

ORDEN POUR LE MÉRITE
FÜR WISSENSCHAFTEN UND KÜNSTE

REDEN UND GEDENKWORTE

FÜNFUNDDREISSIGSTER BAND

2006 – 2007

WALLSTEIN VERLAG

2. VERVIELFÄLTIGUNG UND SELEKTION:
DAS GRUNDPRINZIP DER BIOLOGIE

MANFRED EIGEN

ÜBER DEN BEGRIFF DER INFORMATION
IN PHYSIK UND BIOLOGIE
– SYNOPSIS –

Ein quantitatives Maß für die Information wurde zuerst von Claude Shannon¹ 1949 in die Nachrichtentechnik eingeführt. Tatsächlich hat Shannon selber den Begriff »Information« zunächst nicht benutzt. In seinen ursprünglichen Veröffentlichungen, die sich allein mit der Kommunikation vorgegebener Nachrichten, dagegen nicht mit deren Inhalt oder deren Entstehung befaßten, nannte er die entsprechende Größe »uncertainty«. Es war vor allem der Physiker Leon Brillouin², der (in seinen in den fünfziger Jahren des vorigen Jahrhunderts in den USA gehaltenen Vorlesungen) den Informationsbegriff für die Naturwissenschaften erschloß. Wir sollten uns aber darüber im Klaren sein – was Brillouin auch in seinem Buch hervorhebt – daß Shannons Theorie allein den quantitativen Aspekt der Information berücksichtigt, indem er die Menge der in einer Nachricht enthaltenen Information durch einen Mittelwert ausdrückt.

Mit anderen Worten: Shannon geht von einer durch die Struktur der Sprache vorgegebenen Wahrscheinlichkeitsverteilung für die Verwendung der verschiedenen Sprachsymbole aus und berechnet den Mittelwert aller möglichen sich daraus ergebenden Symbolkombi-

nationen einer Nachricht gegebener Länge. Der Physiker nennt einen solchen Mittelwert die Entropie der Wahrscheinlichkeitsverteilung. So geht die Geschichte, daß der Physiker John von Neumann, dem Shannon seine Ideen vortrug, gleich eingewandt haben soll: Warum nennen Sie Ihre »uncertainty« nicht Entropie. Das gibt Ihnen in jeder Diskussion einen Vorteil, da nur die wenigsten Menschen wirklich wissen, was Entropie ist. Shannons Theorie ist eine Theorie der Kommunikation, und den Nachrichtentechniker interessiert nicht der Inhalt irgendeiner Nachricht. Er muß ein Netz konstruieren, das für alle möglichen Nachrichten geeignet ist. Vor allem muß die Nachricht fehlerfrei übertragbar sein, eine Forderung, die von besonderer Bedeutung für die Benutzung eines Computers ist, der in kürzester Zeit Milliarden von Rechenoperationen ausführen muß, was eine automatische Fehlerkorrektur erfordert.

Ich möchte auf diese interessanten Folgerungen, die den eigentlichen Wert von Shannons Theorie ausmachen, hier nicht näher eingehen, sondern statt dessen die Frage nach dem anderen Aspekt des Informationsbegriffes, nämlich nach dem semantischen Gehalt der Information, aufwerfen. Dieser ist es ja, der in unserer sprachlichen Verwendung des Begriffs die größere Rolle spielt. Wenn ich die Telefonauskunft anrufe und nach einer bestimmten Nummer frage, dann werde ich mit der Antwort: »dreiundzwanzig bits« kaum etwas anfangen können. Damit habe ich zwar meine Ungewißheit (uncertainty) quantifiziert, aber nicht die von mir benötigte »Information« erhalten.

Hier muß ich aber gleich ein mögliches Mißverständnis ausschließen. Eine allgemeine Theorie der Semantik einer gegebenen Sprache gibt es nicht, ja, kann es nicht geben. Mithilfe einer Sprache lassen sich alle nur möglichen Situationen zum Ausdruck bringen oder wenigstens umschreiben, wenn auch nicht immer erklären. Mit einer allgemeinen Theorie der Sinnzuweisung müßte man aber alles – wenigstens im Prinzip – auch erklären können. Eine solche Theorie wäre dann wirklich eine »theory of everything«, die es aber im wörtlichen Sinne gar nicht geben kann. Dagegen sollte es aber sehr wohl möglich sein, eine allgemeine Theorie der Sinnentstehung zu formulieren.

Was sind die allgemeinen Gesetzmäßigkeiten der Informationsentste-

hung? Welche Mechanismen werden hierzu benötigt? Gleich wieder eine Einschränkung: Unsere sprachliche Information entsteht in den Gehirnen von Menschen. Wir benötigten also eine vollständige Kenntnis der Funktionsweise unseres Gehirns. Trotz spektakulärer Erfolge in der Erforschung neuronaler Prozesse sind wir aber von einem solchen Ziel noch weit entfernt. Ich habe dennoch diese Frage aufgeworfen, weil ich damit zu meinem eigentlichen Thema komme, nämlich der Frage nach dem Ursprung des Lebens, die eine Frage nach dem Ursprung der genetischen Information der Lebewesen ist.

Natürlich würde eine erschöpfende Behandlung dieser Frage den Rahmen dieser Mitteilung sprengen. Ich habe sie zum Hauptthema eines Buches gemacht, dessen Bearbeitung nahezu abgeschlossen ist und das in näherer Zukunft erscheinen wird.⁵ Ich will die zugrundeliegende Theorie hier nur kurz anhand von drei Begriffen skizzieren, die in dieser Theorie eine Schlüsselrolle spielen, nämlich Informationsraum, Quasispezies und Hyperzyklus. Diese drei Begriffe möchte ich zunächst kurz umreißen:

Informationsraum: Der Raum, in dem Information entsteht, ist nicht der physikalische Raum, der aus Orts- und Zeitkoordinaten besteht, sondern ein diskreter mathematischer Raum, in dem alle möglichen Symbolkombinationen zum Ausdruck kommen. Und zwar muß jeder (diskrete) Punkt in diesem Raum eindeutig einer bestimmten Symbolkombination zugeordnet und die Nachbarschaft der verschiedenen Punkte eindeutig definiert sein. Es läßt sich leicht zeigen, daß bei Verwendung binärer Symbole (wie dies im Computer geschieht) der Punktraum einer Nachricht ein hochdimensionaler nicht-Euclidischer Raum ist, in dem die Zahl der Dimensionen mit der Zahl aller in der Nachricht erscheinenden Symbole übereinstimmen muß. Da im binären Alphabet jeder Punkt zwei Besetzungsmöglichkeiten (z.B. null oder eins, oder Schalter »auf« oder »zu«) hat, gibt es bei N Symbolen insgesamt 2^N Kombinationen aller Symbole. Das ist für große Werte von N eine riesige Zahl, was bedeutet, daß die Wahrscheinlichkeit für jede zufällige Besetzung einer gegebenen Kombination beliebig klein wird.

Das Problem der Informationsentstehung ist nun, die Wahrscheinlichkeit für eine Trajektorie von einem beliebigen Punkt, der einer sinnlosen Kombination von Symbolen entspricht, bis zu einem Punkt, der einen Sinngehalt repräsentiert, zu bestimmen. Das erfordert zweierlei: die durchlaufene Trajektorie muß sinnvoll angeordnet sein, und es muß einen dynamischen Prozeß geben, der eine Bewegung längs einer solchen Anordnung möglich macht. Genau das muß bei der Entstehung einer sinnvollen genetischen Nachricht geschehen. Die biologische Nachricht macht Gebrauch von einer Alphabet, das anstelle von nur zwei vier Symbole enthält. Das ist notwendig, weil es »von allein« entstehen muß und nicht von intelligenten Wesen gehandhabt wird. Der biologische Raum enthält also 4^N oder 2^{2N} mögliche Punkte entsprechend den möglichen Symbolkombinationen der genetischen Sprache. Die Nachbarschaft aller Punkte ist so geordnet, daß alle Punkte, die sich nur um ein Symbol unterscheiden, den Abstand »eins« voneinander haben. Das läßt sich eben nur in einem N - oder $2N$ -dimensionalen Raum durchführen. Nun müssen wir uns nach Mechanismen umsehen, die eine »sinnvolle Bewegung« im Informationsraum, d.h. das Ansteuern einer sinnvollen Symbolkombination, ermöglichen. Der einfachste Mechanismus führt uns zum Begriff der Quasispezies.

Quasispezies: Bewegung im Informationsraum bedeutet Entstehen und Vergehen von Symbolkombinationen, die jeweils einer Bewertung unterzogen werden. Biologische Information ist materiell in Genen, das bedeutet Sequenzen von Nucleinsäuren, gespeichert. Wir müssen also zunächst die Mechanismen der Synthese und des Zerfalls von Nucleinsäuren untersuchen. Dabei erinnern wir uns an die ungemeine Vielfalt möglicher Symbolkombinationen: z.B. 4^N , für $N = 1000$ (das ist etwa die Größe eines Gens:), gleich ca. 10^{600} , eine eins mit sechshundert Nullen.

Eine so große Zahl wäre viel mehr als das Äquivalent der gesamten Materie des sichtbaren Universums, das »nur« etwa 10^{80} Massen des kleinsten Atoms, des Wasserstoffatoms, entspricht. Die Konsequenz lautet: Die zufällige Entstehung einer bestimmten Symbolkombina-

tion innerhalb der ca. 10^{16} Sekunden, die seit dem Urknall vergangen sind, ist mit an Sicherheit grenzender Wahrscheinlichkeit auszuschließen. Zufällige Synthese bedeutet das Herumirren in einem Raum unermeßlicher Ausdehnung. Um zusammenhängende Trajektorien zu erzeugen, benötigt man den Prozeß der Selbstreproduktion. Völlig fehlerfreie Selbstreproduktion bedeutet Stillstand in einer bestimmten Region – im Idealfall sogar in einem einzigen Punkt – des Sequenzraumes. Doch Fehler, sprich Mutationen, sind allgegenwärtig. Unter den in der Biologie in Frage kommenden Molekülklassen sind nur die Nucleinsäuren, DNA und RNA, mit dieser, einer »inhärenten« Fähigkeit der Selbstreproduktion, begabt. Proteine sind hervorragende Funktionsvermittler. Sie können dank ihrer adaptiven Struktur nahezu jede im Molekularbereich gewünschte katalytische Funktion vermitteln; aber sie können weder »lesen« noch »schreiben«. Genau diese Fähigkeit wird zur Selbstreproduktion benötigt, und sie ist nur den Nucleinsäuren »inhärent«.

Wir haben eine Theorie ausgearbeitet⁴ und diese in zahllosen Experimenten an Nucleinsäuren (und auch vollständigen Viren) überprüft⁵. Ja, wir haben Maschinen entwickelt, die auf dieser Basis automatisch arbeiten und damit eine neue »Evolutive Biotechnologie« begründen, mit deren Hilfe sich neue biologische Katalysatoren entwickeln lassen. Natürlich schließt diese Technologie eine »Übersetzung« der Information in funktionelle Proteinstrukturen ein, wozu das in der Molekularbiologie angehäuften Wissen⁶ über diesen Vorgang herangezogen werden muß. Die Theorie stellt Darwins Überlegungen auf eine solide mathematische Grundlage, sodaß Selektion heute ein klar definierter Begriff ist, der auch für selbst-replizierende Moleküle ähnliche Anwendungen findet wie für Lebewesen, denen die Eigenschaft der Selbstreproduktion, basierend auf den genannten molekularen Mechanismen, inhärent ist. Man mag die Frage stellen, was ist denn eigentlich die Zielstruktur der Selektion. Es geht um Existenz, »to be or not to be«. Die Trajektorie im Informationsraum führt automatisch zu solchen Strukturen, deren Reproduktion am besten an die bestehenden Umweltbedingungen angepaßt ist. Ich will hier nicht auf Einzelheiten der Theorie eingehen. Es handelt sich (wie für die Quan-

tenmechanik) um eine Matrixtheorie. Was überlebt, ist nicht die einzelne Spezies, sondern eine durch einen Eigenwert charakterisierte relativ breit gestreute Verteilung, die den relevanten Teil des Informationsraums abdeckt. Die Mathematik zeigt uns, daß unter den vielen möglichen Eigenwerten nur einer, nämlich der maximale, die Selektion bestimmt. Die Theorie erklärt das von Darwin postulierte Verhalten, aber sie sagt noch nichts über den Ursprung (origin) solchen Verhaltens aus. Dazu brauchen wir noch den »Hyperzyklus«.

Hyperzyklus: Der Hyperzyklus basiert auf der Tatsache, daß wir in der Biologie einen Genotypus und einen Phänotypus benötigen. Der Genotypus stellt die Information dar, während der Phänotypus die durch den Genotypus repräsentierte (abstrakte) Information in konkrete materielle Funktion übersetzt. In der Biologie handelt es sich tatsächlich um eine »sprachliche« Übersetzung der Nucleinsäure in die Proteinstruktur, die auch allgemein als »Translation« bezeichnet wird. In Bezug auf die Selektion geht es um eine Rollenverteilung, die ich als Genotyp-Phänotyp Dichotomie bezeichnet habe. Die Selektion fixiert die in der Nucleinsäure niedergelegte Information, jedoch auf Grund von funktionellen Eigenschaften, die in den Proteinen lokalisiert und auf dieser Ebene dem Selektionstest unterworfen sind. Das gilt im Prinzip auch noch, wenn Genotyp und Phänotyp aus der gleichen Molekülklasse – z.B. den Ribonucleinsäuren (RNA) – rekrutiert werden. Das war wahrscheinlich in der sogenannten »RNA-Welt« der Fall, die der »DNA-RNA-Protein-Welt« vorausging. Auch hier waren Genotyp und Phänotyp zwei individuelle Träger, Genotyp ein RNA-Molekül, das die Information für den Phänotyp enthält und der Replikation unterworfen wird, die von einem zweiten RNA-Molekül, das mit der Struktur der ersten identisch sein kann, katalysiert wird, und als »Ribozym« Träger der Funktion ist. Beiden Fällen gemeinsam ist ein nicht-linearer Ansatz (in dem zum Beispiel das Produkt zweier verschiedener oder das Quadrat einer Variablen erscheinen). Dies ist für den Ursprung einer neuen Eigenschaft eine notwendige Voraussetzung und von uns auch bei der Virusreplikation experimentell untersucht worden.

Was ist nun ein »Hyperzyklus«? Ich habe den Begriff hergeleitet aus dem Reaktionsmechanismus, der eine hierarchische Überlagerung mehrerer Zyklen beinhaltet, sozusagen ein Zyklus von Zyklen. Das Wort ist nicht meine Erfindung, wie ich zunächst dachte. Es findet sich bereits in einem Brief von Carl Friedrich Gauß an seinen ungarischen Freund und Kollegen Wolfgang Bolyai (1832) im Zusammenhang mit seinen Gedanken zu einer nicht-Euclidischen Geometrie. Allerdings ist er dort im abstrakt-mathematischen Sinne gebraucht, indem er eine geometrische Eigenschaft symbolisiert. Unser Hyperzyklus⁷ dagegen, keineswegs im Widerspruch zur Gaußschen Interpretation, beschreibt das dynamische Prinzip einer Rückkopplung bei der Entstehung der genetischen Information, das sich durch das Schaltbild einer hierarchischen Überlagerung von Reaktionszyklen darstellen läßt. Der einfachste Fall eines Hyperzyklus läßt sich in einer RNA-Welt realisieren. Die Replikation von den beiden komplementären Plus- und Minus-Strängen muß zur Erreichung einer hinreichend hohen Erzeugungsrate katalytisch befördert werden. In der RNA-Welt kann der Katalysator ein RNA-Molekül, ein Ribozym, sein. Die Replikation des Ribozym-Gens führt zu seiner eigenen Verstärkung. Der Ratenumsatz enthält einen quadratischen Term der RNA-Konzentration, da die Replikation zwei RNA-Stränge erfordert, von denen der eine als Matrize für die »Kopierung«, der anderen als Ribozym funktioniert. Beide rekrutieren sich aus der gleichen Molekülklasse und haben dadurch eine erhöhte Chance für den Ursprung der Replikation. Einen analogen Mechanismus finden wir bei der RNA-Virusvermehrung in der Wirtszelle wieder. Die Infektion der Wirtszelle geschieht durch ein einzelnes Virusgenom, also ein einzelnes RNA-Molekül. Dieses kann sich nicht von allein vermehren. Es muß erst von den Proteinfabriken der Wirtszelle, den Ribosomen, »übersetzt« werden. Eins der Übersetzungsprodukte ist das Enzym, also in diesem Falle ein Protein-Molekül, das die Replikation der Virus RNA katalysiert. Auf diese Weise konnten wir nicht nur den Mechanismus der Virus-Replikation mit seinen verschiedenen Schritten genau untersuchen, sondern auch das Prinzip der hyperzyklischen Verstärkung in allen Einzelheiten studieren⁸.

Was ist die Quintessenz all dieser Untersuchungen? Natürlich wollen wir »Leben an sich« näher verstehen. Wir haben schon gesehen, daß Leben eng mit dem Begriff der Information verknüpft ist. Aber wir fragen nicht einfach: »Was ist Leben“? Das ist nicht einmal eine gute Frage, und zwar wegen der überaus reichen Fülle von Lebensformen, jede für sich ein Wunderwerk der Natur. Wir wollen vielmehr wissen: Was ist allen Lebensformen gemeinsam, oder was unterscheidet ein belebtes von einem unbelebten System vergleichbarer chemischer Komplexität? Ja, das ist die Information, nicht einfach die Menge, die in der Shannon'schen Entropie zum Ausdruck kommt, sondern der Sinngehalt der in ihr wohnenden Semantik, der in der Bewertung der Information und daher für ihre Selektion eine so große Rolle spielt. Hier finden wir auch einen Teil der Antworten auf die Frage nach unserer eigenen Existenz. Der Ursprung verlangt eine nicht-lineare »Keimbildung«, aber die lange Phase der Optimierung ist quasi-linear und nur dadurch Darwinisch.

So wie die Moleküle, die am Anfang »belebter« Materie stehen, selbstreplikativ sein mußten, sind es ihre Nachfahren, die Lebewesen, von denen Thomas Mann im Zauberberg so treffend sagt: »während kein Lebewesen aufzuweisen war, das nicht einer Elternzeugung sein Dasein verdankt hätte.« Die Theorie der »Quasispezies« hat uns ein quantitatives Verständnis ermöglicht, das weit über zeitgenössische Darstellungen hinausgeht und uns neue tiefe Einblicke beschert hat. Aber es ist im Grunde Darwinisch (nicht Darwinistisch, was ich für eine Ideologie halte). Ihr wichtigstes Resultat ist die Unbegrenztheit, die bis zur menschlichen Existenz geführt hat. Das Leben selbst ist vor ca. 4 Milliarden Jahren entstanden, wie wir aus einer Analyse der Adaptoren des genetischen Codes quantitativ herleiten konnten⁹. Unser Universum existiert seit etwa 15 Milliarden Jahren, das ist ungefähr die Zeit, die seit dem Urknall verstrichen ist. Man nimmt heute an, daß es in einem Multiversum zufälligerweise die Kombination von Naturkonstanten hatte, die ein stabiles Universum, das die Entstehung des Menschen möglich machte, zuließ. Eine solche Apologetik des reinen Zufalls wurde in vergangenen Zeiten auch auf die biologische Evolution angewandt.

Sie ist durch die hier diskutierten phänomenologischen Theorien von Hyperzyklus und Quasispezies (und deren experimentelle Bestätigung) nicht mehr relevant. Theorien einer Evolution des Universums wurden von Physikern ebenfalls bereits vorgeschlagen, doch sind sie mir zu dicht am biologischen Vorbild konzipiert. Die hier genannten phänomenologischen Theorien sind rein physikalischer Natur und müssen daher durch Parameter realisiert werden, die aus einer – noch nicht existierenden – allgemeinen Theorie physikalischer Wechselwirkungen gespeist werden. Eine solche Evolution hätte sich bereits in der ersten Sekunde der Evolution des Universums vollzogen. Das sind über 10^{40} kleinste Zeiteinheiten, die sogenannten »Planck-Zeiten«, und dies sind immerhin mehr »Zeiteinheiten, als für die biologische Evolution zur Verfügung standen«. »Heiße« Perioden waren die, in denen sich die verschiedenen Wechselwirkungen der Materie in »Phasenumwandlungen« selbständig machten, zuerst die Gravitation, dann die »starken Kernwechselwirkungen« und schließlich »schwache Kern-« und »elektromagnetische« Wechselwirkungen: Drei Phasenumwandlungen, die drei fundamentalen Naturkonstanten oder deren invarianten Kombinationen entsprechen. Aber das ist Zukunftsmusik, ein lohnendes Feld für zukünftige Generationen.

Anmerkungen

- 1 C. E. Shannon and W. Weaver, »The Mathematical Theory of Communication«, U. of Illinois Press, Urbana, Ill., 1949.
- 2 L. Brillouin, »Science and Information Theory«, Academic Press, New York, San Francisco, London, 1956, Second Edn., 1962.
- 3 M. Eigen, »From Strange Simplicity to Complex Familiarity«, Oxford University Press, in Vorbereitung.
- 4 M. Eigen, J. McCaskill and P. Schuster, »The Molecular Quasispecies«, Advances in Chemical Physics, Vol. 75, 1989, John Wiley and Sons, Inc, J. Phys. Chem., 1988, 92, 6881.

- 5 C. K. Biebricher, M. Eigen and W. C. Gardiner, »Kinetics of RNA Replication«, *Biochemistry*, 22, 2544, 1983; 23, 3186, 1984; 24, 6550, 1985.
- 6 B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J. D. Watson, »Molecular Biology of the Cell«, Garland Publ., Inc., New York and London, 1983.
- 7 M. Eigen and P. Schuster, »The Hypercycle, A Principle of Natural Self-Organization«, Springer-Verlag, Berlin, Heidelberg, New York, 1979.
- 8 M. Eigen, C. K. Biebricher, M. Gebinoga und W. C. Gardiner, »The Hypercycle. Coupling of RNA and Protein Biosynthesis in the Infection Cycle of an RNA Bacteriophage«. *Biochemistry* 1991, 30, 11005.
- 9 M. Eigen, B. F. Lindemann, M. Tietze, R. Winkler-Oswatitsch, A. Dress, A. von Haeseler, »How old is the Genetic Code? Statistical Geometry of t-RNA Provides an Answer«, *Science* 244, 673, 1989.