



Matthias Maring (dir.)

Fallstudie zur Ethik in Wissenschaft, Wirtschaft, Technik und Gesellschaft

KIT Scientific Publishing

Autonome Systeme und evolutionäre Robotik: neues Paradigma oder Missverständnis?

Mathias Gutmann, Benjamin Rathgeber and Tareq Syed

Publisher: KIT Scientific Publishing
Place of publication: Karlsruhe
Year of publication: 2011
Published on OpenEdition Books: 31 mai 2017
Series: KIT Scientific Publishing
Electronic EAN: 9782821881587



<http://books.openedition.org>

Electronic reference

GUTMANN, Mathias ; RATHGEBER, Benjamin ; und SYED, Tareq. *Autonome Systeme und evolutionäre Robotik: neues Paradigma oder Missverständnis?* In: *Fallstudie zur Ethik in Wissenschaft, Wirtschaft, Technik und Gesellschaft* [Online]. Karlsruhe: KIT Scientific Publishing, 2011 (Erstellungsdatum: 08 septembre 2023). Online verfügbar: <<http://books.openedition.org/ksp/3609>>. ISBN: 9782821881587.

Autonome Systeme und evolutionäre Robotik: neues Paradigma oder Missverständnis?

Mathias Gutmann – Benjamin Rathgeber – Tareq Syed

Autonome Systeme (AS) stellen eine Form technischer Mittel dar, die in wesentlichen Hinsichten von klassischen Mitteln abweichen. Sie lassen sich nach unterschiedlichen Kriterien sortieren, wie Christaller et al. (2001) am Beispiel von Robotern zeigen (etwa nach Autoritäts- und Freiheitsgraden oder nach Nutzungsformen). Die Besonderheit von AS wird deutlich, wenn man – mit Blick auf den Gegensatz zu heteronomen Systemen – die Möglichkeiten der Kontrolle bedenkt, welche sich über verschiedene Formen technischer Systeme ausüben lassen. Diese hängen einerseits von der Struktur, andererseits vom jeweiligen Einsatz des Systems ab. Eine Einteilung technischer Systeme wird – wie bei Taxonomien üblich – zweckrelativ sein und sich auf eine Beschreibung der Systeme beziehen. So ließe sich etwa folgende, durchaus gängige Einteilung angeben (die, wie zumeist in solchen Fällen, fließende Übergänge und Mischformen aufweisen kann):

1. *Instrumentalisierung*: Vollständige oder starke Kontrolle wird dann vorliegen, wenn sowohl die Definition der Zwecke als auch die Verwendung der Mittel vollständig – oder zumindest im Wesentlichen – in den Händen des Nutzers liegt; das Mittel ginge in seinem Gebrauch auf. Exemplarisch kann hier an Hammer, Wagenheber oder ein Auto (älteren Typs) gedacht werden. Mit einem solchen starken Konzept von Kontrolle ist ein entsprechendes Verständnis von Störung verbunden, da die Definition der Zwecke ausschließlich auf Seiten des Nutzers liegt. Dies bedeutet, dass hier der Ausdruck *Störung* im Wesentlichen unter Bezug auf die Zwecke definiert wird.¹

2. *Maschinisierung*: Die Vollständigkeit der Kontrolle kann mit der Komplexität der Systeme abnehmen. Schon bei modernen Autos – welche mit zahlreichen, computergestützten Teilsystemen ausgestattet sind – kann von Kontrolle im ersten Sinn nicht mehr die Rede sein, wiewohl immerhin Überwachung durch den Nutzer stattfindet. Der Nutzer wahrt zwar die Zwecksetzungsautonomie vollständig, gibt aber die Kontrolle über die Zweckrealisierung (und damit über die Mittel) zumindest teilweise auf. Hierher gehören alle Systeme, deren Strukturen hinreichend komplex sind (wie etwa Autos mit Fahrerinformations- oder Warnsystemen). Die Identifizierung von Störungen ist hier grundsätzlich der Situation in 1. ähnlich. Allerdings werden jetzt – der Komplexität der Mittel entsprechend – erhöhte Anforderungen an die „richti-

¹ Zu möglichen Formen der Störung beim Hammer bzw. Hämmern s. Heidegger (1993, § 15).

ge“ Verwendung gestellt. Dies kann sich in einer entsprechenden Erhöhung des Diagnoseaufwandes ausdrücken, welcher selber zu einem Teilaspekt der Optimierung des Systems werden kann. Die Feststellung der Störung des Systems kann dementsprechend nicht mehr unter Bezug auf die Nutzerzwecke allein erfolgen.²

3. *Automatisierung*: Automatisierte Systeme liegen dort vor, wo definierte Funktionen durch ein System exekutiert werden, ohne dass ein Nutzer unmittelbar kontrollierend oder überwachend eingreift. Hierher gehören einfache Automaten wie etwa Aufzüge und Thermostate, aber auch komplexere Systeme wie automatische Schienenfahrzeuge. Am Beispiel des Automobils können hier solche Systeme genannt werden, welche über autonomisierte Teilfunktionen wie etwa automatische Gangschaltung und Temporegler verfügen. Die Zwecksetzungsautonomie zieht sich in solchen Fällen immer stärker auf die Angabe von Steuerungsparametern zurück (abgesehen von obligater Resteingriffsmöglichkeit), während die Realisierung der (Unter-)Zwecke zusehend dem System überlassen bleibt. Die Störungsdefinition wird sich nun von der Beurteilung des Mitteleinsatzes im engeren Sinne entfernen und nur noch auf die Gesamtfunktion des automatisierten Teilsystems beziehen. Bis auf Ausschlusskriterien (etwa Vermeidung von Beschädigungen der Umgebung) ist die Beurteilung der Funktion am Oberzweck orientiert. Ersichtlich hat eine solche Reduktion der Überwachung die Übernahme der Überwachungsfunktion als Teilfunktion des Systems selber zur Folge. Die Überwachung der Teilrealisierung der vorgegebenen Zwecke wird zur Funktion des Systems. Dadurch reduziert sich die Identifizierung von Störungen entweder gänzlich auf die Orientierung am Oberzweck oder wird – nur mit Blick auf die tatsächliche Nutzung – Gegenstand eines Aushandlungsdiskurses.³

4. *Autonomisierung*: Autonome Systeme im engeren Sinne sind schließlich solche, bei welchen nicht nur fixierte Funktionen ohne Intervention des Nutzers ausgeführt werden, sondern das System selbst in gegebenen Kontexten sowohl die auszuführenden Funktionen als auch die Formen ihrer Ausführung bestimmt. Während der „automatische Highway“ – bei welchem der „Fahrer“ eben *diese* Funktion vollständig verliert und nur noch die Zielangabe behält, die gesamte Verkehrsorganisation aber beim System verbleibt – noch in den Bereich der Technikvision gehört, findet der Bau adaptiver Systeme durchaus schon statt. Wiederum handelt es sich um ein Spektrum mit

2 Dabei ist die zunehmend komplexe Natur der Mittel nur eine Ursache; eine weitere besteht in den zunehmenden Möglichkeiten der Mittel-Umdeutung und der Zweckfiktion (dazu Janich 1998, Gutmann 2004).

3 Es lassen sich automatisierte Teilsysteme zu anderen als den ursprünglichen Zwecken einsetzen – wie alle anderen Mittel auch. Nur kommt hier der „sensorische“ Aspekt hinzu, was die Möglichkeiten erheblich erweitert.

unscharfen Grenzen. Dabei kann der Nutzer selber als Element der Systemumgebung angesehen werden, sodass in Interaktionen mit dem System die jeweiligen Zwecke identifiziert und die Mittel der Realisierung festgelegt werden können. Das System wäre dann nicht nur als adaptiv, sondern auch als lernend oder gar planend zu bezeichnen, da es weitgehend auch auf der Mittelwahlseite selbstständig agierte. Solche „entscheidenden“ Systeme bleiben gleichwohl an Zielfunktionen orientiert (etwa die Minimierung der Strecke beim Ausweichen von Hindernissen). Die Autonomie bleibt – durch die Bestimmung des Systems als eines *technischen* – in letzter Konsequenz immer heteronom (hinsichtlich der Zweckbestimmung), da ansonsten der Mittelcharakter aufgehoben würde. Die Steuerung des Systems und seine Kontrolle werden zusehends zur Funktion des Systems selber; dies reduziert die Unterscheidung von gestörten und ungestörten Systemen auf den Bezug zum Oberzweck und gibt damit der Möglichkeit Raum, den Grad der Störung zum Gegenstand der Interaktion mit dem System zu machen.

So avanciert also auch immer die Leistungen technischer Systeme zu sein scheinen, sie bleiben doch als technische in der logischen Grammatik von Mitteln, Werkzeugen und Medien beschreibbar – und in dieser Form wird sich die „ethische“ Bewertung derselben letztlich an der *Verlässlichkeit* solcher Systeme orientieren. Aus dieser Folgerung ergäbe sich eine deutliche Parallele zu „klassischen“ technischen Systemen, die eine Erweiterung der ethischen Betrachtung kaum erforderte.

Eine mögliche Abkehr, ja ein regelrechter Bruch mit dieser vertrauten Form scheint allerdings dort vorzuliegen, wo nicht nur die *Kontrolle*, sondern auch die *Erstellung* des Systems – zusehends oder schließlich vollständig – zu einer Leistung des Systems gemacht werden soll. In diesem Fall nämlich würde – so scheint es – auch die letzte Verbindung mit menschlichem Handeln als Zweckrealisierung unterbrochen, und es stellten sich nun in der Tat ganz und gar neue ethische Fragen. Wir wollen im weiteren einen Vorschlag untersuchen, welcher unter dem Titel der „evolutionären Robotik“ ein besonders anspruchsvolles Feld für die Herstellung von Systemen bietet, die mit biomorphen Metaphern (etwa als *adaptive*, *kognitive*, *evolutionäre* oder *selbstorganisierende*) bezeichnet werden. Während jedoch die logische Grammatik der Askription kognitiver Ausdrücke methodologisch intensiv untersucht wird (zu einigen damit verbundenen Grundproblemen vgl. Gutmann/Rathgeber 2010), liegen Rekonstruktionen solcher biomorphen Askriptionen bisher nur in wenigen Ausarbeitungen vor, so dass eine etwas genauere Darstellung erfolgen muss (dazu im Überblick Bölker/Gutmann/Syed 2010, ferner Gutmann 1996).

1. Das Evolutionäre und das Ingenieursparadigma

Evolutionäre Robotik (ER) gilt als eine der radikalsten Neuerungen im Feld autonomer Systemtechniken.⁴ In ihrer Zielsetzung lässt sie sich wie folgt umreißen:

„The basic idea behind evolutionary robotics goes as follows [...]. An initial population of different artificial chromosomes, each encoding the control system (and sometimes the morphology) of a robot, are randomly created and put in the environment. Each robot (physical or simulated) is then let free to act (move, look around, manipulate) according to a genetically specified controller while its performance on various tasks is automatically evaluated. The fittest robots are allowed to reproduce (sexually or asexually) by generating copies of their genotypes with the addition of changes introduced by some genetic operators (e.g., mutations, crossover, duplication). This process is repeated for a number of generations until an individual is born which satisfies the performance criterion (fitness function) set by the experimenter“ (Nolfi/Floreano 2000, 1).

Ganz offenkundig ist diese Darstellung metaphorischer Natur. Es werden Ausdrücke, die explizit den Biowissenschaften (vor allem der Evolutionsbiologie) entstammen und üblicherweise ihre Anwendung auf belebte Gegenstände finden, zur Charakterisierung der Leistungen von Artefakten genutzt: Denn die „Chromosomen“, von welchen hier die Rede ist, dürften sich der histologischen Anfärbung ebenso widersetzen, wie sie vermutlich auch keinen Phänotyp *qua* Expression eines Genotyps ausbilden; Roboter „handeln“ weder, noch haben sie Sex oder gelten in irgendeiner Weise als geboren. Lassen wir zunächst die Rolle des Experimentators außer Acht, so gibt es grundsätzlich zwei Möglichkeiten, methodologisch auf diese Darstellung zu reagieren:

1. Es könnte auf der bloß konnotativen Funktion des genutzten Vokabulars bestanden werden oder
2. Der Versuch einer Explikation dieser uneigentlichen Redeweisen wäre vorzunehmen.

Während im ersten Fall das gesamte Projekt der ER als schlichtes systematisches (Selbst-) Missverständnis zu gelten hätte, besteht der Reiz der zweiten Variante vor allem darin, verstehen und begreifen zu können, was der Referent der Beschreibungen eigentlich ist. Genau dies würde es überhaupt erst ermöglichen, „ethische“ Beurteilungen der in Rede stehenden Technologien vorzunehmen.

⁴ Eine systematische Rekonstruktion dieses Ansatzes im Vergleich zum Organic Computing s. Gutmann et al. (2011).

Das für ER zentrale neue (eben evolutionäre) Paradigma wird in Abgrenzung zum klassischen (Ingenieurs-)Paradigma charakterisiert. Das klassische (Ingenieurs-)Paradigma ist gemäß Nolfi & Floreano (2000, 6ff.) durch das Bemühen gekennzeichnet, die Lösung anstehender technischer Probleme durch systematische Dekomposition komplexer in einfache Einheiten zu erreichen (ein Verfahren, das schon Bertalanffy als „elementalistic“ bzw. „analytic“ charakterisierte und dessen Schwäche er durch systemtheoretische Integration zu beheben hoffte⁵; s. Bertalanffy 1969). Dies gilt sowohl für materielle wie nicht-materielle Eigenschaften von Robotern (nicht-materielle Eigenschaften wären z.B. „Flucht- und Ausweichverhalten“, „Navigation“ etc.). Sollen Roboter etwa Umgebungsinformationen sammeln, diese verarbeiten und gemäß derselben agieren, so wäre eine dem Ingenieursparadigma gehorchende Reihung jene von äußerer Wahrnehmung, Daten- und Informationsverarbeitung, Modellierung, Handlungs-Planung und Handlungs-Ausführung. Es schließt sich die inkrementelle Implementierung sowohl der notwendigen „behavior modules“ als auch der „control-systems“ an. Regelmäßig erweisen sich die Resultate solcher vom Beobachter gesteuerten Konstruktionen als suboptimal (vgl. Nolfi/Floreano 2000, 6ff.; sowie Thrun/Burgard/Fox 2006).

Demgegenüber ist das evolutionäre Paradigma durch das Fehlen einer zentralen und koordinierten Planung charakterisiert. Die Veränderungen, die den evolvierenden Einheiten unterliegen, sind definitionsgemäß nicht intentional. Zudem sind sie abhängig von einer Reihe von Parametern, die mit den funktionalen Strukturen der Individuen (aus welchen die Einheiten bestehen), deren Verhalten untereinander sowie deren Verhalten zu belebten und unbelebten Umgebungsaspekten zusammenhängen. Die Transformation der evolutiven Einheit wird bezüglich einer Fitness-Funktion selektiv bewertet, wobei die Steigerung der Fitness definitionsgemäß als „Erfolg“ zu werten ist. Nach dieser Auffassung würde die Rolle des Ingenieurs im evolutionären Paradigma wesentlich transformiert. Dessen Funktion bestünde nun vor allem in der Initiierung eines Selbst-Optimierungsprozesses, einer System-Generierung. Eine direkte Kontrolle über die Auszeichnung der Systemparameter oder der zu ihrer Gewinnung notwendigen Transformationsschritte darf hierfür zumindest dann nicht erforderlich sein, wenn an dem Anspruch der Selbsterstellung festgehalten wird. Das System wäre dadurch der Autor seiner eigenen Veränderung (Nolfi/Floreano 2000, 9). Die Ausnutzung – so ließe sich diese Neubestimmung der Rolle des Ingenieurs umschreiben

⁵ Es handelt sich allerdings um ein *methodologisches* und *nicht* um ein empirisches Problem, sodass schon die Resultate der Dekomposition nur zweckrelativ identifiziert werden können; erst eine methodologische Rekonstruktion erlaubt es, die naheliegende Reifizierung der „Basis-Verhalten“ zu vermeiden (dazu Hartmann 1998).

– von „natürlichen Prozessen“ führte zu „Quasi-Organismen“, die ihren konventionell erzeugten (robotischen) Vorläufern deutlich überlegen wären. Der resultierende Prozess der Selbst-Optimierung erlaubte die flexible und zugleich angemessene Veränderung der evolvierenden Einheiten auch unter Bedingungen der Veränderung von Umgebungseigenschaften – und mithin der jeweiligen Fitness-Funktion.

2. Fitness = Fitness?

Das Verständnis dessen, was ER von klassischem Engineering unterscheidet hängt wesentlich am Verständnis des Ausdrucks *Fitness*. Genau an dieser Stelle erweist sich allerdings eine Homonymie als problematisch, die schon innerhalb der biowissenschaftlichen Theoriebildung für einige Verwirrung sorgte (s. Gould/Lewontin 1979): denn die Fitness-Funktion ist im Falle der Erzeugung von Robotern auf „evolutionärem Wege“ mit der Zwecksetzung des Experimentators verknüpft. Dieser definiert notwendigerweise sowohl die eigentlichen Aufgaben (etwa Finden und Durchfahren einer Öffnung) als auch die dazugehörigen Leistungskriterien:

„The fitness function is a performance criterion that evaluates the performance of each individual phenotype. Higher fitness values are better. For example, if one wants to optimize a mathematical function, individuals coding for values that generate higher returns will receive higher fitness. Similarly, one may attempt to maximize the difference between the value returned by the function and a target value“ (Nolfi/Floreano 2000, 20).

Eine solche Zielvorgabe ist aber im Rahmen eines darwinistischen Verständnisses von Evolution *notwendig* ausgeschlossen: Es ist ja eine entscheidende Leistung desselben, evolutionäre Prozesse *ohne* teleologische Annahmen modellieren zu können (Mayr 1979). Es bleibt nur die Möglichkeit, Evolution unumwunden als Optimierungsvorgang zu verstehen, eine These, die zu erhärten durch die Anführung einiger typischer Beispiele unternommen wird⁶ (hierzu etwa Reiter 1995 am Beispiel des Knochens oder Mosbrugger/Roth 1995 für Landpflanzen, vgl. ferner Rechenberg 1973, Wickler/Seibt 1983 und Maynard Smith 1982). Akzeptieren wir zunächst die These, dass Evolution Optimierung *sei*, so stellt sich allerdings nun methodologisch die Frage, woher wir um diese Eigenart von Evolution eigentlich wissen. Signifikanterweise lässt sich „Optimierung“ zunächst im Feld der

⁶ Genau genommen bedeutet Optimierung dabei jeweils unterschiedliches, womit es sich auch hier um einen homonymen Ausdruck handelte. Es besteht eben zunächst nur eine Familienähnlichkeit zwischen der Optimierung eines Phylogrammes und der eines Schraubenschlüssels. Eigentlich wäre die Verwendung des Wortes als Begriff also zu indizieren (weiterführend dazu Gutmann 1996).

Technik selber sehr gut erläutern. Wir verstehen darunter einen mehrstelligen Ausdruck, dem gemäß etwa „x durch A in Hinsicht auf Z mithilfe von y optimiert“ wird; „optimieren“ als Tatprädikat ist dann noch durch den Ausdruck „verändern“ zu ersetzen, wobei Z die Zielfunktion der Veränderung angibt. Grundlage solcher Verbesserungen sind zunächst adäquate Beschreibungen des zu optimierenden Gegenstandes (etwa der Rumpf eines Schiffes), wobei die Adäquatheit durch Bezug auf die relevanten Parameter sicherzustellen ist.⁷ Formal handelt es sich hier um jene Gleichungsfaktoren und Randbedingungen, von welchen die zu optimierende Größe beeinflusst wird (dazu Nachtigall 1994, 1995). Es ist ferner eine Zielfunktion anzugeben, die es erlaubt, die gewünschten Extremwerte für die Optimierung zu bestimmen. Damit ist ein Nicht-Artefakt so beschrieben, als wäre es ein Artefakt:

„Man kann Organismen ja auch parallel setzen zu technischen Gebilden, sozusagen als real existierende Systeme betrachten, die – wie immer sie sich entwickelt haben – erkennbare Struktur-Funktions-Beziehungen ausgebildet haben“ (Nachtigall 1995, 140).

Es ist eine solche „Parallelsetzung“ in der Tat genau dann möglich, wenn als *Tertium* von Lebewesen und Artefakt die Gesetze⁸ in Anschlag gebracht werden, bezüglich deren die Funktion des Artefakts beschreibbar sind. Auf diese Weise lassen sich – um einige einschlägige Beispiele zu nennen – etwa Kriterien, die für die Schmierung technischer Gelenke relevant sind, auf Gelenke im Schulter- und Hüftgelenkbereich höherer Wirbeltiere übertragen⁹.

Am Modell etwa der Ermittlung eines Optimalwertes für gewisse statische Kenngrößen eines Baumes lässt sich diese Betrachtung erläutern. Dabei wird zunächst die geforderte Beschreibung des Baumes angefertigt (unter Nutzung technischer Beschreibungssprachstücke):

„Der Hauptträger eines pflanzlichen Hochbausystems (Halm eines Grases, Schaft eines Bambusgewächses, Stamm eines Baumes) sollte aus Materialgründen und aus energetischen Gründen – Materialanhäufung kostet Bauenergie – nirgendwo einen größeren Radius r erreichen, als zum Abfangen einer zulässigen Spannung σ zu l gerade nötig ist. Er wäre dann, analog zu vielen baustatischen Kon-

7 Evident geht hier empirisches Wissen ein, denn welche Parameter relevant sind, wäre im Rahmen von Vorversuchen zu ermitteln.

8 Es muss sich hier übrigens nicht nur um etwa physikalische Naturgesetze handeln; vielmehr kommen auch technische Konstruktionszusammenhänge in Frage (dazu Bonik/Gutmann 1981, Reuleaux 2007, Hilliges et al. 1992, 1994).

9 Weitere Beispiele: Haft- und Gleitreibungsparameter werden bei der Beschreibung von Reifenbewegungen auf Bewegungen von Hufen und Fußsohlen übertragen; Strömungsparameter von Schiffsrümpfen auf die Gestalt von Fischen.

strukturen der Technik, ein „Körper gleicher Festigkeit“ (Nachtigall 1995, 140f.).

Der Baum wird demnach als Hochbausystem beschrieben, wobei sein Stamm als eigentlicher Hauptträger gilt. Legt man eine solche Beschreibung zugrunde, so ist die Identifizierung von Optimalwerten möglich, welche nach der Theorie für „kreisrunde Träger gleicher Festigkeit“ ermittelt wird (Nachtigall 1995, 141f.). Nimmt man – im Lichte der Beschreibung des Lebewesens „als“ Hochbausystem – empirische Messungen vor, lässt sich die Güte der Modellierung bestätigen.

Nehmen wir nun an, solche Messungen verlaufen zufriedenstellend, so lässt sich der Baum beschreiben, *als ob* er ein – bezüglich des identifizierten Parameters und der Nutzung gewisser Strukturen sowie deren Anordnungen – optimiertes Hochbausystem wäre. Es bleibt das „Als-ob“ aber notwendig erhalten und insofern kann Evolution so verstanden werden, als handele es sich um einen technischen Vorgang. Methodologisch aber rekonstruieren wir Aspekte von Evolution (in diesem Fall etwa funktionale Veränderungen der morphologischen Struktur) durch Nutzung technischer Konstruktionen *als Modelle* für funktionale Transformationen (im Detail dazu Gutmann 1996). Berücksichtigen wir die modellierende Zwecksetzung nicht, so entsteht der Eindruck, die Beschreibung des Lebewesens hätte denselben methodologischen Status wie die Beschreibung des Artefakts; im Gefolge dieser Identifikation würden wir mit lebenswissenschaftlich kaum zu begründenden Vermutungen über den „gestalteten“ Charakter von Lebewesen konfrontiert, wie sie im Rahmen von Konzepten des Intelligent Design (in der Fortsetzung kreationistischer Ansätze) ausgewiesen werden können. Insofern kann Nolfi & Floreano (2000) nur zugestimmt werden, wenn sie auf die Homonymie des Ausdruckes Fitness hinweisen:

„This value (which expresses the fitting between organisms and its environment, the authors) is the number of offspring in the case of natural evolution and the fitness value in the case of artificial evolution“ (Nolfi/Floreano 2000, 154).

Dies führt zur Einsicht, dass die evolutionäre Konstruktion autonomer Systeme nur dann überhaupt *als Konstruktion* angesprochen werden kann, wenn sie zur Realisierung von Zwecken führt. Eine solche Realisierung ist aber mit jenen Theorieteilern, von welchen sich die Vertreter der evolutionären Robotik eine Ersetzung des Ingenieursparadigmas erhoffen, schlicht unverträglich. Dies mündet in die weitergehende Frage ein, was die – für die Zwecke der evolutionären Robotik – relevante Form von Evolutionstheorie überhaupt ist:

„One might wonder, whether Lamarckian evolution [...] could be more effective than Darwinian evolution [...]. Ackley and Littmann [...]

for instance claimed that in artificial evolution, where inherited characters can be easily coded into the genotype, given that the mapping between genotype and phenotype is generally quite simple, there is no reason for not using Lamarckian evolution. Indeed the authors showed that Lamarckian evolution is far more effective than Darwinian evolution in a stationary environment. On the other hand, as shown by Sasaki and Tokyo [...], Darwinian evolution largely outperforms Lamarckian evolution when the environment is not stationary or when individuals are exposed to different learning experiences” (Nolfi/Floreano 2000, 287).

Selbst im Falle der – bei instabilen Umgebungen – als „besser“ angesehenen „Darwinian Evolution“ gilt der Unterschied bezüglich der resultierenden Roboter nur dann, wenn diese *technisch* bessere Lösungen sind. Es zeichnet sich damit ein Dilemma ab, das wir als jenes des *artifex absconditus*¹⁰ bezeichnen können. Dieses Dilemma stellt sich genau dann ein, wenn wir zugleich an technischen Fitness-Kriterien und an darwinistischer Evolution festhalten wollen.

3. Das Dilemma des *artifex absconditus*

Auf der Grundlage der vorgelegten Rekonstruktion lässt sich das oben angekündigte Dilemma ausformulieren, welches der Bewertung autonomer Systeme vorangeht:

1. Entweder wird an dem Anspruch festgehalten, es handele sich bei ER um Engineering; dann wird der Ausdruck „evolutionary“ eine metaphorische Beschreibung von Ingenieurshandeln sein, in dessen Gefolge menschliche Zwecke realisiert werden – und zwar nach Kriterien des Gelingens und Misslingens der entsprechenden Konstruktionen sowie dem Erfolg ihrer Veränderung (sei es als Optimierung, sei es als Differenzierung).
2. Oder es wird am Anspruch festgehalten, es handele sich bei ER um evolutionäre Vorgänge; dann wird der Ausdruck „Engineering“ eine metaphorische Beschreibung der Veränderung von Fitness-values nach Maßgabe einer wie auch immer gearteten Selection-Landscape. Im Gefolge dieser Fitness-Veränderungen lässt sich dann nicht mehr zwischen gelingender und misslingender oder erfolgreicher und erfolgloser Zweckrealisierung unterscheiden (es wäre dies ja nur die durch einen am Ingenieurshandeln orientierten Beobachter vollzogene Bewertung).

Trifft das Erste zu, so ist das evolutionäre Paradigma lediglich eine Art der Realisierung menschlicher Zwecke von technischen Systemen (etwa deren

¹⁰ Die Bezeichnung ‚*artifex absconditus*‘ ist angelehnt an ‚*deus absconditus*‘ – der verborgene Gott – und bezieht sich auf dessen prinzipielle Nichterkennbarkeit.

Optimierung) – und mithin eine Form technischen Handelns. Soll dies zugleich als evolutionär im wörtlichen Sinne gelten, dann müsste es sich um einen nicht-darwinistischen Modus handeln, der unterstellte, dass Evolution faktisch designartig strukturiert ist (etwa hinsichtlich Optimierung).

Trifft das Zweite zu, dann führt das evolutionäre Paradigma nicht mehr zur Realisierung menschlicher Zwecke.¹¹ Es ließe sich am darwinistischen Ansatz festhalten, wobei dann die Bestimmung der Veränderungen am Gegenstand nur noch metaphorisch als Optimierung zu verstehen wäre.

Dieses Dilemma verdeutlicht, dass es unmöglich ist, unter Beibehaltung ingenieurwissenschaftlicher Standards das Ingenieurhandeln als „evolutionär“ zu charakterisieren und *zugleich* unter Beibehaltung lebenswissenschaftlicher Standards die Resultate evolutionärer Transformation als „ingenieursartig“ auszugeben.

4. Rechtfertigbarkeit evolutionärer autonomer Systeme

Nach der systematischen Rekonstruktion zentraler Sprechweisen, die für die Beschreibung von Eigenschaften autonomer technischer Systeme im Zusammenhang der ER verwendet werden, kann nun auch die eingangs angesprochene Frage beantwortet werden, inwieweit Systeme, die nach Maßgabe des evolutionären Paradigmas hergestellt wurden, als „verlässlich“ einzustufen sind. Nehmen wir nämlich an, dass die Verlässlichkeit solcher Systeme eine notwendige Bedingung für die Möglichkeit der Verantwortung ihres Baues und Betriebes ist, so kann unabhängig von der näheren Darstellung von „Verantwortung“ (wir verzichten zugleich ausdrücklich auf die autoritäre Fassung, welche beispielsweise Jonas diesem Begriff gab und lassen eine Einführung hier auf sich beruhen)¹² eine einfache Fallunterscheidung vorgenommen werden:

1. Halten wir an dem wörtlichen Verständnis von „evolutionär“ fest, so haben wir es definitionsgemäß mit Systemen zu tun, deren Verhalten nicht einfachhin prognostizierbar ist. Da die Kriterien des Gelingens von Problemlösungen – im Extremfall vollständig – in der Autorenschaft der Systeme lägen, ist ein Vertrauen in solche Systeme nicht gerechtfertigt.
2. Halten wir an dem metaphorischen Verständnis von „evolutionär“ fest, so steht die Zweckbestimmung selber zur Rechtfertigung an. Da das Ausmaß der Verlässlichkeit des Systems hier vom Ausmaß der Prognosti-

11 Dies gälte übrigens auch dann, wenn die bewertende Beschreibung des Beobachters eine vorliegende Variante als eine solche Problemlösung charakterisierte.

12 Autoritäre Implikationen dieser Art liegen auch bei der Weberschen Konzeption von Verantwortsethik nahe (Weber 1919); immerhin sind sie bei Weber offen reflektiert und aus der Problemstellung der Rechtfertigung von Macht heraus entwickelt. – Vgl. auch Jonas (1979).

zierbarkeit seines Verhaltens¹³ abhängig gemacht wurde, sind nicht nur die Zwecke selber zu rechtfertigen, sondern zudem die Risiken (welche erst nach der Rechtfertigung als zumutbar gelten können).

Wird also das Ingenieursparadigma – aller evolutionären Rhetorik zum Trotz – beibehalten, so lassen sich für „evolutionär“ hervorgebrachte autonome Systeme durchaus Rechtfertigungsbedingungen formulieren. Ob aber und in welchem Ausmaß diese auch als faktisch erfüllt gelten – im Sinne der tatsächlichen Rechtfertigung –, ist dann nicht mehr Gegenstand angewandt-ethischer oder gar philosophischer Erwägungen, als vielmehr die Aufgabe der Technikfolgenabschätzung.

5. Literatur

- Bertalanffy, L. v. (1969): *General System Theory*. New York 1969.
- Bölker, M. – Gutmann, M. – Syed, T. (2010): Existiert „genetische Information“? S. 155–181 in Bölker, M. – Gutmann, M. – Hesse, W. (Hrsg.): *Menschenbilder und Metaphern im Informationszeitalter*. Münster 2010.
- Bonik, K. – Gutmann, W.F. (1981): *Kritische Evolutionstheorie*. Hildesheim 1981.
- Christaller, T. et al. (2001): *Robotik*. Heidelberg – Berlin 2001.
- Gould, S.J. – Lewontin, R.C. (1979): The spandrels of san marco and the panglossian paradigm: A critique of the adaptationist programme. S. 581–598 in *Proceedings Of The Royal Society of London, Series B* 205 (1979) 1161.
- Gutmann, M. (1996): *Die Evolutionstheorie und ihr Gegenstand – Beitrag der Methodischen Philosophie zu einer konstruktiven Theorie der Evolution*. Berlin 1996.
- Gutmann, M. (2004): *Erfahren von Erfahrungen. Dialektische Studien zur Grundlegung einer philosophischen Anthropologie*. 2 Bd. Bielefeld 2004.
- Gutmann, M. – Rathgeber, B. (2010): Kognitive Metaphern. S. 115–139 in Bölker, M. – Gutmann, M. – Hesse, W. (Hrsg.): *Information und Menschenbild*. Berlin – Heidelberg 2010.
- Gutmann, M. – Rathgeber, B. – Syed, T. (2011): Organic Computing – Metaphor or Model? Kap. 1.7 in Müller-Schloer, C. – Schmeck, H. Ungerer, T. (Hrsg.): *Organic Computing – A Paradigm Shift for Complex Systems*. Basel 2011.
- Hartmann, D. (1998): *Philosophische Grundlagen der Psychologie*. Darmstadt 1998.
- Heidegger, M. (1993): *Sein und Zeit*. Tübingen 1993.
- Hilliges, M. et al. (Hrsg.) (1992): *Natural Structures. Proceedings of the SFB 230*. Stuttgart (1992) 8.

13 Es wären hier also graduelle Unterschiede denkbar, sodass in dieser Beurteilungsklasse weitere Kriterien für die Verantwortbarkeit von Bau und Betrieb der Systeme einführbar sind.

- Hilliges, M. et al. (Hrsg.) (1994): Evolution of natural Structures. Proceedings of the SFB 230. Stuttgart (1994) 9.
- Janich, P. (1998): Die Struktur technischer Innovationen. S. 129–177 in Hartmann, D. – Janich, P. (Hrsg.): Die Kulturalistische Wende. Frankfurt a.M. 1998.
- Jonas, H. (1979): Das Prinzip Verantwortung. Frankfurt a.M. 1979.
- Maynard Smith J (1982): Evolution and the Theory of Games. Cambridge – London – New York 1982.
- Mayr, E. (1979): Evolution und die Vielfalt des Lebens. Berlin – Heidelberg – New York 1979.
- Mosbrugger, V. – Roth, A. (1995): Evolution und Optimierung am Beispiel der Wassertransportsysteme der Landpflanzen. S. 155–169 in Kull, U. – Ramm, E. – Reiner, R. (Hrsg.): Evolution und Optimierung. Stuttgart 1995.
- Nachtigall W. (1994): Technische Biologie und Bionik: Procedere – Problem – Perspektiven. S. 285–298 in Maier W. – Zoglauer T. (Hrsg.): Technomorphe Organismuskonzepte. Stuttgart 1994.
- Nachtigall W. (1995): Zum Optimierungsbegriff in der Biologie. Ableitung, Ansatzmöglichkeiten, Aussagegrenzen. S 137–154 in Kull, U. – Ramm, E. – Reiner, R. (Hrsg.): Evolution und Optimierung. Stuttgart 1995.
- Nolfi, S. – Floreano, D. (2000): Evolutionary Robotics. Cambridge 2000.
- Rechenberg, I. (1973): Evolutionsstrategie. Stuttgart 1973.
- Reiter, T.J. (1995) Knochen – eine selbstoptimierende Struktur? S. 121–136 in Kull, U. – Ramm, E. – Reiner, R. (Hrsg.): Evolution und Optimierung. Stuttgart 1995.
- Reuleaux, F. (2007): Theoretische Kinematik (Reprint). Saarbrücken 2007.
- Thrun, S. – Burgard, W. – Fox, D. (2006): Probalistic Robotics. Cambridge, MA – London 2006.
- Weber, M. (1919): Politik als Beruf. München 1919.
- Wickler, W. – Seibt, U. (1983): Männlich, weiblich. München 1983.

6. Fragen

Im Hinblick auf die oben entwickelten Überlegungen lassen sich folgende Fragen formulieren:

- Was unterscheidet metaphorische Redeweisen von Kategorienfehlern¹⁴? Ist die Unterscheidung trennscharf? Bedient sich die *evolutionäre Robotik* bloß metaphorischer Redeweisen oder macht sie einen Kategorienfehler?
- Was ist kennzeichnend für das Ingenieurs-, was für das evolutionäre Paradigma?

14 Ein Kategorienfehler wäre etwa die fehlerhafte Übertragungen bestimmter Ausdrücke aus einem ‚Bereich‘ in einen anderen – hier z.B. die Übertragung bestimmter menschlicher Eigenschaften auf Maschinen und Tiere.

- Ist Evolution identisch mit Optimierung?
- Was bedeutet es, Evolution als Optimierungsprozess zu beschreiben?
- Inwiefern stellt das rekonstruierte Dilemma des *artifex absconditus* die Möglichkeit evolutionärer Robotik in Frage?
- Ist die technische Herstellung autonomer Systeme möglich, welche als *denkend* oder *handelnd* beschrieben werden? Was spräche dafür, was dagegen?
- Unter welchen Bedingungen ist der Bau und Betrieb *autonomer Systeme* im Sinne evolutionärer Robotik möglich?
- Welche Differenz besteht zwischen der Rechtfertigbarkeit und Rechtfertigung des Baues und Betriebes technischer Systeme?
- Wer ist für *autonome Systeme* bzw. *Systeme der evolutionären Robotik* verantwortlich? Der Hersteller, der Nutzer oder die Systeme selbst? Wer haftet gegebenenfalls (rechtlich)?