

## **Anhang**

### **Die Grenzen der Gesamtfitness**

Eine wichtige Rolle bei der Erklärung der biologischen Herkunft von Altruismus und fortgeschrittener sozialer Organisation spielt heute die theoretische Genetik. Da um dieses Thema in jüngerer Zeit mit zahlreichen Publikationen eine große Kontroverse geführt wird, füge ich hier eine neuere Analyse der Gesamtfinesstheorie an, aus der auch ersichtlich wird, warum wir diese Theorie durch die datenbasierte Populationsgenetik ersetzen sollten. Das vorliegende Material stammt aus einem bereits veröffentlichten Forschungsbericht, aus dem freilich die mathematischen Analysen und die Referenzen für unsere Zwecke gestrichen wurden. Der Artikel wurde vor seiner Veröffentlichung umfassend von wissenschaftlichen Experten begutachtet.

Literaturangabe:

Benjamin Allen, Martin A. Nowak und Edward O. Wilson, »Limitations of Inclusive Fitness«, in: *Proceedings of the National Academy of Sciences USA*, Bd. 110, Nr. 50, S. 20135-20139 (2013).

### **Bedeutung**

Die Gesamtfinesstheorie ist die Hypothese, dass der evolutionäre Erfolg eines Merkmals sich berechnen lässt durch die Addition von Finesseffekten, multipliziert mit Verwandtschaftskoeffizienten. Obwohl neuere mathematische Analysen nachweisen, dass dieser Ansatz nur unter erheblichen

Einschränkungen gilt, behaupten seine Anhänger, er sei genauso allgemeingültig wie die Theorie der natürlichen Selektion selbst. Diese Behauptung stützt sich auf eine lineare Regressionsanalyse, mit der die Fitness eines Individuums in Komponenten zerlegt wird, die auf das Individuum selbst beziehungsweise auf andere zurückgehen. Wir weisen nach, dass diese Regressionsmethode die Vorhersage und Interpretation von Evolutionsprozessen nicht leisten kann. Insbesondere kann sie nicht zwischen Korrelation und Kausalität unterscheiden und führt damit zur Fehlinterpretation ganz einfacher Szenarien. Die Schwächen der Regressionsmethode unterstreichen die Einschränkungen der Gesamtfisnesstheorie insgesamt. Bis vor kurzem war die Gesamtfisnesstheorie als allgemeine Theorie zur Erklärung sozialen Verhaltens allgemein anerkannt. Im Anschluss an bereits früher geäußerte Kritik zeigen wir hingegen, dass die Gesamtfisness ein nur eingeschränkt gültiges Konzept ist, das sich nur auf wenige Evolutionsprozesse anwenden lässt. Die Gesamtfisnesstheorie ist die Hypothese, dass sich persönliche Fitness aus additiven Komponenten zusammensetzt, die durch individuelle Handlungen verursacht werden. Diese Hypothese gilt für die meisten evolutionären Prozesse oder Szenarien nicht. Um diese Einschränkung zu umgehen, nutzen Gesamtfisnesstheoretiker die Methode der linearen Regression. Auf Grundlage dieser Methode werden folgende Behauptungen aufgestellt: 1. Die Gesamtfisnesstheorie prognostiziert, in welche Richtung sich die Allelfrequenz verändert. 2. Die Gesamtfisnesstheorie zeigt die Gründe für diese Veränderungen auf. 3. Die Gesamtfisnesstheorie ist so allgemeingültig wie die natürliche Selektion. 4. Die Gesamtfisnesstheorie stellt ein universelles Gestaltungsprinzip der Evolution dar. In dieser

Veröffentlichung bewerten wir diese Behauptungen und widerlegen sie. Um zu analysieren, ob Mutationen, die das Sozialverhalten verändern, von der natürlichen Selektion gefördert oder verhindert werden, ist kein Aspekt der Gesamtfittesstheorie nötig.

Die Gesamtfittesstheorie ist ein Ansatz, mit dem Fitteseffekte der sozialen Evolution erklärt werden. Eingeführt wurde sie 1964 von W. D. Hamilton, der nachwies, dass unter bestimmten Bedingungen die Evolution Organismen selektiert, die die höchste Gesamtfittes besitzen. Dieses Ergebnis wurde als Gestaltungsprinzip interpretiert:

Organismen, die der Evolution unterliegen, verhalten sich so, dass sie ihre Gesamtfittes maximieren.

Hamilton selbst definierte die Gesamtfittes so:

<E1>»Die Gesamtfittes ist vorstellbar als die persönliche Fittes, die ein Individuum faktisch dadurch ausdrückt, dass es Nachkommen hervorbringt, die das Erwachsenenalter erreichen. Sie ergibt sich, indem man erst etwas abzieht und dann etwas hinzufügt. Abgezogen werden all die Komponenten, die auf die soziale Umwelt des Individuums zurückzuführen sind; damit bleibt die Fittes zurück, die es aufweisen würde, wenn es keinem der Nach- und Vorteile dieser Umwelt ausgesetzt wäre. Zu diesem Betrag werden dann in bestimmten Bruchteilen die Nach- und Vorteile hinzugefügt, die dieses Individuum für die Fittes seiner Nächsten erbringt. Diese Bruchteile sind einfach die Koeffizienten des Verwandtschaftsgrads zu den Nächsten, die es beeinflusst: Eins für geklonte Individuen, ein Halb für Geschwister, ein Viertel für Halbgeschwister, ein Achtel für Cousins ... und schließlich Null für sämtliche Nächsten, deren Verwandtschaftsgrad so gering ist, dass er vernachlässigt werden kann.«

Zwar verwenden moderne Formulierungen der Gesamtfittesstheorie andere Verwandtschaftskoeffizienten, aber sämtliche übrigen Aspekte in Hamiltons Definition bleiben unberührt.

Entscheidend ist hier die Hypothese, dass persönliche Fitness sich in additive Komponenten zerlegen lässt, die durch individuelle Handlungen festgelegt werden. Von der persönlichen Fitness eines bestimmten Individuums werden alle Komponenten abgezogen, die auf die »soziale Umwelt« zurückzuführen sind; wir müssen also von der persönlichen Fitness eines Individuums sämtliche Effekte subtrahieren, die von anderen Individuen verursacht werden. Danach müssen wir berechnen, wie unser Individuum die persönliche Fitness aller übrigen Individuen in der Population beeinflusst. In beiden Fällen müssen wir annehmen, dass die persönliche Fitness sich als Summe der Komponenten ausdrücken lässt, die durch individuelle Handlungen verursacht werden. Gesamtfittes ist die Auswirkung der eigenen Handlungen auf den Handelnden, addiert mit den Auswirkungen der Handlungen auf andere, und das jeweils multipliziert mit dem Verwandtschaftskoeffizienten zwischen dem Handelnden und den anderen.

Es ist offensichtlich, dass die Hypothese der Additivität, die für den Begriff der Gesamtfittes ganz grundlegend ist, nicht unbedingt allgemein gültig ist. So kann die persönliche Fitness eines Individuums zum Beispiel eine nicht-lineare Funktion der Handlungen anderer sein. Oder das Überleben eines Individuums könnte gleichzeitiges Handeln mehrerer anderer erfordern; so hängt etwa der Fortpflanzungserfolg der Ameisenkönigin vom koordinierten Handeln mehrerer Gruppen spezialisierter Arbeiterinnen ab. Im Experiment wurde nachgewiesen, dass die Fitteseffekte von kooperativem Verhalten bei Mikroorganismen nicht additiv sind. Ganz

eindeutig kann die Hypothese, Fitnesseffekte seien additiv, keine allgemeine Gültigkeit beanspruchen.

### **Zwei Ansätze zur Gesamtfitness**

In der Literatur zur Gesamtfitness gibt es zwei Ansätze, um mit der Einschränkung der Additivität umzugehen. Der erste Ansatz besteht darin, sich nur mit vereinfachten Modellen zu befassen, bei denen Additivität besteht. So taucht etwa in William D. Hamiltons ursprünglicher Formulierung der Gesamtfisnesstheorie Additivität nur als Annahme auf. Additivität ergibt sich auch aus der Hypothese, dass Mutationen sich nur geringfügig auf den Phänotyp auswirken und dass die Fitness von einem Phänotyp zum anderen geschmeidig variiert.

M. A. Nowak, C. E. Tarnita und E. O. Wilson untersuchten die mathematischen Grundlagen dieses ersten Ansatzes. Sie wiesen nach, dass auch dieser Ansatz neben der Additivität von Fitnesseffekten noch mehrere restriktive Annahmen voraussetzt und damit nur auf eine begrenzte Teilmenge von Evolutionsprozessen anwendbar ist. Daraufhin unterzeichneten über 100 Autoren die Erklärung, die Gesamtfitness sei genauso allgemein gültig wie die genetische Theorie der natürlichen Selektion selbst. Wie erklärt sich dieser offene Widerspruch? Dafür müssen wir erkennen, dass die obige Aussage auf einem zweiten, alternativen Ansatz beruht, der das Problem der Additivität im Rückblick betrachtet. Bei diesem Ansatz muss das Ergebnis der natürlichen Selektion bereits zu Beginn bekannt oder ausdrücklich spezifiziert sein; Ziel der Übung ist es dann, die additiven Kosten- und Nutzelemente zu bestimmen, die zu diesem Ergebnis geführt haben - und zwar

unabhängig davon, ob sie den tatsächlichen biologischen Interaktionen entsprechen. Kosten ( $C$ , *cost*) und Nutzen ( $B$ , *benefit*) werden mit Hilfe der linearen Regression bestimmt. Die Veränderung der Genfrequenz wird dann in der Formel  $BR-C$  gefasst, wobei  $R$  den Verwandtschaftsgrad beziffert. Diese Regressionsmethode führte Hamilton in einer Folgeveröffentlichung nach seinem ersten Aufsatz zur Gesamtfittesstheorie ein; nach und nach wurde sie dann zu einem Rezept verfeinert, nach dem sich Änderungen der Allelfrequenz in Form von Hamiltons Regel umschreiben ließen. Auf der Regressionsmethode fußen viele Behauptungen über Stärke und Allgemeingültigkeit der Gesamtfittesstheorie. So wird zum Beispiel häufig behauptet, dank der Regressionsmethode könne die Gesamtfittes auf die sonst erforderliche Additivität verzichten. Oder die Regressionsmethode erlaube die Vorhersage, in welche Richtung die natürliche Selektion sich entwickle, wodurch sich jede Frequenzveränderung als Folge sozialer Interaktionen zwischen verwandten Partnern quantitativ beziffern lasse.

Wir bewerten diese Behauptungen, indem wir untersuchen, ob und in welchem Umfang die Regressionsmethode Aussagen über eine gegebene evolutionäre Veränderung treffen kann. Wir zeigen, dass die Behauptungen zur Prognose- und Erklärungsmacht der Methode falsch sind; die Behauptung ihrer Allgemeingültigkeit bedeutet nichts, was sich evaluieren ließe. Diese Ergebnisse stellen die Vorstellung in Frage, die Gesamtfittes stelle ein universelles Gestaltungsprinzip der Evolution dar - in Wirklichkeit gibt es gar kein solches Gestaltungsprinzip.

**Die Regressionsmethode liefert keine Vorhersagen**

Bewerten wir also nun die verschiedenen Behauptungen über die Regressionsmethode, und beginnen wir mit der Behauptung, sie prognostiziere die Selektionsrichtung. Diese Behauptung kann gar nicht stimmen, weil die Veränderung der Allelfrequenz über den Untersuchungszeitraum hinweg ja zu Beginn festgelegt wird. Die »Vorhersage« rekapituliert lediglich, was bereits bekannt war, womit die Formel  $BR - C$  mit dem vorbestimmten Ergebnis übereinstimmt.

Genauso wenig prognostiziert die Regressionsmethode, was in anderen Zeiträumen oder unter anderen Bedingungen ablaufen wird. Sobald am untersuchten Szenario oder am Zeitintervall eine Veränderung vorgenommen wird, müssen die Ausgangsdaten neu spezifiziert und die ganze Methode neu angewandt werden, was dann neue, unabhängige Ergebnisse liefert.

Dass die Methode keine Vorhersagen liefern kann, ist keine Überraschung. Es ist faktisch unmöglich, das Ergebnis eines Prozesses zu prognostizieren, ohne zuvor Hypothesen über sein Verhalten aufzustellen. Ganz ohne Annahme eines Modells kann man nur eines: bereits existierende Daten in anderer Form wiedergeben.

Im Experiment trat diese fehlende Prognosefähigkeit zu Tage. Eine neuere Studie wendete die Regressionsmethode auf die kooperative Produktion eines Wirkstoffs an, der für die Antibiotika-Resistenz von *Escheria coli* benötigt wird. Die Autoren kommen zu dem Schluss: »Selbst wenn für ein bestimmtes System von Produzenten und Nichtproduzenten die Werte von  $B$ ,  $C$  und  $R$  gemessen wurden, lässt sich nicht vorhersagen, wie sich eine Veränderung entweder der Populationsstruktur oder der Biochemie der Individuen auswirken wird.«

## **Die Regressionsmethode liefert keine kausalen Begründungen**

Bewerten wir nun die Erklärungsmacht der Regressionsmethode. Die aktuelle Literatur ist sich in diesem Punkt eher uneinig. Einige Autoren behaupten, die Methode liefere kausale Begründungen für Änderungen von Allelfrequenzen, andere dagegen treffen die schon eingeschränktere Aussage, sie liefere nützliche konzeptuelle Hilfestellung. Zudem werden die Werte, die über die Regressionsmethode errechnet werden, üblicherweise in Form von sozialen Verhaltensweisen beschrieben, etwa von Altruismus oder Boshaftigkeit; damit erhalten diese Werte einen »kausalen Glanz«, selbst wenn eine Kausalität direkt nicht postuliert wird.

Die Behauptung, die Regressionsmethode identifiziere die Ursachen für die Änderung von Allelfrequenzen, kann gar nicht korrekt sein, weil die Regressionsmethode lediglich Korrelationen aufdecken kann - und Korrelation heißt nicht Kausalität. Da die Regressionsmethode zudem additive soziale Fitnesseffekte ausfindig machen soll, die zu den vorliegenden Daten passen, ist zu erwarten, dass es zu falschen Ergebnissen kommt, wenn soziale Interaktionen nicht additiv sind oder wenn unterschiedliche Fitness durch andere Faktoren verursacht wird. Auf dieser Grundlage präsentieren wir hier drei hypothetische Szenarien, in denen die Regressionsmethode die Ursachen für die Änderung von Allelfrequenzen falsch darstellt.

In unserem ersten hypothetischen Szenario veranlasst ein »Follower«-Merkmal seine Träger, Individuen mit hoher Fitness zu identifizieren und mit ihnen zu interagieren. Nehmen wir an, diese Interaktionen wirken sich auf die Fitness nicht aus. Dieses Auswahlverhalten führt aber trotzdem dazu, dass zwischen Fitness und der Tatsache, einen »Follower« als

Interaktionspartner zu haben, eine positive Relation besteht; die Regressionsmethode ergibt damit  $B > 0$ . Der vorgeschlagenen Interpretation zufolge müssten »Follower« als kooperativ verstanden werden, die die Fitness ihrer Partner erhöhen. Dabei wird aber natürlich die Kausalität umgedreht: Denn die hohe Fitness führt zur Interaktion und nicht umgekehrt. In verschiedenen Variationen kommt es zu diesem »Follower«-Verhalten vielleicht in vielen biologischen Systemen. So könnte ein Vogel das Nest eines Partners mit hoher Fitness aufsuchen, weil er dieses Nest irgendwann erben möchte. Ebenso könnte eine soziale Wespe dann eher im Nest ihrer Eltern verbleiben, wenn diese eine hohe Fitness aufweisen, und das auch mit dem Ziel, das Nest am Ende zu erben. Wendet man aber auf solche Situationen die Regressionsmethode an, so lässt sich schlicht egoistisches Verhalten leicht als Kooperation fehlinterpretieren.

Als zweites Beispiel betrachten wir das Merkmal »Eifersucht«. Eifersüchtige Individuen identifizieren Partner mit hoher Fitness und greifen sie an, um ihre Fitness zu senken. Wir gehen davon aus, dass diese Angriffe für den Angreifer Kosten verursachen, aber nur wenig erfolgreich sind, so dass die angegriffenen Individuen nach den Angriffen weiterhin eine überdurchschnittlich hohe Fitness aufweisen. Die Regressionsmethode führt zu dem Ergebnis  $B, C > 0$  und suggeriert damit, dass eifersüchtige Individuen unter eigenen Kosten kooperieren. Das ist wieder eine Fehlinterpretation: Die Angriffe schaden beiden, und die positive Fitness-Korrelation beruht auf der Auswahl der Interaktionspartner und dem Misserfolg der Angriffe.

Als drittes Beispiel dient uns das Merkmal »Amme«. Eine Amme identifiziert Individuen mit niedriger Fitness und versucht

unter hohen Kosten, deren Fitness zu erhöhen. Nun gehen wir aber davon aus, dass diese Hilfe nur wenig bewirkt und die unterstützten Individuen weiterhin eine unterdurchschnittliche Fitness aufweisen. Die Regressionsmethode stellt fest  $B < 0$ ,  $C > 0$  und interpretiert damit fälschlicherweise die weiterhin niedrige Fitness als Folge aus den hohen Kosten einer Sabotage durch die Ammen.

### **»Hypothesefreie« Ansätze**

Zum Abschluss kommen wir zu der Behauptung, die Gesamtfisnesstheorie sei »genauso allgemein gültig wie die genetische Theorie der natürlichen Selektion selbst«.

Begründet wird sie so: Daraus, dass die Regressionsmethode sich auf jede beliebige Veränderung der Allelfrequenz anwenden lässt (unabhängig von den tatsächlichen Ursachen dieser Änderung), folgt, dass sich jeder Einzelfall von natürlicher Selektion durch die Gesamtfisnesstheorie erklären lässt.

Doch wie wir gerade gesehen haben, liefert die Regressionsmethode lediglich »als-ob-Geschichten«, die für das gegebene Szenario weder Prognosen noch Ursachen bereitstellt.

Natürlich kann es vorkommen, dass die Regressionsmethode korrekte kausale Begründungen liefert, oder dass die Ergebnisse für ein Szenario mehr oder weniger auch auf ein anderes übertragbar sind. Allerdings liefert die Regressionsmethode keine Kriterien, nach denen diese Fälle identifiziert werden könnten – wer nämlich solche Kriterien formulieren wollte, müsste zusätzliche Hypothesen über die zugrunde liegenden Prozesse aufstellen. Ohne solche Hypothesen aber beantworten die Ergebnisse der Regressionsmethode nicht eine wissenschaftliche Fragestellung zu der untersuchten

Situation. Die Behauptung, die Theorie sei allgemeingültig, ist daher sinnlos.

Dieser fehlende Nutzen ist nicht auf ein technisches Versehen zurückzuführen. Er ergibt sich vielmehr aus dem Versuch, Hamiltons Regel auf jeden Schritt der natürlichen Selektion anwenden zu wollen. Natürlich ist dieser Impuls verständlich, weil Hamiltons ursprüngliche Formel so verlockend klingt. Doch ihre Stärke gewinnt eine Theorie aus ihren Hypothesen, und eine Theorie ohne Hypothesen kann nichts vorhersagen oder erklären. So erklärte auch Wittgenstein in seinem *Tractatus Logico-Philosophicus*, dass eine Aussage, die in allen Situationen wahr ist, keinerlei spezifische Information über eine bestimmte Situation enthält.

### **Es gibt kein universelles Gestaltungsprinzip**

Der Begriff der Gesamtfitness taucht auf, wenn man versucht, die Evolution von Sozialverhalten auf der Ebene des Individuums zu begründen. So soll die Gesamtfisnesstheorie zum Beispiel die Existenz steriler Ameisen-Arbeiterinnen mit dem Verhalten der Arbeiterinnen selbst erklären. Die vorgelegte Erklärung lautet dann, dass die Arbeiterinnen ihre Gesamtfitness insgesamt steigern, indem sie der Königin helfen, statt selbst Nachkommen zu produzieren.

Die Behauptung, die Evolution steigere die Gesamtfitness, wird als universelles Gestaltungsprinzip der Evolution interpretiert. Diese Behauptung beruht auf dem Argument Hamiltons, die Evolution steigere die geringe Gesamtfitness einer Population, und auf dem eigenen, davon getrennten Argument von Alan Grafen, durch Evolution entstandene Organismen verhielten sich so, dass sie ihre Gesamtfitness

möglichst steigern. Diese beiden Behauptungen fußen auf restriktiven Annahmen, unter anderem der Additivität von Fitnesseffekten. Da im Experiment nachgewiesen wurde, dass Fitnesseffekte in realen biologischen Populationen nicht-additiv sind, kann man nicht davon ausgehen, dass diese Ergebnisse allgemeingültig sind. Zudem haben sowohl Theorie als auch Experiment ergeben, dass frequenzabhängige Selektion zu komplexen dynamischen Phänomenen wie multiplen Gleichgewichtszuständen führen kann, zu Grenzzyklen und Seltsamen Attraktoren; und das schließt die Möglichkeit aus, dass es überhaupt einen generell zu maximierenden Faktor gibt. Damit führt Evolution grundsätzlich nicht zur Maximierung der Gesamtfitness oder irgendeiner anderen Quantität.

#### **<U4>Vernünftige Ansätze zur Evolutionstheorie**

Glücklicherweise sind generell zu maximierende Faktoren oder Gestaltungsprinzipien gar nicht nötig, wenn man die Evolution des Sozialverhaltens verstehen will. Stützen wir uns lieber auf einen durchwegs genetischen Ansatz: Betrachten wir einmal Mutationen, die Verhalten verändern. Unter welchen Bedingungen werden diese Mutationen von der natürlichen Selektion begünstigt (oder nicht begünstigt)? Das Selektionsziel ist nicht das Individuum, sondern das Allel oder der Genabschnitt, der das Verhalten betrifft.

Um diese Fragen theoretisch zu untersuchen, benötigt man ein Modell. Die Annahmen können hochspezifisch sein, also nur für ganz bestimmte biologische Situationen gelten, oder allgemein sein, also anwendbar auf ein weites Spektrum von Szenarien. Seit Kurzem gibt es Modelle auf der Grundlage allgemeiner (aber präziser) Hypothesen, mit deren Hilfe sich wirkungsvolle

Untersuchungen über die räumlich (also durch Gruppen oder physiologisch) strukturierte Evolution von Populationen durchführen lassen, über die Evolution kontinuierlicher Merkmale und sogar über die Gesamtfisnesstheorie (in Fällen, in denen Fitnesseffekte tatsächlich additiv und auch andere Anforderungen erfüllt sind). Obwohl diese Modelle sich so einsetzen lassen, dass allgemeine Ergebnisse erzielt werden, ist keines von ihnen universell oder hypothesefrei. Vielmehr nutzen sie ihre Hypothesen für gut definierte, überprüfbare Prognosen über die Systeme, für die sie gelten.

## **Diskussion**

Die Gesamtfisnesstheorie versucht, ein universelles Gestaltungsprinzip für die Evolution aufzustellen, das am Individuum ansetzt. Das Ergebnis ist ein empirisch nicht belegbarer Wert, der nicht allgemein existiert (wenn Additivität verlangt wird) oder keinen Prognose- oder Begründungswert hat (wenn die Regressionsmethode zum Einsatz kommt). Betrachten wir die Sache dagegen aus genetischer Perspektive und fragen, ob die natürliche Selektion Allele fördert oder unterdrückt, die soziales Verhalten verändern, so brauchen wir gar keine Gesamtfisness.

Die Dominanz der Gesamtfisnesstheorie hat Fortschritte auf diesem Gebiet über Jahrzehnte hinweg gehemmt. Sie hat vernünftige Kritik und alternative Ansätze systematisch unterdrückt. Besonders der Versuch, der Notwendigkeit der Additivität zu entgehen, indem man die Regressionsmethode nutzte, führte zu logischer Verwirrung und falschen Behauptungen von Allgemeingültigkeit. Vernünftige Berechnungen zur Gesamtfisness, die die Additivität berücksichtigen, sind

in der Tat eine alternative Methode, um in wenigen beschränkten Fällen Fitnesseffekte zu erklären; doch diese Methode ist nie notwendig und häufig unnötig kompliziert. Kein Problem der Evolutionsbiologie macht eine Analyse auf Grundlage der Gesamtfitness notwendig.

Da die Einschränkungen der Gesamtfitness nun einmal erkannt sind, kann die Soziobiologie jetzt frei nach vorne blicken. Wir unterstützen die Entwicklung realistischer Modelle, die auf solidem naturwissenschaftlichen Verständnis gründen. Mit Hilfe der Populationsgenetik, der evolutionären Spieltheorie und neuen, noch zu entwickelnden analytischen Prozessen wird die soziobiologische Theorie gestärkt und belastbar in die Zukunft gehen.