

## **Theorien und Verwendungsweisen des Informationsbegriffs in der Genetik**

Ulrich Stegmann

Seit den 50er Jahren macht die Genetik von informationellen Metaphern Gebrauch. Es ist vom „genetischen Code“ die Rede oder davon, dass die DNA-Sequenz „dekodiert“, „umgeschrieben“ oder „übersetzt“ wird. Aber solche Ausdrücke gelten nicht nur als Metaphern ohne theoretischen Gehalt. Genetiker und viele Entwicklungsbiologen sind der Auffassung, dass es tatsächlich so etwas wie genetische Information gibt. Sie soll in der DNA gespeichert sein und richtig oder falsch abgelesen werden können. Offenbar schreibt man der DNA also semantische Eigenschaften zu.

Information im Sinne der mathematischen Informationstheorie von Shannon und Weaver kann diese semantische Ebene jedoch nicht erfassen (z.B. Yockey 1992; Sterelny & Griffiths 1999; Maynard Smith 2000). Daraus ergibt sich die Frage, ob es eine alternative Präzisierung dieses Begriffs gibt, die seiner semantischen Verwendungsweise gerecht wird, oder ob der Ausdruck „genetische Information“ nur eine Metapher ist. Kann man Molekülen und rein kausalen, chemischen Abläufen sinnvollerweise Eigenschaften zubilligen, die wir sonst nur unseren mentalen Zuständen zugestehen?

In der gegenwärtigen philosophischen Debatte gibt es im wesentlichen zwei Explanationen von genetischer Information. Der teleosemantischen Theorie zufolge ist die genetische Information eines DNA-Abschnitts dessen biologische Funktion (Sterelny et al. 1996; Maynard Smith 2000). Biologische Funktion wird dabei im ätiologischen Sinn als die Wirkung eines Merkmals verstanden, aufgrund dessen dieses Merkmal in der Vergangenheit selektiert wurde. Wenn ein DNA-Abschnitt also selektiert wurde, weil er ein bestimmtes Enzym produziert, dann enthält er Informationen über dieses Enzym. Dieser Vorschlag hat eine gewisse Nähe zu teleosemantischen Theorien mentaler Gehalte in der Philosophie des Geistes (z.B. Millikan 1984; Papineau 1993), weil er semantische Information als biologische Funktion versteht und biologische Funktion wiederum ätiologisch interpretiert.

Der „kausalen“ Theorie zufolge ist die genetische Information eines DNA-Abschnitts ebenfalls dessen biologische Funktion. Aber die biologische Funktion wird mit der kausalen Rolle identifiziert, die dieser Abschnitt für die Entwicklung spielt (Godfrey-Smith 2000). Wenn ein DNA-Abschnitt die kausale Rolle hat, ein bestimmtes Enzym zu produzieren, dann enthält dieser Abschnitt die semantische Information für dieses Enzym.

Eine Schwierigkeit der teleosemantischen Theorie ist, dass ihr zufolge jede Struktur, die für eine bestimmte Funktion selektiert wurde, die Information für ihre Funktion enthalten sollte (Sterelny et al. 1996; Godfrey-Smith 1999; Griffiths 2001). Eine der absurden Konsequenzen wäre zum Beispiel, dass Beine die Information für Fortbewegung enthielten, weil sie für diese Wirkung selektiert wurden. Nach Sterelny, Smith und

Dickison (1996) konstituieren daher nur die biologischen Funktionen von Entwicklungsfaktoren teleosemantische Information. Diese Autoren akzeptieren die Konsequenz, dass auch nicht-genetische Entwicklungsfaktoren Informationen enthalten können, z.B. die an der Translation beteiligten Enzyme.

Allerdings sind Biologen nicht der Auffassung, dass diese Enzyme Informationen enthalten, zumindest nicht in dem Sinne wie die DNA. Zwar wird auch manchen nicht-genetischen Entwicklungsfaktoren der Besitz von Information zugesprochen, aber das gilt nur für einige wenige Faktoren wie z.B. die Konzentrations-Gradienten von Morphogenen (Wolpert et al. 2002). Meiner Ansicht nach ist die teleosemantische Theorie daher zu unspezifisch.

Die kausale Theorie der genetischen Information steht vor einer ähnlichen Schwierigkeit. Da sie semantische Information mit einer bestimmten kausalen Rolle identifiziert, sollten auch nicht-genetische Entwicklungsfaktoren Informationen enthalten, weil sie diverse kausale Rollen für die Entwicklung haben. Aber das stimmt nicht mit der Verwendungsweise informationeller Begriffe in der Genetik überein.

Godfrey-Smith versucht diese Konsequenz durch Einführung zusätzlicher Kriterien zu vermeiden, die eine kausale Rolle erfüllen muss, um semantische Information zu sein. Eines dieser Kriterien ist, dass die chemischen Bausteine von DNA und Proteine unterschiedlichen chemischen Stoffklassen angehören (Godfrey-Smith 2000). Aber dieses Kriterium macht die Theorie m.E. zu restriktiv. So gilt die Neusynthese eines DNA-Tochterstrangs bei der Zellvermehrung als Information übertragend, obwohl sie es der kausalen Theorie zufolge nicht sein sollte. Denn sowohl der DNA Eltern- als auch der Tochterstrang bestehen aus derselben Art von Bausteinen (Desoxyribonukleotide).

Die Schwierigkeiten der teleosemantischen und der kausalen Theorie der genetischen Information werfen die grundsätzliche Frage auf, unter welchen Bedingungen molekularen Prozessen legitimerweise genetische Information zugesprochen werden kann. Über ein Desideratum besteht Konsens. Eine befriedigende Theorie der genetischen Information muss zeigen, dass bestimmte molekulare Prozesse (oder Moleküle) zwei semantische Eigenschaften besitzen (Godfrey-Smith 1999; Sterelny & Griffiths 1999): Sie müssen Wahrheits- oder Erfüllungsbedingungen haben, damit Fehlrepräsentationen möglich sind, und die durch die Prozesse übertragenen Informationen müssen speicherbar sein. Wenn molekulare Entitäten tatsächlich semantische Informationen tragen, dann müssten sie diese Eigenschaften haben – in einer ihnen eigenen Weise.

Als ein weiteres Desideratum gilt offenbar die Naturalisierbarkeit semantischer Eigenschaften, denn sowohl die teleosemantische als auch die kausale Theorie setzen Naturalisierbarkeit voraus. Sie haben den Charakter reduktiver Analysen, weil sie eine Reihe nicht-semantischer Eigenschaften von molekularen Entitäten angeben, die konstitutiv für die erforderten semantischen Eigenschaften sein sollen. Diesen Ansatz verfolgt z.B. die teleosemantische Theorie bei der Rekonstruktion von Fehlerhaftigkeit. Ein Merkmal ist dann „fehlerhaft“, wenn es nicht der biologischen Funktion eines DNA-Abschnitts entspricht, zur Ausbildung dieses Merkmals beizutragen. Ein DNA-Abschnitt

trägt semantische Information nur über solche Kausalfolgen, die er (biologisch gesprochen) haben soll.

Das allein reicht aber nicht aus. Eine philosophische Theorie der genetischen Information sollte im wesentlichen auf diejenigen Prozesse oder Moleküle zutreffen, die Biologen für tatsächlich informationell halten – und sie sollte nicht auf diejenigen Prozesse zutreffen, die als nicht-informationell gelten. Denn ihre Aufgabe besteht nicht darin, eine stipulative Definition von genetischer Information zu liefern, sondern die Präzisierung eines schon eingeführten Begriffs. Dieses Desideratum kommt z.B. in der bereits erwähnten Kritik an der teleosematischen Theorie zum Tragen: Ihr wird vorgeworfen, dass sie Entwicklungsfaktoren den Besitz von Informationen zuspricht, die nach biologischer Auffassung keine Informationen tragen.

Aus diesen Desiderata ergeben sich drei Aufgaben für die Entwicklung einer Theorie der genetischen Information. Die erste Aufgabe ist die Klärung der Verwendungsweise des Ausdrucks „genetische Information“ in der Genetik und Entwicklungsbiologie. Dadurch könnte zum einen festgestellt werden, worüber die genetische Information eines DNA-Abschnitts sein soll (über Proteine, phänotypische Merkmale etc.?). Zum anderen kann geklärt werden, ob nur die DNA als Informationsträger gilt. Außerdem könnte Aufschluss über die Prozesse gewonnen werden, bei denen Information übertragen oder durch die die gespeicherte Information umgesetzt werden soll.

Die zweite Aufgabe sehe ich in der Charakterisierung der nicht-semantischen Eigenschaften dieser Prozesse und ihre Ursachen. Das Ziel ist zu prüfen, ob die als informationell beschriebenen Prozesse Gemeinsamkeiten nicht-semantischer Art haben, die nur ihnen eigen sind. Vielleicht interagieren in diesen Prozessen nur Moleküle unterschiedlicher Stoffklassen miteinander oder die Interaktionen bestehen immer in direkten, stereospezifischen Wechselwirkungen etc. Zeichnen sich solche Gemeinsamkeiten ab, dann sollte geprüft werden, ob sie für die semantischen Charakteristika konstitutiv sind. Diese dritte Aufgabe besteht in einer Rückführung der semantischen Charakteristika auf die zuvor identifizierten nicht-semantischen Eigenschaften.

Abschließend ein Vorschlag zur Verwendungsweise von „genetischer Information“. Ich denke dieser Ausdruck wird in mindestens drei unterschiedlichen Weisen gebraucht, die sich recht klar voneinander abgrenzen lassen. Sie unterscheiden sich unter anderem in den angenommenen Informationsträgern und in der Art und Weise, wie diese Information in den Informationsträgern enthalten sein soll.

Die vielleicht bekannteste Gruppe von Hypothesen, die mit dem Informationsbegriff operiert, ist das sogenannte „Zentrale Dogma der Molekularbiologie“. Es wurde erstmals von Francis Crick formuliert (Crick 1958) und macht im wesentlichen zwei Typen von Aussagen: zum einen Aussagen über die molekularen Entitäten, die genetische Information enthalten und zum anderen Aussagen über die Richtung von Informationstransfers. Zum Beispiel enthalten dem Zentralen Dogma zufolge sowohl DNA als auch RNA Informationen, denn sie können Quellen von Informationstransfers sein, und die Richtung der Informationstransfers weist von Nukleinsäuren (DNA, RNA) zu Nukleinsäuren oder Proteinen, aber niemals von Proteinen zu Nukleinsäuren.

James Watson hat am klarsten auf die molekulare Basis dieser Informationstransfers hingewiesen (Keyes 1999). In seinem Lehrbuch (Watson 1965) werden Cricks Informationspfeile mit konkreten molekularen Prozessen verbunden, z.B. der Informations-transfer von DNA zu RNA mit dem Prozess der Transkription. Die Transkription und die anderen als informationell angesehenen Prozesse (z.B. Replikation, Translation) haben eine entscheidende Gemeinsamkeit: Es wird immer ein neues Molekül aus seinen Bausteinen synthetisiert, und dabei dient ein bereits vorhandenes Makromolekül als Matrize. Letztere bestimmt, welche Bausteine an welcher Position im Produkt eingebaut werden. Bei der Replikation wird ein neuer DNA-Strang aus freien Nukleotiden, bei der Translation ein neues Peptid aus einzelnen Aminosäuren synthetisiert. Der Informationstransfer von DNA zu DNA soll sich also durch Produktion eines neuen DNA-Strangs aufgrund eines bereits vorhandenen Strangs vollziehen (sog. Matrizen-gesteuerte Synthese).

Die Klärung der Verwendungsweise von „genetischer Information“ im Kontext des Zentralen Dogmas zeigt, dass Informationstransfers eine Reihe von Bedingungen *nicht* voraussetzen: Erstens ist es nicht nötig, dass die Matrizen- und Produkt-Komponenten indirekt miteinander agieren. Zum Beispiel gilt die Replikation als Realisierung eines Informationsflusses, obwohl die Komponenten direkte, stereospezifische Wechselwirkungen miteinander eingehen. Zweitens müssen die Matrizen-Komponenten nicht einer anderen chemischen Klasse als die Produkt-Komponenten angehören. Bei der Replikation zum Beispiel stammen beide aus derselben Stoffklasse (Desoxyribonukleotide). Dieses Ergebnis widerspricht einer These Godfrey-Smiths (2000) über die übliche Verwendung des Ausdrucks „genetische Information“.

Im Zusammenhang mit der Regulation von Transkription und Translation wird „genetische Information“ anders verwendet. Dies kann am Beispiel der Startsignale für die Transkription näher erläutert werden. Die DNA soll z.B. Signale über die Position enthalten, an der mit der Transkription eines bestimmten Gens begonnen und an der sie abgebrochen wird (Alberts et al. 2002). Sie soll Instruktionen wie „Starte mit der Transkription an dieser Stelle!“ übermitteln. Die Signale bestehen aus bestimmten Abschnitten der DNA und liegen in der Kontrollregion eines Gens, dem Promotor. Im Promotor des Bakteriums *E. coli* gibt es zwei charakteristische (sog. Consensus-) Sequenzen, die durch einen Abschnitt getrennt sind, dessen Sequenz stark variiert, aber dessen Länge konserviert ist. Die Signalsequenzen liegen ihrerseits in einem bestimmten Abstand zur Startstelle der Transkription und markieren sie dadurch. An die Signalsequenzen können nun bestimmte Proteine binden. Nachdem z.B. die Untereinheit eines Proteins an die TATA-Box (eine Consensus-Sequenz) von Eukaryoten gebunden hat, können sich weitere Proteine anlagern. Das führt schließlich zu einem Transkriptionskomplex, der auch die DNA-Polymerase II enthält, welche die Transkription in Gang setzt.

Nach biologischer Auffassung wird die Information über die Startposition der Transkription durch die Bindung eines Proteins an eine Signalsequenz übermittelt, weil die Position dieser Sequenz bestimmt, an welcher Stelle die Transkription beginnt. Die

Übermittlung dieser „Signal-Information“, wie ich sie nennen möchte, wird daher offenbar in der spezifischen Bindung eines Proteins an die entsprechende Signalsequenz gesehen. Bei dieser Art der Informationsübertragung wird kein Molekül synthetisiert, weil die RNA erst gebildet wird, nachdem das Signal an die Polymerase übermittelt wurde. Dadurch unterscheidet sich die Signal-Information von der „Matrizen-Information“. Die Übertragung von Matrizen-Information basiert auf der Synthese eines Makromoleküls, die Übertragung von Signal-Information dagegen auf der spezifischen Bindung zwischen zwei Molekülen.

Es gibt weitere Unterschiede: 1. Im Bereich der Signalsequenzen binden die Proteine an die doppelsträngige DNA. Das bindende Protein identifiziert diese DNA-Abschnitte nicht nur anhand ihrer Basensequenz, sondern auch anhand anderer Eigenschaften, z.B. anhand der Länge des DNA-Abschnitts zwischen den beiden Consensus-Sequenzen. Im Gegensatz zur Matrizen-Information ist also die Bedeutung der Signal-Information zumindest teilweise unabhängig von der Basensequenz. 2. Die Realisierung der Signal-Information kann (muss aber nicht) die Synthese eines Moleküls beinhalten. Die Information „Starte mit der Transkription an dieser Stelle!“ wird durch den Start der Transkription an der angegebenen Position umgesetzt, beinhaltet also die Synthese eines neuen Moleküls. Dagegen besteht die Umsetzung der Information von sogenannten Stoppsignalen gerade im Abbruch der Molekül-Synthese.

„Genetische Information“ wird noch auf eine dritte Weise verwendet – genetische Information im Sinne eines genetischen Programms. Molekularbiologen wie James Bonner, Francois Jacob und Sydney Brenner schlugen in den späten 60er Jahren vor, dass die DNA nicht nur Informationen über Proteine, sondern auch über die gesamte Entwicklung eines Organismus enthält (Keller 2000). Im Gegensatz zum Zentralen Dogma war die Existenz so eines Programms von Anfang an umstritten.

Die Vertreter des genetischen Programms lassen vielfach offen, was genau in diesem Programm festgelegt ist. Aber eine Vorstellung ist, dass es die beiden wesentlichen Eigenschaften genregulatorischer Netzwerke kodiert: Die Identität der transkribierten Gene und die Bedingungen, unter denen sie transkribiert werden (Bolouri & Davidson 2002). Die so abgelegte Information soll konditionale Form haben, z.B. „Wenn Bedingungen <A und B>, dann synthetisiere <Gen X>“. Informationen dieser Art sollen durch die Art und Anzahl der regulatorischen Sequenzen eines Gens festgelegt sein (Alberts et al. 2002). Binden bestimmte Moleküle (Transkriptionsfaktoren) an solche spezifischen Bindungsstellen, dann kann die Transkription des entsprechenden Gens gehemmt oder aktiviert werden. Zur Illustration sei angenommen, dass Gen A ein Protein „produziert“, welches an eine Bindungsstelle von Gen C bindet, und Gen C exprimiert daraufhin sein Genprodukt. Wenn Gen B die Bindungsstelle für das Genprodukt von Gen A fehlt, dann wird es (im einfachsten Fall) nicht exprimiert.

Die Beschaffenheit der mit einem Gen verbundenen Regulationssequenzen legt also fest, welcher Transkriptionsfaktor binden muss, damit das Gen transkribiert wird oder nicht – oder welche Kombination von Faktoren für die ein oder andere Wirkung erforderlich ist. Im Falle des fiktiven Gens C legt die Regulationssequenz eine hinreichende

Bedingung für die Expression dieses Gens fest, nämlich dass es exprimiert wird sofern Gen A exprimiert wird.

Diese „Programm-Information“ ist sowohl von der Matrizen- als auch von der Signal-Information zu unterscheiden. Erstens richtet sich die Programm-Information nicht an einen bestimmten Faktor. Denn es gibt nicht einen Faktor, der an alle Bindungsstellen bindet, und deswegen kann diese Information nicht einfach durch die Bindung eines Faktors übermittelt werden. Dagegen richtet sich die Signal-Information an einen Faktor, z.B. an die RNA-Polymerase, die dann auch die signalisierte Anweisung ausführt. Im Falle der Programm-Information fehlt dieser Faktor.

Der zweite, vielleicht noch wichtigere Unterschied betrifft die molekularen Mechanismen, durch welche die enthaltene Information ausgeführt wird. Die Information in obigem Beispiel ist „Wenn Bedingung <Genprodukt A>, dann transkribiere <Gen C>“. Sie wird durch das gemeinsame Vorliegen mehrerer Bedingungen realisiert, d.h. sie wird realisiert, wenn Gen A exprimiert wird *und* das Genprodukt A an die entsprechende regulatorische Bindungsstelle von Gen C bindet *und* Gen C transkribiert wird. Für die Realisierung dieser Information sind damit zwei unterschiedliche Mechanismen erforderlich, die spezifische Bindung von Transkriptionsfaktoren an die DNA und die Matrizen-geleitete Synthese. Im Gegensatz dazu beinhaltet die Ausführung von Signal- und Matrizen-Information nur jeweils einen dieser Prozesse.

Aus dieser Untersuchung der Verwendungsweisen des Ausdrucks „genetische Information“ ziehe ich den Schluss, dass der Informationsbegriff in der Genetik und Entwicklungsbiologie mindestens drei unterschiedliche Gruppen molekularer Prozesse herausgreift: Die Matrizen-gesteuerte Synthese von Makromolekülen, die hochselektive Bindung von Molekülen an bestimmte Abschnitte von Nukleinsäuren und jene Prozesse, die sowohl aus Matrizen-gesteuerter Synthese als auch aus spezifischen molekularen Bindungen bestehen.

Eine wichtige Konsequenz dieser These ist, dass es drei unterschiedliche, aber jeweils adäquate Explikanda von genetischer Information geben kann. Manche der Schwierigkeiten der bisherigen Theorien genetischer Information resultieren möglicherweise aus der unzutreffenden Annahme, dass „genetische Information“ einheitlich verwendet wird.

## Literatur

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. 2002: *Molecular Biology of the Cell*. New York: Garland Science.
- Bolouri, H. & Davidson, E. H. 2002: Modeling transcriptional regulatory networks. *BioEssays* 24, 1118–1129.
- Crick, F. 1958: On Protein Synthesis. *Symposia of the Society for Experimental Biology* 12, 138–167.
- Godfrey-Smith, P. 1999: Genes and Codes: Lessons from the Philosophy of Mind? In *Where biology meets psychology: philosophical essays* (ed. V. G. Hardcastle), pp. 305–331. Cambridge, MA: MIT Press.

- Godfrey-Smith, P. 2000: On the Theoretical Role of “Genetic Coding”. *Philosophy of Science* 67, 26–44.
- Griffiths, P. E. 2001: Genetic Information: A Metaphor in Search of a Theory. *Philosophy of Science* 68, 394–412.
- Keller, E. F. 2000: Decoding the Genetic Program: Or, Some Circular Logic in the Logic of Circularity. In *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives* (eds. P. J. Beurton, R. Falk & H.-J. Rheinberger), pp. 159–177. Cambridge: Cambridge University Press.
- Keyes, M. E. 1999: The Prion Challenge to the ‘Central Dogma’ of Molecular Biology, 1965–1991. Part I: Prelude to Prions. *Studies in the History and Philosophy of Biology and Biomedical Sciences* 30, 1–19.
- Maynard Smith, J. 2000: The Concept of Information in Biology. *Philosophy of Science* 67, 177–194.
- Millikan, R. G. 1984: *Language, Thought and Other Biological Categories*. Cambridge: MIT Press.
- Papineau, D. 1993: *Philosophical Naturalism*. London: Blackwell.
- Sterelny, K. & Griffiths, P. E. 1999: *Sex and Death: An Introduction to Philosophy of Biology*. Chicago: University of Chicago Press.
- Sterelny, K., Smith, K. & Dickison, M. 1996: The Extended Replicator. *Biology and Philosophy* 11, 377–403.
- Watson, J. D. 1965: *Molecular Biology of the Gene*. 1<sup>st</sup> Ed. New York: W. A. Benjamin.
- Wolpert, L., Beddington, R., Jessel, T., Lawrence, P., Meyerowitz, E. & Smith, J. 2002: *Principles of Development*. Oxford: Oxford University Press.
- Yockey, H. P. 1992: *Information Theory and Molecular Biology*. Cambridge: Cambridge University Press.