können von dieser mit Hilfe einer U-Transformation [21, (2.16)] gebildet werden. Die Gesamtmasse (ADM-Energie) des Systems ist

$$M = \frac{RVr_+^2}{2g^2} (\cosh 2\alpha + \cosh 2\gamma + \cosh 2\sigma), \quad (46)$$

die Bekenstein-Hawking-Entropie $S = S_{BH} = \frac{A_{10}}{4G^{(10)}} = \frac{A_5}{4G^{(5)}}$ lautet

$$S = 2\pi\sqrt{Q_1 Q_5 N} = \frac{2\pi RV r_+^3}{g^2} \cosh 2\alpha \cosh 2\gamma \cosh 2\sigma, \quad (47)$$

und die Hawking-Temperatur

$$T = \frac{1}{2\pi r_+ \cosh \alpha \cosh \gamma \cosh \sigma}. \quad (48)$$

Alle thermodynamischen Größen liefern dieselben Werte in zehn oder fünf Dimensionen, z.B. $S = A_{10}/(4l_p^{10}) = A_5/(4l_p^5)$. Mit $\alpha = \gamma = \sigma = 0$, d.h. mit drei verschwindenden Ladungen $Q_1 = Q_5 = N = 0$, erhalten wir die Schwarzschild-Lösung

$$ds^2 = -\kappa dt^2 + \kappa^{-1} dr^2 + r^2 d\Omega_3^2 + dx_5^2 + \dots + dx_9^2. \quad (49)$$

mit $\kappa = f_{-1-\sqrt{2}} = 1 - r_+^2/r^2$. Ist dabei nur die Windungszahl N (also σ) nichtverschwindend, so erhalten wir die Schwarzschild-Metrik mit einem *Boost*¹⁰ in die x_9 -Richtung. Ist dagegen nur die Q_1 -Ladung (also α) nicht Null, so erhalten wir einen 6-dimensionalen schwarzen String, ist nur $Q_5 \neq 0$ (also γ), eine schwarze Fünfbran [21, (2.13–15)].

¹⁰*boost* – „Anhebung“, in der Relativitätstheorie auch „Lorentzbeschleunigung“

3.2.2 Das Informationsparadoxon

Die Entropie eines Schwarzen Lochs ist die Menge an Information, die sich hinter seinem Ereignishorizont befindet. Nun scheint es einen Widerspruch zwischen der klassischen Allgemeinen Relativität und der Quantentheorie zu geben. Nach der Allgemeinen Relativität gilt für ein Schwarzes Loch das „Keine-Haare-Theorem“, d.h. es ist vollständig beschrieben durch die drei Parameter Masse, Rotation und elektrische Ladung. Eine Konsequenz daraus ist, dass jegliche Struktur oder Information, die Energie oder Materie vor Überquerung oder Bildung des Ereignishorizonts noch hatte, für alle Zeiten verloren ist, zumindest für den „Außenraum“, also die Umgebung des Lochs.

Das widerspricht einem Grundsatz der Quantenmechanik, der *Mikroreversibilität*. Sie besagt, dass im Prinzip jeder Prozess umkehrbar sein muss. Mit anderen Worten muss stets ein Maxwellscher Dämon existieren können, der die Entwicklung eines Systems umkehren kann.

Dieses Paradoxon führte zu einer etwa 20 Jahre währenden Debatte zwischen S. Hawking und K. Thorne auf der einen Seite, die Schwarzen Löchern die einzigartige Eigenschaft zuschrieben, Information für alle Zeit zu vernichten, und L. Susskind und G. 't Hooft auf der anderen Seite, die darin eine Verletzung der Quantenmechanik sahen.¹¹

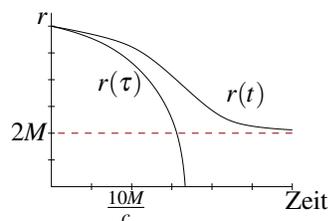


Abbildung 7: Die radiale Entfernung r vom Ursprung eines Schwarzschildischen Schwarzen Lochs, die ein in das Loch fallender Beobachter mit Startpunkt $r = 6M$ abhängig von der Zeit beschreibt. In seiner Eigenzeit τ überquert er nach kurzer Zeit den Ereignishorizont $r_+ = 2M$ und rast auf die Singularität zu, während er in der Zeit t eines Beobachters weit draußen den Horizont *nie erreicht* [10, §19(a)].

Die Auflösung des vermeintlichen Paradoxons liegt in der *Beobachterabhängigkeit der Zeit* gemäß der Allgemeinen Relativität (Abb. 7). Ein erläuterndes Beispiel stammt von Susskind:¹¹ „Just to make the point let’s imagine a whole galaxy falling into a huge black hole with a Schwarzschild radius equal to a billion light years. From the outside, the galaxy and all its unfortunate inhabitants appear to be . . . [slowed down and frozen] at the horizon! On the other hand, the infalling galactic inhabitants glide through perfectly happily. To them the trauma only happens at the singularity a billion years later. But as in the case of certain life-after-death theories, the folks on the other side can never communicate with us.“ Wofür der fallende Beobachter eine endliche Zeit benötigt, erscheint einem Beobachter weit draußen im wahrsten Sinne als Ewigkeit. Trotzdem ist für jeden der beiden Beobachter die Situation physikalisch

¹¹<http://arxiv.org/abs/hep-th/0204027>

konsistent, insbesondere für den Außenraumbeobachter fällt in endlicher Zeit keine Information *in* das Schwarze Loch. Er könnte, ganz als Maxwellscher Dämon, die am Horizont „eingefrorenen“ Stringzustände mit einem Teleskop betrachten und rekonstruieren [46]. Die Tatsache, dass sich die physikalischen Erfahrungen der beiden Beobachter nicht widersprechen wird *Komplementaritätsprinzip* genannt, sie ist eine Konsequenz des Äquivalenzprinzips und damit der klassischen Allgemeinen Relativitätstheorie.

3.3 Information in der Finanzmathematik

Information spielt auch in der Ökonomie eine wichtige Rolle. Aus betriebswirtschaftlicher Sicht beispielsweise erlangt sie als einer der Produktionsfaktoren immer größere Bedeutung. Volkswirtschaftlich wird beispielsweise ein Markt als effizient definiert, wenn die Marktpreise alle verfügbare Information enthalten [42, App. 26]. Die Finanzmärkte sind einige der wenigen realen Märkte, die man als effizient annimmt [22, §10.1].

Dennoch gibt es kaum Ansätze, Information in der Ökonomie zu quantifizieren. Was kostet denn nun Information? Oder auch nicht erhaltene Information? Ein Ansatz verbindet Information mit der erwarteten Rendite eines Portfolios. Der Zusammenhang offenbart sich erst nach zwei Schritten: Zunächst benötigt man die so genannte Moment-Ertrag-Relation (*moment-return relation*) eines beliebigen Assets oder Wertpapiers, und ferner ein Portfolio mit gegebenen Verteilungen des Investitionskapitals und der stochastischen Momente, jeweils auf einzelne Subportfolios.

3.3.1 Die Moment-Ertrag-Relation

In der Finanzmathematik wird angenommen, dass die zeitliche Entwicklung des Ertrags eines Wertpapiers (einer Aktie, einer Anleihe, aber auch eines Derivats wie einer Option) auf einem so genannten *stochastischen Prozess* X_τ [2, 39] beruht. (Üblicherweise ist X_τ eine Brownsche Bewegung, und der Wertpapierkurs ein Ito-Prozess [12, 36, 44, 60].) Unter der milden Bedingung, dass für diesen Prozess ein stochastisches Integral überhaupt nur definiert ist, kann gezeigt werden [52], dass der zu erwartende Ertrag R des in das Wertpapier investierten Kapitals W über den Zeithorizont t durch die *Moment-Ertrag-Relation*

$$R = \frac{W}{t} \ln s, \quad \text{mit } s = s(t) = 1 + \mu t + \sigma x(t) \quad (50)$$

gegeben ist. Hierbei ist μ die erwartete Rendite, σ die Volatilität des Wertpapierkurses und $x(t)$ das stochastische Integral („nulltes Moment“) des Prozesses. Die Funktion s ist die Momentfunktion des Wertpapiers. Interessant ist folgende elementar beweisbare Folgerung: Die Ertragsfunktion genügt dem Gesetz des sinkenden Grenzertrags sowohl bei steigendem erwarteten Ertrag μ als auch bei steigendem Risiko σ .

3.3.2 Zusammengesetzte Portfolios

Betrachten wir nun ein Portfolio, das aus n Subportfolios besteht, in denen jeweils der Anteil p_i des zur Verfügung stehenden Gesamtkapitals investiert wird (also $p_i \geq 0$ und $\sum_i p_i = 1$). Bezeichnen wir ferner die Momentfunktion des i -ten Subportfolios mit s_i , und die Summe der Momentfunktionen des gesamten Portfolios mit \tilde{s} . Die Verteilung der Momentfunktionen über die Subportfolios ist dann darstellbar durch den Vektor $q = (q_1, \dots, q_n)$, wobei für jedes i gilt:

$$q_i = \frac{s_i}{\tilde{s}} \quad \text{mit } \tilde{s} = \sum_{j=1}^n s_j. \quad (51)$$

Es lässt sich zeigen, dass der zu erwartende Ertrag des gesamten Portfolios mit den Informationen der beiden Verteilungen zusammenhängt, und zwar durch die Formel [52]:

$$r = \frac{1}{t} [\ln \tilde{s} - H(p) - K(p; q)]. \quad (52)$$

Zudem kann man zeigen, wie der zu erwartende Ertrag bei verschiedenen Randbedingungen maximiert wird:

- Gibt es keine Restriktionen, so wird der erwartete Ertrag r maximal, wenn das Investitionskapital und die Momente gleichverteilt sind. Dieses Ergebnis, also die Optimalität der Gleichverteilung, stimmt vollständig mit dem Resultat der klassischen Portfoliotheorie überein, dass ein risikoaverser Investor sein Portfolio auf möglichst viele Risikopositionen diversifiziert („*Don't put all your eggs in one basket*“) [42, 44, 60]. Der maximale zu erwartende Ertrag ist gegeben durch die Formel

$$r_* = \frac{1}{t} \ln[\tilde{s} - H(p)] = \frac{1}{t} \ln \frac{\tilde{s}}{n}. \quad (53)$$

- Sei $\tilde{s} = \text{const}$ gegeben („Gesamtmoment-Bedingung“). Dann ist der erwartete Ertrag r maximal, wenn die beiden Wahrscheinlichkeitsdichten p und q gleich sind: Geometrisch bedeutet das, dass der „Investmentvektor“ (p_1, \dots, p_n) die gleiche Richtung hat wie der „Momentenvektor“ (s_1, \dots, s_n) .

Ein weiterer Aspekt ergibt sich für den Fall unvollständiger Information. Oft wird die Verteilung der Momente nur unvollständig bekannt sein, sei es durch prinzipielles Unwissen oder bewusstes Zurückhalten von Information der Manager der Subportfolios. Man sieht sofort, dass in diesem Fall die Wahl der optimalen Kapitalverteilung misslingt, denn die Informationsdifferenz drückt den Gewinn. Der Ertrag wird also nicht maximal.

Ferner kann man nun Information monetär bewerten. Denn mit Gleichung (52) lautet der maximale zu erwartende Ertrag $r_* = \ln \tilde{s} - H(p)$. Ändert sich nun die Risikoverteilung, beispielsweise durch Markt- oder Portfolioveränderungen, so ist $p \neq q$,

und der Ertrag r ist durch (52) gegeben. Bei einer Änderung des Risikos entspricht die Informationsdifferenz damit

$$r_* - r = \frac{K(p; q)}{t}. \quad (54)$$

Anders ausgedrückt lautet diese Gleichung:

$$\text{Renditeverlust} = \frac{\text{Informationsdifferenz}}{\text{Zeit}}. \quad (55)$$

Hierbei bemisst sich der Renditeverlust aus der Differenz der optimalen zur tatsächlich erwarteten Rendite, und die Informationsdifferenz der beiden Entropien ist mit Gleichung (54) durch die Kullback-Leibler-Information (3) gegeben. Information kostet also, nämlich pro Zeiteinheit den Preis der relativen Ertragsminderung.

3.4 Information in der Biologie

Leben. Das Leben ist auf unserer Erde vor drei bis vier Milliarden Jahren entstanden. Was jedoch ist Leben? Interessanterweise lässt sich ähnlich wie für Information bislang keine physikalisch-chemische Definition geben. Möglich ist jedoch, strukturelle und dynamische Merkmale des Lebendigen zu beschreiben.

Allem Lebendigen gemeinsam sind die Notwendigkeit des Stoffwechsels (*Metabolismus*), die Fähigkeit zur Vermehrung (*Reproduktion*) und die Möglichkeit der Veränderung des Erbguts (*Mutationsfähigkeit*). Leben ist immer mit Individualisierung verbunden, das heißt, lebende Organismen sind immer von der Umwelt gut abgegrenzte Gebilde, deren kleinste Einheit die Zelle ist. Die hohe Komplexität auch primitivster Organismen ist mit einem hohen Ordnungsgrad verbunden, dessen Aufrechterhaltung nur möglich ist, wenn der Organismus als energetisch *offenes System* in ständigem Stoff- und Energieaustausch (*Fließgleichgewicht*) mit seiner Umgebung steht. Hierbei bleiben die Systemeigenschaften und damit auch die Individualität durch ein komplexes Regelsystem (*Homöostase*) weitgehend erhalten. Die Funktionsträger für alle Stoffwechselprozesse sind energiereiche Makromoleküle (Proteine und Ribonukleinsäuren, aber auch Polysaccharide, Lipide, Phosphate), die z.B. als Enzyme die Stoffwechselreaktionen ermöglichen.

Die Fortpflanzungsfähigkeit des lebenden Organismus beruht auf der Fähigkeit bestimmter Makromoleküle, der Nukleinsäuren, zur identischen Verdoppelung. Ihre Funktion als Träger des genetischen Materials sichert den Informationsfluss über Generationen hinweg, die Bedingung für den Fortbestand des Lebens über das Individuum hinaus. Letztlich an die Nukleinsäuren gebunden ist auch die Veränderlichkeit des Erbguts, einmal durch Mutation, zum andern bei der sexuellen Fortpflanzung durch Neukombination. Charakteristisch für alle lebenden Organismen ist weiters die Fähigkeit zu Wachstum und Differenzierung, wobei die zugrunde liegenden Prozesse durch die vorgegebene genetische Information gesteuert werden.

Evolution der Informationsverarbeitung. Nach einem Gedanken von Ebeling et al. [13, §2.3] kann man die Entwicklung des Lebens als eine Entwicklung der Informationsverarbeitung betrachten. Es werden entsprechend der Evolution des Lebens drei Stufen identifiziert, die chemische, die biologische und die soziale Phase.

Lebewesen zeichnet aus, dass sie physikalische Systeme sind, die Entropie exportieren, also insbesondere offene Systeme im Fließgleichgewicht sind [16]. Wie konnte es zu solch komplexen Systemen kommen? Wie ist das Leben entstanden, wo es doch „in Wahrheit nichts gibt als Atome und Leere“, wie Demokrit zu Recht vor 2500 Jahren feststellte?

Die Erde wurde zunächst durch die *chemische Evolution* geprägt, unter Bedingungen fern vom thermodynamischen Gleichgewicht. Der ständige Zufluss von Energie in Form von Sonnenlicht, starke Gradienten der chemischen Konzentrationen und der Temperatur, sowie Diffusion und turbulente Prozesse führten zu vielen parallel ablaufenden chemischen Prozessen. Es entstand eine eindrucksvolle, wenngleich karge leblose Welt mit Meeren, Atmosphäre und Gesteinsformationen — alles aufgebaut aus einfachen anorganischen Verbindungen mit weniger als 15 Atomen.

Am Ende dieser Phase entstanden vor etwa 3 bis 4 Milliarden Jahren *Enzyme*¹² [11], biochemische Katalysatoren, die durch Herabsetzen der zu überwindenden Aktivierungsenergie [8, §5.5.8] spontane Reaktionen eines Substrats bei Normaltemperatur und Normaldruck um das Milliardenfache beschleunigen. Das Enzym geht mit den umzusetzenden Stoffen eine vorübergehende Verbindung (*Enzym-Substrat-Komplex*) ein, wird aber durch die Reaktion am Ende nicht verändert. Die meisten Enzyme sind ihrer chemischen Natur nach Proteine, doch auch Ribonukleinsäuren (RNA) können als *Ribozyme* katalytisch wirksam sein. In den Frühzeiten der chemisch-biologischen Evolution waren einfach gebaute RNA-Moleküle, die den heutigen Polymerasen ähnelten, offenbar die einzigen Biokatalysatoren.

Ohne Enzyme könnte kein irdisches Leben existieren. Geschwindigkeit und Umfang spontaner Reaktionen würden nicht ausreichen, um die etwa 36 Millionen Stoffwechselprozesse pro Minute ablaufen zu lassen. Die Enzymaktivität ist am höchsten etwa bei Körpertemperatur, bei Hitze oder Kälte kommen die Reaktionen zum Stillstand. Diese Hitzeempfindlichkeit enzymkatalysierter Reaktionen ist der eigentliche Grund für die Temperaturabhängigkeit des Lebens [11, S. 70].

Enzyme hoben die Evolution auf der Erde auf eine neue qualitative Stufe. Trieben bis dahin zwar starke, aber ungerichtete Kräfte die unbelebte Materie über thermische, elektromagnetische oder Gravitationsenergie an, so konnten mit Enzymen plötzlich neue und komplexe Strukturen in kurzer Zeit aufgebaut werden, bis hin zu Zellen aus Tausenden von Atomen. „Enzyme haben das Leben aus nichts als Atomen und Leere geschaffen.“ [11, S. 15] So bewirkten sie den abrupten Phasenübergang zur *biologischen Evolution*.

Aus zweierlei Gründen kann man diesen Phasenübergang auch als erste Stufe einer

¹²griech. *ενζυμο* (*enzymo*) – „in Hefe“. Historisch erklärte die erste Entdeckung eines Enzyms (damals „Ferment“), der *Zymase* durch die Brüder Hans und Eduard Buchner 1897, den Gärungsprozess.

Evolution der Informationsverarbeitung sehen. Einerseits tritt mit der Zelle erstmals das Konzept des *Individuums* mit einer Schnittstelle zur Umwelt, zunächst Membranen, später Zellwände und Häute, die den selektiven Austausch mit der Umgebung, zunächst Stoffwechsel, ermöglichen und die Basis für Informationsaustausch legen. Andererseits wird in der DNA offenbar erstmals in der (irdischen) Natur *Information gespeichert und weiter gereicht*, vor allem die Information zur Konstruktion von Enzymen. Die Kombination und Verkettung von nur vier Aminosäuren ergeben den Bauplan für ein komplexes System, können also als durch Symbole codierte Information betrachtet werden.

Auf diese Weise konnten Zellsysteme immer größer werden, einzelne Zellen sich auf gewisse Arbeiten spezialisieren und so mehrzellige Organismen wie Mikroorganismen und Pflanzen bilden. Solche Organismen haben stets Sensorzellen mit speziellen Fähigkeiten für die Signalverarbeitung. Daraus entwickelten sich später bei Tieren die Neuronen, die miteinander vernetzt das Gehirn bildeten.

Spätestens mit dem Zusammenleben individueller Tiere in Rudeln oder Herden entstand die *soziale Evolution*. Mit ihr entsteht die *Kommunikation* zwischen Individuen, die spezielle vereinbarte Signale oder Laute verwendet und so dem Informationsaustausch komplexer Systeme dient. Auf diese Weise bildeten sich insbesondere beim modernen Menschen schließlich Sprachen. Mit dem bewussten Gedächtnis ergibt sich eine dem genetischen Code parallele Möglichkeit, Information über das individuelle Leben hinaus zu speichern und weiter zu geben, in Form von Mythen, Sagen oder Höhlenmalereien. *Kultur* entstand. Durch die Erfindung der Schrift wurde die Informationsspeicherung effizienter, auch die Menge an verarbeitbarer Information wurde größer, es konnten kompliziertere Zusammenhänge gedacht und festgehalten werden.

Bei mindestens einer Spezies dieser Erde entstand individuelles *Bewusstsein*. Auffallend aus Sicht der Informationsverarbeitung ist, dass von den etwa 11 MBit/s Informationsfluss der Sinneswahrnehmung weniger als 40 Bit/s vom menschlichen Bewusstsein verarbeitet werden [38, S. 191].

Weitere wichtige Meilensteine aus Sicht der Informationsverarbeitung waren die Erfindung des Buchdrucks vor einem halben Jahrtausend, Telegrafie und Telekommunikation in den 1830er Jahren, und die Entstehung des Internets Ende der 1960er Jahre. Durch jede dieser Erfindungen konnte die *Verbreitung* von Information jeweils dramatisch beschleunigt werden, während die jeweiligen Medien, insbesondere Bücher und Computer, die *Speicherkapazität* erhöhten. Vom Standpunkt der Informationsverarbeitung können wir also sechs Evolutionsstufen erkennen:

Durch die Brille der Informationsverarbeitung erhält damit die Geschichte des irdischen Lebens eine auffallende Schlüssigkeit, so dass man versucht ist zu fragen: Ist nicht vielleicht die ständige Verbesserung der Verarbeitung von Information die Antriebsfeder für Leben?

System	Zeichen	Medium
Internet	Algorithmus	Computer
Kultur	Schrift	Buch
soz. Gruppe	Sprache	Mensch
Gehirn	el. Impuls	Neuron
Organismus	DNA	Zelle
Stoffwechsel	freie Enthalpie	Enzym

Abbildung 8: Die Evolution der Informationsverarbeitung auf der Erde

4 Erste Vereinheitlichung: Rényi

4.1 Axiomatischer Zugang

Eine wesentliche Aufgabe der Informationstheorie besteht darin, Maße für den syntaktischen Informationsgehalt einer Nachricht zu definieren. Eine *Nachricht* ist hierbei eine Folge von n Buchstaben c_1, \dots, c_n eines Alphabets Σ , die von einer Quelle gesendet wird. Die Wirkung von Information besteht dann in einer Verringerung der Unsicherheit des Empfängers der Nachricht. Die Mutmaßungen oder Erwartungen sind dabei verknüpft mit Wahrscheinlichkeiten, welche die „Erfahrung“, also die statistische Vorkenntnis reflektieren. Nach Empfang der Nachricht hat der Empfänger Information gewonnen.

Das Prinzip jeder Wette, etwa beim Pferderennen, basiert auf dieser Sicht auf Information. Der Informationsgehalt ist hoch, wenn ein nichtwahrscheinliches Ereignis eintritt, und niedrig, wenn ein sehr wahrscheinliches Ereignis eintritt. Entsprechend wird ein Außenseitersieg bei einer Wette hoch belohnt, ein Favoritensieg dagegen niedrig.

Ein Informationsmaß soll aber nicht nur eine einzelne Nachricht bewerten, sondern die Quelle insgesamt. Es wird also ein Funktional der Wahrscheinlichkeitsverteilung aller Nachrichten sein.

Ein axiomatischer Zugang zu einem Informationsmaß wurde in den 1940er und 1950er Jahren von verschiedenen Theoretikern vorgenommen, insbesondere Shannon, Khinchin, Faddejew und Rényi [13, §3.3]. Das übliche Axiomensystem für ein endliches Informationsmaß H lautet wie folgt. Sei $A = \{a_1, \dots, a_n\}$ eine endliche Menge von Elementarereignissen, und seien

$$\Delta_n = \{\mathbf{p} \in [0, 1]^n : \sum_{i=1}^n p_i = 1\}, \quad \Delta = \bigcup_{n=1}^{\infty} \Delta_n. \quad (56)$$

(Δ_n ist eine n -dimensionale Hyperfläche im ersten „Quadranten“ des \mathbb{R}^n .) Dann ist jeder Vektor $\mathbf{p} \in \Delta_n$ eine Wahrscheinlichkeitsverteilung¹³ auf A mit $P(\{a_i\}) = p_i$. Sei dann $H = H[A] : \Delta \rightarrow [0, 1]$ eine reellwertige Funktion, für die folgende Axiome gelten:

¹³D.h. $(A, \mathcal{P}(A), P)$ ist ein Wahrscheinlichkeitsraum [2, §1].

1. H ist stetig. Damit bewirkt eine kleine Änderung der Wahrscheinlichkeiten p_i nur eine kleine Änderung des Wertes von H .
2. Für die Gleichverteilung ist H maximal. D.h. für ein beliebiges $\mathbf{p} \in \Delta_n$ gilt

$$H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \geq H(p_1, \dots, p_n). \quad (57)$$

3. Hinzufügen eines unmöglichen Ereignisses ändert nicht den Wert von H , also $H(p_1, \dots, p_n, 0) = H(p_1, \dots, p_n)$.
4. Für unabhängige Ereignisräume genügt H der Produktregel. Ist AB der Produktraum zweier unabhängiger Ereignisräume A und B , so gilt

$$H[AB] = H[A] + H[B]. \quad (58)$$

Explizit besagt diese Gleichung, dass, wenn A aus m Elementarereignissen und B aus n Elementarereignissen besteht, gilt

$$H(p_{11}^{AB}, \dots, p_{mn}^{AB}) = H(p_1^A, \dots, p_m^A) + H(p_1^B, \dots, p_n^B) \quad (59)$$

Hierbei bedeutet die Produktwahrscheinlichkeit p_{ij}^{AB} die Wahrscheinlichkeit, im Ereignisraum A das Elementarereignis a_i und im Ereignisraum B das Elementarereignis b_j zu erhalten.

Axiom 4 impliziert die Additivität der Information.

Folgerungen. Eine Klasse von Funktionen, die die Axiome erfüllen, hängt von einem reellen Parameter $\alpha > 0$ ab und ergibt sich [13, §3.8] als

$$H_\alpha(p_1, \dots, p_n) = \frac{k}{1-\alpha} \ln \left(\sum_{i=1}^n p_i^\alpha \right), \quad (60)$$

wo $k = 1/\ln s$ und $s = |\Sigma| \geq 2$ den Umfang des zugrunde liegenden Alphabets Σ bedeuten. (Für das binäre Alphabet $\Sigma = \{0, 1\}$ beispielsweise ist $s = 2$, und daher die Einheit für die Information das Bit.) H_α heißt *Rényi-Information* der Ordnung α , [41, Anhang (45c)]. Sie ist eine nichtnegative, monoton fallende Funktion von α , d.h.

$$H_\alpha(\mathbf{p}) \geq H_\beta(\mathbf{p}) \geq 0 \quad \text{für } \alpha < \beta \quad (61)$$

($\alpha, \beta > 0$) bei gegebenem Wahrscheinlichkeitsvektor \mathbf{p} . Ferner ist $H_0(\mathbf{p}) = k \ln n$ für $\mathbf{p} \in \Delta_n$. Insbesondere mit (61) gilt also $0 \leq H_\alpha(\mathbf{p}) \leq k \ln n$ für $\alpha > 0$. Im Grenzübergang $\alpha \rightarrow \infty$ ergibt $H_\infty(\mathbf{p}) = -k \ln p_{\max}$.

Sehr wichtig sind die zwei speziellen Werte $\alpha = 1$ und 2 , denn H_1 ist die Shannon'sche Information und H_2 ist eng verknüpft mit der in der Quantentheorie diskutierten Brukner-Zeilinger-Information, wie im Folgenden dargelegt wird.

4.2 Shannonsche Information

Durch Grenzwertbetrachtung sieht man, dass für $\alpha \rightarrow 1$ die Rényi-Information der Shannonschen Entropie gleicht,

$$H_1(p_1, \dots, p_n) = -k \sum_{i=1}^n p_i \ln p_i. \quad (62)$$

Interessanterweise war bei der historisch ersten Axiomatisierung des Informationsmaßes von Shannon 1948 das letzte Axiom wesentlich schärfer [43]. Es lautete wie folgt.

4'. H genügt der Produktregel („Rekursionspostulat“). Ist AB der Produktraum zweier Ereignisräume A und B , so gilt

$$H[AB] = H[A] + H[B|A] = H[B] + H[A|B] \quad (63)$$

Explizit heißt die erste Gleichung, wenn A aus m Elementarereignissen besteht B aus n Elementarereignissen, dass

$$H(p_{11}^{AB}, \dots, p_{mn}^{AB}) = H(p_1^A, \dots, p_m^A) + \sum_{i=1}^m p_i^A H\left(\frac{p_{i1}^{AB}}{p_i^A}, \dots, \frac{p_{in}^{AB}}{p_i^A}\right) \quad (64)$$

Der Quotient $p_{ij}^{A|B} = p_{ij}^{AB} / p_i^A$ ist die bedingte Wahrscheinlichkeit, das Ereignis b_j zu erhalten unter der Voraussetzung, dass a_i bereits eingetreten ist.

Da $\frac{p_{ij}^{AB}}{p_i^A} = p_i^A p_j^B$ für unabhängige Ereignisräume, sieht man sofort, dass Axiom 4 eine Konsequenz von Axiom 4' ist, aber die Umkehrung gilt nicht.

Um die Bedeutung des vierten Axioms verständlich zu machen, stellen wir kurz folgende Überlegungen an. Beim Würfeln gibt es sechs Elementarereignisse. Sei $A = \{\{2, 4, 6\}, \{1, 3, 5\}\}$ der Ereignisraum „gerade oder ungerade Augenzahl“, und $B = \{\{1\}, \{3, 5\}\}$. Dann sind die Ereignisse von $AB = \{\{1\}, \{3, 5\}, \{2,4,6\}\}$ (Abb. 9). Demnach besagt Axiom 4, dass die Information unabhängig davon ist, ob man erst die

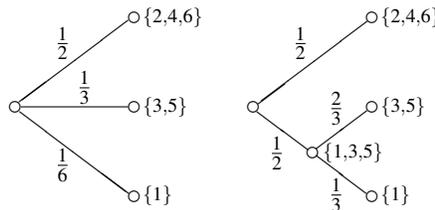


Abbildung 9: $H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(0) + \frac{1}{2}H(\frac{2}{3}, \frac{1}{3})$.

Frage nach ungerader Augenzahl stellt und danach die Zugehörigkeit zu einer der zwei Zahlkategorien $\{1\}$ oder $\{3, 5\}$, oder ob man in einem Schritt nach den Zahlkategorien von AB fragt.

Folgerungen. Für den Fall, in dem das System A das System B eindeutig definiert, gilt $H[B|A] = 0$, denn das Ergebnis aus B ist sicher, d.h. es gibt keinen Informationsgewinn durch B . Damit ist $H[AB] = H[A]$. Sind im anderen Extremfall die Systeme A und B unabhängig voneinander, dann ist $H[B|A] = H[B]$ und $H[A|B] = H[A]$, d.h.

$$H[AB] = H[A] + H[B]. \quad (65)$$

Im allgemeinen Fall gibt es mehr oder weniger starke Abhängigkeiten zwischen beiden Fällen, also $0 \leq H[B|A] \leq H[B]$ (und entsprechend A und B vertauscht), und damit

$$\max(H[A], H[B]) \leq H[AB] \leq H[A] + H[B]. \quad (66)$$

Es kann nun gezeigt werden [43], dass die Axiome 1, 2 und 3 und 4' die Funktion H_1 , also die Shannonsche Entropiefunktion, implizieren.

4.3 Brukner-Zeilinger-Information

Für $\alpha = 2$ lautet die Rényi-Information $H_2(\mathbf{p}) = -k \ln(\sum p_i^2)$, d.h. für die Funktion

$$Z(\mathbf{p}) = e^{-H_2(\mathbf{p})} - \frac{1}{n} \quad (67)$$

gilt $Z(\mathbf{p}) = \sum_{i=1}^n (p_i - \frac{1}{n})^2$, wegen $\sum_1^n (p_i - \frac{1}{n})^2 = \sum_1^n p_i^2 - \frac{1}{n}$. Diese Funktion $Z(\mathbf{p})$ ist die Brukner-Zeilingersche Zustandssumme [7]. Sie wurde ad-hoc von Brukner und Zeilinger als Informationsmaß für Quantensysteme eingeführt, da die Shannonsche Information und ihr quantenmechanisches Pendant, die von-Neumannsche Entropie, den Informationsgewinn bei Messungen von Quantensystemen, insbesondere mit nicht-kommutierenden Observablen, nicht wiedergibt. Die Ursache ist darin begründet, dass das Shannonsche Maß nicht invariant unter unitären Transformationen des Quantensystems ist. Diese Eigenschaft wiederum ist Konsequenz des Axioms 4', des „Rekursionspostulats“ oder auch „grouping axiom“. Brukner und Zeilinger argumentieren, dass es im Zusammenhang mit Messungen von Quantensystemen fallen gelassen werden muss.

Im Sinne der Axiome 1 bis 4 jedoch ist nicht die Funktion (67) als Informationsmaß zu betrachten, sondern H_2 , also

$$H_2(\mathbf{p}) = -k \ln \left(Z(\mathbf{p}) + \frac{1}{n} \right). \quad (68)$$

Daher sollte H_2 als die *Brukner-Zeilinger-Information* definiert werden. Mit (61) gilt für eine gegebene Wahrscheinlichkeitsverteilung $\mathbf{p} \in \Delta_n$ stets

$$H_2(\mathbf{p}) \leq H_1(\mathbf{p}). \quad (69)$$

Die Brukner-Zeilinger-Information ist also nie größer als die Shannon-Information, d.h. man kann aus einer Messung eines Quantensystems weniger Information erlangen als man in einem entsprechendem klassischen System erwarten würde.

Die Größe $r = e^{-H_2(\mathbf{p})} = \sum p_i^2$ ist bei quantenmechanischen Systemen als *Reinheit* (*purity*) bekannt [64], da r maximal ist für einen reinen Quantenzustand, d.h. $p_i = 1$ für ein i , und somit $r = 1$.

5 Die Relativität der Information

Eine verschlüsselte Nachricht ist für jemanden, der den Schlüssel nicht kennt, lediglich eine chaotische Aneinanderreihung von Zeichen. Die Information hat *für ihn* keinen Wert. Oder betrachten Sie die folgende Funktion $f: \mathbb{R} \rightarrow \mathbb{R}$,

$$f(x) = \sum_{n=0}^{\infty} \frac{(-1)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots \quad (70)$$

Eine wichtige Frage mag sein, welche Periode die Funktion hat. Sollten Sie diese unendliche Reihe das erste Mal sehen, so wird die Information „ $f(x)$ hat die Periode 2π “ für Sie neu und daher wertvoll sein. (Es gilt übrigens $f(x) = \cos x$.)

5.1 Relative Wahrscheinlichkeitsverteilungen

Satz 5.1 (Bayes 1763) Für zwei Ereignisse A und B mit $P(B) > 0$ gilt

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A).$$

Hierbei ist $P(A)$ die A-Priori-Wahrscheinlichkeit für das Ereignis A und $P(B|A)$ die Wahrscheinlichkeit für ein Ereignis B unter der Bedingung, dass A auftritt, die so genannte *bedingte Wahrscheinlichkeit*. Sie erfüllt die Beziehung

$$P(A \cap B) = P(A)P(B|A). \quad (71)$$

Die Korrektheit des Bayesschen Satzes folgt unmittelbar aus der Definition der bedingten Wahrscheinlichkeit [32, §2.2].

Der Satz von Bayes erlaubt gewissermaßen die Umkehrung von Schlussfolgerungen: Die Berechnung von $P(\text{Ereignis} \mid \text{Ursache})$ ist häufig einfach, aber oft ist eigentlich $P(\text{Ursache} \mid \text{Ereignis})$ gesucht, also ein Vertauschen der Argumente. Für das Verständnis können der Entscheidungsbaum und die A-Priori-Wahrscheinlichkeit helfen.

Beispiel 5.2 (Die Informationen der Eltern) Ein Vater behauptet: „Ich habe zwei Kinder, eines ist ein Junge.“ Mit welcher Wahrscheinlichkeit ist das andere Kind ein Junge, wenn die Wahrscheinlichkeit für die Geburt von Jungen und Mädchen als gleich angenommen wird?

Sei A das Ereignis, dass beide Kinder Jungen sind, und sei B das Ereignis, dass eines der Kinder ein Junge ist. Notieren wir zunächst die vier gleichwahrscheinlichen Möglichkeiten JJ, JM, MJ und MM, wobei J=„Junge“ und M=„Mädchen“ bedeuten soll, sowie XY = „X Erstgeborenes, Y Zweitgeborenes“. Dann ist $P(A) = \frac{1}{4}$, denn unter den vier Möglichkeiten ist nur eine (JJ) mit zwei Jungen. Entsprechend ist $P(B) = \frac{3}{4}$, denn drei Möglichkeiten (JJ, JM, MJ) gibt es für Ereignis B . Außerdem ist $P(B|A) = 1$,

denn wenn beide Kinder Jungen sind, ist das Ereignis B sicher. Damit ergibt der Satz von Bayes

$$P(A|B) = \frac{4}{3}P(A) = \frac{4}{3} \cdot \frac{1}{4} = \frac{1}{3}.$$

Das zweite Kind ist also nur mit der Wahrscheinlichkeit $\frac{1}{3}$ ein Junge.¹⁴

Nehmen wir nun an, die Mutter sei etwas beredsamer und behauptete: „Ich habe zwei Kinder. Das ältere der beiden ist ein Junge.“ Mit welcher Wahrscheinlichkeit ist nun das zweite Kind ein Junge? Mit der zusätzlichen Information ist es jetzt nicht mehr das Ereignis A von oben, dessen Wahrscheinlichkeit wir suchen, sondern das Ereignis A_2 , dass das zweite Kind ein Junge ist. Dessen A-Priori-Wahrscheinlichkeit ist $P(A_2) = \frac{1}{2}$, denn von den vier Möglichkeiten JJ, JM, MJ, MM sind nur zwei günstig für A_2 . Entsprechend wissen wir durch die Information der Mutter, dass nicht das Ereignis B von oben eingetreten ist, sondern das Ereignis B_1 „das Erstgeborene ist ein Junge“. Dessen Wahrscheinlichkeit ist $P(B_1) = \frac{1}{2}$ (wegen JJ, JM), und $P(B_1|A_2) = \frac{1}{2}$, denn wenn A_2 eingetreten ist, gibt es nur zwei Möglichkeiten JJ und MJ, von denen nur eine günstig für B_1 ist. Daher gilt

$$P(A_2|B_1) = P(A_2) = \frac{1}{2}.$$

Das zweite Kind ist also mit der Wahrscheinlichkeit $\frac{1}{2}$ ein Junge.

Die Entropie der Verteilung $\mathbf{p} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ für die vier Ereignisse JJ, JM, MJ, MM ist $H = \log_2 4 = 2$ Bit. Der Informationsgehalt der ersten Aussage ist $-\log_2 P(B) = \log_2 \frac{4}{3} \approx 0,415$ Bit, der Informationsgehalt der zweiten $-\log_2 P(B_1) = 1$ Bit. \square

Beispiel 5.3 (*Das Ziegenproblem*) [40] In einer TV-Show zeigt der Showmaster einem Kandidaten drei verschlossene Türen. Hinter einer der Türen befindet sich ein Auto, hinter den beiden anderen jeweils eine Ziege. Errät der Kandidat die Tür mit dem Auto, so hat er es gewonnen. Er wählt eine der Türen aus, sie wird aber noch nicht geöffnet. Der Showmaster, der weiß, wo sich das Auto befindet, öffnet daraufhin eine der beiden anderen Türen, hinter der eine Ziege ist. Der Kandidat hat nun die Wahl, bei seiner Tür zu bleiben oder zur anderen zu wechseln. Was sollte er tun?

Die für viele verblüffende Lösung¹⁵ ergibt sich einfach aus der Anwendung des Bayesschen Satzes. Nennen wir

- A das Ereignis, dass der Kandidat die Tür mit dem Auto gewählt hat, und
- B das Ereignis, dass der Showmaster eine Tür mit einer Ziege öffnet.

¹⁴ Man kommt natürlich auch ohne den Satz von Bayes auf das Ergebnis, denn die Überlegungen zur Ermittlung der bedingten Wahrscheinlichkeit $P(B|A)$ gelten auch für $P(A|B)$: Ist Ereignis B eingetreten, so haben wir einen der drei Fälle JJ, JM und MJ, wovon nur JJ günstig für A ist.

¹⁵ Das Ziegenproblem, im englischen Sprachraum als *Monty Hall Problem* bekannt, schlug Anfang der 1990er Jahre hohe Wellen. In ihrer regelmäßigen Kolumne des US-amerikanischen Magazins *Skeptical Inquirer* stellte die amerikanische Autorin Marilyn vos Savant es 1990 vor und löste damit eine weltweite Debatte aus, siehe <http://www.zeit.de/2004/48/N-Ziegenproblem>.

Die A-Priori-Wahrscheinlichkeiten sind $P(A) = \frac{1}{3}$ und $P(B) = 1$, denn der Kandidat wählt zufällig eine Tür von dreien, der Showmaster *weiß* aber, wo sich eine Ziege befindet. Daher gilt auch für die bedingte Wahrscheinlichkeit $P(B|A) = 1$, denn der Showmaster weiß immer noch eine Tür mit einer Ziege, auch wenn der Kandidat richtig geraten hat. Mit dem Bayesschen Satz gilt dann einfach

$$P(A|B) = P(A) = \frac{1}{3}.$$

Entsprechend ist $P(\bar{A}|B) = \frac{2}{3}$, d.h., mit der Wahrscheinlichkeit von $\frac{2}{3}$ liegt der Kandidat falsch, er sollte also wechseln!

Die Entropie der Wahrscheinlichkeitsverteilung $\mathbf{p} = (P(A), P(\bar{A})) = (\frac{1}{3}, \frac{2}{3})$ ist

$$H = \frac{1}{3} \log_2 3 + \frac{2}{3} \log_2 \frac{3}{2} = \log_2 3 - \frac{2}{3} \approx 0,919 \text{ Bits,}$$

der Informationsgehalt des Ereignisses B für den Kandidaten ist $\log_2 \frac{2}{3} \approx 0,585$, denn mit der Wahrscheinlichkeit $P(\bar{A}|B) = \frac{2}{3}$ offenbart es die verbliebene Tür als diejenige mit dem Gewinn. \square

Beispiel 5.4 (*Das Ziegenproblem mit einem unwissenden Showmaster*) Nehmen wir die gleiche Situation wie im obigen Ziegenproblem, nur dass jetzt auch der Showmaster *nicht* weiß, wo sich das Auto befindet, sondern dass er einfach eine der beiden verbliebenen Türen aussucht, aber nicht öffnet. Sollte der Kandidat wechseln?

Sei wieder A das Ereignis, dass der Kandidat die Tür mit dem Auto gewählt hat, und B das Ereignis, dass der Showmaster eine Ziegentür wählt. Wieder ist $P(A) = \frac{1}{3}$, aber nun ist $P(B) = \frac{2}{3}$, denn auch der Showmaster rät, und $P(B|A) = 1$, denn wenn der Kandidat richtig liegt, kann er nur noch Ziegentüren wählen. Das ergibt

$$P(A|B) = \frac{3}{2}P(A) = \frac{3}{2} \cdot \frac{1}{3} = \frac{1}{2}.$$

Seine Chancen stehen nun 50:50, dass er die Tür mit dem Auto gewählt hat, er kann also ebenso gut bei seiner Wahl bleiben wie wechseln. \square

Die obigen Beispiele zeigen den folgenden zentralen Punkt.

These 1 *Jede zusätzliche Information verändert die Wahrscheinlichkeitsverteilung.*

5.2 Wo ist die Objektivität der Information?

Information ist Entropie. Damit führt die Aussage, dass Information relativ ist, automatisch zu der gleichen Aussage über Entropie. Bedeutet das also, dass Entropie nicht objektiv sei?

Für eine Antwort müssen wir kurz ausholen und grundlegende Begriffe der Statistischen Mechanik betrachten. Ein *Mikrozustand* ist die vollständige Beschreibung der

individuellen elementaren Komponenten eines physikalischen Systems. Beispielsweise besteht ein Mikrozustand eines idealen Gases aus den genauen Orts- und Impulsdaten jedes einzelnen Gasteilchens. Ein *Makrozustand* dagegen ist die „vergrößerte“ (*coarse-grained*) Beschreibung des Systems anhand statistischer Größen, der *makroskopischen Variablen*. Für ein ideales Gas beispielsweise sind solche makroskopischen Größen die Temperatur (als Durchschnittswert der kinetischen Energie aller Teilchen), der Druck (als $\frac{2}{3}$ der mittleren kinetischen Energie mal Teilchendichte) oder das Volumen (als Aufenthaltsorte der Teilchen). So ist „ein ideales Gas der Temperatur $T = 293 \text{ K}$ “ der Makrozustand, der durch einen Mikrozustand mit dem entsprechenden makroskopischen Durchschnittswert „mittlere kinetische Energie = T “ gegeben ist. Meist gibt es mehrere Mikrozustände, die einen Makrozustand ergeben, jeder einzelne aber mit einer bestimmten Wahrscheinlichkeit. Wahrscheinlichkeitstheoretisch sind also die Mikrozustände die Elementarereignisse, während die Makrozustände den Teilmengen der σ -Algebra entsprechen.

Wie wird in der Statistischen Mechanik nun die Wahrscheinlichkeitsverteilung der Mikrozustände zu einem gegebenem Makrozustand bestimmt? Sie ist die eindeutige Lösung eines Optimierungsproblems, und maximiert die Entropie unter den zu Grunde liegenden physikalischen Nebenbedingungen [61, §5.17.3]. Das ist das auf Gibbs zurückgehende „Prinzip der maximalen Entropie“. Auf diese Weise kann man z.B. die Verteilungen verschiedener Quantengase (Bose-Einstein- oder Fermi-Dirac-Statistik) oder des klassischen idealen Gases (Maxwell-Boltzmann-Statistik) herleiten. Gibt es keine besonderen physikalischen Nebenbedingungen, so wird die Entropie maximiert durch die Gleichverteilung, d.h. alle möglichen Mikrozustände haben dieselbe Wahrscheinlichkeit. Diese Verteilung von Mikrozuständen wird in vielen Lehrbüchern der Statistischen Physik sogar als Postulat eingeführt, z.B. [14, §5].

These 2 *In einem physikalischen System ergeben sich die Wahrscheinlichkeiten aller möglichen Mikrozustände bei gegebenem Makrozustand durch das Prinzip der maximalen Entropie. Die Wahrscheinlichkeiten hängen also von den durch den Makrozustand festgelegten physikalischen Nebenbedingungen ab.*

Das rettet die Objektivität der Entropie. Entropie ist damit die Information über den exakten Mikrozustand, die man nicht hat, wenn man nur den Makrozustand kennt. Ändern sich die Nebenbedingungen, die man kennt, so ändern sich die Wahrscheinlichkeiten und damit auch die Entropie.

6 Diskussion

In den ersten beiden Abschnitten dieses Beitrags wurden verschiedene gängige Konzepte und mathematische Definitionen von Information vorgestellt, von denen die meisten sich als Spezialfall der Rényi-Information darstellen lassen, insbesondere eine Modifikation der kürzlich eingeführten und in der Quantenmechanik relevanten Brukner-Zeilingerschen Information. Ausnahmen sind jedoch die Fisher-Information

und die algorithmische Information, die ganz anderen Konzepten entspringen. Die Fisher-Information (11) wird definiert als eine differenzierbare Funktion der Likelihood eines Parameters und drückt die Information einer Stichprobe, z.B. von Messwerten, in Bezug auf ihn aus. Die algorithmische Information (8) bezieht sich stets auf einen Ausgabestring x einer universellen Turing-Maschine U und könnte als Grad der Überraschung, ihn durch ein zufällig eingespeistes Programm zu erhalten, also als Informationsgehalt im Shannonschen Sinne, betrachtet werden. Dennoch lässt sich nicht erkennen, wie algorithmische Information als ein Rényi-Maß aufgefasst werden könnte.

Im letzten Abschnitt wurde die Relativität von Information betrachtet. Anhand von Beispielen, insbesondere dem berühmten Ziegenproblem, wurde der direkte Einfluss von Information und Wissen auf gegebene Wahrscheinlichkeiten gezeigt. In Anerkennung der physikalischen Notwendigkeit eines objektiven Informations- und damit Entropiebegriffs wurde die Rolle von Makrozuständen sowie des Prinzips der maximalen Entropie zur Bestimmung der Wahrscheinlichkeitsverteilung der Mikrozustände eines physikalischen Systems verdeutlicht.

Der Weg zu einer mathematischen Definition von Information dauerte bislang über ein halbes Jahrhundert, und er ist noch nicht zu Ende. Auch wenn der semantische Aspekt über den Bayesschen Wahrscheinlichkeitsbegriff mathematisch formulierbar ist, so bleibt der Aspekt der *Algorithmisierung*, also der Komprimierung von Information beispielsweise in Gesetze und Formeln, mit den bisherigen Ansätzen unvereinbar, ebenso wie die Bewertung unbekannter Parameter über Likelihoods durch Stichproben, also die *indirekte Informationsgewinnung* durch Messungen. Intuitiv verwenden wir für dies alles das Wort Information. Sollte eine Vereinheitlichung der Information, vielleicht mit weiteren Untersuchungen zu ihrer Relativität, uns tatsächlich nicht möglich sein?

Literatur

- [1] BARNDORFF-NIELSEN, O. ; GILL, R. D. ; JUPP, P. E.: ‘Quantum Information’. In: ENGQUIST, B. (Hrsg.) ; SCHMID, W. (Hrsg.): *Mathematics Unlimited — 2001 and Beyond*. New York : Springer-Verlag, 2001, S. 83–107
- [2] BAUER, H. : *Wahrscheinlichkeitstheorie*. 4. Aufl. Berlin New York : Walter de Gruyter, 1991
- [3] BEKENSTEIN, J. D.: ‘Black holes and entropy’. In: *Phys. Rev. D* 7 (1973), Nr. 8, S. 2333–2346
- [4] BENNETT, C. H.: ‘The thermodynamics of computation — a review’. In: *Int. J. Theor. Phys.* 21 (1982), Nr. 12
- [5] BENNETT, C. H.: ‘Logical Depth and Physical Complexity’. In: HERKEN, R. (Hrsg.): *The Universal Turing Machine. A Half-Century Survey*. Wien : Springer-Verlag, 1994, S. 207–235
- [6] BRANDT, S. : *Datenanalyse*. 4. Aufl. Heidelberg Berlin : Spektrum Akademischer Verlag, 1999
- [7] BRUKNER Āaslav ; ZEILINGER, A. : ‘Conceptual inadequacy of the Shannon information in quantum mechanics’. In: *Phys. Rev. A* 63 (2001), S. 022113. – <http://arxiv.org/abs/quant-ph/0006087>
- [8] C. GERTHSEN, H. V.: *Physik*. 18. Edition. Berlin Heidelberg : Springer Verlag, 1995

- [9] CHAITIN, G. J.: ‘Incompleteness theorems for random reals’. In: *Advances in Applied Mathematics* 8 (1987), S. 119–146. – <http://www.cs.auckland.ac.nz/CDMTCS/chaitin/aam.pdf>
- [10] CHANDRASEKHAR, S. : *The Mathematical Theory of Black Holes*. Oxford : Oxford University Press, 1983
- [11] DRESSLER, D. ; POTTER, H. : *Katalysatoren des Lebens. Struktur und Wirkung von Enzymen*. Heidelberg Berlin New York : Spektrum Akademischer Verlag, 1992
- [12] DUFFIE, D. : *Dynamic Asset Pricing*. Princeton : Princeton University Press, 1996
- [13] EBELING, W. ; FREUND, J. ; SCHWEITZER, F. : *Komplexe Strukturen: Entropie und Information*. Stuttgart Leipzig : B.G. Teubner, 1998
- [14] FLIESSBACH, T. : *Statistische Physik. Lehrbuch zur theoretischen Physik IV*. 2. Aufl. Heidelberg Berlin Oxford : Spektrum Akademischer Verlag, 1995
- [15] GELL-MANN, M. : *The Quark and the Jaguar. Adventures in the Simple and the Complex*. New York : W.H. Freeman, 1994
- [16] GRANDY, W. T.: *Foundations of Statistical Mechanics. Volume II: Nonequilibrium Phenomena*. Dordrecht : D. Reidel, 1988
- [17] GREENE, B. : *Das elegante Universum. Superstrings, verborgene Dimensionen und die Suche nach der Weltformel*. Berlin : Siedler, 2000
- [18] GRUSKA, J. : ‘Quantum Computing Challenges’. In: ENGQUIST, B. (Hrsg.) ; SCHMID, W. (Hrsg.): *Mathematics Unlimited — 2001 and Beyond*. New York : Springer-Verlag, 2001, S. 529–563
- [19] HEUSER, H. : *Funktionalanalysis. Theorie und Anwendung*. Stuttgart : B.G. Teubner, 1986
- [20] HEUSLER, M. : ‘Stationary black holes: uniqueness and beyond’. In: *Living Rev. Relativity* 1 (1998), Nr. 6. – <http://www.livingreviews.org/lrr-1998-6>
- [21] HOROWITZ, G. T. ; MALDACENA, J. M. ; STROMINGER, A. : ‘Nonextremal black hole microstates and U-duality’. In: *Phys. Lett.* B383 (1996), S. 151–159. – <http://arxiv.org/abs/hep-th/9603109>
- [22] HULL, J. C.: *Options, Futures & Other Derivatives*. 4 Edition. Upper Saddle River, NJ : Prentice-Hall International, 2000
- [23] JAYNES, E. T.: ‘Information theory and statistical mechanics’. In: *Phys. Rev.* 106 (1957), S. 620–630
- [24] JAYNES, E. T.: ‘Information theory and Statistical Mechanics’. In: FORD, K. (Hrsg.): *1962 Brandeis Summer Institute in Theoretical Physics*. New York : W.A. Benjamin, 1963
- [25] JAYNES, E. T.: ‘Where do we stand on maximum entropy?’. In: LEVINE, R. (Hrsg.) ; TRIBUS, M. (Hrsg.): *The Maximum Entropy Formalism*. Cambridge, Ma : M.I.T Press, 1978
- [26] KAKU, M. : *Introduction to Superstrings and M-Theory*. 2nd Edition. New York : Springer-Verlag, 1999
- [27] KOLDE, K. : *Astronomie*. Frankfurt : Diesterweg / Otto Salle, 1973
- [28] LANDAUER, R. : ‘Irreversibility and heat generation in the computing process’. In: *IBM J. Research and Development* 3 (1961), S. 183–191
- [29] LEFF, H. S. ; REX, A. F.: *Maxwell’s Demon. Entropy, Information, Computing*. Princeton : Princeton University Press, 1990
- [30] LUBBE, J. C. A. d.: *Information Theory*. Cambridge : Cambridge University Press, 1997
- [31] LYRE, H. : *Quantentheorie der Information*. Wien New York : Springer-Verlag, 1998

- [32] MACKAY, D. : *Information Theory, Inference, and Learning Algorithms*. Cambridge : Cambridge University Press, 2003
- [33] MALDACENA, J. M.: ‘Black holes and D-branes’. In: *Nucl. Phys. B Proc. Suppl.* 61 (1998), S. 428–442. – <http://arxiv.org/abs/hep-th/9705078>
- [34] MERTENS, P. : *Lexikon der Wirtschaftsinformatik*. Berlin Heidelberg : Springer-Verlag, 1997
- [35] MOROWITZ, H. J.: *Foundations of Bioenergetics*. New York : Academic Press, 1978
- [36] NEFTCI, S. : *An Introduction to the Mathematics of Financial Derivatives*. San Diego : Academic Press, 1996
- [37] NIELSEN, M. A. ; CHUANG, I. L.: *Quantum Computation and Quantum Information*. Cambridge : Cambridge University Press, 2000
- [38] NØRRETRANDERS, T. : *Spüre die Welt. Die Wissenschaft des Bewusstseins*. Reinbek bei Hamburg : Rowohlt Verlag, 1997
- [39] ØKSENDAL, B. : *Stochastic Differential Equations. An Introduction with Applications*. 5th Edition. Berlin Heidelberg New York : Springer-Verlag, 1998
- [40] RANDOW, G. von: *Das Ziegenproblem - Denken in Wahrscheinlichkeiten*. Reinbek : Rowohlt, 1992
- [41] RÉNYI, A. : *Wahrscheinlichkeitstheorie. Mit einem Anhang über Informationstheorie*. Berlin : VEB Deutscher Verlag der Wissenschaften, 1977
- [42] SAMUELSON, P. A. ; NORDHAUS, W. D.: *Economics*. 15th Edition. New York etc. : McGraw-Hill, 1995
- [43] SHANNON, C. E. ; WEAVER, W. : *The Mathematical Theory of Communication*. Urbana : University of Illinois Press, 1949
- [44] STEINER, M. ; BRUNS, C. : *Wertpapiermanagement*. Stuttgart : Schäffer-Poeschel Verlag, 1994
- [45] STROMINGER, A. ; VAFA, C. : ‘Microscopic origin of the Bekenstein-Hawking entropy’. In: *Phys. Lett.* B379 (1996), S. 99–104. – <http://arxiv.org/abs/hep-th/9601029>
- [46] SUSSKIND, L. : ‘Das Informationsparadoxon bei Schwarzen Löchern’. In: *Spektrum der Wissenschaft* 6 (1997)
- [47] SZILARD, L. : ‘Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen’. In: *Zeitschr. f. Physik* 53 (1929), S. 840–856. – Engl. transl. in [29]
- [48] UNSÖLD, A. ; BASCHEK, B. : *Der neue Kosmos. Einführung in die Astronomie und Astrophysik*. 6. Aufl. Berlin Heidelberg New York : Springer Verlag, 1999
- [49] VRIES, A. de: *Über die Beschränktheit der Energienorm bei der Evolution der Dirac-, Weyl- und Maxwellfelder in gekrümmten Raumzeiten*. Bochum : Brockmeyer, 1994. – <http://haegar.fh-swf.de/homepage/publikationen/diss.pdf>
- [50] VRIES, A. de: ‘The evolution of the Weyl and Maxwell fields in curved space-times’. In: *Math. Nachr.* 179 (1996), S. 27–45
- [51] VRIES, A. de: ‘The apparent shape of a rotating charged black hole, closed photon orbits and the bifurcation set A_4 ’. In: *Class. Quantum Grav.* 17 (2000), Nr. 1, S. 123–144. – <http://stacks.iop.org/0264-9381/17/123>
- [52] VRIES, A. de: ‘How to price information by Kullback-Leibler entropy and a moment-return relation for portfolios’. In: *Int. J. Theor. Appl. Finance* 4 (2001), Nr. 3, S. 535–543

- [53] VRIES, A. de: 'Im Schatten Schwarzer Löcher'. In: *Physik in Unserer Zeit* 35 (2004), S. 128–133. – <http://dx.doi.org/10.1002/piuz.200401040>
- [54] VRIES, A. de: 'Shadows of rotating black holes approximated by Dürer-Pascal limaçons'. In: MUNO, R. W. (Hrsg.): *Jahresschrift der Bochumer Interdisziplinären Gesellschaft eV 2003*. Stuttgart : ibidem-Verlag, 2005, S. 1–20
- [55] WAERDEN, B. L. d.: *Mathematische Statistik*. 2. Aufl. Berlin Heidelberg New York : Springer-Verlag, 1965
- [56] WALD, R. M.: *General Relativity*. Chicago London : University of Chicago Press, 1984
- [57] WEIZSÄCKER, C. F.: *Aufbau der Physik*. München Wien : Carl Hanser Verlag, 1985
- [58] WEIZSÄCKER, C. F.: *Zeit und Wissen*. München Wien : Carl Hanser Verlag, 1992
- [59] WHEELER, J. A.: 'Time today'. In: WITT, C. D. (Hrsg.) ; WITT, B. D. (Hrsg.): *Physical Origin of Time Asymmetry*. Cambridge : Cambridge University Press, 1994
- [60] WILMOTT, P. : *Derivatives. The Theory and Practice of Financial Engineering*. Chichester : John Wiley, 1998
- [61] ZEIDLER, E. : *Applied Functional Analysis. Applications to Mathematical Physics*. New York : Springer-Verlag, 1995
- [62] ZEIDLER, E. (Hrsg.): *Teubner Taschenbuch der Mathematik. Teil 1*. Leipzig : B. G. Teubner, 1996
- [63] ZWIEBACH, B. : *A First Course in String Theory*. Cambridge : Cambridge University Press, 2004
- [64] ŻYCZKOWSKI, K. : 'Rényi extrapolation of Shannon entropy'. In: *Open Sys. & Information Dyn.* 10 (2003), S. 297–310. – <http://arxiv.org/abs/quant-ph/0305062>