Definitionen von *Information*

R. Schaback

Version 8. Dezember 2017

In den Wissenschaften finden sich vier Definitionen von *Information*, die gegeneinander einigermaßen klar abgrenzbar sind:

1. Der *mathematische* Informationsbegriff nach *Shannon*. Er basiert auf den *Kolmogoroff*'schen Axiomen der Wahrscheinlichkeitstheorie, definiert Information als negativen Logarithmus einer Wahrscheinlichkeit, führt zur *Entropie* und ist deshalb auch in der Physik von zentraler Bedeutung.

2. Verwendet man stattdessen einen bayesianistischen Zugang zu Wahrscheinlichkeiten als *degree of belief*, so ist *Information* das, was einen *change of beliefs* bewirkt. Dieser Informationsbegriff sei hier *bayesianistisch* genannt.

3. Die *algorithmische* Information, die in einem Text steckt, ist nach *Solomonoff/Kolmogoroff/Chaitin* die minimale Länge eines Programms, das diesen Text produziert. Dieser Begriff ist an ein Maschinenmodell gebunden und liefert ein Maß für *Komplexität*.

4. Der *semantische* Informationsbegriff definiert Information als den *semantischen Gehalt von strukturierten Daten*.

Das Ziel dieses Beitrags ist, diese Informationsbegriffe etwas genauer darzustellen, gegeneinander abzugrenzen und eventuelle Gemeinsamkeiten aufzuzeigen. Eine besondere Tiefe wird dabei nicht angestrebt. Sie bleibt anderen Beiträgen in diesem Band vorbehalten. Und der Initiator der Informationstheorie, C.E. Shannon, stellte schon 1953 fest:

> "It is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field." [25, p. 180]

1 Der mathematische Informationsbegriff

Der Informationsbegriff in der Mathematik ist grundlegend für die von *Claude Shannon* begründete *Informationstheorie*. Man kann die These vertreten, letztere sei eher eine Theorie der *Kommunikation*, und dies wird sogar durch die Titel von *Shannons* Originalarbeiten [22, 23, 24] gestützt. Hier kann kein kompletter Überblick über die heutige Informationstheorie [12] gegeben werden, insbesondere nicht mit allen mathematischen Details. Stattdessen wird auf diejenigen Gesichtspunkte fokussiert, die für die Querverbindungen zu den anderen Informationsbegriffen wichtig sind. Weil der mathematische Informationsbegriff auf Wahrscheinlichkeitstheorie zurückgeht, ist es nötig, diese kurz darzustellen.

1.1 Kolmogoroff'sche Axiomatik

Man betrachtet eine Menge M im Sinne der mathematischen Mengenlehre, deren Elemente man *Elementarereignisse* nennt. Ein typischer Fall ist die Menge $M := \{1, 2, 3, 4, 5, 6\}$ der möglichen Ergebnisse des herkömmlichen Würfelns. Man betrachtet aber auch *Mengen* E von Elementarereignissen, und nennt sie *Ereignisse*. Eine gerade Zahl zu würfeln, wird dann durch das Ereignis $E = \{2, 4, 6\}$ dargestellt. Das *Eintreten eines Ereignisses* ist mathematisch nichts anderes als die Auswahl einer Teilmenge von M, und das ist gleichbedeutend mit der Auswahl eines Elementes der *Potenzmenge* von M, nämlich der Menge Ω aller Teilmengen von M. Die Mathematik kennt keinen Zufall, aber sie modelliert ihn dadurch, daß sie jedem Ereignis $E \in \Omega$ bzw. $E \subseteq M$ eine gewisse Zahl $p(E)$ zwischen Null und Eins zuordnet, die *Wahrscheinlichkeit des Eintretens von* E genannt wird. Dabei wird $p(\Omega) = 1$ gefordert, weil das Eintreten eines beliebigen Elementarereignisses die Wahrscheinlichkeit 1 haben soll. Zwei Ereignisse E_1 und E_2 sind *inkompatibel*, wenn sie als Mengen disjunkt sind, d.h. ihr Durchschnitt $E_1 \cap E_2$ ist gleich der leeren Menge $\emptyset$. Dieser wird die Wahrscheinlichkeit Null zugewiesen, während sich für die Vereinigung $E_1 \cup E_2$ zweier inkompatibler Ereignisse E_1 und E_2 die Wahrscheinlichkeiten addieren sollen, d.h.

$$p(E_1 \cup E_2) = p(E_1) + p(E_2).$$

Diese Regeln bilden (in etwas vereinfachter Form) das *Kolmogoroffsche Axiomensystem der Wahrscheinlichkeitstheorie*. Beim Modell des Würfelns wird jedem Elementarereignis die Zahl $1/6$ zugeordnet, und weil das Ereignis $E = \{2, 4, 6\}$ aus drei inkompatiblen Elementarereignissen besteht, wird ihm die Zahl $p(E) = 1/2$ zugeordnet, und das ist ein Modell für das Würfeln einer geraden Zahl.

Der Begriff der mathematischen Wahrscheinlichkeit wird nur über die Modellbildung mit 'Zufall' verbunden. Die Mathematik würfelt nicht, und sie kennt keinen Zufall im umgangssprachlichen Sinne. Mathematische Wahrscheinlichkeiten sind Zahlen, sonst nichts. Die Diskussion über 'Zufall' geht an ihnen vorbei, ebenso die 'frequentistischen' oder 'subjektiven' Interpretationen von 'Wahrscheinlichkeit', etwa als 'relative Häufigkeit' oder 'Grad der Gewißheit'. Erst bei der außermathematischen Anwendung kommen solche Gesichtspunkte ins Spiel.

1.2 Information von Ereignissen

Man definiert dann die (mathematische) *Information* eines *Ereignisses* $E \in \Omega$ als

$$I(E) := -\log_2 p(E),$$

d.h. die Information eines Ereignisses ist der negative Zweierlogarithmus der (mathematischen) Wahrscheinlichkeit dieses Ereignisses.

Dieser Informationsbegriff hat die folgenden Eigenschaften:

1. Ein Ereignis, das mit Wahrscheinlichkeit Eins eintritt, hat Information Null.

2. Ein Ereignis, das mit Wahrscheinlichkeit Null eintritt, hat positiv unendliche Information.

3. Ein Ereignis liefert umso mehr Information, je unwahrscheinlicher es ist.

4. Der Informationsgehalt unabhängiger und gleichzeitig eintretender Ereignisse kann summiert werden. Genauer: Wenn zwei Ereignisse E_1 und E_2 unabhängig voneinander eintreten können, hat nach den Regeln der Wahrscheinlichkeitstheorie das Und–Ereignis (E_1 *und* E_2) die Wahrscheinlichkeit $p(E_1)p(E_2)$, und deshalb gilt für die Information

$$I(E_1 \textit{ und } E_2) = I(E_1) + I(E_2). \tag{1}$$

1.3 Vergleich mit anderen Informationsdefinitionen

Weil mathematische Information der negative Logarithmus einer mathematischen Wahrscheinlichkeit ist, und weil durch Anwendung von Logarithmus- und Exponentialfunktion Wahrscheinlichkeit und Information ineinander umrechenbar und somit theoretisch vollkommen gleichwertig sind, hat bei diesem Zugang auch der Informationsbegriff keinerlei Bedeutung, die über mathematische Wahrscheinlichkeiten hinausgeht.

Wenn man den mathematischen Wahrscheinlichkeitsbegriff durch einen anderen ersetzt, nämlich z.B. den, daß Wahrscheinlichkeiten für einen ‘Grad der Gewißheit’ stehen, ändert sich das Bild, und das wird im nächsten Kapitel dargestellt. Dort wird die Information nicht fest an die Wahrscheinlichkeit eines Einzelereignisses gekoppelt, sondern die Information bewirkt eine Veränderung, sie wird *differentiell* und bleibt nicht statisch.

Für die Abgrenzung zum semantischen Informationsbegriff ist noch einmal festzuhalten, daß es keine ‘Bedeutung’ von Ereignissen im Sinne einer Semantik gibt. Und in Bezug auf den algorithmischen Informationsbegriff des Kapitels 3 spielt der Entstehungsprozeß der Ereignisse keine Rolle.

Das schränkt die Anwendbarkeit des mathematischen Informationsbegriffs erheblich ein, denn wenn man z.B. natürlichsprachigen Texten oder biologischen Signalen oder Abschnitten aus dem Genom Information beimessen will, müßte man alle diese als zufällig ansehen. Konstruiert man für einen außermathematischen Sachverhalt ein probabilistisches mathematisches Modell, z.B. in der statistischen Mechanik oder der Thermodynamik, so kann der mathematische Informationsbegriff innerhalb des Modells angewendet werden. Sobald aber ein Prozeß nicht durch ‘Zufall’ dominiert wird, sind der mathematische Informationsbegriff und die Informationstheorie eine Sackgasse, nicht nur für die Kognitionspsychologie [14].

1.4 Mathematische Entropie

Der mathematische Entropiebegriff bezieht sich nicht auf Einzelereignisse und ihren mathematischen Informationsgehalt, d.h. den negativen Logarithmus ihrer Wahrscheinlichkeit, sondern immer auf *Mengen* möglicher Ereignisse. Es sei $\mathcal{E} \subseteq \Omega$ eine Menge von Ereignissen E, die jeweils eine Wahrscheinlichkeit $p(E)$ und eine Information $I(E) := -\log_2 p(E)$ haben. Jetzt bildet man den *Erwartungswert* der Information, d.h.

$$\mathbb{E}_{\mathcal{E}}(I) := \sum_{E \in \mathcal{E}} p(E) I(E) = -\sum_{E \in \mathcal{E}} p(E) \log_2 p(E).$$

Auf Anraten von *John von Neumann* nennt *Shannon* diesen Ausdruck *Entropie*. Er gibt den *mittleren Informationsgehalt* aller Ereignisse aus $\mathcal{E}$ wieder, wenn diese mit ihren Wahrscheinlichkeiten gewichtet werden.

Eine tiefergehende Darstellung des Zusammenhangs zwischen Entropie und Information findet sich im Beitrag [19] in diesem Band.

1.5 Spezial– und Grenzfälle

Beim Modell des Würfelns hat man also die Entropie

$$I\!\!E_\Omega(I) = -6 \cdot \frac{1}{6} \log_2 \left(\frac{1}{6}\right) = \log_2(6),$$

und man sieht daran auch, daß die Entropie einer Menge von n gleichwahrscheinlichen Ereignissen immer $\log_2 n$ ist.

Die Entropie eines Systems ist Null, wenn nur ein einziges Ereignis möglich ist. Fragt man bei fester Ereignismenge nach der Wahrscheinlichkeitsverteilung, die zu maximaler Entropie führt, so bekommt man die Gleichverteilung heraus. Mit anderen Worten: das Entropiemaximum ist gegeben, wenn sich das System am wenigsten gut vorhersagen läßt, weil alle Ereignisse gleichwahrscheinlich sind. Kurz, lax und ungenau formuliert:

> Ein System hat umso mehr Entropie, je weniger leicht es vorhersagbar ist, d.h. je 'chaotischer' es ist.

Die letzte Aussage kann man erheblich klarer und tiefer fassen, siehe [19].

1.6 Entropie und Codierung

Um den Zusammenhang zwischen Entropie und Codierung von Nachrichten herzustellen, betrachten wir das Beispiel eines Lokals, das 8 Speisen anbietet, die von den Gästen mit gleicher Wahrscheinlichkeit $1/8$ gewählt werden. Um eine fehlerfreie primitive Kommunikation mit der Küche zu realisieren, kann man die 8 Gerichte mit den 3 Binärziffern kodieren, die man für die Zahlen $0, 1, \ldots, 7$ braucht, nämlich

$$000, 001, 010, 011, 100, 101, 110, 111.$$

Man kann in der Küche Lampen anbringen, die das Bitmuster simultan zeigen, und dafür braucht man $3 = \log_2 8 = I\!\!E_\Omega(I)$ Lampen, d.h. die Entropie des Systems gibt die Anzahl der nötigen Lampen an.

Die obige Codierung ist aber unpraktisch, wenn einzelne Gerichte sehr viel häufiger bestellt werden als andere, d.h. wenn das System eine niedrigere Entropie als 3 hat. Nehmen wir an, das Gericht 0 würde mit einer großen Wahrscheinlichkeit p bestellt, während die 7 übrigen Gerichte sich die Restwahrscheinlichkeit $1 - p$ gleichmäßig teilen. Dann codieren wir

Gericht 0	durch	0
Gericht 1	durch	1001
Gericht 2	durch	1010
Gericht 3	durch	1011
Gericht 4	durch	1100
Gericht 5	durch	1101
Gericht 6	durch	1110
Gericht 7	durch	1111

und berechnen den Erwartungswert der Länge der zu übertragenden Bitfolge:

$$p \cdot 1 + (1 - p) \cdot 4 = 4 - 3 \cdot p$$

und das lohnt sich gegenüber der 3–Lampen–Lösung schon ab $p > \frac{1}{3}$. Die Entropiedefinition liefert direkt das Prinzip der entropieoptimalen Codierung:

> Man sollte die Länge der Binärcodierung eines Ereignisses mit Wahrscheinlichkeit p etwa wie $-\log_2 p$ wählen, um zu erreichen, daß der Erwartungswert der Codierungslänge etwa gleich der Entropie wird.

Besser geht es nicht, wie man beweisen kann. Insbesondere ist also die mathematische Information eines Ereignisses ein Maß für die Länge einer optimalen Binärcodierung dieses Ereignisses.

Würde die mathematische Informationsdefinition auf Sprachen uneingeschränkt zutreffen, und würden die Sprachen alles optimal kodieren, so müßten die am häufigsten eintretenden Sachverhalte durch die kürzesten Wörter beschrieben werden. Inwieweit dies auf natürliche Sprachen oder das Genom zutrifft, bleibt hier offen.

1.7 Konsequenzen

Der mathematische Entropiebegriff modelliert den Informationsgehalt einer Menge von Ereignissen mit bekannten Einzelwahrscheinlichkeiten, und gibt die zu erwartende mathematische Information bei zufälliger Auswahl eines Ereignisses an. Die mathematische Entropie ist nur für vollständig bekannte Systeme sauber definiert, und das hat Folgen für den Entropiebegriff in der statistischen Physik [19]. Wie beim mathematischen Informationsbegriff gibt es keinen Bezug zu irgendeiner Form von Semantik. Aber es gibt einen Bezug zu Sprachen: die Entropie gibt den Erwartungswert der Länge einer optimalen Binärcodierung der Ereignisse des Systems an.

2 Bayesianischer Informationsbegriff

Der mathematische Informationsbegriff rekurriert auf Wahrscheinlichkeiten im Sinne der üblichen *Kolmogoroff*schen Axiomatik. Sobald außermathematische Wahrscheinlichkeitsdefinitionen ins Spiel kommen, sieht die Lage anders aus. Der *objektive* oder *frequentistische* Wahrscheinlichkeitsbegriff definiert Wahrscheinlichkeiten als relative Häufigkeiten, und der als *subjektiv* oder *bayesianisch* bezeichnete Wahrscheinlichkeitsbegriff interpretiert Wahrscheinlichkeit als *Grad persönlicher Überzeugung* (*degree of belief*) oder *Grad der Gewißheit*. Man kann dann die Frage stellen, wie Wahrscheinlichkeit und Information zusammenhängen, und das findet sich sehr ausführlich im Beitrag [20] in diesem Band. Hier wird nur ein Überblick gegeben.

2.1 Information als *change of* rational *beliefs*

Geht man vom subjektiven bzw. *Bayes*'schen Wahrscheinlichkeitsbegriff aus, so besteht *Information* aus allem, was den *Grad der Gewißheit* oder den *degree of belief* ändert. An dieser Stelle muß man fragen, was *Gewißheit* oder *belief* bedeutet, und das führt auf ein weites Feld, u.A. in Richtung auf Logik [11] und Kognition [14]. Der Informationsbegriff wird differentiell, denn Information bewirkt eine Änderung eines mentalen oder kognitiven Zustands. Das entspricht der in Abschnitt 4.6 und in [18] genauer dargestellten Rolle der Information als Zustandsänderung eines interpretierenden Systems und erlaubt viele fächerübegreifende Anwendungen, die hier zugunsten von [18] nicht weiter verfolgt werden.

Schränkt man sich ein auf

> "Information is whatever forces a change of *rational* beliefs" [1, S. 4],

so geht es nicht mehr um den Grad *persönlicher* Überzeugung, sondern um etwas *rational* mathematisch oder quantitativ Faßbares. Versuche, den *degree of belief* direkt zu formalisieren, gehen auf Cox [4] zurück, mit dem Ergebnis, daß jede vernünftige Formalisierung wieder zur Axiomatik der Wahrscheinlichkeitstheorie zurückführt [2]:

> "Probability theory is the unique method of rational, quantitative and consistent inductive inference that can claim to be of general applicability. It focuses on degrees of rational belief and not on other qualities such as simplicity, explanatory power, degree of confirmation, desirability, or amount of information. The reason the method is unique is not because we have succeeded in formulating a precise and rigorous

> definition of rationality. Rather, the method is unique for the more modest reason that it is the only one left after obvious irrationalities - such as inconsistencies - have been weeded out."

Dieser Rückgang auf die klassische Wahrscheinlichkeitstheorie ist durchaus strittig unter Philosophen, die rationales Schließen unter Unsicherheit erforschen, aber man kann bayesianistisches Schließen auch mit dem üblichen Formalismus verdeutlichen. Deshalb wird hier eine sehr vereinfachte Darstellung gewählt, die Information als *change of rational beliefs* erklärt.

Im Sinne des mathematikorientierten Bayesianismus ist *rational belief* eine a–priori– Annahme gewisser Wahrscheinlichkeitsaussagen (*priors*) im Sinne des Abschnitts 1, und durch zusätzliche Informationen, z.B. Messungen, können sich diese Wahrscheinlichkeitsaussagen zu *posteriors* verändern, wodurch man einen *change of beliefs* hat. Durch Information verändern sich die *priors* in *posteriors*. *Caticha* erklärt das in [1, S. 4] so:

> S. 4: "It may be worthwhile to point out an analogy with Newtonian dynamics. The state of motion of a system is described in terms of momentum - the 'quantity' of motion - while the change from one state to another is explained in terms of an applied force. Similarly, in Bayesian inference a state of belief is described in terms of probabilities - the 'quantity' of belief - and the change from one state to another is due to information. Just as a force is defined as that which induces a change in motion, so information is that which induces a change of beliefs."
>
> S. 7: "When there is no new information there is no reason to change one's mind."

Man kann letzteres umdrehen: Wenn sich die *beliefs* nicht ändern, war auch keine Information im *Bayes*'schen Sinne da.

2.2 Bedingte Wahrscheinlichkeiten

Der einfachste Fall einer Änderung von Wahrscheinlichkeiten durch Hinzunahme einer neuen Information als Resultat einer Beobachtung wird beschrieben durch *bedingte Wahrscheinlichkeiten*. Dazu finden sich im Beitrag [20] in diesem Band einige Beispiele.

Wenn in der Notation des Abschnitts 1.1 zwei Ereignisse E_1 und E_2 vorliegen, so betreffen die Wahrscheinlichkeiten $p(E_1)$ und $p(E_2)$ eine Situation ohne jedes Vorwissen. Tritt aber das Ereignis E_1 *zuerst* ein, so ist das Vorwissen verändert, und das Eintreten des Ereignisses E_2 hat danach die *bedingte* Wahrscheinlichkeit

$$p(E_2|E_1) = \frac{p(E_2 \cap E_1)}{p(E_1)}$$

als *Wahrscheinlichkeit von* E_2 *vorausgesetzt* E_1. An dieser Stelle wird klar, warum man zwei Ereignisse E_1 und E_2 als *unabhängig* definiert, falls $p(E_1 \cap E_2) = p(E_1) \cdot p(E_2)$ gilt: es folgt dann nämlich $p(E_1|E_2) = p(E_1)$ und $p(E_2|E_1) = p(E_2)$, d.h. das Vorab-Eintreten des einen Ereignisses ändert die Wahrscheinlichkeit des anderen nicht.

Im allgemeinen Fall ändern sich die a-priori-Wahrscheinlichkeiten $p(E)$ aller Ereignisse $E \in \Omega$ durch Eintreten eines speziellen Ereignisses $E_0 \in \Omega$ in a-posteriori-Wahrscheinlichkeiten $p(E|E_0)$. Der *rational change of beliefs* durch die Zusatzinformation des Eintretens von E_0 ist dann der Übergang von $p(E)$ in $p(E|E_0)$. Während die mathematische Information des Ereignisses E_0 gleich $-\log_2(E_0)$ ist, hat man bei der bayesianistischen Information die Gesamtheit der Änderungen aller Wahrscheinlichkeiten $p(E)$ auf $p(E|E_0)$ für alle Ereignisse E zu betrachten. Wie man dann dem Informationsgehalt von E_0 einen einzigen Zahlenwert zuordnen kann, wird im Abschnitt 2.3 diskutiert.

Zur Illustration soll hier das mathematische Modell des Würfelns mit zwei Würfeln beschrieben werden. Wenn Würfel 1 die Zahl i und Würfel 2 die Zahl j zeigt, modellieren wir dieses Elementarereignis mit dem Zahlenpaar (i, j) und lassen die Werte 1 bis 6 für i und j zu. Die Grundmenge der Elementarereignisse ist dann $M = \{E_{i,j} := (i, j) \ : \ 1 \leq i, j \leq 6\}$ mit $p(E_{i,j}) = 1/36$ für alle Elementarereignisse (i, j). Die außermathematische Interpretation ist natürlich, daß alle Ergebnisse des Würfelns mit zwei Würfeln gleichwahrscheinlich sind und die Wahrscheinlichkeit 1/36 haben. Die Elementarereignisse sind inkompatibel, und deshalb hat jedes Ereignis E als Teilmenge von M die Wahrscheinlichkeit $|E|/36$, wenn $|E|$ die Anzahl der Elemente von E ist.

Die Ereignisse für das Eintreten der Würfelsumme k sind

$$S_k := \{(i, j) \ : \ 1 \leq i, j \leq 6, \ i + j = k\} \text{ für } k = 2, \ldots, 12$$

und man bekommt durch Abzählen der Elemente die Wahrscheinlichkeiten

$$\begin{aligned} & \{p(S_2), \ldots, p(S_{12})\} \\ = \ & \{1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36\}. \end{aligned} \tag{2}$$

Die wahrscheinlichste Würfelsumme ist 7. Solange keine weitere Information vorliegt, haben wir damit einen *rational belief* oder *prior*, und der bezieht sich auf dieses Modell.

Der *change of beliefs* tritt ein, wenn durch irgendeinen außermathematischen Umstand bekannt ist, daß z.B. einer der beiden Würfel eine Drei trägt. Das ist eine Zusatzinformation und sie ändert die Lage grundlegend. Das betreffende Ereignis ist die Menge

$$T_3 := \{(1,3),(2,3),(3,3),(4,3),(5,3),(6,3),(3,1),(3,2),(3,4),(3,5),(3,6)\},$$

die alle Würfelergebnisse mit mindestens einer Drei modelliert. Jetzt ändern sich die Wahrscheinlichkeiten $p(S_k)$ aus (2) auf die bedingten Wahrscheinlichkeiten

$$p(S_k|E_3) = \frac{p(S_k \cap E_3)}{p(E_3)},\ k = 2, \ldots, 12$$

mit den Werten

$$0, 0, 2/11, 2/11, 1/11, 2/11, 2/11, 2/11, 0, 0, 0 \tag{3}$$

gemäß der Formel für die bedingten Wahrscheinlichkeiten unter Benutzung von $p(T_3) = 11/36$. Die neue Information hat also die *beliefs* wesentlich verändert. Die Würfelsummen 4 bis 9 sind möglich und bis auf die Würfelsumme 6 auch gleichwahrscheinlich.

Die obige Darstellung braucht nicht das *Bayes-Gesetz*

$$p(E_2|E_1) = p(E_1|E_2)\frac{p(E_2)}{p(E_1)},$$

das sich aus der Formel für bedingte Wahrscheinlichkeiten sofort ergibt. Denn es geht im *Bayes*-Gesetz um die *Umkehrung* der jeweiligen Bedingtheiten, was hier noch keine Rolle spielt. Erst bei den Beispielen im Artikel [20] in diesem Band treten solche indirekten Probleme auf. Im obigen Beispiel könnte man die Frage umkehren und nach der Wahrscheinlichkeit für *mindestens eine Drei* bei der Vorabinformation *Würfelsumme ist vier* fragen. Dann liefert das *Bayes*-Gesetz

$$p(T_3|S_4) = \frac{p(T_3 \cap S_4)}{p(S_4)} = p(S_4|T_3)\frac{p(T_3)}{p(S_4)} = \frac{2}{11}\frac{\frac{11}{36}}{\frac{4}{36}} = \frac{1}{2},$$

aber es ist auch direkt klar, daß es dann nur die zwei gleichwahrscheinlichen Möglichkeiten (1,3) und (3,1) gibt. Auch die direkte Rechnung über bedingte Wahrscheinlichkeiten

$$p(T_3|S_4) = \frac{p(T_3 \cap S_4)}{p(S_4)} = \frac{\frac{2}{36}}{\frac{4}{36}}$$

ist einfacher.

2.3 Quantitative Bayesianistische Information

Wie oben dargestellt, ändert das Eintreten eines Ereignisses E_0 alle prior-Wahrscheinlichkeiten $p(E)$ von möglichen Ereignissen E in posterior-Wahrscheinlichkeiten $p(E|E_0)$ ab, und das ist der *change of rational beliefs*, der den bayesianistischen Informationsgehalt von E_0 ausmacht. Will man diesen in eine einzige Zahl pressen (was durchaus fragwürdig ist), so muß man einen Zahlwert nehmen, der umso größer ist, je drastischer die Änderung der Wahrscheinlichkeitsverteilung ist.

Das läßt sich durch einen der vielen möglichen Abstandsbegriffe für Wahrscheinlichkeitsverteilungen auch quantifizieren, aber dazu ist hier kein Platz. Besonders einfach ist es, die durch E_0 bewirkte Reduktion der Entropie auszurechnen, und das ergibt im obigen Würfelbeispiel

$$0.7241 = 3.2744 - 2.5503$$

als eine mögliche Quantifizierung des bayesianistischen Informationsgehalts von E_0. Der mathematische Informationsgehalt von E_0 wäre

$$-\log_2(11/36) = 1.7105.$$

Der bayesianische Informationsgewinn hängt aber von der betrachteten Fragestellung ab (bei fester Ausgangsmenge von Elementarereignissen), und kann deshalb nicht in eine einzige Zahl komprimiert werden. Man kann leicht Beispiele angeben, bei denen die zugehörige Entropie beim Eintreten des Ereignisses E_0 sogar auf Null geht [21].

3 Algorithmischer Informationsbegriff

Nach *Gregory Chaitin* [3] ist die algorithmische Informationsthorie "*... the result of putting Shannon's information theory and Turing's computability theory into a cocktail shaker and shaking vigorously*".

3.1 Texte als Zeichenketten

Gegenüber dem mathematischen Informationsbegriff aus Abschnitt 1 geht es hier nicht um Wahrscheinlichkeiten, sondern wie beim semantischen Informationbegriff des Abschnitts 4 um den Informationsgehalt von *Texten* oder *Nachrichten*. Diese sind *Zeichenketten*, d.h. Folgen oder mehrdimensional strukturierte Anordnungen von Zeichen eines Alphabetes A. Weil man Satzzeichen und Steuerzeichen (z.B. Zeilenvorschub, Seitenanfang) mit zum Alphabet rechnet, fallen alle natürlichsprachlichen Texte unter diese Definition, ebenso wie einige Spezialsprachen:

- $a^2 + b^2 = c^2$ in der Mathematik,
- `for (i=0; i<j; i++)` in der Informatik,
- CTTATTCATCTGGTGATTTGGCTACTTCTTAA im Genom,
- LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMG...
  im FASTA-Format zur Beschreibung der Primärstruktur von Proteinen,
- *An jenem Tag im blauen Mond September...* in der Literatur,
- Partituren in der Musik, siehe Abb. 3
- Strukturformeln in der Chemie, siehe Abb. 1

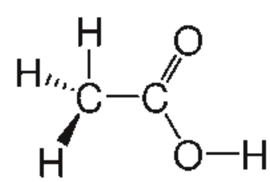

Abbildung 1: Strukturierte Daten aus der Chemie [1]

3.2 Syntax und Semantik von Texten

Texte im obigen Sinne haben eine Struktur, die gewissen Regeln genügen muß, die die *Syntax* des Textes formal beschreiben. Sie haben auch eine *Semantik*, die im Abschnitt 4 eine noch zentralere Rolle spielen wird und im weitesten Sinne die 'Bedeutung' des Textes ausmacht.

[1] https://commons.wikimedia.org/wiki/File:Essigsäure_Struktur.svg

3.3 Komplexitätsdefinition

Der Informationsgehalt eines Textes wird aber beim algorithmischen Informationbegriff nicht über Syntax oder Semantik eines Textes definiert, sondern nur durch dessen *Komplexität*. Letztere wird exakt definiert durch die *minimale Länge eines Algorithmus, der in einem vorgegebenen Maschinen– oder Sprachbeschreibungsmodell diesen Text produziert*, unter Einhaltung der Syntaxregeln. In der Informatik gibt es diverse Maschinenmodelle, und für jedes Modell ist klar definiert, was ein Algorithmus ist und welche Länge er hat. Die allgemeine *Komplexitätstheorie* der Informatik ist sehr viel weiter gefaßt und wird hier ignoriert.

Zum Standardlehrstoff der Informatik gehört auch der Zusammenhang zwischen Maschinenmodellen und Klassen formaler Sprachen, wobei man allerdings nicht direkt die Komplexität einzelner Texte behandelt, sondern danach fragt, welches Maschinenmodell erforderlich ist, um beliebige Texte aus einer durch eine Klasse von generativen Grammatiken definierte Klasse von Sprachen syntaktisch zu analysieren. Das stiftet einen Zusammenhang zwischen Maschinenmodellen und der Komplexität von generativen Grammatiken.

3.4 Konsequenzen

Der obige Komplexitätsbegriff ist quantitativ nicht praktikabel, weil man ziemlich leicht beweisen kann, daß es kein Programm auf einer Turingmaschine gibt, die zu einem gegebenen Text dessen Komplexität berechnet. Er liefert dennoch einige nützliche Einsichten zum Thema Information:

- Wenn die biologische Evolution als Evolution der Genome als Zeichenketten gesehen wird, ist der Zusammenhang zwischen der Komplexität des Textes mit der Struktur der generierenden 'Maschine' interessant.

- Auch die Evolution natürlicher Sprachen kann unter dem Gesichtspunkt der Komplexitätssteigerung gesehen werden, ebenso wie der Verlauf des Spracherwerbs bei Kindern.

- Effiziente technische Übertragungstechniken für Audio– und Videodaten versenden nicht die Daten selber, sondern Anweisungen zu deren Synthetisierung. Man versendet nicht den Kuchen, sondern das Rezept. Der algorithmische Informationsgehalt des Kuchens ist gleich der Länge des kürzestmöglichen Rezepts für diesen Kuchen in einer festgelegten 'Rezeptsprache'.

- Bei der Notenschrift, der Sprache des Genoms und den Syntheseanweisungen für Proteine ist ebenfalls die Nachricht gleich dem Synthesezept.
- Das sind jeweils Texte, die durch *Ausführung* interpretiert werden, siehe Abschnitt 4.5.

4 Semantischer Informationsbegriff

4.1 Information durch Interpretation von Nachrichten

Nach dem klassischen Informatik-Lehrbuch [**?**] von F.L. Bauer und G. Goos entsteht *Information* durch *Interpretation* einer *Nachricht*. Dabei sind Nachrichten als *strukturierte Daten* oder *Signale* zu sehen, die einem physikalischen *Medium*, z.B. Papier, Schall, elektromagnetischen Wellen als *Struktur* aufgeprägt sind. *Information* entsteht erst, wenn diese Daten durch irgendeinen Prozeß, der z.B. in Menschen, Tieren, Zellen oder Maschinen ablaufen kann, *interpretiert* werden. Die entstehende Information ist von diesem Interpretationsprozeß abhängig. Die Frage nach der 'Natur der Information' wird dadurch reduziert auf die Analyse der verschiedenen Interpretationsprozesse. Man wird deshalb sehr verschiedene Antworten bekommen, je nachdem welchen Interpretationsprozeß man ins Auge faßt.

4.2 Information als *semantic content*

Diese der Philosophie zuzurechnende Definition ist auf den ersten Blick nicht wesentlich anders. Sie hat eine interessante Geschichte [17] und wird von *Floridi* [8] formuliert als

> *σ is an instance of information, understood as semantic content, if and only if:*
>
> – *σ consists of one or more data;*
> – *the data in σ are well-formed;*
> – *the well-formed data in σ are meaningful.*

4.3 Vergleich

Der Informationsbegriff nach *Floridi* setzt *semantic content* und *meaning* voraus, was immer das auch sei, während der Informationsbegriff nach 4.1

nur von einem *Interpretationsprozeß* spricht, dessen Ergebnis *Information* ist. Bei *Floridi* wird nicht spezifiziert, was *Semantik* oder *meaning* ist und durch welchen Vorgang sie entsteht, während in 4.1 ein unspezifizierter *Interpretationsprozeß* postuliert wird, dessen Ergebnis *Information* ist. Begriffe wie *semantic content* oder *meaning* sind nachgeordnet und werden nicht gebraucht.

Bei *Floridi* gehören die *well–formed data* mit zum Informationsbegriff, während *Daten* oder *Nachrichten* bei der Definition in Abschnitt 4.1 als potentiell uninterpretierte Objekte nichts mit Information zu tun haben, solange sie nicht interpretiert werden. Diese Diskrepanz sieht Mingers [17] so, daß der Interpretationsprozeß die Information in 4.1 *subjektiv* macht, während sie bei vollständiger Bindung an die Daten *objektiv* genannt werden kann und auch ohne Interpretation existiert.

Wenn die Information an die Daten gebunden ist, und nicht erst durch einen Interpretationsprozeß entsteht, muß man einen wie auch immer gearteten, jedenfalls aber sehr allgemeinen und objektiven Semantikbegriff voraussetzen, der seinerseits der Klärung bedarf, ebenso wie der Begriff des Interpretationsprozesses, wenn man der subjektiven Definition den Vorzug gibt.

Wie man sich auch immer entscheidet: es ist in beiden Fällen klärungsbedürftig, was *Semantik* ist. Entweder als Zugangsvoraussetzung zum Informationsbegriff oder als Ergebnis von Interpretationsprozessen. Der zweite Weg bietet die Chance, gleichzeitig *Information* und *Semantik* zu untersuchen, zumal nicht klar ist, wie man *Semantik*, ohne den Begriff der *Information* zu haben, überhaupt verstehen und an den Anfang einer Begriffsbildung von *Information* stellen kann.

Es erweist sich deshalb für das Folgende als pragmatisch vorteilhaft, Interpretationsprozesse genauer unter die Lupe zu nehmen, um dem semantischen Informationsbegriff und damit auch der Semantik näher zu kommen.

Im Gegensatz zu den anderen Informationsbegriffen geht es beim semantischen Informationsbegriff weder um irgendwelche Zufallsprozesse noch um die Komplexität eines datenerzeugenden Verfahrens. Insofern ist keine Überschneidung mit dem mathematischen und dem algorithmischen Informationsbegriff aus 1 bzw. 3 erkennbar. Ein Zusammenhang mit dem bayesianischen Informationsbegriff aus 2 findet sich in Abschnitt 4.6.

4.4 Beispiele

Ein typischer Fall sind die strukturierten Daten, die uns der Anblick des Sternenhimmels bietet. Bei *Floridi*'s Definition sind die Daten und die Information eine Einheit. Bei der Sichtweise aus Abschnitt 4.1 gewinnen wir diesen Daten erst durch Interpretation Informationen ab, beginnend bei den Tierkreiszeichen bis hin zu Spektraldaten von Galaxien. Die Informationen hängen vom Interpretationsprozeß ab und können sehr verschiedenartig sein, obwohl die Daten dieselben sind und unabhängig vom Interpretationsprozeß existieren. Es wird kein 'Sender' angenommen, der Information in die Nachricht packt und diese dann absichtsvoll verschickt. Information läßt sich nur aus der Perspektive des Interpretierenden oder des 'Empfängers' sehen [6]. Der Satz *Semantic information is the propositional content of data* [17, S. 391] aus der 'objektiven' Variante der semantischen Informationsdefinition ist beim Anblick des Sternenhimmels problematisch.

Strukturierte Texte wie in Abschnitt 3.1 auf Seite 12 und in den Abbildungen 2, 3 und 1 sind die gängigsten Beispiele für Nachrichten bzw. strukturierte Daten. Sie haben eine syntaktische Struktur und einen semantischen Gehalt, der sich bei geeigneter Interpretation erschließt. Syntax und Semantik werden üblicherweise weder den strukturierten Nachrichten noch der eventuell daraus ableitbaren Information zugeordnet, sondern den formalen oder natürlichen *Sprachen*, in denen die Texte abgefaßt sind.

Die Programmiersprachen der Informatik werden syntaktisch exakt definiert [10] und ihre Syntaxanalyse wird vom Computer nach den Syntaxregeln ausgeführt. Bei modernen Sprachen wie JAVA findet allerdings lediglich eine Übersetzung in eine andere Sprache statt, die von einer 'virtuellen Maschine' durch Aktion interpretiert wird. Deshalb befindet sich die Spezifikation der Semantik der Sprache in der Definition der zugehörigen virtuellen Maschine [13]. Die Interpretation durch Aktion wird in Abschnitt 4.5 genauer behandelt.

Als ein weiteres Beispiel strukturierter Daten betrachten wir Abb. 2. Auch wenn keinerlei semantisches Verständnis vorliegt, ist die syntaktische Strukturierung durch eine Art Reimschema und gewisse Symmetrien klar erkennbar. Das ist allerdings nur eine Syntaxanalyse und keine Interpretation. Sie könnte auch von einem Computer durchgeführt werden. Eine Interpretation[2]

[2] Láo-zi: dào dé jing (Lao-tse: Tao-Te-King)
In der strukturtreuen Übersetzung von Günter Debon:

道可道
非常道
名可名
非常名

Abbildung 2: Strukturierte Daten

setzt sehr viel mehr voraus und ist offenbar unter Sinologen und Philosophen bis heute heftig umstritten.

An dieser Stelle kann man einwenden, daß Eigenschaften der *Struktur* der Daten, z.B. Symmetrien, bereits Information sind und objektiv den Daten angehören. Eigenschaften von Strukturen sind aber nicht gleich den Strukturen, sondern Aussagen über Strukturen, während Strukturen nur dadurch Strukturen sind, daß sie bewirken, daß sich strukturierte Daten überhaupt als etwas Gegebenes vom Nicht-Gegebenen unterscheiden lassen. Insofern gehen Eigenschaften von Strukturen über strukturierte Daten hinaus. Sie sind selbst nicht unmittelbar gegeben, sondern erfordern einen Interpretationsprozeß. Wenn dieser auf die Struktureigenschaften fokussiert, kommt er mit Notwendigkeit zu 'objektiven' Ergebnissen. Das kann der subjektive semantische Informationsbegriff durchaus konzedieren.

In gewissen Fällen kann eine starke und dem üblichen Interpretationskontext entsprechende Strukturierung eines Textes dazu verleiten, eine nicht existente Semantik vorzugaukeln. Zum Beispiel ist es relativ leicht, für wissenschaftliche Spezialsprachen Texte zu generieren, die syntaktisch korrekt und semantisch absolut sinnlos sind. Mit MathGen [5] kann man sich eine

Könnten wir weisen den Weg, es wäre kein ewiger Weg;
Könnten wir nennen den Namen, es wäre kein ewiger Name [26]

hochtrabend klingende computergenerierte mathematische Arbeit in Sekundenbruchteilen schreiben lassen. Mehrfach sind solche Produkte zur Publikation angenommen worden, wenn auch nur in dubiosen Journalen [9].

Abbildung 3 zeigt ein Beispiel[3] für versteckte Information in strukturier-

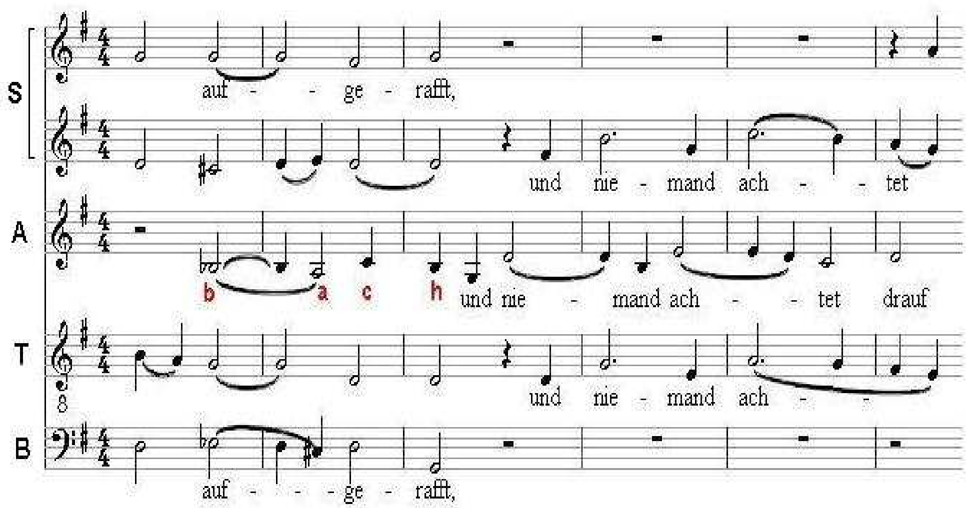

Abbildung 3: Strukturierte Daten

ten Daten. Musiker interpretieren diese strukturierten Daten durch Aktion und stellen ein Tonsignal her, das ebenfalls aus strukturierten Daten besteht und das von Hörenden neu interpretiert werden muß. Es ist in diesem Falle allerdings fraglich, ob alle Hörer die versteckte B-A-C-H- Information herausinterpretieren können, denn ".... niemand achtet drauf".

4.5 Interpretation durch Aktion

In vielen Fällen erfolgt die Interpretation strukturierter Daten durch eine konkrete Aktion des Interpretierenden:

- auf dem Kasernenhof durch unmittelbares und unreflektiertes Befolgen sprachlicher Befehle,
- in der Informatik ebenfalls durch Befehlsausführung, z.B. `z=3*x-y;`

[3]https://commons.wikimedia.org/wiki/File:Musik.beispiel.b-a-c-h.jpg

- durch Ausführen von biochemischen Synthesen bei Interpretation von Genomdaten,
- durch unmittelbare Reaktionen von Organismen auf Reize jedweder Art, z.B. Warn– oder Locksignale.

Wenn die Ausführung durch die Nachricht eindeutig definiert ist, kann man mit gutem Recht die Information, die aus der Nachricht durch Interpretation entsteht, mit der wohldefinierten und im Idealfall auch stets eintretenden Aktion des Interpretierenden gleichsetzen. In diesem Falle liegt die 'Natur der Information' auf der Handlungsebene, und die Information ist eine Aktion. Sofern diese objektiv beobachtbar ist, wie etwa in der Verhaltensforschung kann man Information dieser Art exakt beschreiben.

4.6 Information als Zustandsänderung

Liegen keine Interpretationsprozesse vor, die als Aktionen äußerlich und objektivierbar manifest werden, kann man zumindest eine potentielle Zustandsänderung betrachten, die im Interpretierenden durch die Interpretation erfolgt, und diese als Interpretationsergebnis und damit als Information ansehen. Das trifft zu bei der Lektüre von Büchern oder dem Hören von Sprache oder Musik.

Wenn man eine Aktion eines Organismus, eines Bewußtseins, einer Maschine oder eines biochemischen Reaktionssystems als Konsequenz einer Zustandsänderung des Systems ansieht, ist die in Abschnitt 4.5 beschriebene Interpretation durch Aktion eine Konsequenz einer Interpretation durch Zustandsänderung, also ein Spezialfall dieses Abschnitts.

Diese Argumentationslinie berührt den sehr allgemeinen Begriff von Kognition

> *If a living system enters into cognitive interaction, its internal state is changed in a manner relevant to its maintenance, and it enters into a new interaction without loss of its identity* [16, S.13]

bei Maturana [15]. Dort wird jede Interaktion eines lebenden Organismus mit der Umwelt als kognitiver Prozeß gesehen. In diesem Sinne kann man das Umfeld eines Organismus als eine permanent vorhandene und sich zeitlich und räumlich ändernde Nachricht sehen, die ständig vom Organismus interpretiert wird und dann zu lebenserhaltenden Aktionen des Organismus führt. Der kognitive Prozeß ist in ständigem Wandel durch *Lernen* aus der

Beobachtung der Umwelt, und er führt dann zur Interpretation von Signalen durch Aktion:

> *In animal communication, 'meaning' is generated when the recipient has learned that the occurrence of signal A reliably predicts event B* [7, S. 133].

Dieses Beispiel stützt die These, eher Semantik aus Interpretationsprozessen abzuleiten als umgekehrt Semantik vorauszusetzen, um Information zu definieren.

Auch den Informationsbegriff als *change of beliefs* aus Abschnitt 2.1 kann man hier einordnen, denn es geht um eine Zustandsänderung, in diesem Falle von *beliefs*. Diese können einerseits innerhalb des kognitiven Bereichs liegen, andererseits aber auch quantitativ durch Veränderungen wahrscheinlichkeitstheoretischer *priors* erfaßt werden.

4.7 Konsequenzen

Der semantische Informationsbegriff hat zur Folge, daß Information aus einer sehr allgemein zu fassenden *Zustandsänderung des eine Nachricht interpretierenden Systems* besteht. Die Frage nach der 'Natur der Information' erfordert dann eine spezifische Betrachtung solcher Systeme und ihrer typischen Zustandsänderungen.

5 Fazit

Die vier hier knapp vorgestellten Begriffe von 'Information' decken ein weites Spektrum ab, ohne sich wesentlich zu überschneiden. Der mathematische und der rational-bayesianistische rekurrieren auf Wahrscheinlichkeiten, der algorithmische auf Komplexität, und alle drei haben einen ziemlich eingeschränkten Hintergrund im Gegensatz zum semantischen Informationsbegriff. Letzterer rekurriert aber auf Semantik, die ihrerseits klärungsbedürftig ist, und man hat ein Henne-Ei-Problem: soll man Information über Semantik erklären oder lieber Semantik durch Information?

Der Beitrag [18] nimmt den letztgenannten Standpunkt ein und versteht Information als das Agens, das Änderungen von Prozessen bewirkt. Prozesse, die Semantik 'generieren', sind dabei ausdrücklich zugelassen, neben vielen anderen, die man aus anderen Wissenschaftsdisziplinen heranziehen kann.

Literatur

[1] A. Caticha. Information and entropy. *AIP Conference Proceedings*, 954(1):11–22, 2007.

[2] A. Caticha. Quantifying rational belief. arXiv:0908.3212, 2009.

[3] Gregory J. Chaitin. Zahlen und Zufall. Algorithmische Informationstheorie. Neueste Resultate über die Grundlagen der Mathematik. In *Naturwissenschaft und Weltbild*, pages 30–44. Hölder-Pichler-Tempsky, Vienna, 1992. Translated from the English by Andreas Wenzl and Hans-Christian Reichel.

[4] R.T. Cox. Probability, frequency, and reasonable expectation. *Americal Journal of Physics*, 14:1–13, 1946.

[5] N. Eldredge. http://thatsmathematics.com/mathgen.

[6] J. Fischer. Where is the Information in Animal Communication? In J. Menzel, R. amd Fischer, editor, *Animal Thinking: Contemporary Issues in Comparative Cognition*, pages 151–161. MIT Press, Cambridge, 2011.

[7] J. Fischer and K. Hammerschmidt. Information and Influence in Animal Communication. In A.D.M. Smith, editor, *The Evolution of Language*, pages 129–136. World Scientific, 2010.

[8] Luciano Floridi. Semantic Conceptions of Information. In E.N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. 2014. http://plato.stanford.edu/archives/spr2014/entries/information-semantic.

[9] S. Friedl. http://thatsmathematics.com/blog/archives/185.

[10] Gosling.J., G. Steele, G. Bracha, and A. Buckley. *The Java Language Specification*. Java SE 8 edition, 2014. http://docs.oracle.com/javase/specs/jls/se8/jls8.pdf.

[11] I. Hacking. *Probability and Inductive Logic*. Cambridge University Press, 2001.

[12] Günter Hotz. *Algorithmische Informationstheorie*. Teubner-Texte zur Informatik, 25. B. G. Teubner Verlagsgesellschaft mbH, Stuttgart, 1997.

[13] T. Lindholm, F. Yellin, G. Bracha, and A. Buckley. *The Java Virtual Machine Specification*. Java SE 8 edition, 2014. http://docs.oracle.com/javase/specs/jvms/se8/jvms8.pdf.

[14] G. Lüer and U. Lass. Informationsverarbeitung in der Kognitionspsychologie und in den kognitiven Neurowissenschaften. In diesem Band, 2014.

[15] H. Maturana. Cognition. In P.M. Hejl, W.K. Köck, and G. Roth, editors, *Wahrnehmung und Kommunikation*, pages 29–49. Frankfurt: Peter Lang, 1978.

[16] H.R. Maturana and F.J. Varela. *Autopoiesis and Cognition: The Realization of the Living*, volume 42 of *Boston Studies in the Philosophy of Science*. Reidel, 1980.

[17] J. Mingers. Prefiguring Floridi's Theory of Semantic Information. *tripleC*, 11:388–401, 2013.

[18] R. Schaback. Daten, Prozesse und Information. in diesem Band, 2014.

[19] K. Schönhammer. Der Entropiebegriff in der Physik und seine Beziehung zum Konzept der Information. in diesem Band, 2014.

[20] K. Schönhammer. Wahrscheinlichkeit und Information - wie Henne und Ei ? in diesem Band, 2014.

[21] K. Schönhammer. Persönliche Mitteilung, 2016.

[22] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.

[23] Claude E. Shannon and Warren Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, Ill., 1949.

[24] Claude E. Shannon and Warren Weaver. *Mathematische Grundlagen der Informationstheorie*. R. Oldenbourg Verlag, Munich-Vienna, 1976. Übersetzt aus dem Englischen von Helmut Dressler, Scientia Nova.

[25] N.J.A. Sloane and A.D. Wyner. *Claude E. Shannon: Collected Papers*. Wiley-IEEE Press, 1993.

[26] Lao Tse. *Tao-Tê-King. Das Heilige Buch vom Weg und von der Tugend*. Philipp Reclam jun., Stuttgart, 1979. Übersetzung, Einleitung und Anmerkungen von Günther Debon.