

6. Evolution

Jean Baptiste de Lamarck (1744-1829)

- prägte den Begriff "Biologie"
- biologische Arten sind über lange Zeiträume veränderlich: Theorie einer allmählichen "Evolution" (Entwicklung)
- die Veränderungen haben mit funktionaler Anpassung zu tun
- die Veränderungen hängen ab von der Intensität des Gebrauchs bestimmter Organe (Bsp. Giraffenhals) – Weitergabe erworbener Eigenschaften

die letztgenannte Annahme wurde von späteren Evolutionstheorien verworfen

(kann aber in künstlichen Systemen realisiert werden: Systeme mit "Lamarckscher Evolution")

Charles Darwin (1809-1882)

London 1859: "The Origin of Species"

(Alfred R. Wallace war unabhängig zu ganz ähnlichen Ergebnissen gekommen)

- biologische Arten durch generationenübergreifenden Entwicklungsprozess aus anderen biologischen Arten entstanden
- die Veränderung steht in Zusammenhang mit der Anpassung (*adaptation*) der Individuen einer Art an die Erfordernisse ihrer Umwelt
- Darwin: "This principle of preservation or the *survival of the fittest*, I have called Natural Selection" (natürliche Selektion / Auslese)
- Paradigmenwechsel der Biologie: von beschreibender und ordnender zu erklärender und nach Ursachen forschender Wissenschaft

beachte: Darwin wusste nichts über Genetik, Mutationen oder DNA!

Was stellt das "Prinzip der natürlichen Auslese" dar?

- eine Tautologie (es überleben die, die am besten überleben können)?
 - einen Erfahrungssatz über eine Eigengesetzlichkeit speziell der *lebenden* Systeme, wie wir sie kennen?
 - eine Gesetzmäßigkeit, die lediglich bei Lebewesen bevorzugt in Erscheinung tritt, aber nicht auf diese beschränkt ist?
- Darwin hat sich nicht klar geäußert
 - heute vertritt man die 3. Position: Darwinsches Selektions- und Evolutionsverhalten lässt sich im Labor und im Computer reproduzieren

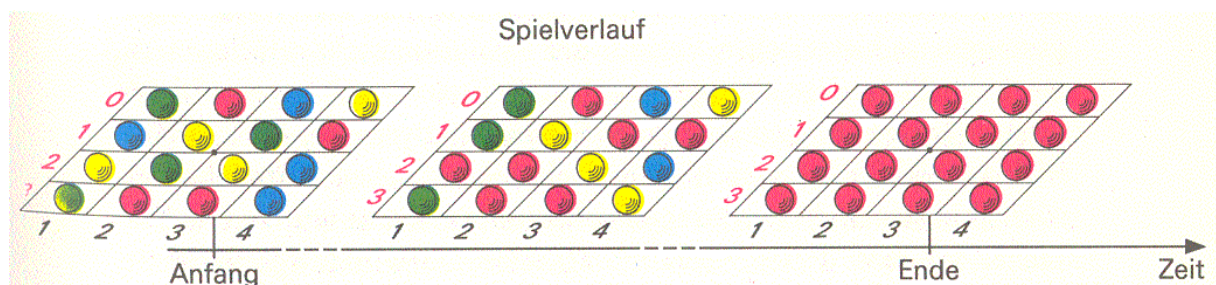
Aber: die Bedingungen

- für Auftreten von Selektion
- für den Zusammenhang von Anpassung und Selektion sind nicht trivial

Beispiel:

Kugelspiel "Selektion" (nach Eigen & Winkler 1979)

- gespielt wird auf einem endlichen quadratischen Gitter
- auf jedes Gitterfeld passt genau eine Kugel
- es gibt eine endliche Zahl von Kugelfarben (z.B. 4)
- zu Beginn sind alle Farben in gleicher Menge auf dem Gitter vertreten und regellos verteilt; alle Felder sind besetzt
- dann werden – strikt abwechselnd – die beiden folgenden Regeln angewandt:
- es wird zufällig eine Position im Gitter bestimmt; die dort befindliche Kugel wird entfernt
- es wird eine weitere Position bestimmt; die dort befindliche Kugel wird verdoppelt, ihr Duplikat kommt auf den leeren Platz vom Schritt vorher.



Ergebnis: es kommt immer zur Selektion *einer* Kugelfarbe, die schließlich das gesamte Gitter füllt.

⇒ in diesem Fall wäre der Begriff "fittest" für die selektierte Kugelfarbe tatsächlich tautologisch (es hängt nur vom Zufall ab, welche Farbe "gewinnt" – kein Einfluss von Umwelt und Anpassung!)

Variante des Kugelspiels: wenn man bei jedem "Reproduktionsschritt" mit gewisser Wahrscheinlichkeit eine "Mutation" (zufällige Änderung der Farbe) zulässt, findet *keine* eindeutige Selektion mehr statt!

Merke:

- nicht immer hängt Selektion mit Anpassung zusammen
- nicht immer begünstigt das Vorkommen von Mutationen die Selektion

zu Darwins Zeit Unklarheit über die Vererbung

- Annahme einer Vermischung der (quantitativen) Eigenschaften bei der Vererbung (*blending inheritance*)
- dann müssten sich mit der Zeit alle Unterschiede nivellieren

Neodarwinismus (Beginn 20. Jh.):

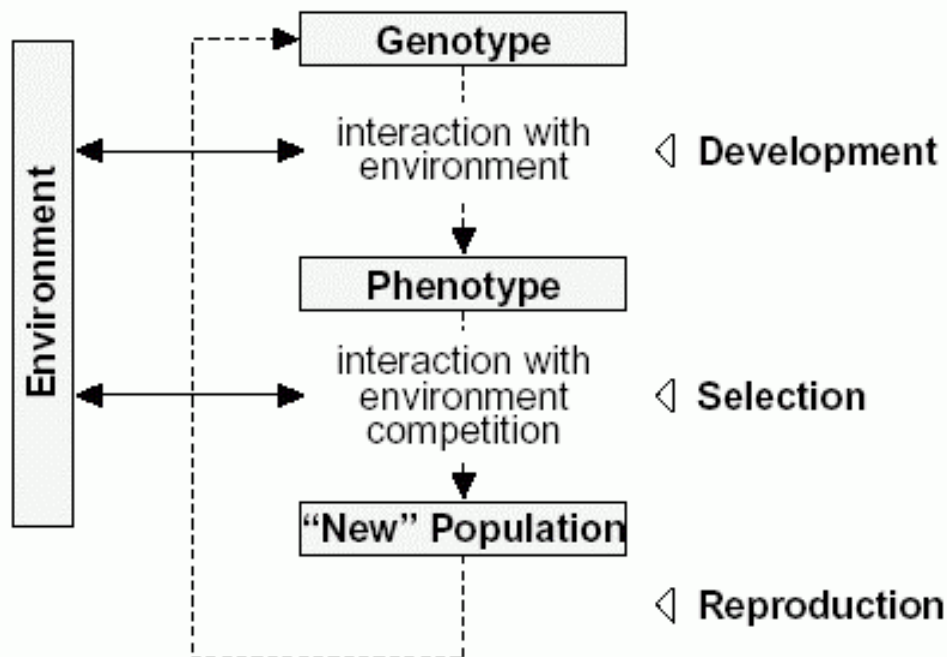
- Integration der Mendelschen Genetik (Vorliegen diskreter Allele, Dominanz etc.), der Zellbiologie und Embryologie in die Darwinsche Theorie
- *Weissmann-Prinzip*: der Genotyp wirkt auf den Phänotyp, nicht umgekehrt (Ausschluss Lamarck'scher Vererbung)

"Synthetische Theorie" (bis heute):

- Integration von Populations- und Molekulargenetik
- Erkennen der Rolle von Mutationen und crossing over
- Evolution wird definiert als die Änderung der Häufigkeit von Genen innerhalb des "Genpools" einer Population von Individuen
- John B. S. Haldane, Ronald A. Fisher, Sewall Wright: exakte Herleitung der Wirkung natürlicher Selektion mit math. und populationsgenetischen Mitteln

Angenommen, ein einzelnes Gen in einer Population hat zwei Allele A und a , und A verschafft seinem Träger einen geringfügigen Selektionsvorteil. Kann die auf die Population einwirkende natürliche Selektion dann die Häufigkeit des Allels A erhöhen? – Ja (Haldane, Fisher, Wright)
d.h.: Darwinsche Evolution "kann funktionieren".

Der Prozess der Evolution im Verständnis der "synthetischen Theorie" (nach Pfeifer et al. 2002):



Aber: vieles bleibt unklar!

Stuart Kauffman:

"Fast 140 Jahre nach Darwins zukunftsweisendem Buch kennen wir immer noch nicht die Leistungsfähigkeit und die Grenzen der natürlichen Selektion, wissen immer noch nicht, welche Arten komplexer Systeme durch einen Evolutionsprozess zusammengebaut werden können..."
(Kauffman 1995)

Nichttriviale Fragen:

- unter welchen Bedingungen führt Darwinsche Evolution zu "verbesserten" (optimalen?) Lösungen?
- was ist unter der selektierenden "Umwelt" zu verstehen? (für eine Zelle innerhalb eines Organismus etwas anderes als für den Organismus selbst...) – mehrere Skalenebenen
- welche Rolle spielen selektionsneutrale Änderungen (genetische Drift)? Nach der Theorie von Motoo Kimura eine sehr wichtige...
- "wieviel" kann in der Evolution auf einmal verändert werden? Ist die Evolution stetig oder schubweise ("Gradualismus" versus "Punktualismus")? Macht die Evolution "Sprünge" ("Saltationismus")?
- wie "zufällig" sind Mutationen? Es ist bekannt, dass sich bestimmte Teile des Genoms häufiger verändern als andere, und dass es Schutzmechanismen gegen Mutationen gibt – "Evolution der Mutationsraten"? Steuert die Evolution ihren eigenen "Motor"?
- was ist die "Zielscheibe" der Evolution (das Angriffsstück der Veränderung)?
 - das einzelne Gen
 - das Genom eines Individuums
 - der Genpool einer Population
 - eine biologische Art
 - ein ganzes Ökosystem
 - die ganze Erde als "Super-Organismus"

die beiden Extrempositionen sind unter den Bezeichnungen "das egoistische Gen" (*The selfish gene*; R. Dawkins) und "Gaia-Hypothese" (J. Lovelock) bekannt geworden.

Richard Dawkins: "Das egoistische Gen" (1976, dt. 1978): die Theorie legt nahe, dass die Evolution primär auf Gene wirkt

– dies äußert sich u.a. in Verhaltensphänomenen wie dem Töten von Löwenjungen durch ein Männchen, das ein Löwenrudel als neues Alpha-Männchen übernommen hat (so wird die maximale Ausbreitung der eigenen Gene sichergestellt – entgegen den "Interessen" der Population oder der Art)

aber wie kommt dann "altruistisches Verhalten" zustande, von dem es auch Beispiele im Tierreich gibt?

- Evolution von kooperativen Strategien (Axelrod et al.) – siehe Kapitel 9

Die *selfish gene* - Hypothese bietet hier Erklärungen an

→ Soziobiologie

→ der Organismus (auch der Mensch) ist im hohen Maße "Sklave seiner Gene", ja, lediglich deren "Verbreitungsmaschine"

aber: es gibt auch Kritik an dieser einseitigen Interpretation (vgl. Hemminger 1994):

- Gene steuern sich gegenseitig und bilden Netzwerke
- es gibt in der Natur auch kooperative Systeme, deren Untereinheiten nicht dasselbe genetische Material teilen (wie es das Erklärungsmodell der Soziobiologie fordert)
- der Selektionsvorgang ist ein dynamisches Gesamtgeschehen, an dem sowohl die Gene, als auch die Organismen, als auch deren Umwelt teilhaben
- ebenso einseitig könnte man vom "Egoismus der Organismen" sprechen, dem die Gene dienen müssen (denn Gene, die zu geringer Fitness der Organismen führen, werden in der Population eliminiert)
- Dawkins wertet den kurzlebigen Organismus als "weniger wichtig" als das langlebige Gen – metaphysische Wertung
- "Universalitätsbedürfnis" einiger Biologen? ("Wir können alles erklären – Psychologie, Soziologie etc. werden überflüssig")

Immerhin postuliert Dawkins in der menschlichen Kultur "Meme", die, analog zu Genen in der Biologie, eine eigene Ebene der Replikation konstituieren (und ebenfalls die Menschen für ihre Ausbreitung funktionalisieren: durch Imitationslernen).

- "Egoistisches Gen" vs. "Gaia-Hypothese": Variante des Konflikts zwischen Reduktionismus und Holismus

Wichtig ist:

- die grundlegenden Begriffe "genetische Variabilität" und "Selektion" machen nur Sinn in der Anwendung auf *Populationen*, d.h. wenn mehrere Individuen vorliegen:
- Parallelismus der Individuen einer Population als Voraussetzung Darwinscher Evolution
- Vererbung als Form zeitlicher Persistenz in diskreten Iterationsschritten (des Vorgangs der Replikation)

⇒ sowohl *Parallelismus* als auch *Iteration* sind essenziell für Darwinsche Evolution (aus algorithmischer Sicht)

- Begrenzung der Ressourcen, allgemeiner gesagt: *Endlichkeit* ist Voraussetzung für Möglichkeit der Selektion (u. damit für Evolutionsfähigkeit)

auch die *Sterblichkeit* der höheren Organismen ist aus dieser Perspektive zu sehen!

"genomische" vs. "nichtgenomische" Evolution:

- genomische Evolution: lässt sich anhand von Populationen von Zeichenketten (Nukleotidsequenzen) beschreiben

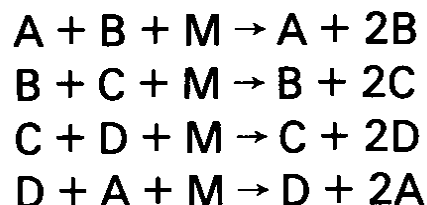
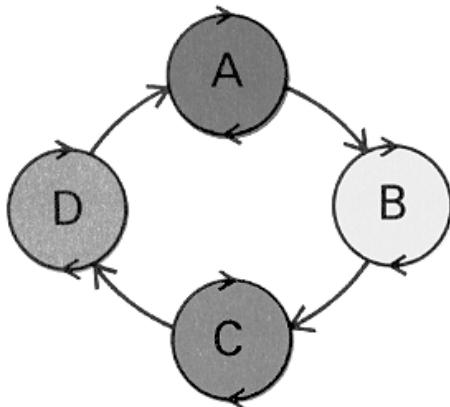
Trennung von Genotyp und Phänotyp, Weissmann'sches Prinzip

- nichtgenomische Evolution: keine Trennung von Genotyp und Phänotyp möglich

Beispiele für lebensnahe Systeme, bei denen Genotyp = Phänotyp ist: *autokatalytische Netzwerke* (siehe Kapitel 3)

Einfache Beispiele solcher Netzwerke:
"Hyperzyklen" (Eigen, Schuster)

jeder Reaktand katalysiert die eigene Entstehung und die des nächsten Reaktanden im Zyklus



(aus Eigen & Winkler 1979)

Solche Systeme können leicht mit "Kugelspielen" auf einem CA-artigen Gitter simuliert werden

(Kugel darf nur dann verdoppelt werden, wenn auf einem der Nachbarfelder eine Kugel der im Zyklus "vorangehenden" Farbe liegt)

– es kommt typischerweise zu Oszillationen (vgl. Belousov-Zhabotinsky-Reaktion)

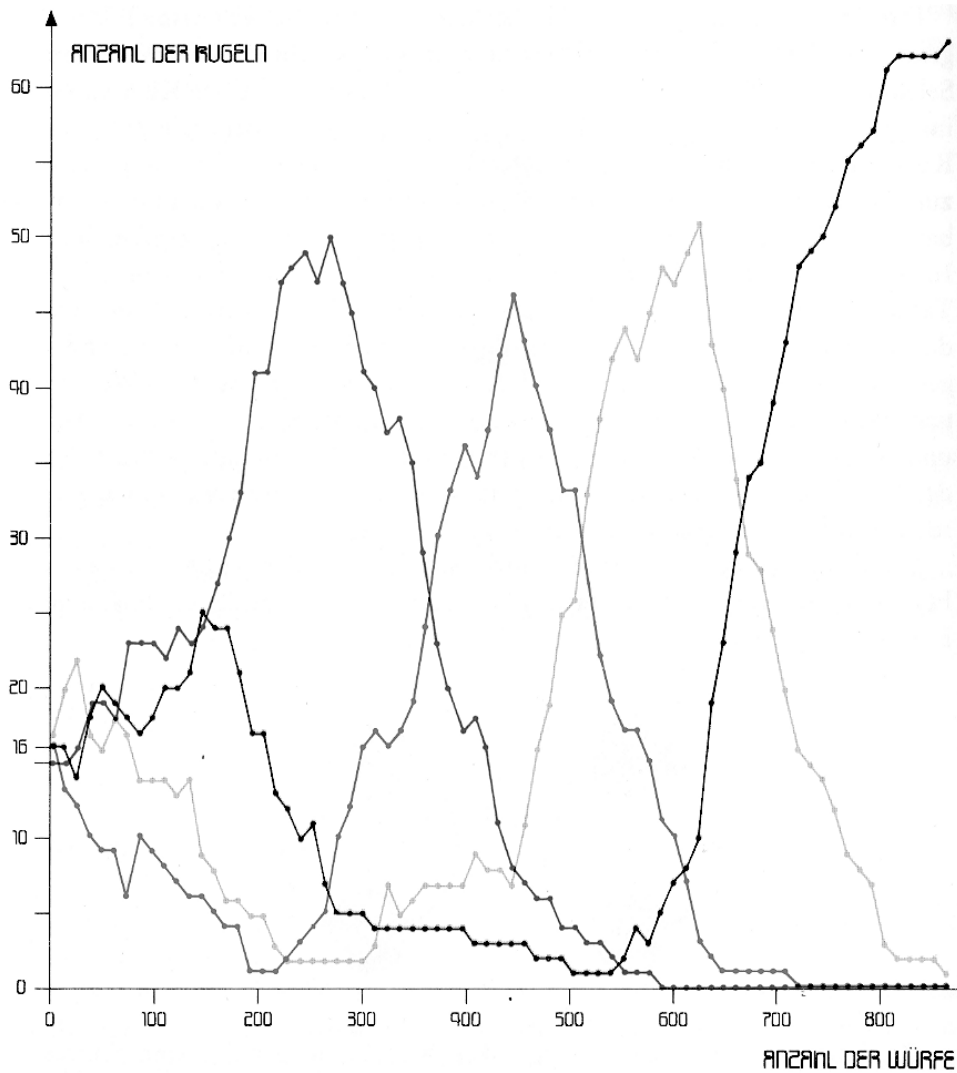
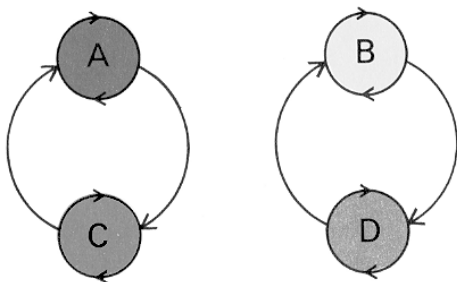


Abb. 52 Computeraufzeichnung eines »Hyperzyklus«-Spiels. Das Diagramm zeigt, wie sich die einzelnen Populationen von A (Rot), B (Gelb), C (Grün) und D (Blau) in Abhängigkeit von der Zahl der Würfe verändern. Der Zyklus stirbt aufgrund einer Fluktuationkatastrophe aus.

Eigenschaften von Hyperzyklen:

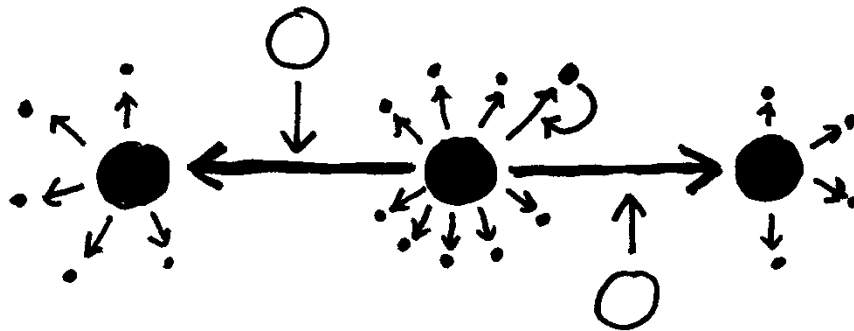
- die miteinander verkoppelten "Individuen" sind aufeinander angewiesen und damit koexistent. Im Existenzbereich des gesamten Zyklus ist jedes einzelne Mitglied stabil.
- Der Zyklus als ganzes ist nach außen hin "wettbewerbsfreudig" und fordert eine irreversible "Alles-oder-Nichts"-Entscheidung heraus



→ nur (A,C) oder (B,D) "überlebt"

Wie können autokatalytische Netzwerke evolvieren?
("Verdrängungswettbewerb" allein würde nichts Neues aufbauen können)

"chemischer Schatten" eines Substrats:
durch "zufällige", nichtkatalysierte (langsame) Reaktionen treten zu jedem Reaktanden des Netzwerks weitere Moleküle (Reaktionsprodukte) auf

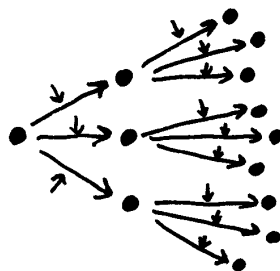


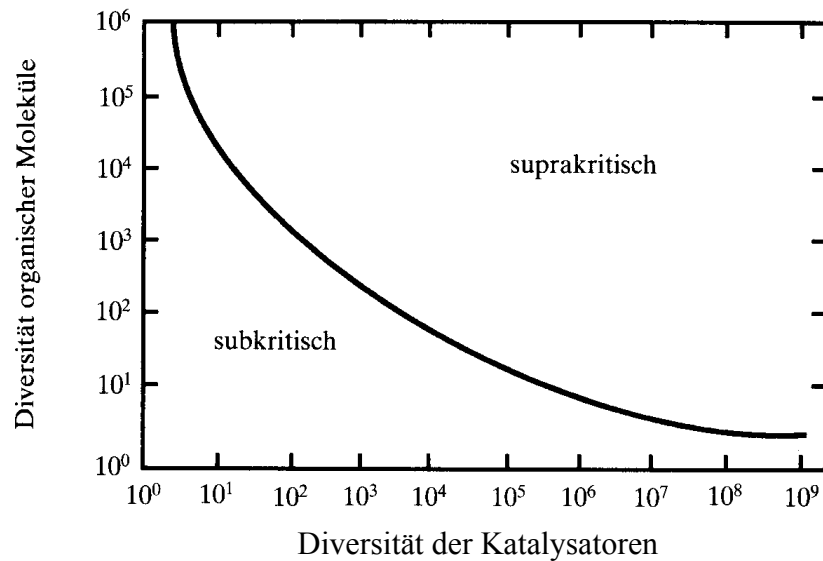
wenn eines davon die eigene Entstehung (direkt oder indirekt) katalysiert, kann es zu neuem Mitglied des Netzwerks werden – "Mutation"

Zahl der möglichen Reaktionen steigt stärker (quadratisch?) als die Zahl der vorhandenen Reaktionspartner (vgl. Kap. 3)
⇒ "kritische Diversität" der Molekülarten ist nötig, damit autokatalytische Abgeschlossenheit spontan entsteht

Regimes des dynamischen Verhaltens:

- "subkritisches Verhalten" einer "Molekül-Suppe": Diversität zu gering, es wird zu wenig katalysiert, kaum neue Moleküle entstehen
- "suprakritisches Verhalten": immer neue Moleküle werden durch katalysierte Reaktionen erzeugt, diese wirken ihrerseits katalytisch... lawinenartig entstehen neue Moleküle!





- Die Zellen der lebenden Organismen müssen subkritisch sein (sonst wäre kein geregelter Zellstoffwechsel möglich)
- Vermutung (St. Kauffman): die Biosphäre der Erde als Ganzes ist suprakritisch! (damit: Evolution "offen")

Stuart Kauffman:

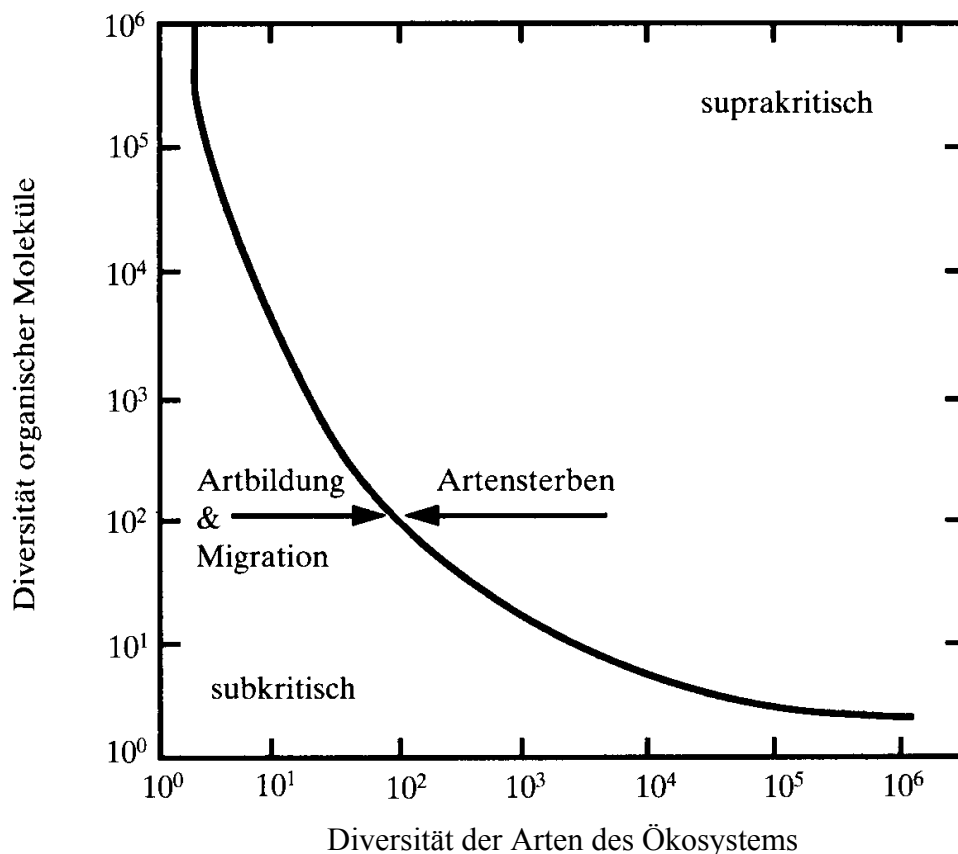
"Die Tatsache, dass wir unsere Speisen essen und nicht mit ihnen verschmelzen, weist... auf einen grundlegenden Sachverhalt hin: Die Biosphäre selbst ist suprakritisch. Unsere Zellen dagegen sind subkritisch. Würden wir mit dem Salat verschmelzen, dann würde die molekulare Diversität, die eine solche Verschmelzung in unseren Zellen erzeugt, eine umwälzende suprakritische Explosion auslösen. Die explosionsartige Zunahme neuer Molekülarten wäre für die unglückseligen Zellen, in denen dieser Prozess stattfände, schon nach kurzer Zeit tödlich. Die Tatsache, dass wir essen, ist kein Zufall, nicht einfach eine der zahlreichen möglichen Methoden, auf die die Evolution gekommen ist, um neue Moleküle in unsere metabolischen Netzwerke einzuspeisen. In den Vorgängen Essen und Verdauen spiegelt sich... die Notwendigkeit wider, uns vor der suprakritischen molekularen Diversität der Biosphäre zu schützen." (Kauffman 1995, S. 189)

- Bedeutung der Zellmembranen (vergleichbar den Graphitstäben in einem Kernreaktor, die ungebremste Kettenreaktion verhindern!)
- Bedeutung des Immunsystems

- subkritisch-suprakritische Grenze gibt obere Schranke für die molekulare Diversität einer Zelle \Rightarrow die Komplexität der Zelle ist begrenzt

Hypothesen (Kauffman, S. 198):

- Ökosysteme evolvieren zur subkritisch-suprakritischen Grenze und verharren dann dort, weil sie durch evolutionäre "Gelegenheiten" aus dem subkritischen Regime und durch Extinktionsdruck aus dem suprakritischen Regime herausgedrängt werden.



- Während jedes einzelne Ökosystem an der Grenze zwischen Sub- und Suprakritizität angesiedelt ist, erzeugen die Ökosysteme in ihrer Gesamtheit aufgrund ihrer stofflichen Wechselwirkungen eine suprakritische Biosphäre, die zwangsläufig immer komplexer wird.

genomische Evolution:

- hat sich als Spezialfall der nichtgenomischen entwickelt
- RNA, DNA und Proteine als spezielle Molekülklassen, die Spezialfunktionen in einem autokatalytischen Netzwerk übernehmen
- auch diese Separierung der Funktionen schützt die Zelle vor Abgleiten in die Suprakritizität
- Herausbildung dieser Aufgabenteilung vermutlich sehr früh in der Geschichte des Lebens
- Selektion in dieser Phase nach "Alles-oder-Nichts"-Prinzip: Universalität des genetischen Codes!

Beschreibung der Erbinformation hier in Form von Strings möglich

→ vereinfacht Beschreibung und theoretische Behandlung solcher Systeme

- thermodynamische Beschreibung
- einfache Manipulierbarkeit der "Genome" in ALife-Experimenten
- Beschreibung der genet. Struktur konkreter Populationen mittels Distanzmaßen zwischen Zeichenketten
- Fitness als direkte "Funktion" des Genom-Strings
- Genetische Algorithmen

wir müssen uns in den folgenden Abschnitten klar sein, dass wir nur eine Teilmenge möglicher evolvierbarer Systeme betrachten!

Beschreibung von Genom-Räumen und Populationen in thermodynamischen Begriffen

Thermodynamik: Versuch, Ensembles vieler, einfacher Teile durch aggregierte (statistische) Größen zu beschreiben (Temperatur, Druck, Entropie...)

Dynamisches System:

- *Phasenraum* = Menge aller möglichen Zustände des Systems
- *Regel* = Rechenvorschrift, die angibt, wie sich der Zustand des Systems zeitabhängig verändert (Dynamik des Systems)

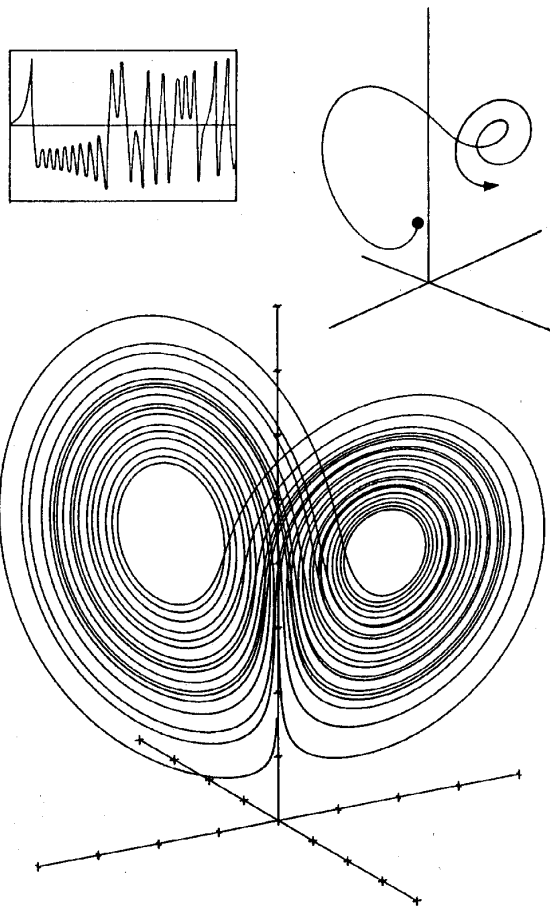
als Regeln häufig verwendet: Differentialgleichungen (kontinuierliches Modell); Funktion, die iteriert angewandt wird (diskretes Modell: Zustände werden nur an abzählbar vielen Zeitpunkten betrachtet).

Beispiel:

Pendel. Zustand: beschrieben durch Position und Geschwindigkeit (2D-Phasenraum (x, v)). Regel: Pendelgesetz aus der Mechanik.

implizite Annahme in der Theorie dynamischer Systeme:
wenn ein beliebiger Anfangszustand *exakt* bekannt ist, lassen sich *prinzipiell* alle zukünftigen Zustände des Systems berechnen.

Orbit eines Punktes im Phasenraum: Menge aller Folgezustände, die sich durch Anwendung der Regel ergeben. (bei iterierter Funktion f , also diskretes dyn. System: Orbit = Iteriertenfolge $x, f(x), f(f(x)), \dots$; bei kontinuierlichem Modell: stetige *Trajektorie* des Punktes.)

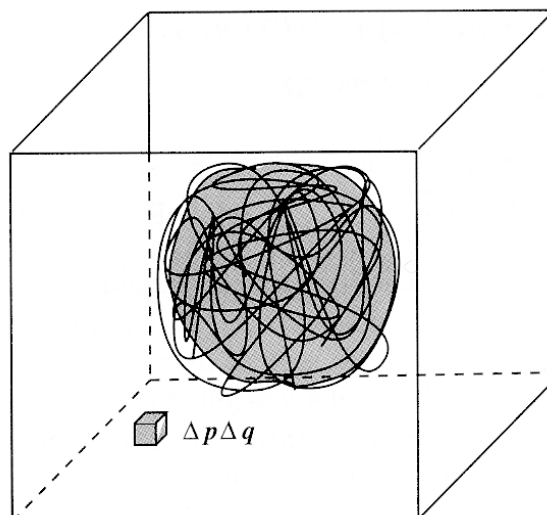


oben links:
Diagramm des zeitlichen
Verlaufs einer System-
eigenschaft, z.B. der Ge-
schwindigkeit

oben rechts:
Orbit des entsprechenden
Punktes im Phasenraum

unten: Attraktor, der den Punkt
schließlich "einfängt"

Die "Dichte" der Orbits im Phasenraum wird quantifiziert durch die Wahrscheinlichkeitsdichtefunktion (statist. Verteilungsfunktion) ρ des Systems:
die Wahrscheinlichkeit, das System in einem Zustand innerhalb einer gegebenen, messbaren Teilmenge des Phasenraums vorzufinden, ist das Integral von ρ über diese Teilmenge.



Mittelung einer Quantität f über die Zustände des Systems:

$$\langle f \rangle = \int f(p, q) \rho(p, q) dp dq$$

Ein System heißt *ergodisch*, wenn jeder Orbit den gesamten Teil des Phasenraums, wo ρ nicht verschwindet, erfüllt.

In diesem Fall kann man den Mittelwert einer Größe f auch durch Mittelwertbildung über die Zeit gewinnen:

$$\langle f \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(p(t), q(t)) dt$$

Raum der Genomzustände:

wird identifiziert mit dem Raum aller Nukleotidsequenzen (einer gegebenen maximalen Länge)

beispielsweise für Sequenzen der Länge 10^9 : 4^{10^9} mögliche Zustände \Rightarrow weit mehr als das Alter des Universums (und in diesem Zustandsraum können *nicht* a priori große Äquivalenzklassen gebildet werden)

\Rightarrow die Evolution kann nur einen verschwindend kleinen Teil des Genomraumes durchlaufen:

nichtergodischer Charakter der Evolution

\Rightarrow der Zustand des Systems hängt in essenzieller Weise von seiner *Geschichte* ab (anders als bei ergodischen Systemen)

Darwinsche Evolution hat "historischen" Charakter!

"Kontingenzprinzip": Eigenschaften des real vorkommenden Lebens können vom historischen "Ur-Genotyp" (Anfangszustand) abhängen

Fortführung der Übertragung thermodynamischer Begriffsbildungen:

Genotyp (Sequenz) – entspr. den Ortskoordinaten und Momenten in der klassischen Thermodynamik

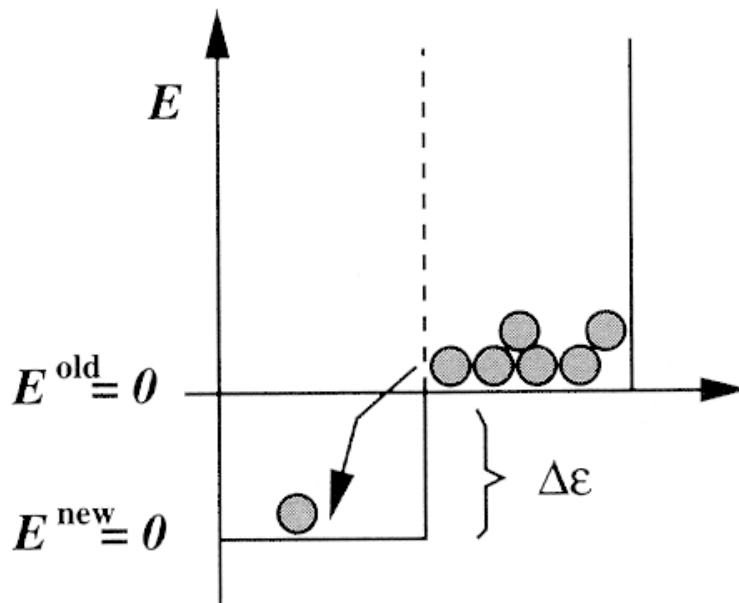
Inferiorität $I(s)$ eines Genotyps s : $I(s) = \log \varepsilon_{best} - \log \varepsilon(s)$, wobei $\varepsilon(s)$ die Replikationsrate von s und ε_{best} die beste in der Population vorkommende Replikationsrate ist

- bei Abwesenheit von Mutationen tendiert die Population durch den Selektionseffekt zu einem Zustand mit durchschnittlicher Inferiorität 0

Inferiorität entspr. der *Energie* eines Systems

Eine Mutation, die einen String mit höherer Replikationsrate liefert als die bisher beste, macht den alten Nullzustand "metastabil" und führt zu einer "Renormalisierung" des Inferioritätsmaßes (da fast alle Elemente der Population plötzlich eine höhere Inferiorität als zuvor haben):

$$\varepsilon_{best}^{neu} = \varepsilon_{best}^{alt} + \Delta\varepsilon$$



dieser Übergang kann als *Phasenübergang* mit latenter Wärme $\Delta\varepsilon$ angesehen werden

Die *Entropie* eines Systems ist ein Maß für die Unsicherheit über den Zustand des Systems

– nimmt zu mit der Anzahl der verschiedenen Zustände, die das System annehmen kann

im Genomraum: Entropie $S = \log N_g$,

wobei N_g die Anzahl der verschiedenen Sequenzen in der Population ist (Maß für die Diversität der Population)

Wenn die relativen Häufigkeiten ρ_i der Genotypen bekannt sind:

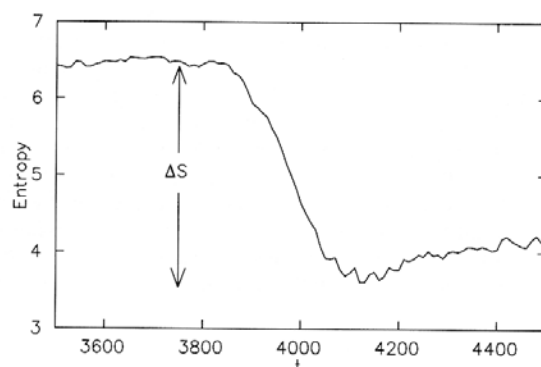
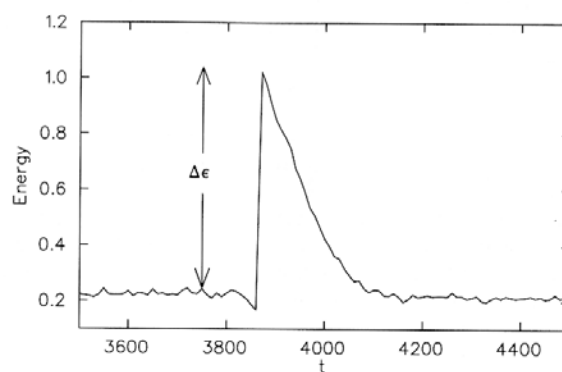
bedingte Entropie
$$S = -\sum_{i=1}^{N_g} \rho_i \log \rho_i$$

Die *Temperatur* eines thermodynamischen Systems ist def. als die inverse Änderungsrate der Entropie mit der Energie:

$$\frac{1}{T} = \frac{dS}{dE}$$

im Genomraum ist diese Größe proportional zur *Mutationsrate*.

Phasenübergang beim Auftreten einer neuen "bestangepassten" Mutante in einem ALife-Experiment (aus Adami 1998):

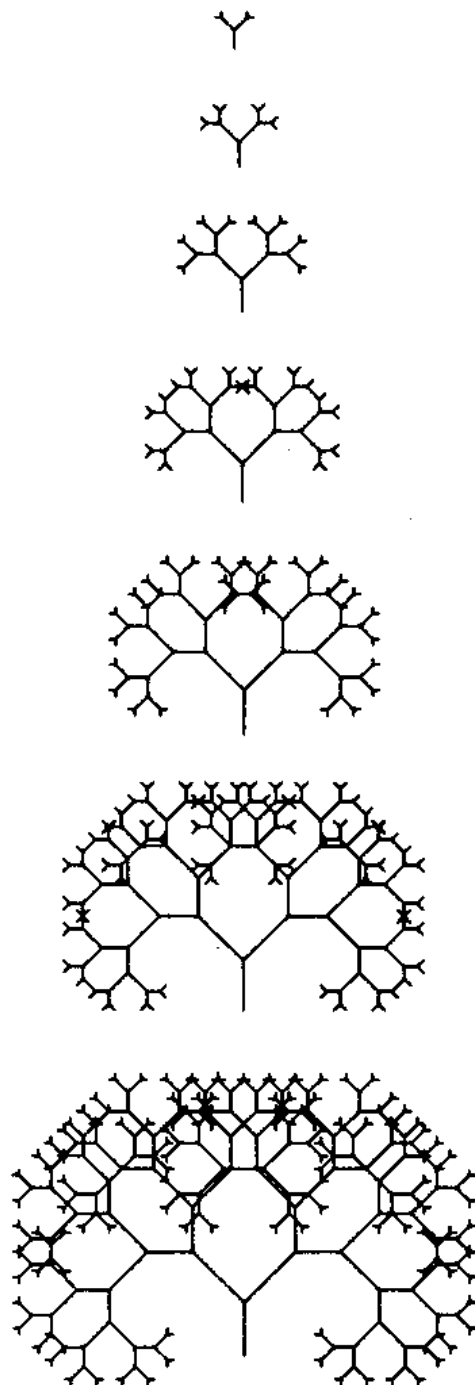


Beispiel eines einfachen Genomraumes, versehen mit morphologischen Phänotypen:

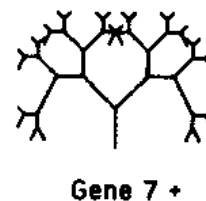
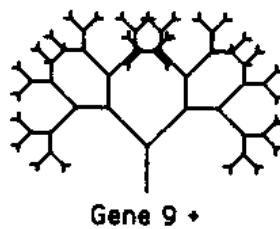
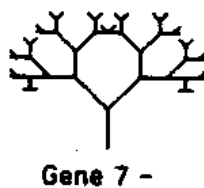
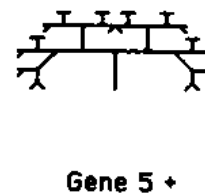
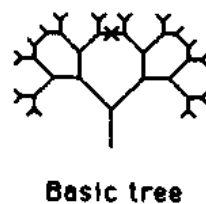
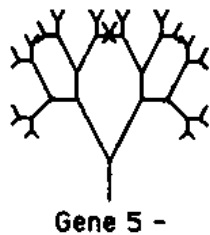
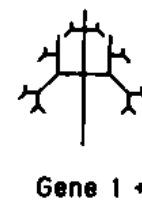
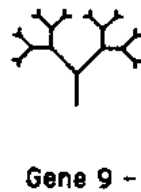
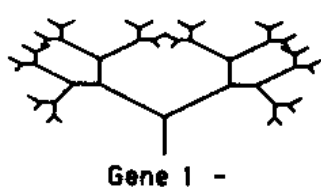
Richard Dawkins' "*Biomorphe*" (Begriff von Desmond Morris)

Morphologie: Baum mit dichotomer Verzweigung

9 Gene, Allele jeweils durch ganze Zahlen codiert
eines der Gene steuert die Rekursionstiefe:



die anderen Gene beeinflussen unterschiedliche metrische Größen (Längenverhältnisse, Winkel):



- *Mutationen* addieren +1 oder -1 zum Wert eines der Gene hinzu
- keine geschlechtliche Vermehrung (Rekombination)

zu einem "Eltern-Biomorph" wird eine Menge mutierter "Nachkommen" generiert

die *Selektion* erfolgt interaktiv durch den *menschlichen Benutzer* (*animal breeder selection*, vgl. Tier- und Pflanzenzüchtung)

– Wirkung: (ästhetische) Entscheidungen des Benutzers werden kumuliert (*cumulative selection*)

interaktiv:

<http://islwww.epfl.ch/~moshes/biomorphs/evolution.html>
(dort mit Rekombinationsmöglichkeit erweitert)

Dawkins schreibt:

Als ich das Programm schrieb, kam mir niemals der Gedanke, daß es etwas anderes entwickeln könnte als eine Varietät von baumähnlichen Gestalten. Ich hatte Trauerweiden erhofft, Libanonzedern, Pyramidenpappeln, Meeresalgen, vielleicht Hirschgeweihe. Nichts in meiner Intuition als Biologe, nichts in meiner 20jährigen Erfahrung im Programmieren von Computern, und nichts in meinen verrücktesten Träumen hatte mich auf das vorbereitet, was tatsächlich auf dem Bildschirm erschien. Ich kann mich nicht daran erinnern, an welchem Punkt der Sequenz es mir zu dämmern begann, daß eine durch Evolution entstandene Ähnlichkeit mit einem Insekt möglich war. Voller Argwohn begann ich zu züchten, Generation auf Generation – und von jedwedem Kind, das am meisten wie ein Insekt aussah. Mein ungläubiges Erstaunen wuchs in gleichem Maße wie die sich entwickelnde Ähnlichkeit ... Ich spüre immer noch das Triumphgefühl, das mich erfüllte, als ich diese sonderbaren Kreaturen zum ersten Mal vor meinen Augen entstehen sah. Ganz deutlich hörte ich die triumphierenden ersten Akkorde von *Also sprach Zarathustra* (dem Thema zum Film *2001*) in meinem Geist. Ich konnte nicht essen, und in jener Nacht schwärmten »meine« Insekten hinter meinen Augenlidern herum, als ich zu schlafen versuchte.⁹

(Dawkins 1990, S. 77)

Später: Schwierigkeit, die "Insekten" im Genomraum "wiederzufinden" (erste Version des Computerprogramms speicherte nur die Phänotypen ab!)

– nichtergodischer Charakter wird schon in diesem einfachen Beispiel deutlich

Bei meinen Wanderungen durch die entfernten Gefilde des Landes der Biomorphe erhielt ich Feenkrabben, Aztekentempel, gotische Kirchenfenster, Eingeborenenzeichnungen von Känguruhs und, bei einer bemerkenswerten, aber nicht wiederholbaren Gelegenheit, eine passable Karikatur des Wykeham-Professors für Logik.¹⁰

Beispiele:



Schwabenschwanz



Mann mit Hut



Mondlandefähre



Präzisionswaage



Köcherfliege



Skorpion



Fadenspiel



Laubfrosch



Spitfire



Gekreuzte Säbel



Bienenorchidee



Kopffüßer



Insekt



Fuchs



Lampe

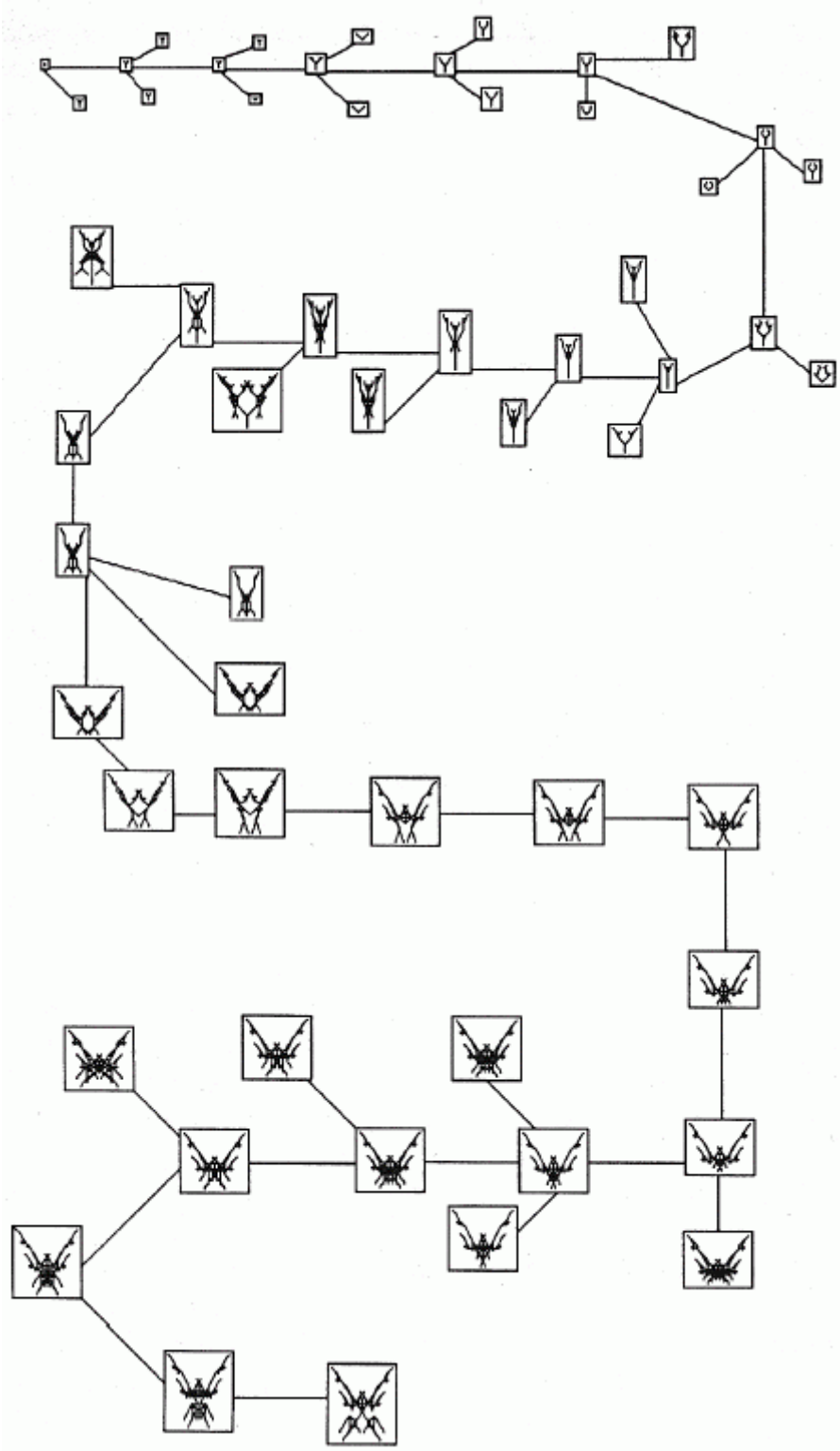


Springende
Spinne



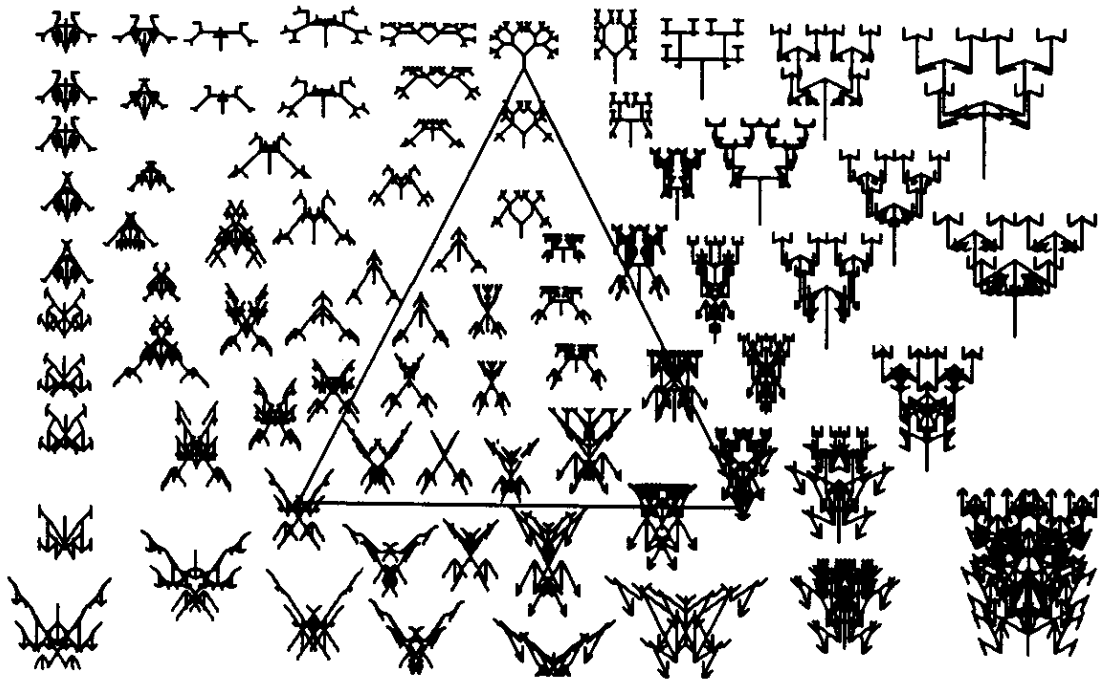
Fledermaus

Beispiel der Entwicklung eines "Insekts" in 29 Generationen:



(Dawkins 1990)

Visualisierung einer Ebene im Genomraum (aufgespannt durch die 3 Ecken des Dreiecks) anhand der Phänotypen:



Beschreibung der Struktur von Populationen

Diversität (bzw. Entropie) als Anzahl verschiedener Genome: sehr grobes Maß, berücksichtigt nicht die Abstände im Genomraum

Abstandsmaße für Zeichenketten:

- *Hamming-Distanz* (für gleichlange Strings; Anzahl der Positionen mit unterschiedlicher Belegung)
- *Levenshtein-Distanz* (Editierabstand): minimale Zahl von Editieroperationen (bzw. Mutationen), die erforderlich ist, um den einen String in den anderen zu transformieren (Editier-Operationen: Überschreiben, Einfügen und Löschen je eines einzelnen Zeichens)
- *relativer Editierabstand*: Editierabstand, geteilt durch die Länge der längeren von beiden Zeichenketten

Taxonomische Struktur:

Einteilung einer Population in *taxonomische Kategorien*

Kategorien der biologischen Systematik:

Art = Spezies, Gattung, Familie, Ordnung, Klasse, Abteilung, Stamm, Reich

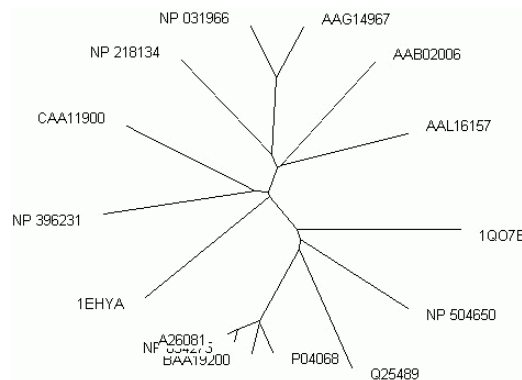
(Carl von Linné)

Mitglieder derselben Kategorie haben untereinander geringere Distanz als zu Mitgliedern anderer Kategorien der gleichen Hierarchie-Ebene

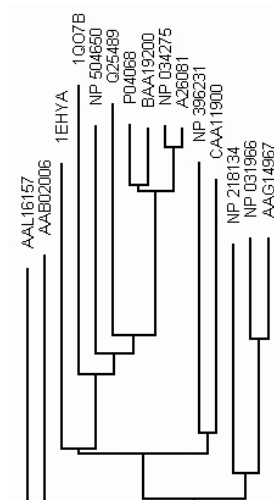
Visualisierung in *phylogenetischen Bäumen* (Verwandtschaftsbäumen):

Kantenlängen sollen zu Genom-Distanzen proportional sein

- ohne Auszeichnung eines Wurzelknotens:



- mit Wurzelknoten; Längen werden nur auf den Vertikalen gewertet:

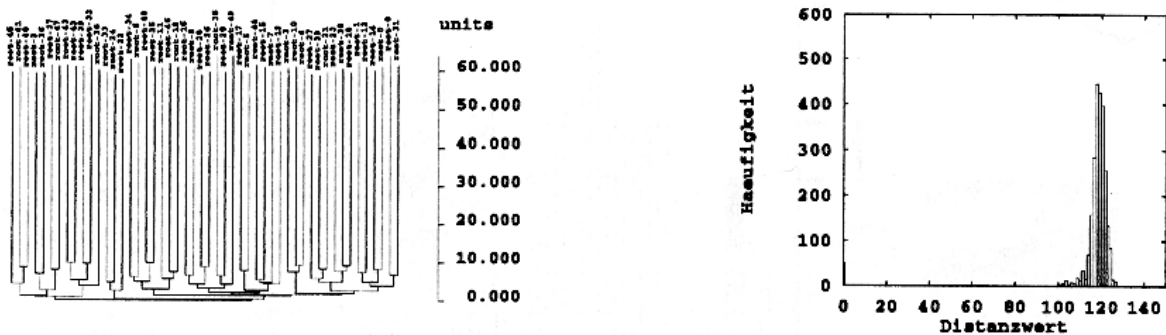


Distanzverteilung

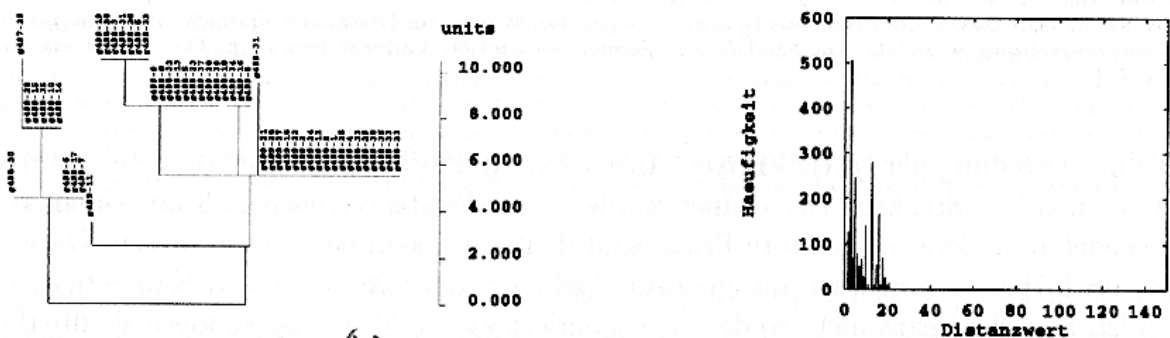
Sei P eine Population aus n Individuen und d ein diskretes Distanzmaß auf P . Dann ist die Distanzverteilung F def. durch

$$F(k) = |\{ (x, y) \mid x, y \in P, d(x, y) = k \}| / 2.$$

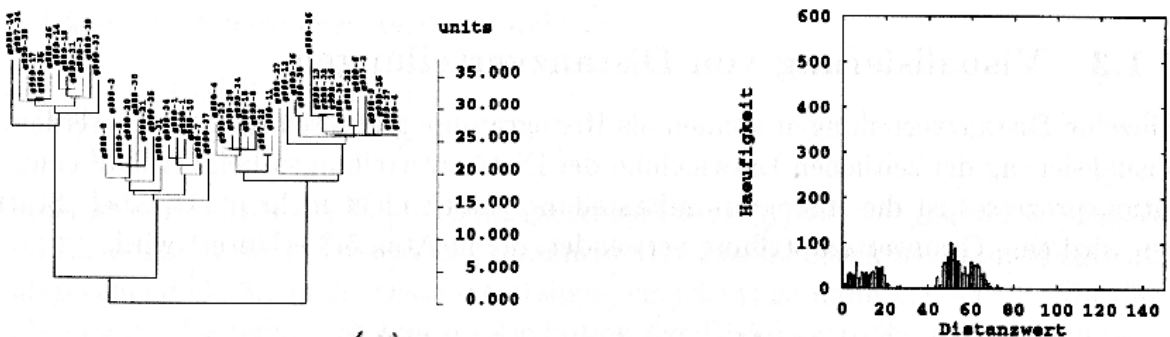
Beispiele für Populationen mit unterschiedlicher taxonomischer Struktur:



(a) Randomisierte Taxonomie



(b) Konvergierte Taxonomie



(c) Komplexe Taxonomie

(aus Kim 1996)

Distanzverteilungskomplexität (DVK):

Shannon-Entropie der normalisierten Distanzverteilung.

Sei f die Verteilung der relativen Häufigkeiten der Distanzwerte:

$$f(k) = \frac{F(k)}{\sum_i F(i)} = \frac{2 \cdot F(k)}{N(N-1)}$$

dabei sei N die Größe der Population.

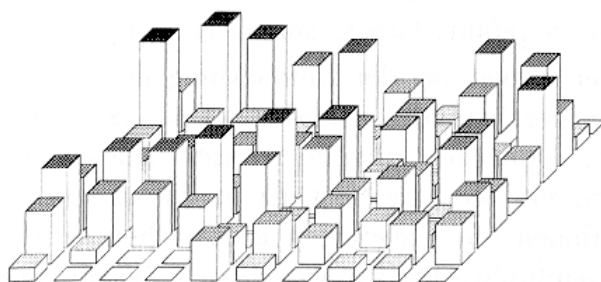
Dann ist die Distanzverteilungskomplexität

$$C = -\sum_k f(k) \log(f(k))$$

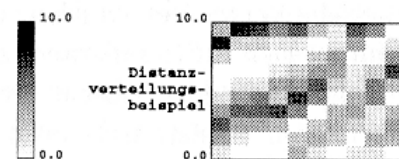
(Kim 1996).

Hohe DVK: gleichmäßige Verteilung der Distanzwerte,
komplexe taxonomische Struktur!

Visualisierung der dynamischen Entwicklung von
Distanzverteilungen (z.B. im Verlauf einer künstlichen
Evolution):



(a)



(b)

Serie von 10 Verteilungen (Zeitachse = horizontal),
(a) dreidimensional, (b) mit Grauwertcodierung
globale Strukturen treten in der Grauwertcodierung deutlicher hervor

(Kim 1996)

Explizitmachen der Fitness

im "Biomorph"-Beispiel fand die Selektion benutzergesteuert (interaktiv) statt

Annahme jetzt: jedem möglichen Genom kann eine Fitness (= Wahrscheinlichkeit, genetisch zu überdauern) *fest zugeordnet* werden

Achtung: das ist Idealisierung!

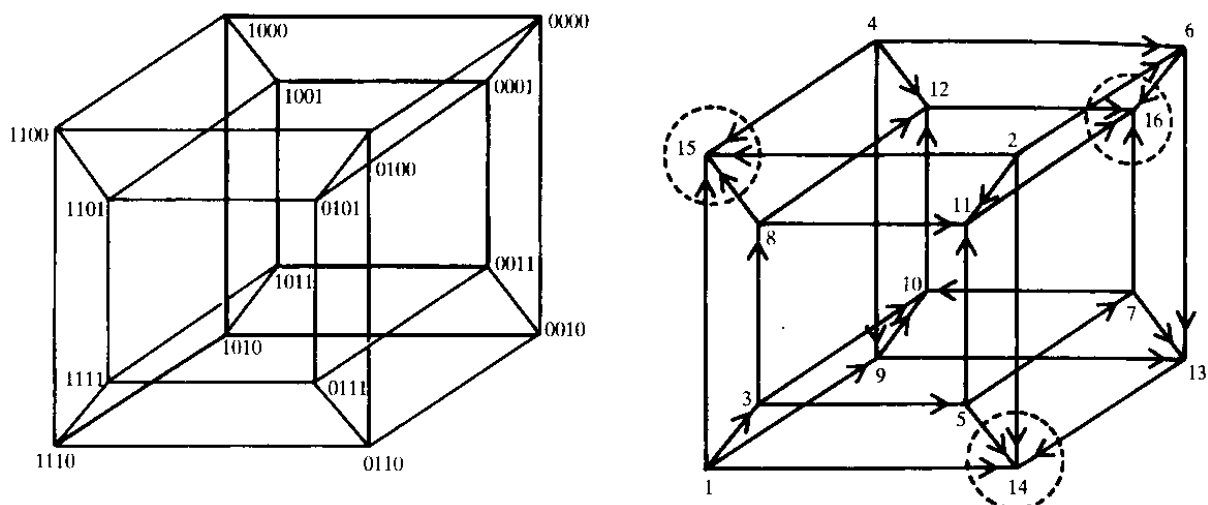
- Überleben und Fortpflanzung kann auch in der Natur von zufälligen Ereignissen abhängen
- Fitness kann von Konkurrenten (und damit von der Zusammensetzung der Population) abhängen
- die gleiche Genom-Zeichenkette kann unter verschiedenen Bedingungen verschiedene Fitness haben

Unter der obigen Annahme kann Evolution als Lösungsversuch einer *Optimierungsaufgabe* (Extremalproblem) angesehen werden:

Finde den Ort (im Genom-Raum) mit maximaler Fitness!

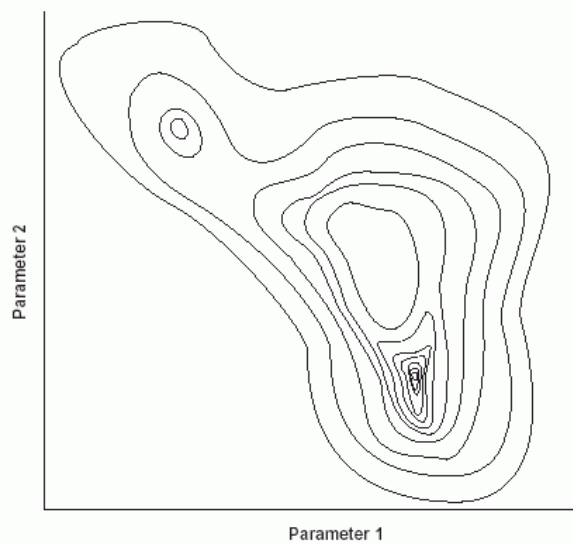
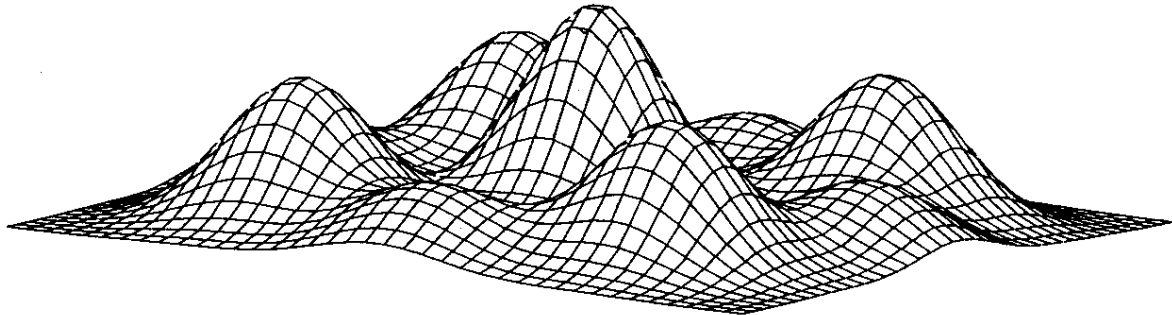
"Fitnesslandschaft" (Begriff nach Sewall Wright 1932):

jedem Punkt des (hochdimensionalen) Sequenzraumes wird seine Fitness als Zahlenwert zugeordnet (*Fitnessfunktion*)



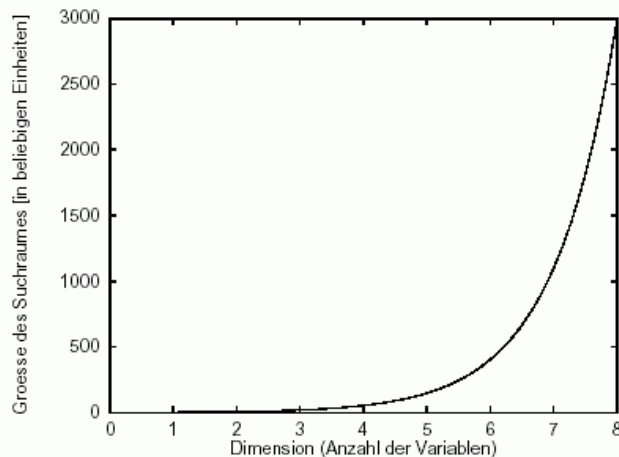
alle Genome der Länge 4 (links); rechts: Belegung mit Fitness-Werten; eingekreist: lokale Maxima ("Gipfel") der Fitnesslandschaft (Kauffman 1995)

2D-Analogen:



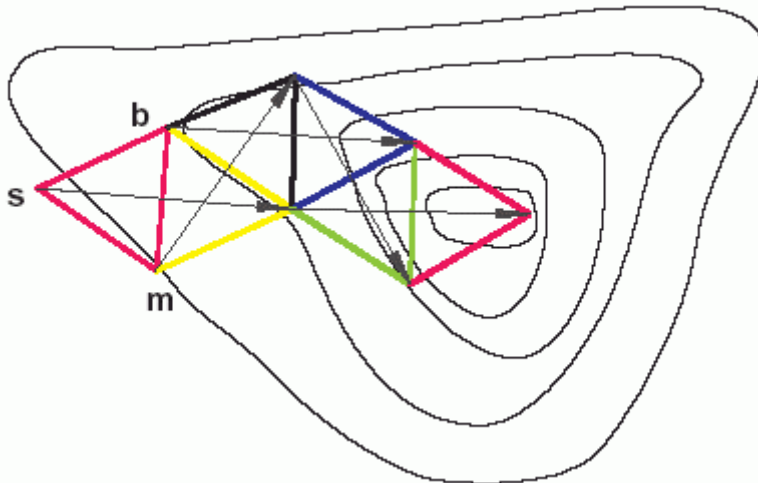
beachte: wegen der Hochdimensionalität des Genom-Raumes kann die Landschafts-Analogie irreführend sein!

insbesondere wächst die Größe des Genom-Raumes exponentiell mit der Dimension



Herkömmliche Optimierungsverfahren:

- analytische Lösung – nur in wenigen Spezialfällen möglich
- "Hill Climbing"-Methode (Gradientenmethode): Man folgt immer der Richtung der größten Steigung
spezielle Variante: *Simplexmethode*
- iteriertes Spiegeln von Simplexes in Richtung der größten Steigung



Nachteil: man kann auf lokalem Maximum "steckenbleiben", ohne dem globalen Maximum auch nur nahezukommen

- *simulated annealing* (simuliertes Tempern / Ausglühen):
 - Schritte in ansteigender Richtung
 - mit gewisser Wahrscheinlichkeit akzeptiert man auch Schritte in die "falsche" Richtung
 - die "Temperatur" (Analogie zum physikalischen Tempern!) gibt die Wahrscheinlichkeit an, mit der man einen Schritt akzeptiert, der den Wert der Zielfunktion verschlechtert
 - im Verlauf der Optimierung wird die "Temperatur" gesenkt

eines der besten Optimierungsverfahren

Nachteil: Auffinden guter Lösungen erfordert sehr langsames "Abkühlen" – Zeitaufwand

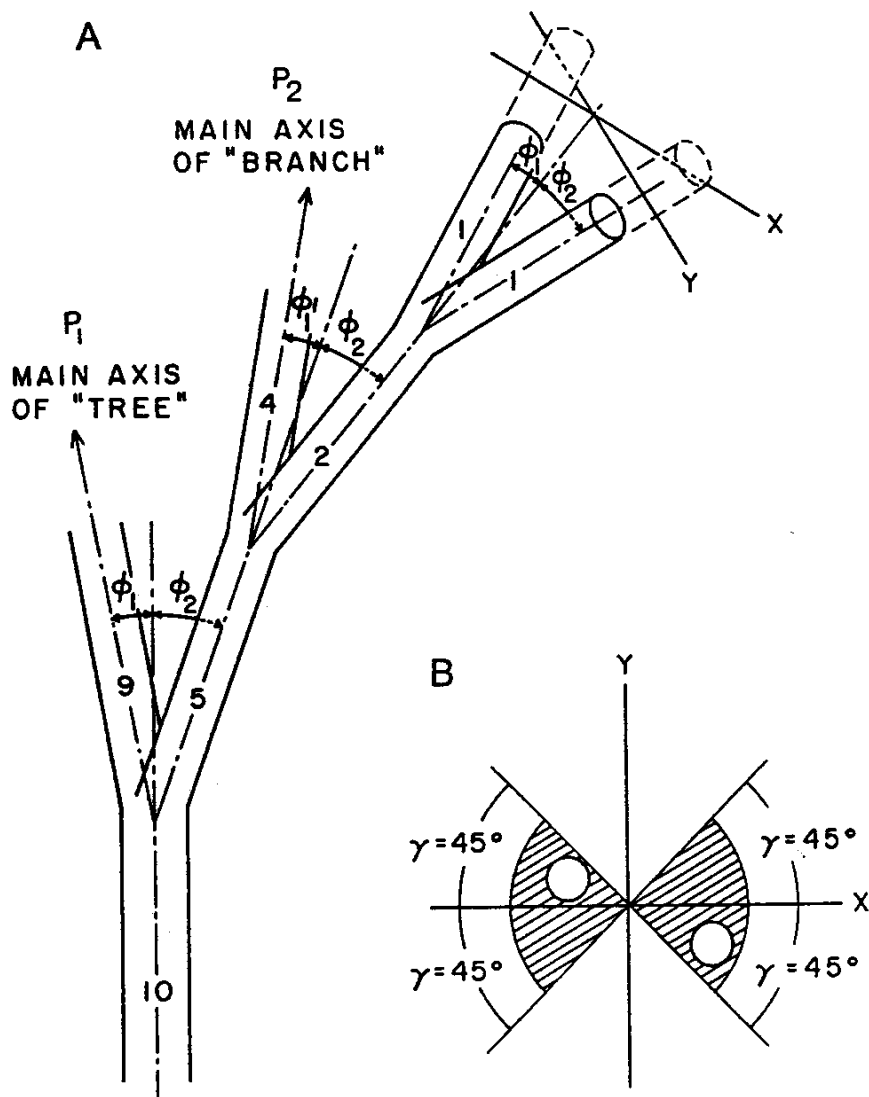
multikriterielle Optimierung

in der Natur: nicht nur eine (skalare) Zielgröße

Karl J. Niklas untersuchte in Simulationsexperimenten die Evolution der Pflanzenmorphologie

einfache, dichotome Verzweigungsstrukturen beschrieben durch 6 Parameter:

- Wahrscheinlichkeit der Verzweigung
- Verzweigungswinkel
- Rotationswinkel der Verzweigung in der Ebene senkrecht zur Mutterachse
- diese Parameter jeweils für Hauptachse und Seitenachse



damit 6-dimensionaler Raum
3-dim. Unterraum der symmetrischen Pflanzen

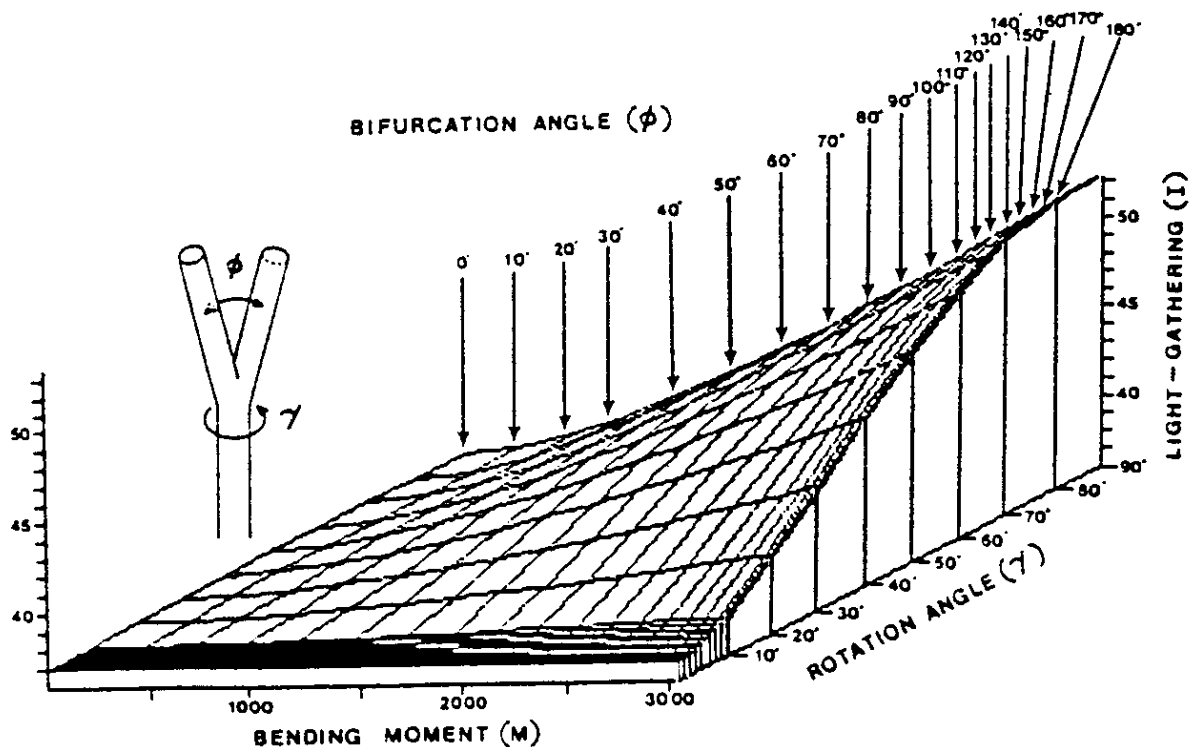
funktionale Anforderungen (vgl. Kapitel 2):

- Lichtinterzeption
- mechanische Stabilität (geringe Biegespannung)
- Verteilung möglichst vieler Sporen oder Samen in möglichst weiter Entfernung

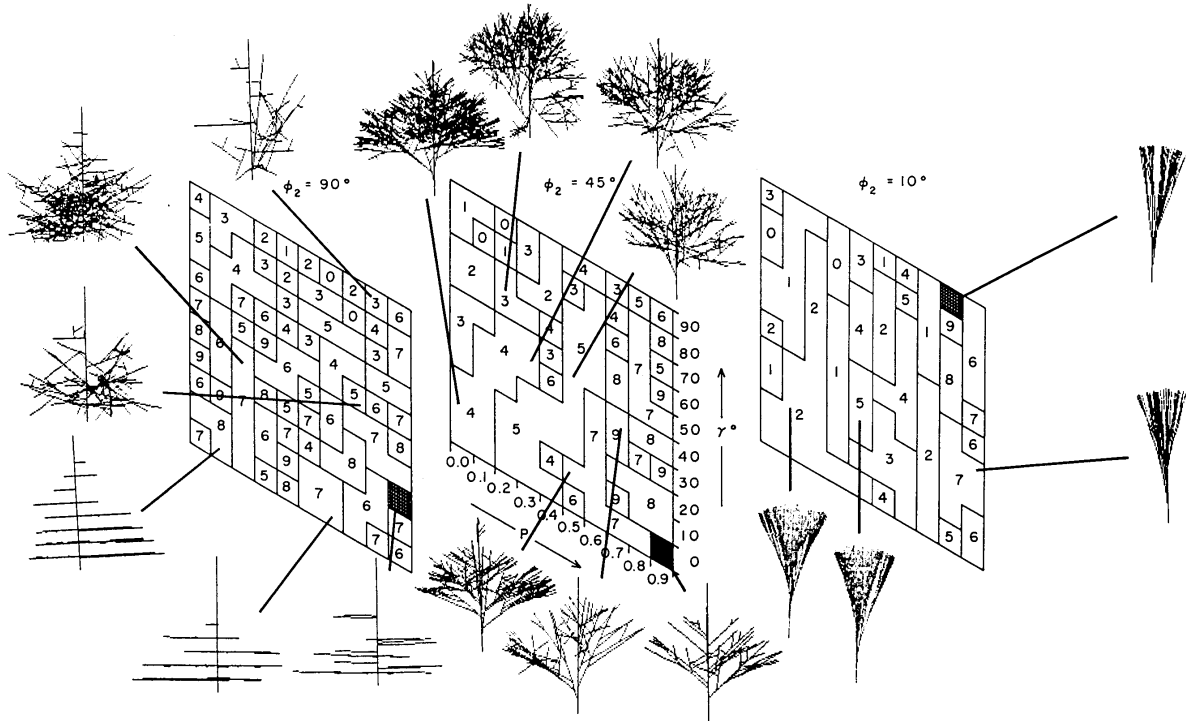
die 3 Anforderungen können im Modell durch 3 reellwertige Größen quantifiziert werden

- für jede der Optimierungsgrößen entsteht eine "Fitnesslandschaft"
- die 3 Größen sind nicht unabhängig voneinander

z.B. für die Lichtinterzeption bei symmetrischen Pflanzen:



Auswertung verschiedener "Schnitte" mit festem Verzweigungswinkel durch den Raum der Phänotypen hinsichtlich Lichtinterzeptionsfähigkeit (Optima = schwarze Felder):

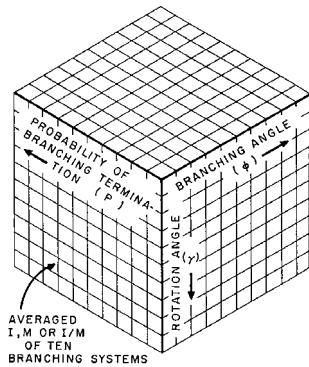


die 3 "Schnitte" entsprechen (grob) realen Pflanzen-Taxa:
Pteridophyten, Angiospermen, Gymnospermen (von rechts nach links)

Optimierungsexperimente:

- zugrundegelegt: diskretisierter 6-dim. "Morphospace" aus 200 000 Phänotypen
- P variiert zwischen 0 und 1 in 0,01-Schritten
- beide Winkel variieren zwischen 0° und 180° in 1°-Schritten
- die 3 Fitness-Faktoren werden einzeln, oder multiplikativ gemeinsam betrachtet
- Optimierung durch hill-climbing (erschöpfende Suche in der lokalen Nachbarschaft nach bestem Nachbarn)
- wenn zwei oder mehr Phänotypen mit gleicher Fitness in der Nachbarschaft gefunden werden, gabelt sich der Optimierungslauf ("adaptive walk")
- Startpflanze: einfachste, Y-förmige Architektur; ähnlich Fossilien aus dem Silur (*Cooksonia*) – *ancestral phenotype* ("Urpflanze")

Visualisierung der Optimierungsläufe verwendet Reduktion auf 3 Dimensionen

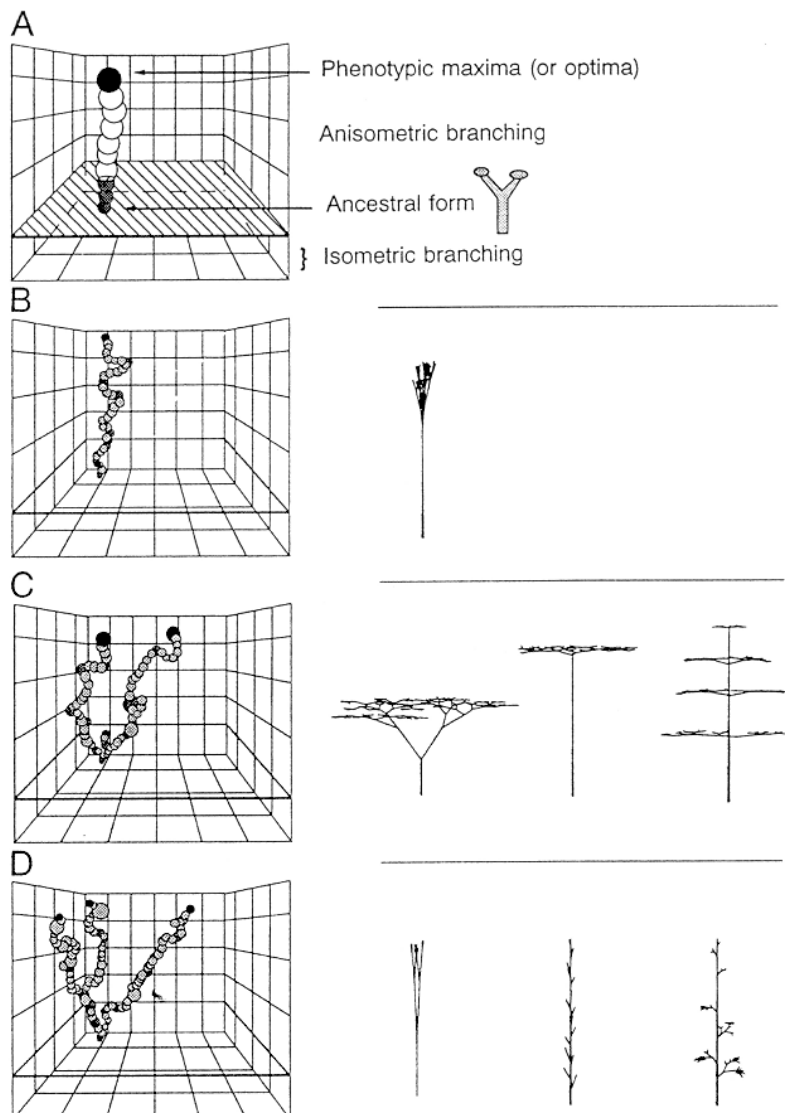


Simulierte adaptive Läufe durch den Morphospace mit monokriterieller Optimierung:

B: nur Sporenverbreitung

C: nur Lichtinterzeption

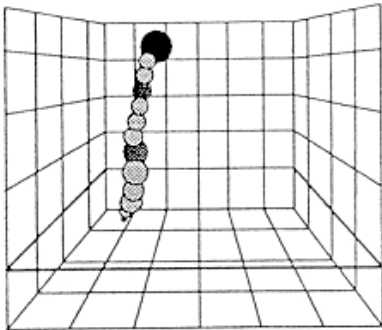
D: nur mechanische Stabilität



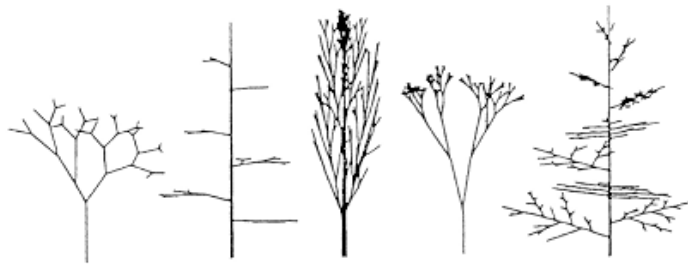
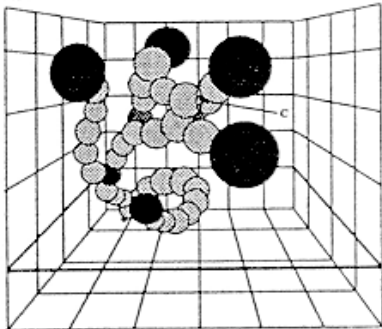
Läufe mit multikriterieller Optimierung:

- A: mech. Stabilität und Sporenausbreitung
- B: Lichtinterzeption und mech. Stabilität
- C: Lichtinterzeption und Sporenausbreitung
- D: alle 3 Kriterien kombiniert

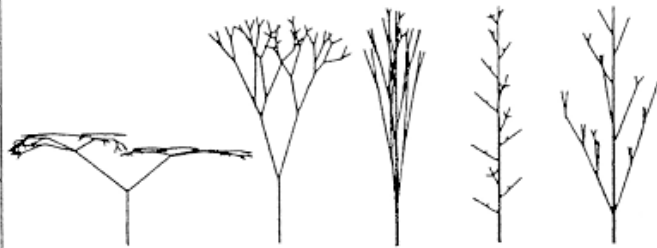
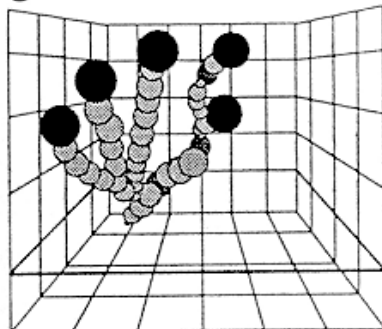
A



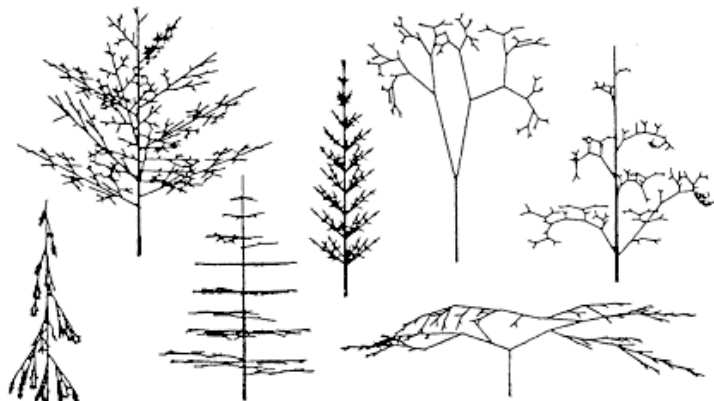
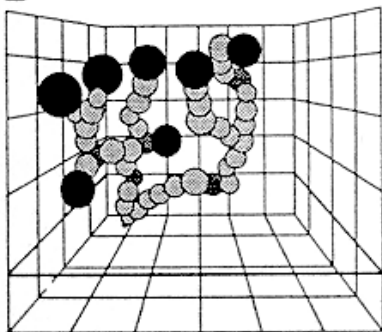
B



C



D

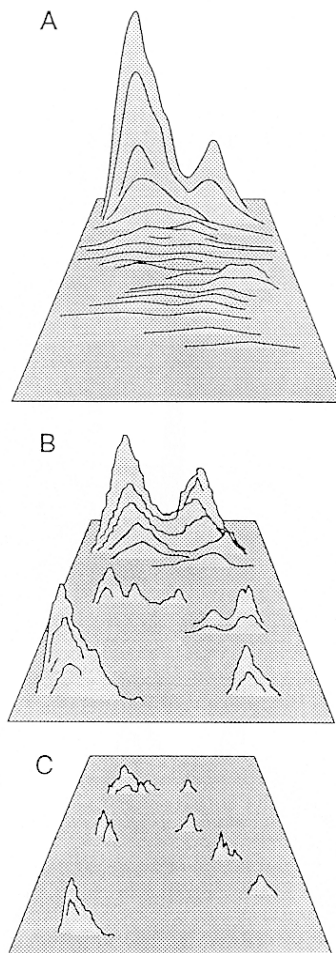


Beobachtungen:

- Es ereignen sich in den Optimierungsläufen morphologische Veränderungen, wie sie auch anhand von Fossilfunden in der tatsächlichen Evolution der Landpflanzen belegt werden können: Übergang von diffusem zu lokalisiertem (apikalem) Wachstum, orthotrope Orientierung der Achsen, Tendenz zur Herausbildung einer Hauptachse (Stamm), gleichmäßige Raumfüllung.
- Die Optimierungsläufe mit multikriterieller Optimierung liefern eine größere Diversität von Phänotypen.
- Die Fitness-Maxima bei den multikriteriellen Ansätzen sind nicht so ausgeprägt wie bei den monokriteriellen Ansätzen.

Veranschaulichung der beiden letztgenannten Punkte im Fitnesslandschaft-Bild:

von A nach C Einbeziehung von immer mehr Fitness-Faktoren



dies wird gestützt durch Befunde aus dem Ingenieur-Bereich:

Die Anzahl der *möglichen Design-Lösungen*, die widerstreitende Anforderungen an eine Konstruktion in Einklang zu bringen suchen, nimmt zu mit der Anzahl der Anforderungen.

in der natürlichen Evolution kommt hinzu:

Umweltbedingungen sind ständiger Veränderung unterworfen

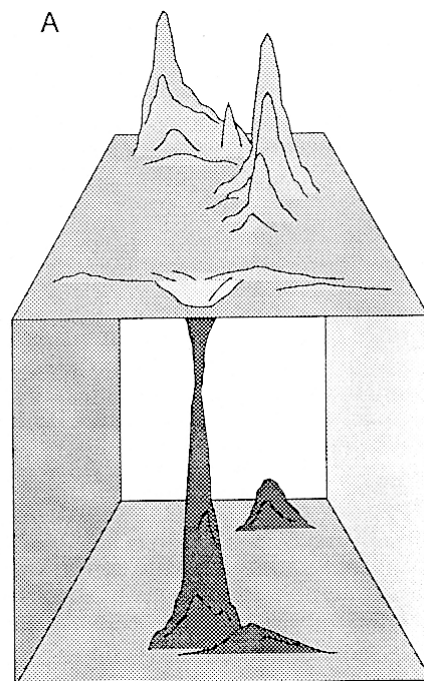
⇒ die Gewichtsfaktoren, mit denen die Optimierungsziele in die Fitnessfunktion einfließen, können sich verändern

⇒ Fitnessfunktion ist komplexe Größe, deren Komponenten in Skalierung und Richtung relativ zueinander (Korrelation / Konflikt) ständig variieren können

⇒ dies fördert in der Evolution die Diversität

Weitere Komplikationen (Niklas 1994, Farnsworth & Niklas 1995):

- Bei Erreichen einer bestimmten Fitness können qualitative Sprünge erfolgen, die neue Fitness-Landschaften erschließen (z.B. bei Eroberung eines neuen Lebensraumes) bzw. die Struktur der Fitnessfunktion grundlegend verändern



die Evolution erobert von der unteren die obere Landschaft

- einige Eigenschaften, die die Fitness steigern, beziehen sich auf Information und Interaktion und sind nicht in physikalischen Größen zu quantifizieren (Beisp.: Fähigkeiten des Kuckucks)

"We believe evolution is not a matter of optimization, but one of finding efficient designs under historic and environmental constraints. ...

We therefore do not believe in the simple idea of an optimal design or optimal fitness. We see organisms as representing a multitude of solutions to a multi-goal optimization problem which has no definitive answer and so leads only to a set of more or less efficient designs."

(Farnsworth & Niklas 1995)

Dennoch schränken wir uns im Folgenden zunächst wieder auf die Sichtweise der Evolution als Hilfsmittel für die Optimierung ein.

Evolutionäre Algorithmen

Grundidee:

- Setze eine Population von Individuen in den Genom-Raum (der mit einer *festen* Fitness-Funktion ausgestattet ist)
- lasse diese evolvieren:
durch Anwendung von Evolutions-Operatoren
 - Mutation
 - Rekombination
 - (evtl. weitere)
 - und Selektion (Überleben von Individuen, bevorzugt solcher mit hoher Fitness)
- Ziel: die Population soll sich zum globalen Fitness-Maximum hinbewegen

Verschiedene Varianten:

- Genetische Algorithmen (John Holland; Mitte 60er Jahre)
- Evolutionsstrategien (Rechenberg, Schwefel, Bieri; 60er Jahre)
- Evolutionary Programming (Fogel 1960)
- genetisches Programmieren (Koza 1987)

John Holland:

- erster Inhaber eines Informatik-Dokortitels in den USA
- Studenten seines Kurses "Theory of Adaptive Systems" waren maßgeblich an der Entwicklung der ersten GA beteiligt
- Holland machte keine Tests und Experimente mit dem Computer: "Ich bin immer ein Papier- und Bleistift-Mensch gewesen..." (Levy 1993, S. 199)
- GA-Experimente seines Doktoranden David Goldberg auf Apple-II-PC, verwendete GA zur Optimierung des Auffindens von Erdgasvorkommen

Genetische Algorithmen

- arbeiten auf Strings (oft Binärstrings)
- verwenden Rekombination ("mating" von 2 Individuen und crossing-over der Genome) und meistens auch Mutation
- Selektion der zu paarenden Individuen nach einer Selektionsregel, die die Fitness der Individuen einbezieht

Basic algorithm:

- Initialize population P
- Repeat for some length of time
 - Create an empty population, P'
 - Repeat until P' is full:
 - Select two individuals from P based on some fitness criterion
 - Optionally "mate", and replace with offspring
 - Optionally mutate the individuals
 - Add the two individuals to P'
 - Let P now be equal to P'

Verschiedene Selektionsregeln:

- Auswahl nur aus den besten 50% (oder 20%, etc.)
 - Nachteil: Verringerung der genetischen Diversität
- Zufallsauswahl, wobei die Wahrscheinlichkeit zur Fitness proportional ist ("*roulette wheel selection*") – gibt auch "den Schlechten" eine Chance
- "Wettkampf-Selektion": 2 Individuen werden völlig zufällig ausgewählt, das fittere von beiden wird verwendet
- Rang-Selektion: Wahrscheinlichkeit der Auswahl ist proportional zur Rangposition (bei Sortierung nach Fitness)
- bestes Individuum überlebt "sicher" (um Verminderung der Fitness der Population zu verhindern), die Auswahl der übrigen erfolgt nach einer der obigen Regeln
- ...

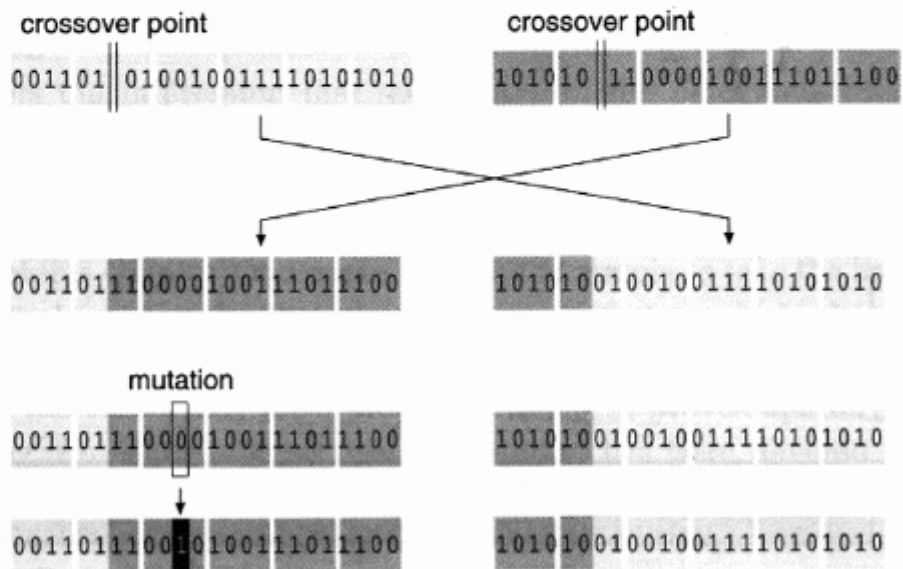
Selektion und Reproduktion (hier mit einer weiteren Selektionsstrategie):

a. selection

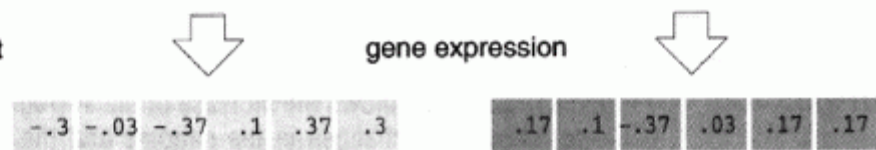
1. take the individual with the highest fitness
2. choose another individual from the population at random, irrespective of fitness, for sexual reproduction
3. add the fittest individual to the new population

| | fittest individual (highest rank) | | | | | | other individual | | | | | |
|-----------------|-----------------------------------|-------|-------|------|------|------|------------------|------|------|-----|------|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| initial genome | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| encoded weights | -0.3 | -0.17 | -0.37 | 0.03 | 0.17 | 0.17 | 0.17 | 0.23 | -0.5 | 0.1 | 0.37 | 0.3 |

b. reproduction

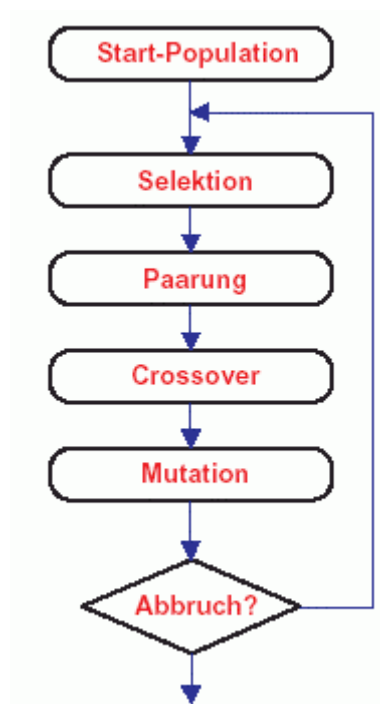


c. development



(Schritt c, "development", entfällt bei den meisten GA)

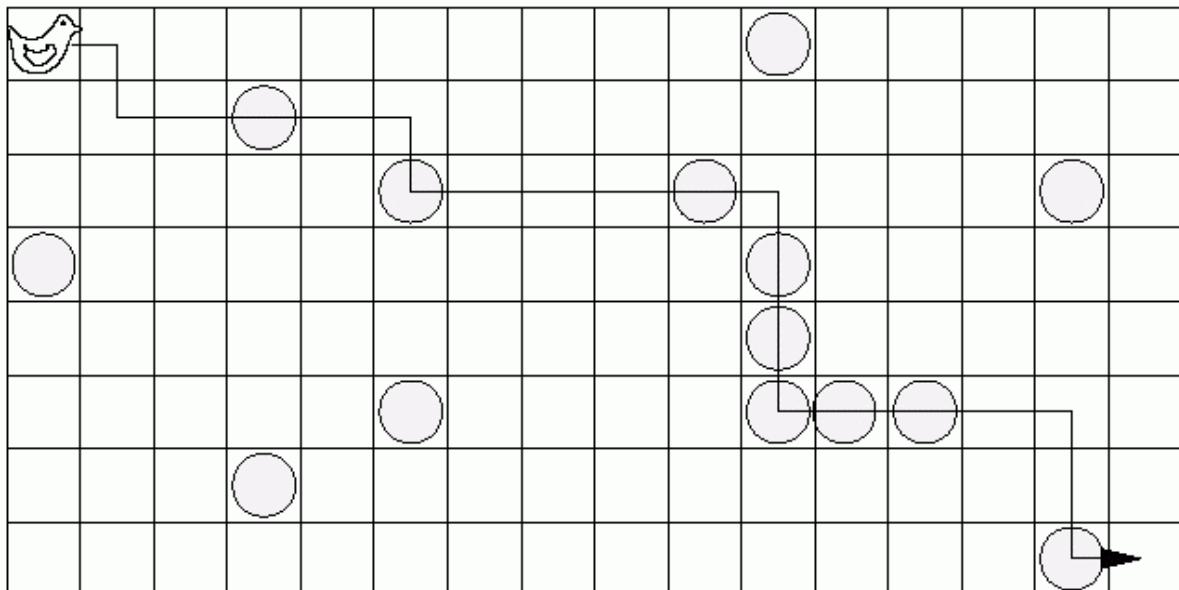
Zusammenfassung: Flussdiagramm des GA



Codierung realer Probleme für die Lösung mittels GA:

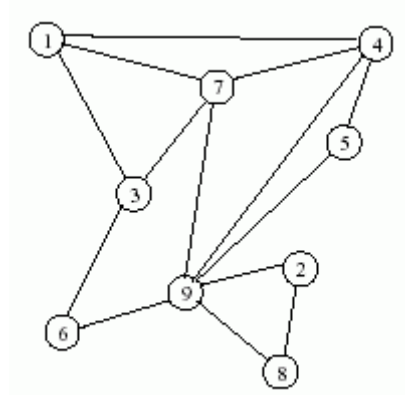
- Beispiel "Hungrige Ente"

Eine Ente, die nur nach rechts und unten schwimmen kann, soll einen Weg durch den Teich finden, bei dem sie möglichst viele Futterbrocken erreicht



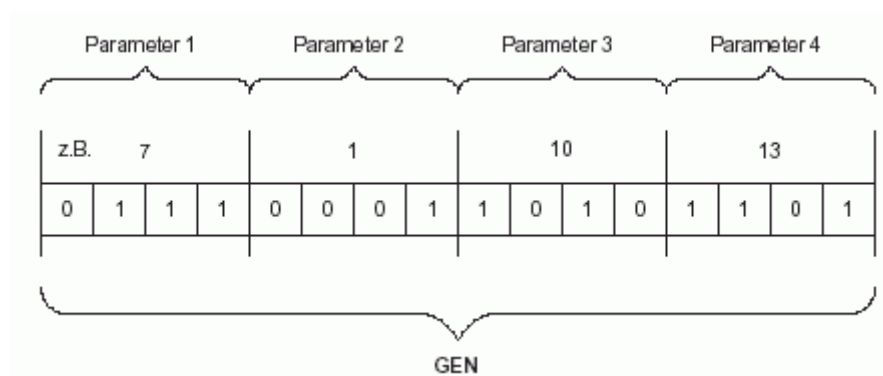
Codierung als Binärstring (0 = nach rechts, 1 = nach unten)

- Problem des Handlungsreisenden (*travelling salesman problem*): Auffinden der kürzesten Rundreise durch alle Ecken eines gegebenen Graphen



Codierung: Permutation der Ecken-Labels

- Maximum einer Funktion mit 4 Variablen:
Binärcodierung der 4 Parameter in Blöcken fester Länge, diese werden aneinandergehängt



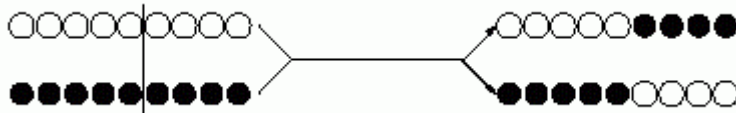
- Bei Zahlencodierung ist der Gray-Code angemessener als die gewöhnliche Binärcodierung (bessere Entsprechung der Differenz zwischen den Zahlen zur Hamming-Distanz der Genome)

| Binärer Code | Gray Code | Dezimalzahl |
|--------------|-----------|-------------|
| 000 | 000 | 0 |
| 001 | 001 | 1 |
| 010 | 011 | 2 |
| 011 | 010 | 3 |
| 100 | 110 | 4 |
| 101 | 111 | 5 |
| 110 | 101 | 6 |
| 111 | 100 | 7 |

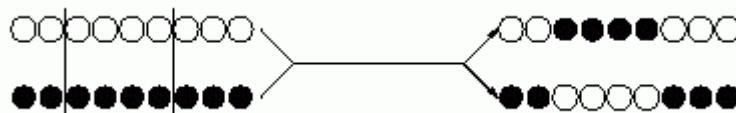
Evolutions-Operatoren:

Crossing over (Kreuzung, Rekombination)

einfaches crossing over: die Zeichenketten werden an einer (zufällig bestimmten) Stelle getrennt und es werden 2 Nachkommen erzeugt, indem der hintere Teilstring vertauscht wird



doppeltes crossing over:

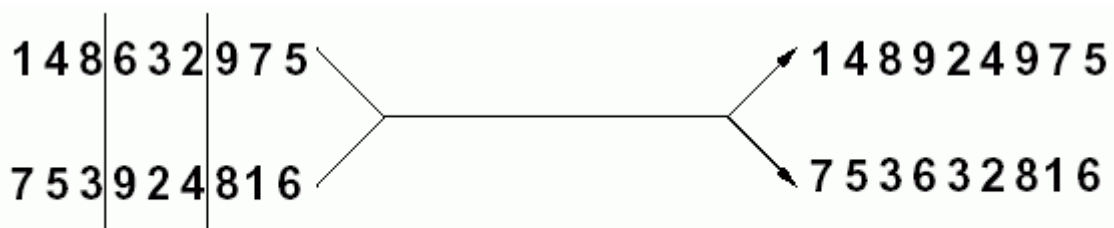


mehrfaches crossing over:



Variante für Permutationsstrings:

ersetze in den Nachkommen doppelt auftretende Zahlen so lange durch andere Zahlen, bis keine Zahlen mehr doppelt auftreten



9 ↔ 6 3 ↔ 2 2 ↔ 4

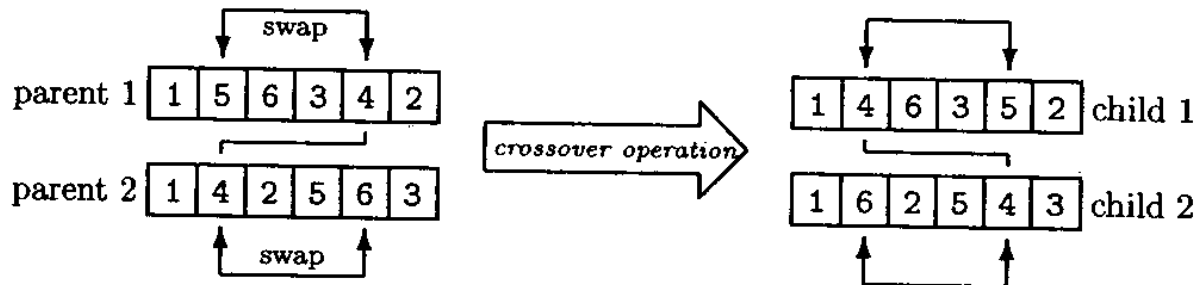
3
2 6
1 4 8 9 2 4 9 7 5

1 3 8 6 2 4 9 7 5

4
2 9
7 5 3 6 3 2 8 1 6

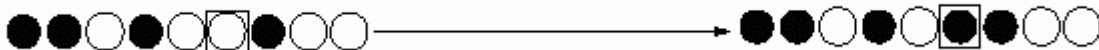
7 5 4 9 3 2 8 1 6

andere crossover-Variante für Permutationsstrings:
 Position zufällig auswählen, 2. Pos. ergibt sich durch
 entsprechenden Eintrag im 2. String, zwischen beiden Pos. wird
 in beiden Strings (einzeln) getauscht



Mutation

Zufallsauswahl einer Position im Genom
 bei Binärstring: dort das Bit "flippen"



bei Zahlenstring:

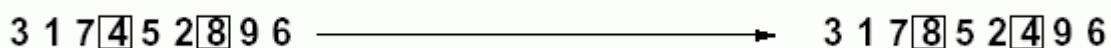
Wert x durch $(Max+1)-x$ ersetzen



oder durch zufällig gezogene Zahl

bei Permutationsstring:

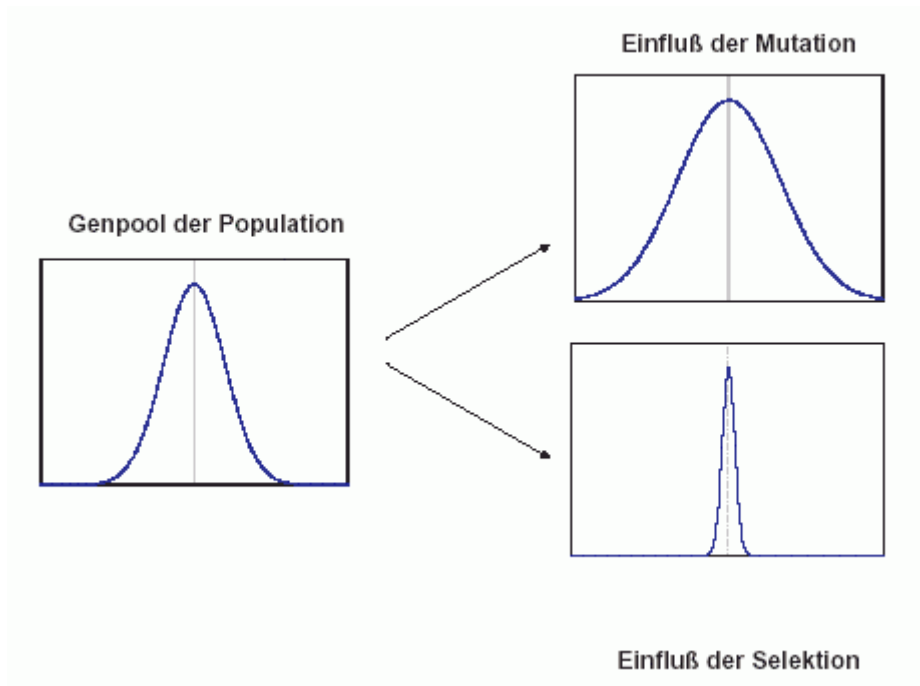
zwei Stellen auswählen und die dortigen Einträge vertauschen



zu beachten:

Mutation und Selektion müssen vernünftig "ausbalanciert" sein

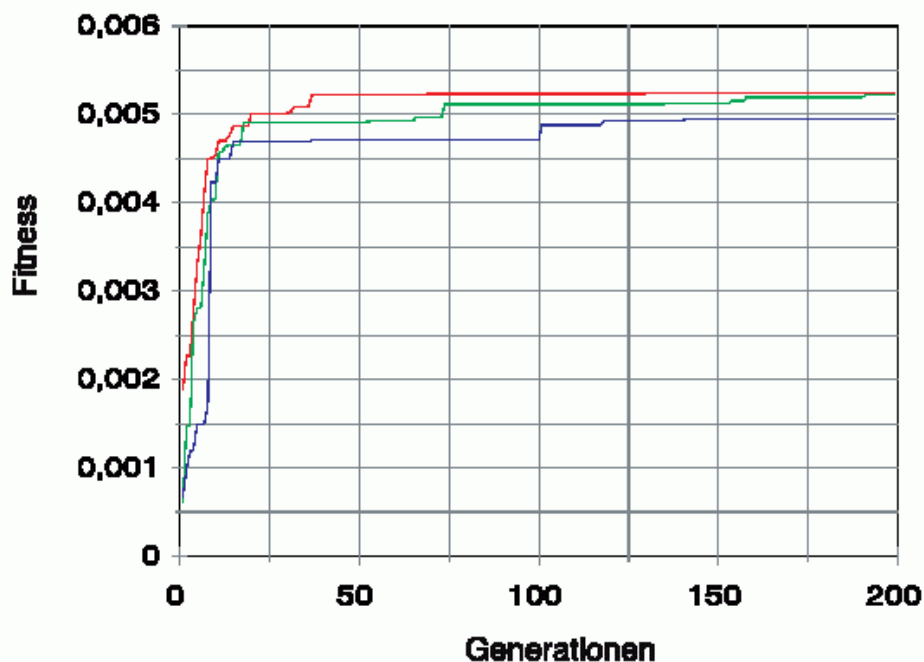
- Mutation erhöht die Diversität des Genpools (eventuell zu stark: Algorithmus wird ungerichtet, Population diffundiert von Maximalstellen wieder fort)
- Selektion vermindert sie (wenn zu stark: "Festsitzen" auf lokalen, nicht-globalen Maxima)



Iteration des GA:

Oft werden mehrere GA-Läufe mit unterschiedlichen Startpopulationen hintereinander durchgeführt, um die Wahrscheinlichkeit zu vergrößern, tatsächlich das globale Maximum der Fitness zu finden (Beachte: keine Garantie, dieses zu finden! – GA sind *heuristische* Algorithmen)

Typische Fitness-Kurven des jeweils besten Individuums einer Population in 3 GA-Läufen:



Faustregeln für GA (nach Güde 2001):

- Populationsgrößen sollten ≥ 100 sein
- schwierigere Probleme erfordern größere Populationsgrößen
- Anwendungshäufigkeit der Operatoren:
crossing over ca. 80%, unveränderte Übernahme in die nächste Generation ca. 20%, Mutationsrate ca. 0,1%

Beispiel für Anwendung eines GA:

- Individuen sind Strings aus Buchstaben
- Fitness misst die Nähe zum Zielstring "furious green ideas sweat profusely"
- Populationsgröße 500
- crossover-Rate 75%, Mutationsrate 1%
- → Zielstring wird in 46 Generationen erreicht

| Time | Average Fitness | Best Fitness | Best String |
|------|-----------------|--------------|--|
| 0 | 0.035314 | 0.200000 | "pjrmrubynrksxiidwctxfodkodjjzfunpk " |
| 1 | 0.070000 | 0.257143 | "pjrmrubynrksxiidnybvsqcq piisyexdt" |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 25 | 0.708686 | 0.771429 | "qurmous gresn idnasvsweqt prifuseky" |
| 26 | 0.724286 | 0.800000 | "qurmous green idnasvsweqt prifuseky" |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 36 | 0.806514 | 0.914286 | "uurious green idnas sweqt profusely" |
| 37 | 0.820857 | 0.914286 | "qurmous green ideas sweqt profusely" |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 41 | 0.895943 | 0.942857 | "uurious green idnas sweat profusely" |
| 42 | 0.908457 | 0.971429 | "qurrious green ideas sweat profusely" |
| ⋮ | ⋮ | ⋮ | ⋮ |
| 45 | 0.927714 | 0.971429 | "qurrious green ideas sweat profusely" |
| 46 | 0.936800 | 1.000000 | "furious green ideas sweat profusely" |

(aus Flake 1998)

weitere Varianten evolutionärer Algorithmen:

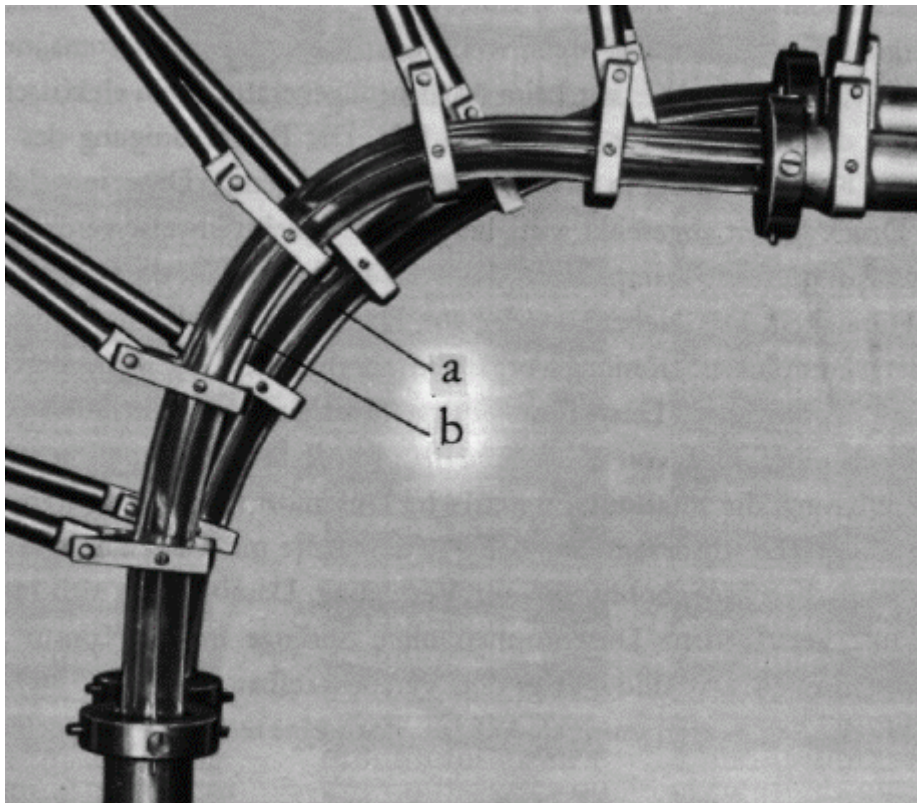
Evolutionsstrategien (ES)

- anfänglich eine Methode zum Finden optimaler Parameterkombinationen in Laborexperimenten (damals noch nicht rechnergestützt), Populationsgröße dann = 1
- Genom besteht aus Vektoren reeller Zahlen
- Mutation, Rekombination
- alle Komponenten eines Vektors werden simultan mutiert
- "Gen-Expression": Vektoren beschreiben reale Objekte
- 2 Vektoren je Individuum: Objekt-Parameter und Strategie-Parameter
- die Strategie-Parameter kontrollieren die Mutationen der Objekt-Parameter (und werden selbst optimiert – "Selbst-Adaptation")

Anwendungsbeispiel: Minimierung des Flusswiderstandes in einer Rohrverbindung (Rechenberg)

a: naive Lösung (viertelkreisförmige Verbindung)

b: optimale Lösung, durch ES gefunden



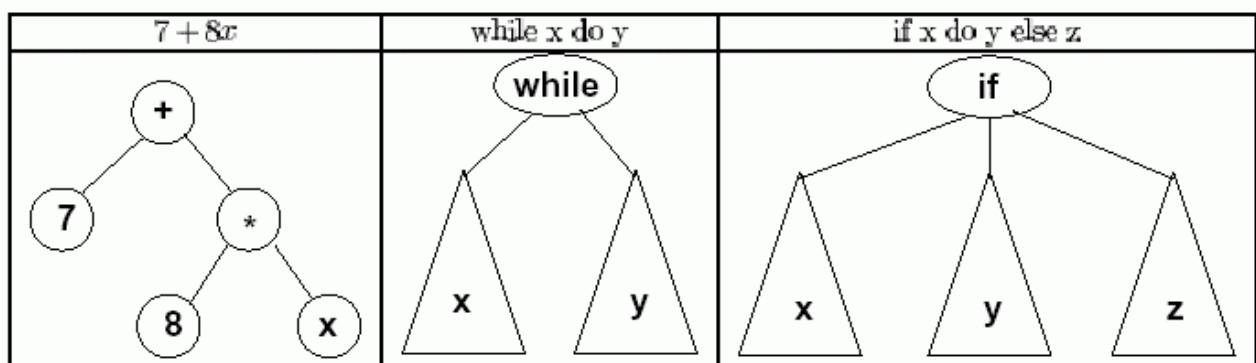
Evolutionäres Programmieren (EP)

- wie bei ES werden Vektoren anstelle von Strings evolviert
- kein crossing over
- Mutationen mittels normalverteilter Störterme
- Mutationsrate wird im Laufe eines Optimierungslaufes vermindert (entspr. Senkung der Temperatur)
- Verwandtschaft zum simulated annealing

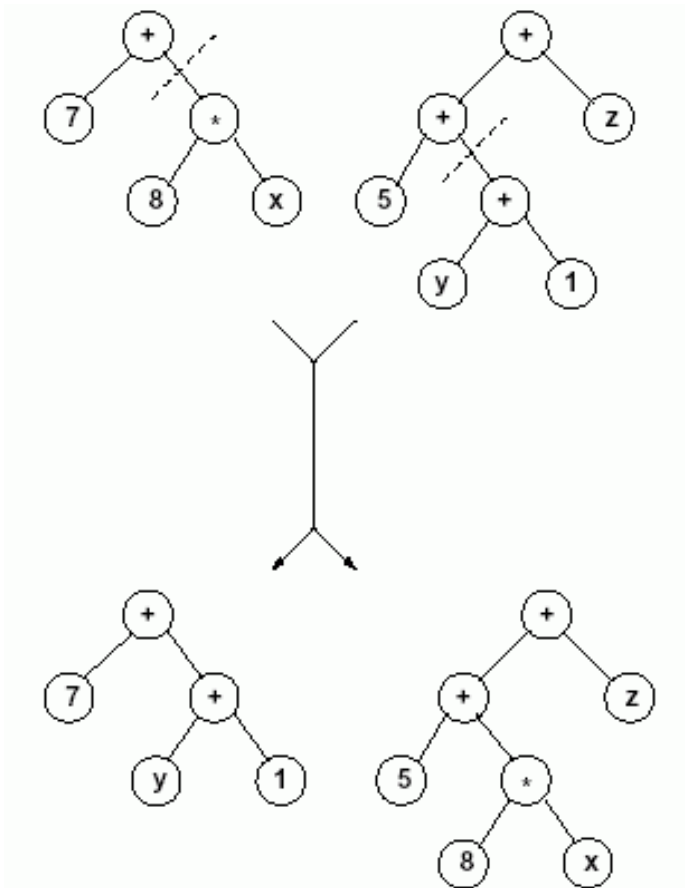
Genetisches Programmieren (GP; Koza 1987)

- komplette Computerprogramme werden evolviert
- Individuen = Terme (oft Lisp-Ausdrücke)
- Blattknoten = Konstanten, Variablen, Register oder externe Inputs
- Genome (Terme) haben variable Länge
- gegen Gefahr des unbegrenzten Wachstums: oft "Straf-Term" für Länge bei der Fitnessberechnung
- Koza verwendete keine Mutationen, nur crossing over!

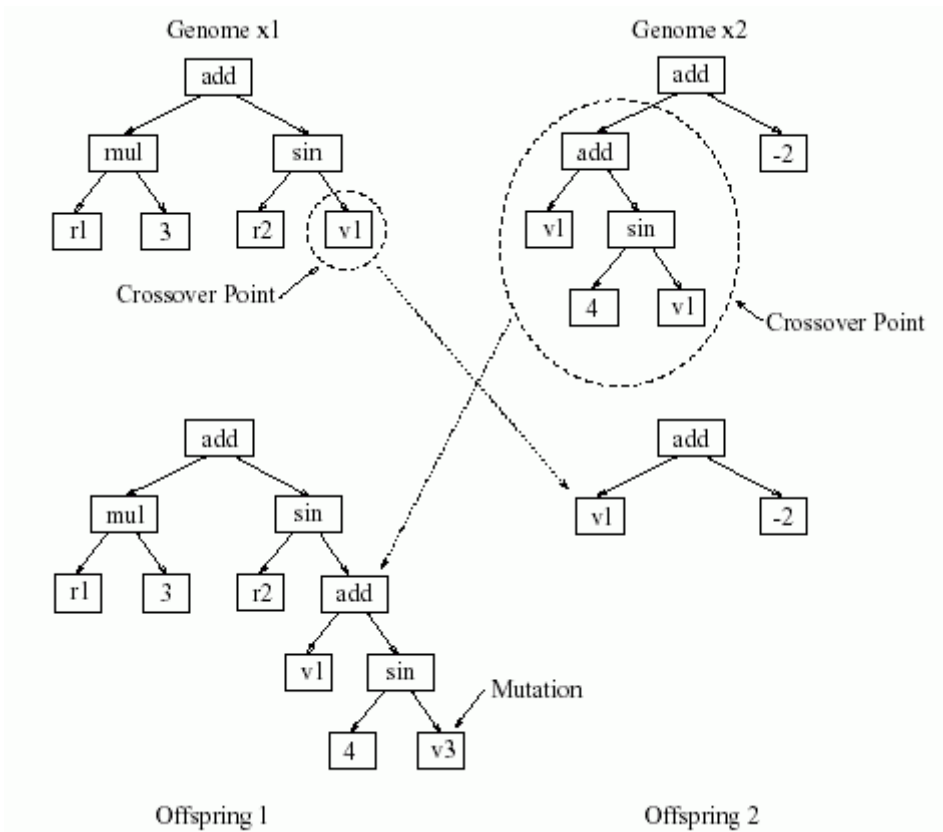
Beispiele für Terme:



Beim *crossing over* werden Subterme ausgetauscht:



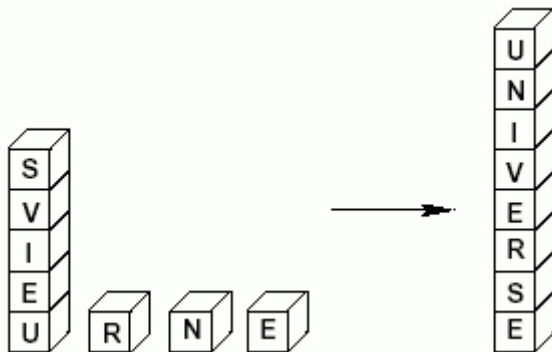
anderes Beispiel:



Anwendungsbeispiel für GP (aus Güde 2001):

wieder sollen Buchstaben so sortiert werden, dass ein vorgegebenes Wort entsteht

– aber diesmal evolviert nicht das Wort, sondern ein Programm, das Buchstaben von einem Vorrat ("Tisch") auf einen Stack legen kann (oder umgekehrt):



Befehlsvorrat:

Terminalzeichen:

| | |
|----|----------------|
| CS | current stack |
| NN | next necessary |

Operatorzeichen:

| | |
|--------|---------------|
| MS x | move to stack |
| MT x | move to table |
| EQ x y | equal |
| NOT x | not |
| DU x y | do y until x |

Mit dieser Sprache wurde eine Zufallspopulation von Programmen erstellt und auf 166 verschiedene Buchstabenkonfigurationen angesetzt. Nach nur 10 Generationen gab es ein Programm, welches in der Lage war, alle 166 Aufgaben zu meistern:

`(EQ (DU (MT CS)(NOT CS)) (DU (MS NN)(NOT NN)))`