Was hat Entropie mit Information zu tun?

PETER C. HÄGELE

1 Entropie in der Statistischen Thermodynamik

Was sagt die Statistische Thermodynamik über die Entropie von Systemen im Gleichgewicht aus (siehe z.B. [REI75], [KIT01])? Man unterscheidet bei einem System *Makrozustände* und *Mikrozustände* (richtiger wäre: Makro- und Mikrobeschreibung).

Ein Makrozustand ist durch wenige Parameter (Zustandsvariable) charakterisiert (Druck, Volumen, Temperatur u.a.), die man messen bzw. vorgeben kann. Ein Mikrozustand ist durch die detaillierte Angabe aller Orts- und Impulskoordinaten der Teilchen bzw. durch die vollständige Charakterisierung des (fast) stationären Quantenzustandes bestimmt.

Ein System hat im Allgemeinen eine unvorstellbar große Zahl von möglichen Mikrozuständen. Ein bestimmter Makrozustand wird nun durch eine im Allgemeinen immer noch sehr große Zahl von Mikrozuständen realisiert. Diese Zahl nennt man *Komplexionenzahl* (oder etwas irreführend *thermodynamische Wahrscheinlichkeit*). Während einer Messung des Makrozustandes durchläuft das System die vielen, den Makrozustand realisierenden Mikrozustände. Die ermittelte Messgröße ist also ein zeitlicher Mittelwert.

Für den weiteren Aufbau der Theorie müssen nun Postulate eingeführt werden, welche die teilweise Unkenntnis über das System formulieren. Nach GIBBS bildet man gedanklich ein *Ensemble* von sehr vielen, makroskopisch gleich präparierten Systemen und postuliert:

1. Zur Berechnung einer makroskopischen Größe (Messwert) kann der zeitliche Mittelwert durch einen Mittelwert über die Systeme (Ensemblemittelwert) gebildet werden.
2. In einem Ensemble aus (thermisch) isolierten Systemen sind alle Mikrozustände (i) gleichwahrscheinlich (für alle i gilt $p_i = p = \text{const.}$).

Diese Postulate sind letztlich durch den Erfolg der Theorie gerechtfertigt.

Eine wesentliche Aufgabe der Theorie ist es, die Wahrscheinlichkeiten der Mikrozustände anderer Systemtypen zu berechnen. So ergibt sich für den wichtigen Fall (geschlossener) Systeme im Wärmekontakt mit einer Umgebung konstanter Temperatur T die fundamental wichtige BOLTZMANN-Verteilung:

$$\boxed{p_i = \frac{\mathrm{e}^{-\frac{E_i}{k_B T}}}{\sum\limits_i \mathrm{e}^{-\frac{E_i}{k_B T}}}}$$

(E_i : Energien des Systems; k_B : BOLTZMANN-Konstante)

Kennt man die p_i, so lassen sich Mittelwerte von Messgrößen berechnen. Durch Vergleich mit Beziehungen der phänomenologischen Thermodynamik ergibt sich für den Mittelwert der *Entropie* folgender Mittelwert:

$$\boxed{S = -k_B \,\overline{\ln p_i} = -k_B \sum_i p_i \ln p_i}$$

Diese Formel vereinfacht sich für (thermisch) isolierte Systeme, bei denen also alle p_i gleich sind ($p_i = p = \frac{1}{K}$ mit $i = 1, \ldots, K$) zu

$$\boxed{S = k_B \ln K}$$

In einem isolierten System ist die Entropie also (bis auf die BOLTZMANN-Konstante) der Logarithmus der Zahl der möglichen Mikrozustände, die einen Makrozustand realisieren. Möglich heißt: verträglich mit der festgelegten Gesamtenergie und Stoffmenge.

Je größer K und damit auch die Entropie ist, desto wahrscheinlicher liegt der zugehörige Makrozustand vor: Beim ständigen „Durchspielen" aller (gleichwahrscheinlichen) Mikrozustände kommt derjenige Makrozustand mit dem größten K am häufigsten vor, er ist am wahrscheinlichsten. Bei realen Systemen mit vielen Teilchen ist im Gleichgewicht das Maximum von K extrem scharf ausgeprägt, so dass praktisch fast alle Mikrozustände zu demselben Makrozustand (dem Gleichgewichtszustand) gehören. Kleine Abweichungen vom Gleichgewicht (Schwankungen, Fluktuationen) sind aber möglich. Die statistische Thermodynamik schwächt den Satz von der Entropiezunahme (Spezialfall des 2. Hauptsatzes) ab, er gilt nur *im statistischen Mittel*:

> **Ein abgeschlossenes System neigt dazu, den wahrscheinlichsten Zustand (und damit ein Maximum der Entropie) anzunehmen.**

2 Information

Die statistische Interpretation des Entropiebegriffs verknüpft Entropie und Wahrscheinlichkeit: Ein makroskopischer Zustand hoher Entropie hat eine höhere Wahrscheinlichkeit als ein Zustand niedriger Entropie. Als Veranschaulichung dieses Sachverhaltes wird häufig angeführt, dass zunehmende Entropie zunehmende Unordnung bedeute. Diese Zuordnung ist aber unpräzise und gilt keineswegs allgemein. Eine präzise und durchaus anschauliche Verknüpfung ist hingegen mit dem Begriff der *Information* möglich.

Information spielt heutzutage in der Nachrichtentechnik, der Entscheidungs- und Lerntheorie, der Psychologie, der Biologie, der Mathematik und nicht zuletzt in der statistischen Physik eine wichtige Rolle. Dieser Begriff wurde im Rahmen einer Kommunikationstheorie von SHANNON und WEAVER quantifiziert (siehe z.B. [PET67]). Im Hinblick auf technische Anwendungen bei der Übertragung von Nachrichtenflüssen werden hier Probleme der Codierung, Redundanz, Übertragungsgenauigkeit, Kanalkapazität usw. untersucht.

In der Umgangssprache steht meist der Aspekt der Bedeutung („semantische Ebene") und des Zwecks („pragmatische Ebene") einer Nachricht im Vordergrund. In der SHANNONschen Informationstheorie wird dagegen der Begriff *Information* auf den Aspekt des „Neuigkeitswertes" oder „Überraschungswertes" einer Nachricht eingeengt. Dieser Aspekt ist allein mit der Eintrittswahrscheinlichkeit („statistische Ebene") verknüpft und nicht etwa mit Bedeutungen, die vom Empfänger einer Nachricht beigemessen werden.

Wie kann man den Vorgang des Würfelns oder des Münzwerfens, das Schießen auf eine Scheibe, wie kann man eine Nachrichtenquelle, eine politische Wahl, ein physikalisches

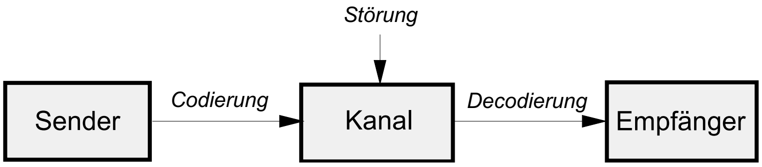

Abbildung 1: Sender-Empfänger-Schema

Experiment unter dem einheitlichen Gesichtspunkt der Informationsübermittlung beschreiben? Man geht dazu von folgender idealisierten Situation aus (vgl. Abb. 1): Ein Sender übermittelt eine Folge von Zeichen (oder Ereignissen) $a_1, a_2, \ldots, a_n$ aus einem Vorrat von n Zeichen, die mit den Wahrscheinlichkeiten $p_1, p_2, \ldots, p_n$ auftreten, über einen Kanal (Übertragungsmedium) an einen Empfänger. Man spricht auch von einem *endlichen Zufallsexperiment* und charakterisiert es durch die Paare $\{(a_1, p_1), (a_2, p_2), \ldots, (a_n, p_n)\}$. Für die Wahrscheinlichkeiten p_i gilt wie üblich

$$0 \leq p_i \leq 1 \qquad \text{und} \qquad \sum_{i=1}^{n} p_i = 1.$$

Damit ist die Situation auf die mathematische Wahrscheinlichkeitstheorie abgebildet. Der Empfänger muss diese Wahrscheinlichkeiten kennen. Die Zeichenfolge wird für die Übertragung vom Sender codiert und dann vom Empfänger decodiert. Sprachliche Laute (Schallwellen) müssen z. B. in elektromagnetische Wellen umgesetzt und dann wieder in Schall zurückverwandelt werden. Die Übertragung ist i. a. Störungen ausgesetzt, welche die Information verändern.

Wie lässt sich nun ein wahrscheinlichkeitstheoretisches Maß der übertragenen Information angeben? Vor dem Auftreten eines Zeichens besteht eine Ungewissheit; sein Auftreten hat einen Überraschungswert und ist in diesem Sinne eine Information für den Empfänger. Welche Menge an Ungewissheit wird durch das tatsächliche Auftreten des Zeichens beseitigt? Lässt man sich von den umgangssprachlichen Begriffen leiten, dann könnte man für ein Zeichen a_i das Maß $1/p_i$ ansetzen: Je geringer die Wahrscheinlichkeit des Auftretens, desto größer der Überraschungswert, der Neuigkeitswert[1], die Information. Für den Aufbau einer Informationstheorie erweist sich aber folgende Definition als zweckmäßiger (z. B. wird dann für mehrere unabhängige Zeichen die Information additiv):

Der Neuigkeitswert eines einzelnen Zeichens a_i aus n möglichen Zeichen ist

$$H_i^{(n)} = \log \frac{1}{p_i} = -\log p_i.$$

Wenn von den Zeichen eines mit Sicherheit eintritt (Wahrscheinlichkeit 1), so wird der Neuigkeitswert 0. Sehr selten vorkommende Zeichen haben dagegen einen entsprechend hohen Neuigkeitswert.

Bei Foldgen von Zeichen begnügt man sich meist mit dem Mittelwert (dem Erwartungswert) der einzelnen $H_i^{(n)}$, also mit der mittleren Information. Gemäß der Regel für die

[1]Mit dem Begriff *Neuigkeitswert* soll nicht gesagt sein, dass Information neu entsteht. Bei der Shannon-Weaverschen Informationstheorie geht es um die *Übertragung* von Information.

Mittelwertbildung ergibt sich

$$H^{(n)}(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{n} p_i H_i^{(n)} = -\sum_{i=1}^{n} p_i \log p_i.$$

Dieses wichtige Maß wurde von SHANNON und WEAVER eingeführt und – wegen seiner formalen Übereinstimmung mit der Entropie der Statistischen Thermodynamik – als Entropie bezeichnet. Zur klaren Unterscheidung soll hier von *Informationsentropie* gesprochen werden.[2]

Da man sinnvollerweise nur das Eintreten künftiger Ereignisse mit Wahrscheinlichkeiten charakterisieren kann (vergangene Ereignisse liegen ja fest!), ist auch die Informationsentropie nicht ein Maß für eine vorhandene (aktuelle), sondern für eine künftige Information. Sie ist ein Maß dessen, was man durch das Eintreffen des Zeichens (oder: nach Ausführen eines Experiments) erfahren kann; sie ist Maß für eine beseitigbare Ungewissheit, sie ist *potentielle Information* H_p, nicht aktuelle Information H_a (C. F. V. WEIZSÄCKER).

Die Situation ist besonders übersichtlich im Falle eines Zufallsexperimentes mit nur zwei Ereignissen, welche die Wahrscheinlichkeiten p und $1-p$ haben. Hier ergibt sich die Informationsentropie zu $H^{(2)} = -p \log p - (1-p)\log(1-p)$. Dies wird in Abb. 2 verdeutlicht. Für $p = \frac{1}{2}$ (z. B. Wurf einer idealen Münze) wird $H^{(2)} = 1$ bit und ist maximal. Im Fall

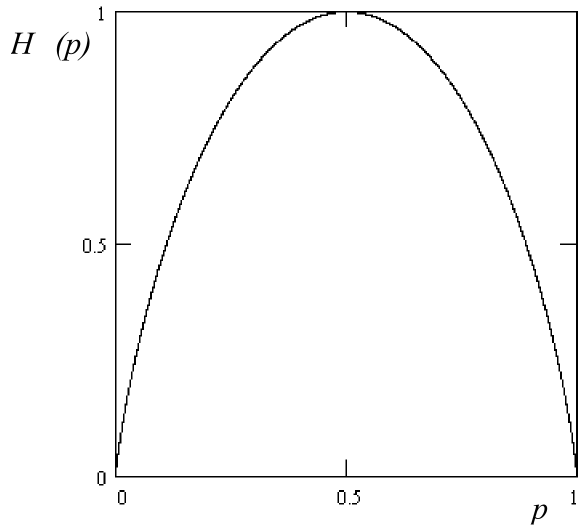

Abbildung 2: Informationsentropie eines Zufallsexperimentes mit zwei Ausgängen

der Gleichwahrscheinlichkeit der Ausgänge wird durch das Eintreten des Ereignisses die maximale Unsicherheit beseitigt, die potentielle Information ist am größten. Der Zusammenhang zwischen potentieller und aktueller Information ist:

$$H_p^{(2)} = \log 2 - H_a^{(2)} \qquad \text{und allgemein} \qquad H_p^{(n)} = \log n - H_a^{(n)}.$$

[2] Der Zahlenwert der Informationentropie hängt von der gewählten Basis des Logarithmus ab. Die Maßeinheit ist 1. Wählt man speziell den Logarithmus zur Basis 2 (binärer Logarithmus, lb oder ld), so fügt man zur Klarheit an die so gewonnenen Zahlenwerte die Pseudomaßeinheit „bit“ an.

Treten alle n möglichen Ereignisse mit derselben Wahrscheinlichkeit auf, so ist (für alle i) $p_i = \frac{1}{n}$, und die Informationsentropie vereinfacht sich zu $H^{(n)}(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) = \log n$. Diese Größe wird auch *Entscheidungsgehalt* genannt. Gemessen in bit ist sie die Mindestanzahl von ja/nein-Entscheidungen in einem Frage-Antwort-Spiel zur Beseitigung einer Unsicherheit. Eine Nachricht bestehe z. B. im Aufleuchten *einer* Anzeigelampe in einem quadratischen Feld von 16 Lampen, welche alle mit derselben Wahrscheinlichkeit aufleuchten können (vgl. Abb. 3). Wie groß ist der Entscheidungsgehalt?

Nach Definition ergibt sich $H^{(16)} = \mathrm{lb}\,16$ bit $= 4$ bit. Das Feld mit der leuchtenden Lampe kann durch (minimal) vier ja/nein-Antworten auf Alternativfragen (Rechts oder links? Oben oder unten?) lokalisiert werden.

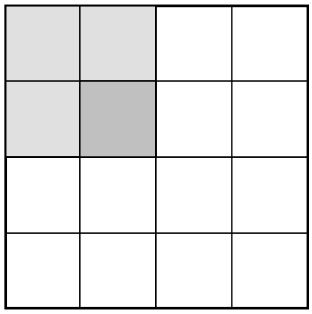

Abbildung 3: Entscheidungsgehalt beim Aufleuchten *einer* Lampe (von 16)

3 Entropie und Information

Betrachtet man nun den Ausdruck für die thermodynamische Entropie, so fällt sofort die weitgehende formale Übereinstimmung mit der potentiellen (nicht der aktuellen) Information auf:

$$\begin{aligned} \text{thermodynamische Entropie } S &= -k_B \sum_i p_i \ln p_i \\ \text{Informationsentropie } I_{\text{pot}} &= - \sum_i p_i \,\mathrm{lb}\, p_i \end{aligned}$$

Eben wegen dieser Übereinstimmung hat SHANNON sein Informationsmaß als Entropie bezeichnet.

Kann man nun also Entropie und Information (bis auf den Proportionalitätsfaktor) einfach gleichsetzen? Dazu muss man die Bedeutung der p_i betrachten: In der thermodynamischen Entropie sind die p_i Wahrscheinlichkeiten von Mikrozuständen, genauer: Die Wahrscheinlichkeitsverteilung der Energiezustände eines materiellen Systems. Bei der Informationsentropie bedeuten diese p_i dagegen die Eintrittswahrscheinlichkeiten beliebiger, inhaltlich nicht spezifizierter Ereignisse. Soweit diese Ereignisse nicht den mechanischen Grundgesetzen gehorchen, braucht die Informationsentropie auch nicht dem 2. Hauptsatz zu genügen!

Die Informationsentropie (und der damit festgelegte Informationsbegriff) ist also allgemeiner als die thermodynamische Entropie. Erst wenn man die p_i im Sinne der statistischen

Thermodynamik festlegt, kann man gleichsetzen:

$$\boxed{S = k_B \ln 2 \cdot I_{\text{pot}}}$$

Für Entropie- und Informations*änderungen* gilt

$$\Delta S = k_B \ln 2 \cdot I_{\text{pot}} \quad \text{und} \quad \Delta S = -k_B \ln 2 \cdot I_{\text{akt}}.$$

Damit ist nun eine informationstheoretische Interpretation der thermodynamischen Entropie möglich: Die Entropie misst die potentielle Information des Experimentators. Sie misst, wie viel derjenige, der den Makrozustand kennt, noch wissen könnte, wenn er auch den Mikrozustand kennen lernte (C. F. v. Weizsäcker, [Wei74, Lyr02]). Bei zunehmender Entropie nimmt diejenige Menge an Wissen *zu*, die der Kenner des jeweiligen Makrozustandes nicht hat, aber durch Messung des jeweiligen Mikrozustandes (prinzipiell) gewinnen könnte.

Oft spricht man auch von Entropie*zu*nahme = Informations*ab*nahme. Hier ist die *aktuelle* Information gemeint! Die Verwechslung von aktueller und potentieller Information hat seit Brillouin einige Verwirrung und Vorzeichen-Unklarheiten in der Literatur gestiftet.

Der Satz von der Entropiezunahme lautet in informationstheoretischer Formulierung:

> **Mit fortschreitender Zeit wird mit überwiegender Wahrscheinlichkeit die aktuelle Information des zu dieser Zeit vorliegenden Makrozustandes abnehmen, seine potentielle Information zunehmen.**

Solche Formulierungen klingen recht subjektivistisch. C. F. v. Weizsäcker hat aber klargestellt, dass der Informationsbegriff „in objektiver Weise subjektbezogen“ ist. Er schreibt [Wei72]:

> „Der Informationsbegriff ist nämlich etwas, was sich auf ein wissendes Subjekt bezieht, auf die Fragen, die dieses Subjekt hat, auf die Antworten, die es dafür gewinnt, aber er ist in objektiver Weise subjektbezogen, und für alle Subjekte, die dasselbe Wissen oder dieselben Methoden haben, Wissen zu erwerben, ist auch das Resultat dasselbe, und dieses ist der objektive Gehalt.“

Version: 3.07.2004 29.07.2005

Literatur

[Kit01] Kittel, Ch., Krömer, H.: Thermodynamik. Oldenbourg 2001.

[Lyr02] Lyre, H.: Informationstheorie. Eine philosophisch-naturwissenschaftliche Einführung. UTB 2289. München: Wilhelm Fink Verlag 2002. (Ein sehr empfehlenswerter Überblick!))

[Pet67] Peters, J.: Einführung in die allgemeine Informationstheorie. Berlin, Heidelberg, New York: Springer 1967.

[Rei75] Reif, F.: Grundlagen der Physikalischen Statistik und Physik der Wärme. (Hrsg. v. W. Muschik). Berlin, New York: de Gruyter 1975.

[Wei72] v. Weizsäcker, C. F.: Vorbereitete Diskussionsbemerkung. Nova Acta Leopoldina **37/1** (206), 503. Leipzig: Joh. Ambrosius Barth 1972.

[Wei74] v. Weizsäcker, C. F.: Evolution und Entropiewachstum. In: E. v. Weizsäcker (Hrsg.): Offene Systeme I. Beiträge zur Zeitstruktur von Information, Entropie und Evolution. Stuttgart: Klett 1974.

Ergänzende und weiterführende Literatur:

Hägele, P. C.: Strukturbildung, Evolution und die Hauptsätze der Thermodynamik. In: Gutsche, E., Hägele, P. C., Hafner, H. (Hrsg.): Zur Diskussion um Schöpfung und Evolution. Gesichtspunkte und Materialien zum Gespräch. Marburg: SMD 1998 (4. Aufl.).

Kornwachs, K., Jacoby, K. (Ed.): Information. New Questions to a Multidisciplinary Concept. Berlin: Akademie Verlag 1996.