

Astronomische Verschwendung: Die Opportunitätskosten verzögerten technologischen Fortschritts

Nick Bostrom
Universität Oxford

[Ursprünglich publiziert in: *Utilitas*, Vol. 15, No. 3 (2003): pp. 308-314]

Aus dem Englischen von Stephan Dalügge.

www.nickbostrom.com

ABSTRACT. Mit sehr fortschrittlicher Technologie könnte eine sehr große Population von Personen, die glückliche Leben leben, in der erreichbaren Region des Universums aufrechterhalten werden. Daraus ergeben sich entsprechende Opportunitätskosten für jedes Jahr, um das sich die Entwicklung solcher Technologien und die folglich Kolonialisierung des Universums verzögert: Ein potenzielles Gut, nämlich das lebenswertere Leben, wird nicht realisiert. Unter plausiblen Annahmen sind diese Kosten extrem groß. Jedoch lautet die Lehre für den Standardutilitaristen nicht, dass wir die Geschwindigkeit des technologischen Fortschritts maximieren sollten, sondern dass wir seine Sicherheit maximieren sollten, d. h. die Wahrscheinlichkeit, dass die Kolonisierung des Weltalls tatsächlich stattfinden wird. Dieses Ziel hat eine so hohe Utilität, dass Standardutilitaristen *all* ihre Energie darauf verwenden sollten. Utilitaristen der „personenbezogenen“ Sorte sollten eine modifizierte Version dieser Schlussfolgerung akzeptieren. Manche andere ethische Sichtweisen, welche utilitaristische Erwägungen mit anderen Kriterien kombinieren, werden zu einem ähnlichen Fazit kommen.

I. DIE RATE DES VERLUSTES POTENZIELLER LEBEN

Während ich diese Worte schreibe, erleuchten und erhitzen Sonnen leere Räume und ungenutzte Energie wird schwarze Löcher heruntergespült. Das fantastische kosmische

Geschenk von Negentropie wird irreversibel und in einem kosmischen Ausmaß zu Entropie degradiert. Dabei handelt es sich um Ressourcen, die eine fortschrittliche Zivilisation hätte nutzen können, um Wertestrukturen zu errichten, empfindungsfähige Wesen etwa, die lebenswerte Leben führen.

Der Verstand kann die Rate dieses Verlustes kaum erfassen. Eine kürzlich veröffentlichte Studie unternahm lose theoretische Überlegungen zur Rate der Zunahme von Entropie und schätzte, dass der Verlust menschlicher Leben in unserem eigenen Supercluster mindestens $\sim 10^{46}$ Menschenleben pro Jahrhundert verzögerter Kolonisierung beträgt.¹ Eine Annahme dieser Schätzung ist, dass alle verloren gegangene Entropie für produktive Zwecke hätte genutzt werden können, obschon keine derzeit bekannten Technologien ansatzweise dazu im Stande wären. Da wir eine Untergrenze abschätzen wollen, ist diese unkonventionelle Annahme unerwünscht.

Wir können jedoch zu einer Untergrenze gelangen, indem wir einfach die Sterne in unserem Supercluster zählen. Diese Zahl kombinieren wir dann mit der Rechenleistung, die mithilfe der Ressourcen eines jeden Sterns und unter Verwendung von Technologien, deren Realisierbarkeit bereits gut begründet worden ist, generiert werden könnte. Wir können dieses Produkt dann durch unseren Schätzwert für die Rechenleistung, die es für die Simulation eines Menschenlebens braucht, teilen.

Nehmen wir als grobe Schätzung an, dass das Virgo-Supercluster 10^{13} Sterne umfasst. Die geschätzte extrahierbare Rechenleistung eines Sterns mit einer assoziierten Rechenstruktur in Planetengröße unter Nutzung fortschrittlicher molekularer Nanotechnologie² ist 10^{42} Operationen pro Sekunde.³ Die Rechenleistung des menschlichen

¹ M. Cirkovic, 'Cosmological Forecast and its Practical Significance' („Eine kosmologische Prognose und ihre praktische Relevanz“), Journal of Evolution and Technology, xii (2002), <http://www.jetpress.org/volume12/CosmologicalForecast.pdf>.

² K. E. Drexler, Nanosystems: Molecular Machinery, Manufacturing, and Computation („Molekulare Maschinen, Produktion und Rechenleistung“), New York, John Wiley & Sons, Inc., 1992.

³ R. J. Bradbury, 'Matrioshka Brains' („Matrjoschka-Gehirne“), Manuskript, 2002, <http://www.aeiveos.com/~bradbury/MatrioshkaBrains/MatrioshkaBrains.html>.

Gehirns wird typischerweise auf maximal 10^{17} Operationen pro Sekunde geschätzt.⁴ Wenig mehr scheint nötig, um die relevanten Teile der Umwelt so detailliert zu simulieren, dass simulierte Wesen Erfahrungen haben, die von typischen Erfahrungen heute lebender Menschen nicht zu unterscheiden sind.⁵ Diesen Schätzungen zufolge verlieren wir für jedes Jahrhundert, um das sich die Kolonisierung unseres Superclusters verzögert, ungefähr das Potenzial für 10^{38} menschliche Leben; das entspricht in etwa 10^{29} potenziellen menschlichen Leben pro Sekunde.

Obschon diese Schätzung insoweit konservativ ist, als dass sie nur Berechnungsmechanismen annimmt, deren Implementierung in der wissenschaftlichen Literatur bereits mindestens skizziert worden ist, ist es nützlich, eine noch konservativere Schätzung aufzustellen, die nicht von nicht-biologischen Instanziierungen der potenziellen Personen ausgeht. Nehmen wir an, dass circa 10^{10} biologische Menschen um einen durchschnittlichen Stern herum leben können. Dann könnte das Virgo-Supercluster 10^{23} biologische Menschen beherbergen. Das entspricht einem Verlust von 10^{13} potenziellen Menschenleben pro Sekunde verzögerter Kolonialisierung.⁶

II. DIE OPPORTUNITÄTSKOSTEN VERZÖGERTER KOLONIALISIERUNG

Aus utilitaristischer Sicht konstituiert dieser ungemeine Verlust potenzieller menschlicher Leben einen entsprechenden Verlust potenziellen Wertes. Ich nehme hier an, dass menschliche Leben, die hätten kreiert werden können, lebenswert gewesen wären. Da für gewöhnlich

⁴ N. Bostrom, 'How Long Before Superintelligence?' („Wie lange bis zur Superintelligenz?“), *International Journal of Futures Studies* ii (1998); R. Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence* („Die Intelligenz der Evolution“), New York, Viking, 1999. Die niedrigere Schätzung ist zu finden in H. Moravec, *Robot: Mere Machine to Transcendent Mind* („Roboter: Von bloßer Maschine zum transzendenten Bewusstsein“), Oxford, 1999.

⁵ N. Bostrom, 'Are You Living in a Simulation?' („Leben wir in einer Simulation?“), *Philosophical Quarterly*, liii (211). See also <http://www.simulation-argument.com>.

⁶ Das Virgo-Supercluster beinhaltet nur einen kleinen Teil der kolonialisierbaren Ressourcen im Universum, aber es ist groß genug, damit die vorgebrachte Argumentation funktioniert. Je größer die betrachtete Region, desto unsicherer müssen wir sein, dass signifikante Teile davon noch nicht von einer Zivilisation nicht-terrestrischen Ursprungs kolonialisiert worden sein werden, bevor wir dorthin gelangen können.

angenommen wird, dass selbst derzeitige Menschenleben typischerweise wertvoll sind, ist dies eine schwache Annahme. Jede Zivilisation fortschrittlich genug, das lokale Supercluster zu kolonisieren, hätte vermutlich ebenso die Fähigkeit, die wenigstens minimal nötigen günstigen Bedingungen zu schaffen, damit künftige Leben lebenswert sind.

Der Effekt auf den absoluten moralischen Wert scheint daher für Handlungen, die technologischen Fortschritt beschleunigen, größer als für praktisch jede andere mögliche Handlung. Ein Vorantreiben des technologischen Fortschritts (oder ein Vorantreiben seiner ermöglichenden Faktoren, etwa ökonomischer Produktivität) sei es auch noch so winzig, als dass es die Kolonialisierung des lokalen Superclusters um nur *eine Sekunde* beschleunigt, würde dem Herbeiführen von 10^{29} menschlichen Leben (oder mindestens 10^{13} menschlichen Leben, wenn wir bei der konservativsten Schätzung blieben), die sonst nicht existiert hätten, entsprechen. Wenige andere philanthropische Bereiche können auf diesen Grad utilitaristischen Ertrags hoffen.

Utilitaristen sind nicht die einzigen, die diese astronomische Verschwendung entschieden ablehnen sollten. Es gibt viele Sichtweisen darüber, was Wert hat, die mit der Bewertung übereinstimmen, dass die derzeitige Rate von Verschwendung einen enormen Verlust potenziellen Wertes darstellt. Beispielsweise können wir eine umfassendere Vorstellung menschlichen Wohlergehens zugrunde legen als die der Utilitaristen (seien sie nun von der hedonistischen, erfahrungsbasierten oder Präferenz erfüllungs-orientierten Sorte): etwa eine Vorstellung, die Wert in menschlichem Wohlergehen, bedeutsamen Beziehungen, einem noblen Charakter, individueller Entfaltung, ästhetischer Wertschätzung usw. sieht. Solange die Evaluierungsfunktion aggregativ ist (das Wohlergehen einer Person also nicht weniger zählt, weil viele andere Menschen existieren, die auch glückliche Leben führen) und nicht relativ zu einem bestimmten Zeitpunkt betrachtet wird (also keine Zeitpräferenz besteht), wird die gleiche Schlussfolgerung zutreffen.

Diese Bedingungen können weiter gelockert werden. Selbst wenn die Utilitätsfunktion nicht komplett aggregativ ist (vielleicht weil eine Komponente des Guten Diversität ist und die marginale Produktionsrate davon mit zunehmender Bevölkerungsgröße abnehmen könnte), kann dennoch eine ähnliche Schlussfolgerung zutreffen, insofern zumindest eine signifikante Komponente des Guten hinreichend aggregativ ist. Ebenso könnte ein gewisses Maß an Zeitabzinsung künftiger Güter berücksichtigt werden, ohne dass sich die Schlussfolgerung ändert.⁷

III. DAS HAUPTZIEL VON UTILITARISTEN SOLLTE DIE REDUKTION EXISTENZIELLER RISIKEN SEIN

Angeichts der obigen Diskussion mag es scheinen, als müsste eine Utilitaristin ihre Anstrengungen auf die Beschleunigung technologischen Fortschritts richten. Selbst die Rendite eines nur sehr kleinen Erfolgs in diesem Unterfangen ist so enorm, dass sie die Bemühungen fast jeder anderen Aktivität in den Schatten stellt. Es scheint, als hätten wir einen utilitaristischen Beweis für die höchste Dringlichkeit technologischer Entwicklung.

Die wahre Pointe liegt jedoch woanders. Wenn wir uns mit (etwas wie) der Maximierung der erwartbaren lebenswerten Leben, die wir kreieren werden, befassen, dann müssen wir zusätzlich zu den Opportunitätskosten verzögerter Kolonialisierung das Risiko des Versagens, das Weltall überhaupt zu kolonialisieren, in Betracht ziehen. Wir könnten einem existenziellen Risiko zum Opfer fallen – einem, bei dem ein Scheitern die Auslöschung von von der Erde stammendem intelligenten Leben oder die drastische Einschränkung seines Potenzials bedeutet.⁸ Da die Lebensspanne von Galaxien in Milliarden von Jahren gemessen

⁷ Utilitaristen halten die Zeitabzinsung moralischer Güter im Allgemeinen für unangemessen (siehe etwa R. B. Brandt, *Morality, Utilitarianism, and Rights* („Moralität, Utilitarismus und Rechte“), Cambridge, 1992, pp. 23 f.). Jedoch ist unklar, ob Utilitaristen es im Angesicht der Möglichkeit, dass unsere Handlungen Konsequenzen für eine unendliche Anzahl von Personen haben könnten, vermeiden können, Abstriche bei diesem Prinzip zu machen.

⁸ N. Bostrom, ‘Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards’ („Existenzielle Risiken: Analyse von Szenarien des menschlichen Aussterbens und verwandter

wird, die Verzögerungen, die wir realistischerweise beeinflussen können, jedoch in Jahren oder Dekaden, trumpft die Berücksichtigung des Risikos die Berücksichtigung der Opportunitätskosten. Zum Beispiel wäre die Reduktion des Risikos um einen einzigen Prozentpunkt (aus utilitaristischer Sicht) eine Verzögerung um 10 Millionen Jahre wert.

Wenn unsere Handlungen also nur den geringsten Effekt auf *die Wahrscheinlichkeit* schlussendlicher Kolonialisierung haben, so wird das ihren Effekt auf das *Wann* der Kolonialisierung übertreffen. Für Standardutilitaristen sollten Prioritäten eins, zwei, drei und vier daher die Reduktion existenzieller Risiken sein. Der utilitaristische Imperativ „Maximiere die erwartbare aggregierte Utilität“ kann vereinfacht werden zur Maxime „Minimiere existenzielles Risiko!“.

IV. IMPLIKATIONEN FÜR AGGREGATIVE, PERSONENBEZOGENE SICHTWEISEN

Das obige Argument setzt voraus, dass unser Anliegen das Maximieren der Gesamtmenge von Wohlergehen ist. Nehmen wir an, dass wir stattdessen eine „personenbezogene“ Version des Utilitarismus anwenden, nach der unsere Verpflichtungen primär derzeit existierende Personen und Personen, die gewiss existieren werden, betreffen.⁹ Laut solch einer personenbezogenen Sicht wäre das Aussterben der Menschheit nur insofern schlecht, als dass es vergangene oder derzeitige Leben schlechter macht, nicht aber, weil es einen Verlust potenzieller lebenswerter Leben darstellt. Was sollte jemand, der diese Doktrin akzeptiert, tun? Sollte er Schnelligkeit, Sicherheit oder etwas anderes drittes betonen?

Um eine Antwort darauf zu treffen, müssen wir einige weitere Fragestellungen betrachten. Nehmen wir an, dass die Wahrscheinlichkeit, dass irgendeine existierende Person

Gefahren“), *Journal of Evolution and Technology*, ix (2002),
<http://www.jetpress.org/volume9/risks.html>.

⁹ Diese Formulierung der Position ist nicht unbedingt die bestmögliche, aber sie ist simpel und wird für die Zwecke dieses Artikels ausreichen.

lange genug überleben wird, um einen signifikanten Teil der Ressourcen des erreichbaren Universums zu nutzen, vernachlässigbar ist. Ressourcen, welche, wie im ersten Abschnitt dieses Artikels beschrieben, nach und nach degradieren. Unter dieser Annahme ist der Grund zur Minimierung existenzieller Risiken, dass das plötzliche Aussterben der Menschheit das Leben eines jeden der (circa 6 Milliarden) existierenden Menschen um im Durchschnitt etwa 40 Jahre verkürzen würde.¹⁰ Obschon dies sicherlich ein großes Desaster wäre, ist es zumindest ungefähr in der gleichen Größenordnung wie andere laufende Tragödien wie beispielsweise weltweite Armut, Hunger und Krankheiten. Unter dieser Annahme sollte ein personenbezogener Utilitarist die Reduktion existenzieller Risiken als ein sehr wichtiges, aber nicht alles andere dominierendes Anliegen betrachten. Es gäbe in diesem Fall keine einfache Antwort darauf, was er tun sollte. Wofür er sich einsetzen sollte, hinge von detaillierten Berechnungen darüber ab, in welchem Bereich philanthropischer Aktivität er am besten in der Lage wäre, Einfluss zu haben.

Es lässt sich jedoch argumentieren, dass wir der Möglichkeit, dass einige derzeit existierende Personen ausreichend lange überleben werden, um von einer kosmischen Diaspora zu profitieren, eine nicht vernachlässigbare Wahrscheinlichkeit zuordnen sollten.

Eine sogenannte technologische „Singularität“ könnte innerhalb unserer natürlichen Lebensspannen passieren.¹¹ Ebenso könnte es einen Durchbruch in Sachen Lebensverlängerung geben, der vielleicht das Resultat maschinengestützter Nanotechnologie wäre, die uns unübertroffene Kontrolle über die biochemischen Prozesse in unseren Körpern gäbe und uns befähigen würde, den Alterungsprozess zu stoppen und umzukehren.¹² Viele führende Technologen und Futuristen schreiben dem Eintreten dieser Entwicklungen in den nächsten paar Dekaden eine recht hohe Wahrscheinlichkeit zu.¹³ Selbst wer gegenüber diesen

¹⁰ Oder was auch immer die Bevölkerung zum Zeitpunkt des Weltuntergangs sein mag.

¹¹ Siehe etwa V. Vinge, 'The Coming Technological Singularity' („Die kommende technologische Singularität“), Whole Earth Review, Winterausgabe (1993).

¹² R. A. Freitas Jr., Nanomedicine („Nanomedizin“), Vol. 1, Georgetown, Landes Bioscience, 1999.

¹³ Z. B. Moravec, Kurzweil, and Vinge op. cit.; E. Drexler, Engines of Creation („Motoren der Schöpfung“), New York, Anchor Books, 1986.

Prognosen skeptisch ist, sollte die schwache Erfolgsbilanz bei der Prognose technologischer Entwicklungen in Betracht ziehen. Aufgrund der wohlbekannten Unzuverlässigkeit vieler solcher Prognosen scheint es nicht gerechtfertigt, der eigenen Prognose gegenüber so zuversichtlich zu sein, dass die erforderlichen Durchbrüche nicht zu unserer Zeit passieren werden, als dass man der Hypothese, dass sie passieren werden, eine Wahrscheinlichkeit von weniger als, sagen wir, 1 % geben sollte.

Die erwartete Utilität einer Chance von 1 % auf ein astronomisch großes Gut könnte immer noch astronomisch sein. Aber wie gut wäre es für (eine substanzielle Teilmenge) derzeit lebender Menschen, Zugang zu astronomischen Mengen von Ressourcen zu haben? Die Antwort darauf ist nicht eindeutig. Einerseits könnte man sagen, dass in der heutigen Welt die marginale Utilität für materielle Ressourcen für ein Individuum rapide abnimmt, sobald seine Grundbedürfnisse gedeckt sind. Bill Gates scheint nicht ein dramatisch höheres Level an Wohlbefinden zu genießen als viele Personen mit bescheideneren Mitteln. Andererseits könnten fortschrittliche Technologien der Art, die sehr wahrscheinlich zu der Zeit der Kolonisierung des lokalen Superclusters eingesetzt würden, neue Wege zur Konversion von Ressourcen in Wohlbefinden bereitstellen. Insbesondere könnten materielle Ressourcen unsere mentalen Kapazitäten ungemein ausweiten und unsere subjektive Lebensspanne unbegrenzt lang erweitern. Es ist dabei keineswegs klar, dass die marginale Utilität einer verlängerten Gesundheitsspanne und erweiterter mentaler Kapazitäten ab einem bestimmten Level stark abnehmen müsste. Gibt es keine solche Abnahme in marginaler Utilität, so müssen wir schlussfolgern, dass die erwartete Utilität der erfolgreichen Kolonisierung unseres Superclusters für *derzeit lebende* Individuen¹⁴ astronomisch groß ist. Diese Schlussfolgerung

¹⁴ Nach der personenbezogenen Sicht ist die zeitliche Position einer Person nicht *per se* der relevante Faktor. Im Prinzip könnte der Einfluss auf das Wohlbefinden einer vergangenen oder künftigen Person ein geeignetes Ziel moralischer Betrachtungen sein. In der Praxis ist es jedoch so, dass die Effekte auf das Wohlergehen vergangener Personen wahrscheinlich eher klein oder sogar null sind (je nachdem, von welcher Konzeption von Wohlergehen man ausgeht) und die Effekte (eines jeden Aktes, der einen signifikanten Effekt auf die Zukunft hat) auf noch nicht existente Personen ist sehr wahrscheinlich der Art, dass im Wettbewerb stehende alternative Handlungen zu einem separaten Set an möglichen Personen, die existieren werden, führt. Das ist zum Beispiel das Resultat daraus, dass unterschiedliche

hält selbst bei recht niedrigen Wahrscheinlichkeiten dieser Möglichkeit Bestand. Es mag unwahrscheinlich sein, aber für einen Maximierer erwarteter Utilität kann der Nutzen eines Lebens von Milliarden Jahren subjektiver Länge selbst eine schmale Aussicht auf Erfolg mehr als kompensieren.

Welche Handlungsanweisung für einen personenbezogenen Utilitaristen folgt also aus diesen Annahmen? Natürlich ist das Vermeiden existenzieller Katastrophen wichtig. Und zwar nicht nur, weil eine existenzielle Katastrophe die natürliche Lebensspanne von etwa sechs Milliarden Menschen verkürzen würde, sondern auch – und unter unseren Annahmen ist das ein noch gewichtigerer Punkt – weil es die Chance, die existierende Personen auf die Ernte der enormen Früchte der schlussendlichen Kolonialisierung haben, erlöschen würde. Jedoch müsste, anders als der alles aggregierende Utilitarist, die personenbezogene Utilitaristin dieses Ziel gegen ein anderes ebenso wichtiges Desideratum abwägen, nämlich das Maximieren der Wahrscheinlichkeit, dass derzeit lebende Personen lange genug leben, um von der Kolonialisierung zu profitieren. Für die personenbezogene Utilitaristin reicht es nicht aus, dass die Menschheit für die Kolonialisierung überlebt; sondern es ist entscheidend, dass existierende Personen gerettet werden. Dies sollte sie dazu veranlassen, die Geschwindigkeit technologischen Fortschritts zu betonen, denn das schnelle Eintreten fortschrittlicher Technologien wäre sicher nötig, um existenten Personen zu helfen, so lange am Leben zu bleiben, bis die Früchte der Kolonialisierung geerntet werden können. Wenn das Ziel der Geschwindigkeit dem Ziel der globalen Sicherheit entgegensteht, sollte der alles aggregierende Utilitarist sich immer für das Maximieren der Sicherheit entscheiden. Die personenbezogene Utilitaristin müsste das Risiko des altersbedingten Sterbens von Personen abwägen gegen das Risiko, in einer Spezies-zerstörenden Katastrophe umzukommen. Andere

Spermien unterschiedliche Eizellen befruchten würden und daher gäbe es typischerweise keine künftige Person, deren Level an Wohlbefinden von heutigen Handlungen beeinflusst würde. Unser Handeln mag beeinflussen, welche künftigen Personen es geben wird, aber nach der personenbezogenen Sicht geben uns solche Resultate keine moralischen Gründe für unser Handeln. Eine Ausnahme sind Föten und befruchtete Eizellen, die noch nicht Personen sind, aber die womöglich in der Zukunft Personen sein werden und deren Leben unsere heutigen Taten beeinflussen können.

ethische Sichtweisen, die auch nicht-utilitaristische Elemente beinhalten, mögen oder mögen auch nicht zu einer dieser Schlussfolgerungen kommen, je nach dem, was hinzugefügt wird.¹⁵

¹⁵ Ich bin dankbar für Kommentare von John Broome, Milan Cirkovic, Roger Crisp, Robin Hanson und James Lenman und für die finanzielle Unterstützung eines British Academy Postdoctoral Award.