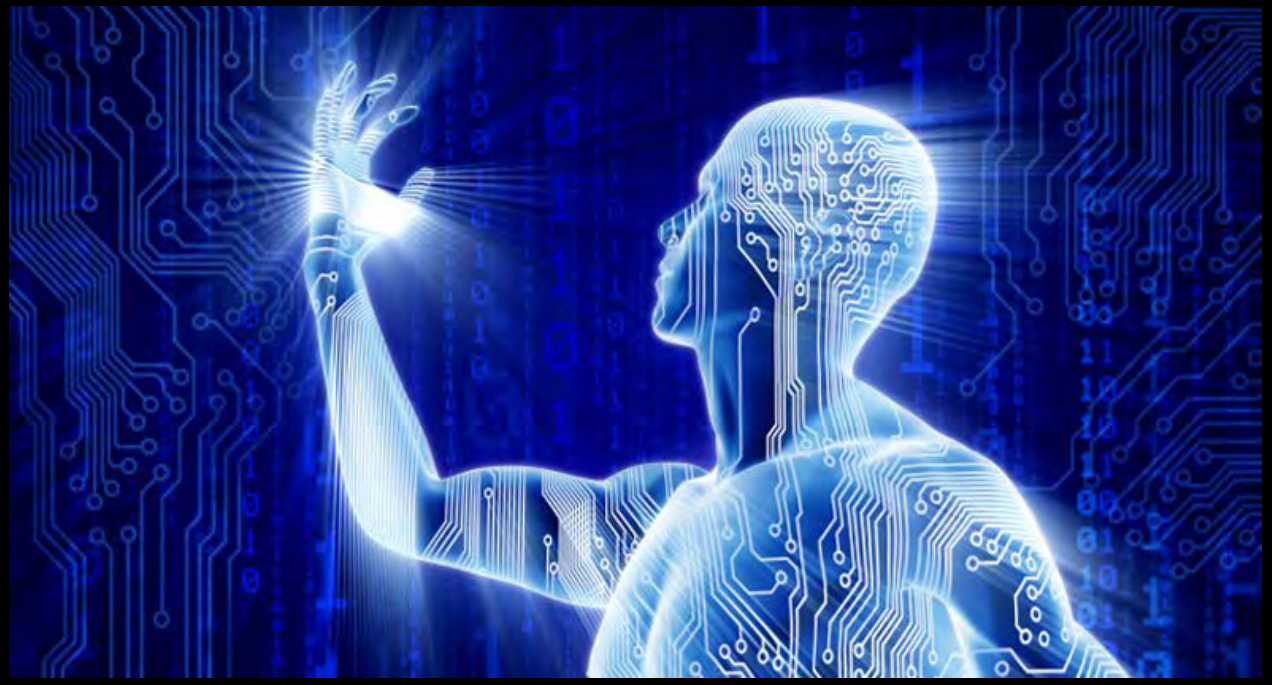


# Die Grenzen der technischen Optimierung

## Transhumanismus und Nick Bostrom

Ralf Stapelfeldt

MoMo Vortrag 18.9.2023

# Vorstellung: Ralf Stapelfeldt

- Wohnen in Berlin (Mitte)
- Studium Betriebs- und Volkswirtschaft in den 1990ern, 25 Jahre Bankenmanagement
- Seit 2021 Leiter der Zentralen Servicebereiche (Finanzen, HR, Facility Management und IT) am Pestalozzi-Fröbel-Haus in Berlin Schöneberg

- 2014 – 2019: Studium der Philosophie berufsbegleitend (FernUni Hagen)
- Seit 2020: Promotionsstudium, Thema: Transhumanismus und Nick Bostrom
- 2023: Bis Ende des Jahres ist die Abgabe der Doktorarbeit geplant

# Inhalt meines heutigen Vortrages

I. Kurze Vorstellung des Transhumanismus

II. Nick Bostrom

III. Bostroms Metaphysik: Computer-Funktionalismus

IV. Bostroms Ethik: Radikaler Utilitarismus

V. Kombination zu einer problematischen Theorie

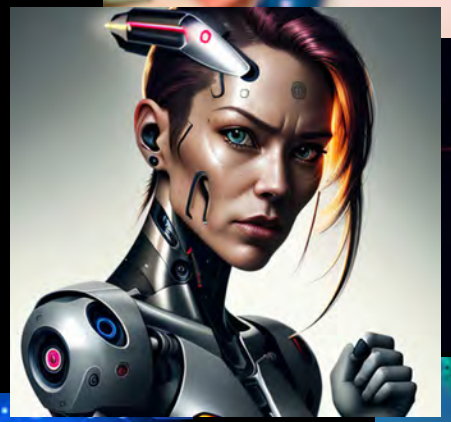
# I. Kurze Vorstellung des Transhumanismus

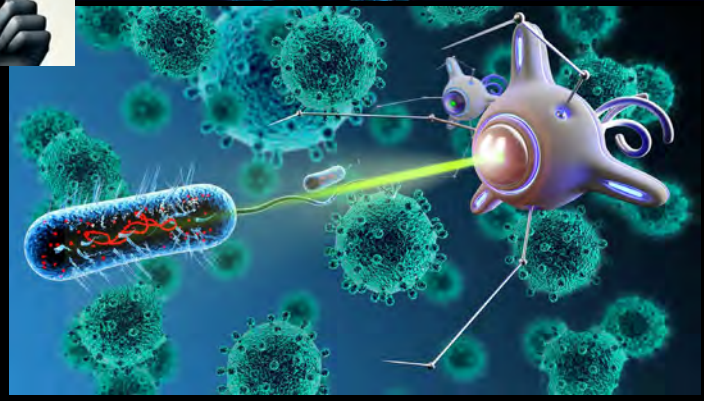
# Transhumanismus auf einen Blick

- Technische Optimierung des Menschen
- Überwindung der biologischen Mängel
- Ziele:
  - Körperliches und geistiges Enhancement
  - Überwindung von Alter, Krankheit und Tod
  - Transzendierung des Menschen durch den transhumanen hindurch zum posthumanen Zustand

# Optimierungsmethoden des Transhumanismus

- Kryonik
- Pharmakologisches Enhancement
- Gentechnik (Selektion und Veränderung)
- Cyborgisierung (Technik-Mensch-Hybride)
- Molekulare Nanotechnologie
- Mind Upload

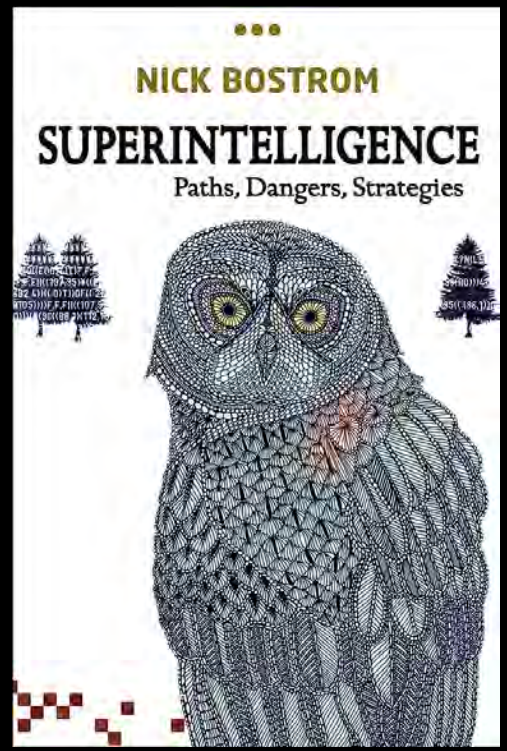




# II. Nick Bostrom

# Wer ist eigentlich Nick Bostrom?

- Bekennender Transhumanist
- Mitbegründer der World Transhumanist Association (heute humanity+)
- Professor für Philosophie an der University of Oxford
- Director des ‚Future of Humanity Institute‘
- Autor des Buchs „Superintelligence“
- Kunde beim Kryonik-Anbieter „Alcor“
- 2009 in den Top 100 der globalen Denker (Magazin ‚Foreign Policy‘, #73)
- Gefragter Berater, hoch einflussreich




# Kernaussage bei Bostrom:
# Wir müssen *alles* dafür tun, um die Posthumanität zu erreichen

- Höchster transhumanistischer Wert:
  „Explorieren des posthumanen Zustands“

- Posthumaner Zustand =
  virtuelles Leben im Computer;
  Genießen von unvorstellbaren Freuden als digitaler Posthumaner

Deshalb (instrumentelle Ziele):

- Optimierung des Menschen,
  insb. kognitives Enhancement über Genveränderung

- Existenzielle Risiken der Menschheit abwehren

# III. Nick Bostroms Metaphysik

# Die Metaphysik Bostrom entschlüsselt

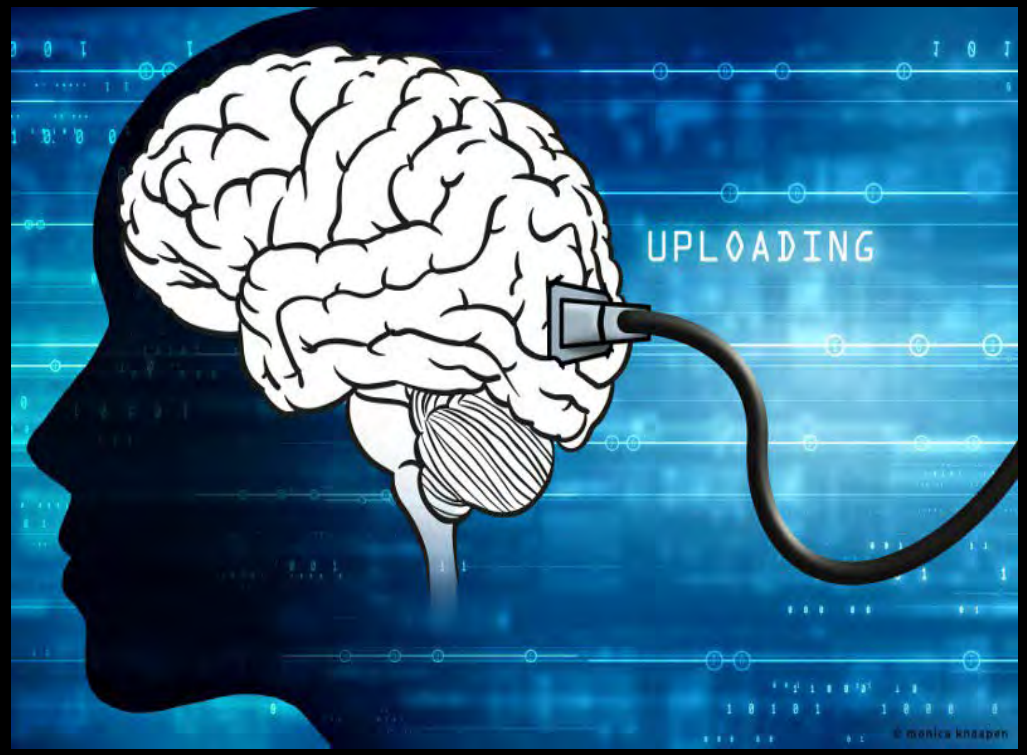
- *Starker Computer-Funktionalismus*
  Der Funktionalismus und der Computationalismus sind wahr.
- *Substratunabhängigkeit des Geistes*
  Mentale Zustände sind substratunabhängig.
  Silizium ist ein für Bewusstsein geeignetes Substrat.
- *Hypothese der technologischen Vollendung*
  Was technologisch prinzipiell möglich ist, wird irgendwann auch praktisch erreicht werden.
- *Mind Upload*
  Mind Uploading ist prinzipiell möglich, wird in einer posthumanen Zukunft ebenso wie die Kreation künstlicher Wesen mit Bewusstsein im Computer praktisch umsetzbar sein und realisiert.
- *Virtuelle posthumane Wesen*
  Wenn die Menschheit das posthumane Stadium erreicht, wird es zukünftig eine gewaltige Zahl an virtuellen Personen mit Bewusstsein geben, die in Computern simuliert werden oder als Originale, Kopien oder optimierte Versionen hochgeladener Personen existieren.

# Mind Upload als ultimatives Enhancement und Fundament der transhumanistischen Theorie Bostroms

- Mind Upload = Mind Transfer = Whole Brain Emulation
- Spekulatives Verfahren, um den Geist einer Person auf einen Computer zu übertragen.
- Der Computer emuliert das Gehirn und simuliert Körper und Umwelt in einem adäquaten Programm.
- Im Zuge der These des starken Computer-Funktionalismus wird angenommen, dass dies auch zu einem virtuellen Bewusstsein führt.
- Wenn der Scan des Gehirns exakt genug ist, soll die Person unter Wahrung ihrer Identität übertragen werden.

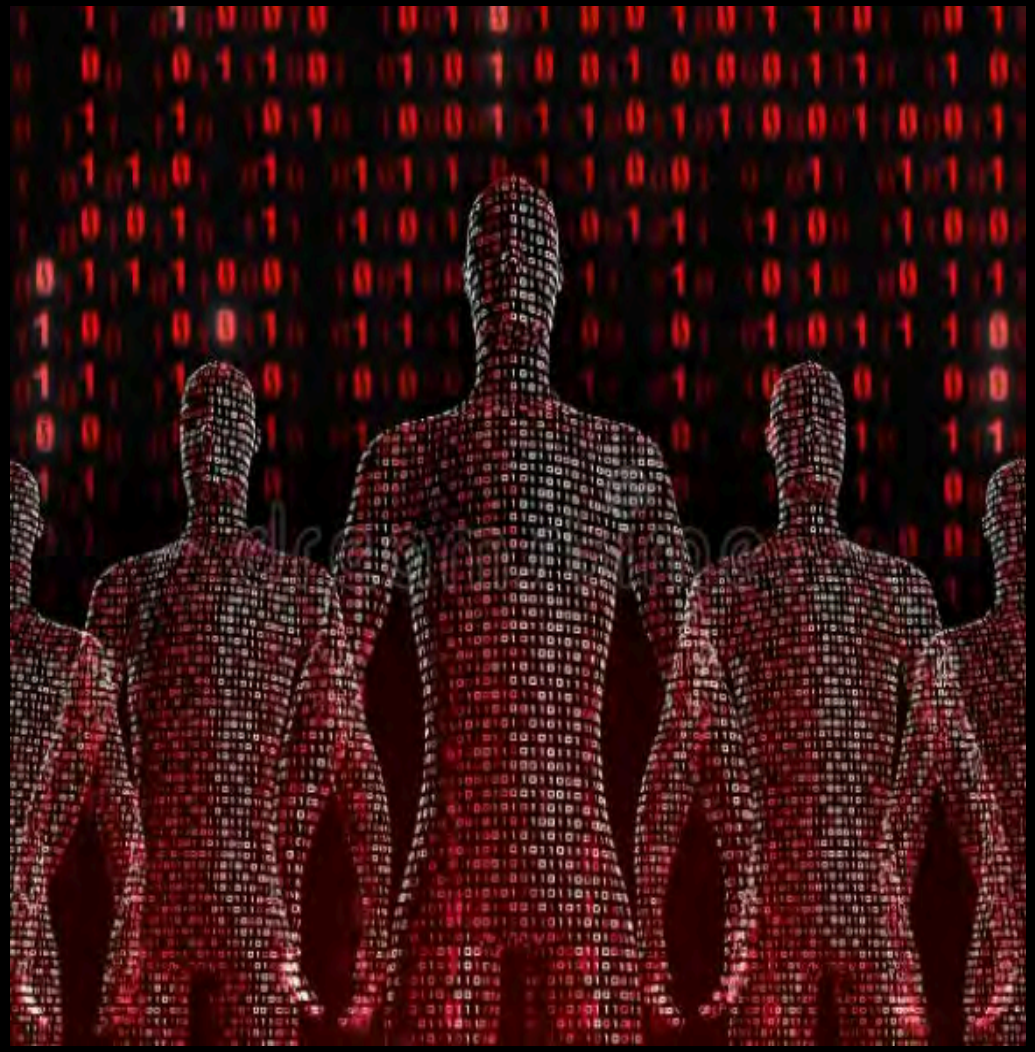
# Die problematische Basis des Mind Uploads

- Prinzipielle Grenzen
  - starker Computerfunktionalismus
  - Substratunabhängigkeit
  - starke Supervenienz
  - schwache Emergenz
  - Abstrakter Geist statt Embodiement
  - Detail-Scan und Erfassungsebene
- Praktische Probleme
- Personale Identität
  (qualitativ und numerisch)
- Vom Albtraum eines gelingenden Mind Uploads



# Ist die Person erst abstrahiert, kopiert es sich ganz ungeniert

- Das abstrakte Modell eines ‚Ich' im Gehirn wird zur Person.
- Ein exakter Scan kann *diese* Art von Person erfassen und auf einen leistungsstarken Computer übertragen.
- Ein geeignetes Programm prozessiert die Person.
- Sie kann jederzeit kopiert werden.
- Sie kann sich mit Lichtgeschwindigkeit ausbreiten.
- Sie kann durch Reprogrammierung ohne die Fesseln des Fleisches und die Grenzen der Biologie optimiert werden.
- Optimierung heißt: Quasi ewiges Leben, keine Krankheiten, Superintelligenz, unverstellbare posthumane Freuden, ….



# Wenn wir als Person in einem Computer leben könnten – ist es dann vielleicht schon der Fall?

Leben wir in einer Computersimulation?

# Das posthumane Utopia

- Intelligentes Leben wird vom biologischen ins digitale Substrat transferiert.
- Mind Uploads, die quasi unendlich leben, werden weiter zu Posthumanen optimiert, die unvorstellbare Freuden ohne Leid empfindenden.
- Diese Posthumanen existieren in astronomischer Zahl in virtuellen, im Computer generierten Welten.
- Die Masse von Sonnen und Planeten wird in gigantische planetenschwere Computern umgewandelt (= digitale Posthumane = Wertstrukturen).
- Digitales Leben kolonisiert über selbstreplizierende von-Neumann-Sonden das ganze erreichbare Universum.
- Die Macht über diesen kosmischen Cyberspace liegt bei einem sog. ‚Singleton‘: *Eine* Entität, in der sich alle Macht der Welt bündelt.
- Die Rolle des Singleton übernimmt eine mit den ‚richtigen‘ Werten gefütterte Superintelligenz.

# IV. Die Ethik Bostroms hergeleitet

# Utilitarismus als Basis

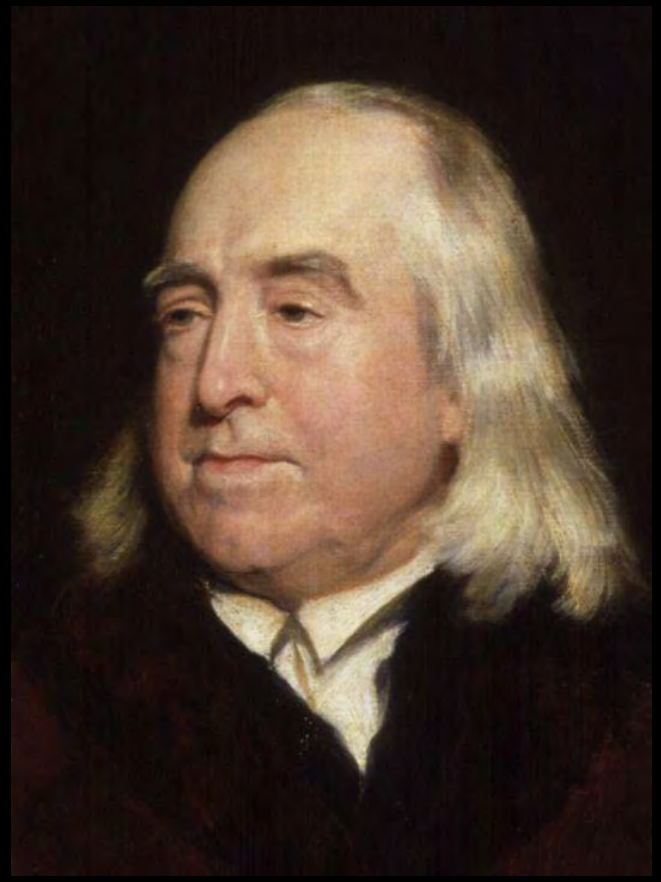
- ***Konsequentialismus***
  Moralische Bewertung einer Handlung anhand der sich aus ihr ergebenden Konsequenzen.

- ***Hedonismus***
  Die Mehrung von Lust und Freude, sowie die Vermeidung von Leid (Saldo = Glück)

- ***Utilitätsprinzip***
  Der Nutzen bestimmt den moralischen Wert einer Handlung, der sich aus deren Beitrag zum Glück bemisst

  Der Gesamtnutzen ergibt sich aus der Addition der Einzelnutzen der Betroffenen.
  „Greatest happiness principle“

- ***Neutralität und Altruismus***
  *Alle* von einer Handlung Betroffenen werden *in gleicher* Weise einbezogen.



# Besondere Aspekte im utilitaristischen Denken

- ***Einfügen einer qualitativen Komponente***
  Kulturelle, intellektuelle und spirituelle Befriedigungen haben einen höheren qualitativen Wert als die rein körperliche Befriedigung.

- ***Anti-Speziezismus***
  Absage an ein Denken, dass dem Menschen qua Menschsein einen besonderen, herausgehobenen moralischen Wert zuweist.

- ***Populationsethik***
  Forschung nach Bedingungen, die zu einer besseren oder schlechteren Bewertung einer Population führen. Sie untersucht die ethischen Fragen, die sich ergeben, wenn wir die Konsequenzen unseres Handelns für die zukünftige Population in qualitativer und quantitativer Hinsicht in den Blick nehmen.

# Zwei fundamental unterschiedliche Varianten des Utilitarismus

Maximierung der Summe des Nutzens *aller* Beteiligter

## Total-Ansicht

In die zu maximierende Nutzensumme fließen die Zustände *aller* in der Gegenwart und in der Zukunft lebenden, auch potenziell existierenden empfindungsfähigen Wesen ein. Das Schaffen zusätzlicher Wesen mit einem Lust-Leid-Saldo > 0 erhöht die Gesamtsumme.

## Vorherige-Existenz-Ansicht

In die zu maximierende Gesamtsumme fließen nur die Zustände jener Wesen ein, die unabhängig von der in Frage stehenden Handlung existieren oder existieren werden. Dieser Ansatz leugnet einen moralischen Wert, der darin bestünde, Freude dadurch zu vermehren, dass zusätzliche Wesen geschaffen werden.

# Das Destillat bei Bostrom: Radikaler Utilitarismus

- ***Anti-Speziezismus***

Nicht nur biologisches, sondern auch digitales Leben fließt in die Nutzenfunktion ein.

- ***Qualitativer Hedonismus***

Die Arten der Freude und der Lust werden in ihrer Qualität, und in diesem Sinne in ihrer Wertigkeit der Berücksichtigung in der zu maximierenden Summe der Nutzenfunktion unterschieden: Posthumane Freuden wiegen schwerer als humane.

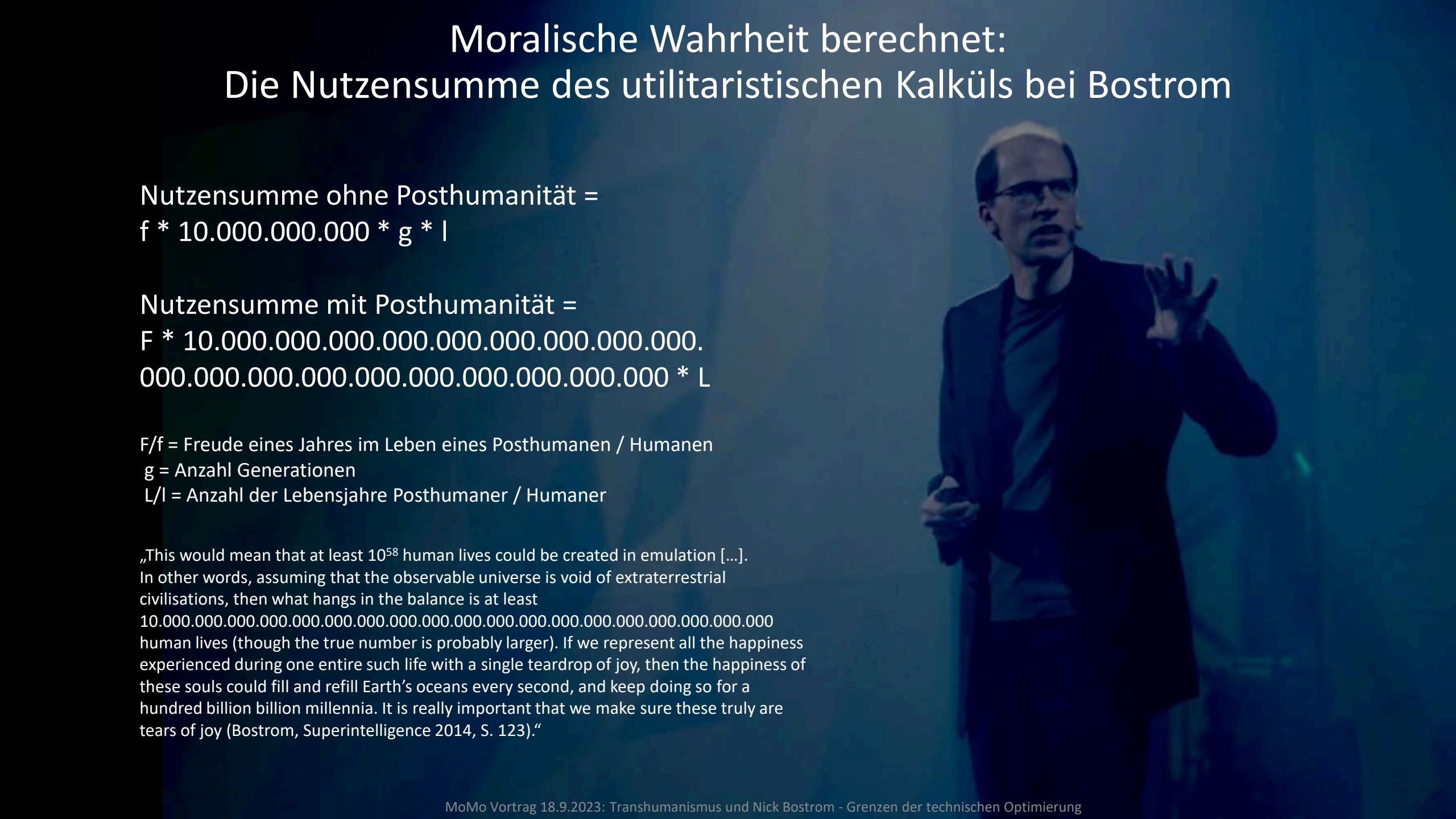
- ***Totalansicht***

Ziel moralischen Handelns ist die Vermehrung der Gesamtnutzensumme, in die das Glück aller in der Gegenwart und in der Zukunft lebenden empfindungsfähigen Wesen einfließt – digitale Posthumane eingeschlossen.

- ***Maximale Erweiterung in Raum und Zeit***

Es werden alle potenziellen Wesen einbezogen, die in der langfristigen Zukunft in der Dimension von Milliarden von Jahren im erreichbaren Universum existieren könnten.

# Moralische Wahrheit berechnet:
# Die Nutzensumme des utilitaristischen Kalküls bei Bostrom



Nutzensumme ohne Posthumanität =
f * 10.000.000.000 * g * l

Nutzensumme mit Posthumanität =
F * 10.000.000.000.000.000.000.000.000.000.
000.000.000.000.000.000.000.000.000.000 * L

F/f = Freude eines Jahres im Leben eines Posthumanen / Humanen
g = Anzahl Generationen
L/l = Anzahl der Lebensjahre Posthumaner / Humaner

„This would mean that at least $10^{58}$ human lives could be created in emulation […]. In other words, assuming that the observable universe is void of extraterrestrial civilisations, then what hangs in the balance is at least 10.000.000.000.000.000.000.000.000.000.000.000.000.000.000.000.000.000.000 human lives (though the true number is probably larger). If we represent all the happiness experienced during one entire such life with a single teardrop of joy, then the happiness of these souls could fill and refill Earth's oceans every second, and keep doing so for a hundred billion billion millennia. It is really important that we make sure these truly are tears of joy (Bostrom, Superintelligence 2014, S. 123)."

# V. Kombination zu einer problematischen Theorie

# Der moralische Imperativ bei Bostrom: Die posthumane Zukunft MUSS erreicht werden

- „This could be a wonderful development: lives free of pain and disease, bubbling over with happiness, enriched with superhuman awareness and understanding and all manner of higher goods (Shulmann / Bostrom 2021).“
- „Minimize existential risk (Bostrom 2003).“ 311 f.)!
- „We thus see that while some possible vulnerabilities can be stabilized with preventive policing alone, and some other vulnerabilities can be stabilized with global governance alone, there are some that would require both. […] ubiquitous-surveillance-powered preventive policing and effective global governance would be sufficient to stabilize most vulnerabilities, making it safe to continue scientific and technological development (Bostrom 2019).”
- “A non-existential disaster causing the breakdown of global civilization is, from the perspective of humanity as a whole, a potentially recoverable setback: a giant massacre for man, a small misstep for mankind (Bostrom 2009).”
- Zu Krieg, Genozid, Pandemie oder klimabedingte Überflutungen und Dürren:
  „But tragic as such events are to the people immediately affected, in the big picture of things – from the perspective of humankind as a whole – even the worst of these catastrophes are mere ripples on the surface of the great sea of life (Bostrom 2002).”
- „Unrestricted altruism is not so common that we can afford to fritter it away on a plethora of feel-good projects of suboptimal efficacy. If benefiting humanity by increasing existential safety achieves expected good on a scale many orders of magnitude greater than that of alternative contributions, we would do well to focus on this most efficient philanthropy (Bostrom 2013).”

Das Paradies der Posthumanität darf nicht riskiert werden.

Alle anderen Probleme und Katastrophen sind im Vergleich dazu moralisch unbedeutend.

# Ein gefährliches Denken mit enormem Einfluss

- Future of Humanity Institute (FHI), Global Priorities Institute, Future of Life Institute, etc…
- Zu den größten Geldgebern des FHI zählt z.B. Elon Musk, ein Fan der Theorie und Werke Bostroms.
- Philosophischer, moraltheoretischer Unterbau für globale Tech-Aktivitäten.
- Enge Verknüpfungen und Überlappungen mit den Bewegungen Effective Altruism und Longtermism.

Effective Altruism

The Longtermism Fund

Vielen Dank für Eure Aufmerksamkeit.

–

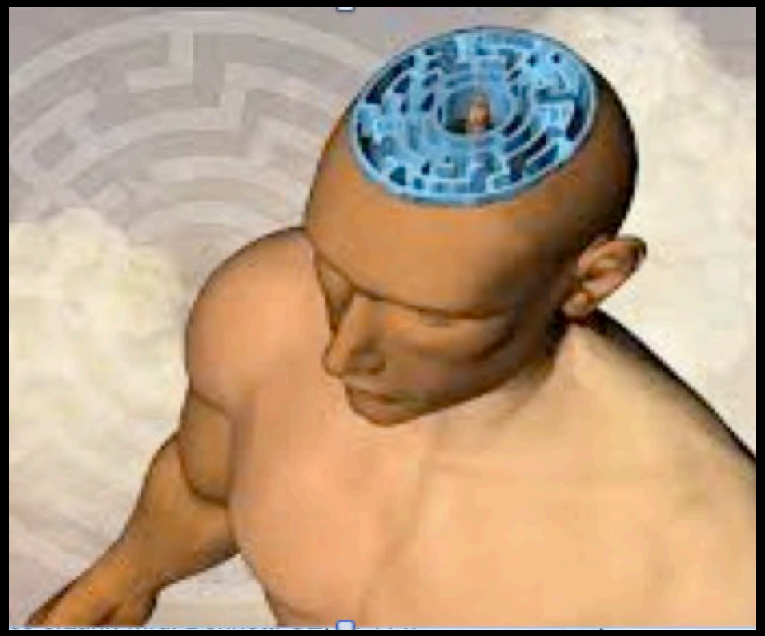
Ich freue mich auf den Austausch.

# Back Up

# Versuch einer Definition des Transhumanismus

Der Transhumanismus versteht den Menschen als ein defizitäres Zwischenergebnis einer unabgeschlossenen *natürlichen* Evolution, das dieser selbst mit Hilfe von Wissenschaft und Technik über die Grenzen seiner Natur hinaus weiter optimieren kann und soll, um durch den Zustand des Transhumanen hindurch zum Posthumanen zu werden. Mit der Verbesserung soll eine höhere Wahrscheinlichkeit auf eine längere Gesundheitsspanne, auf erweiterte geistige und physische Fähigkeiten und ein gutes Leben einhergehen. Der Transhumanismus ergänzt die traditionellen humanistischen Methoden von Bildung, Erziehung und Kultur um technische Mittel und verändert die Stoßrichtung: Anders als dem Humanismus geht es ihm nicht um eine Optimierung des Menschen *im Rahmen seiner Natur*, sondern um Eingriffe, die ihn *aus seiner Natur heraus* entwickeln, um das Humane zu transzendieren.

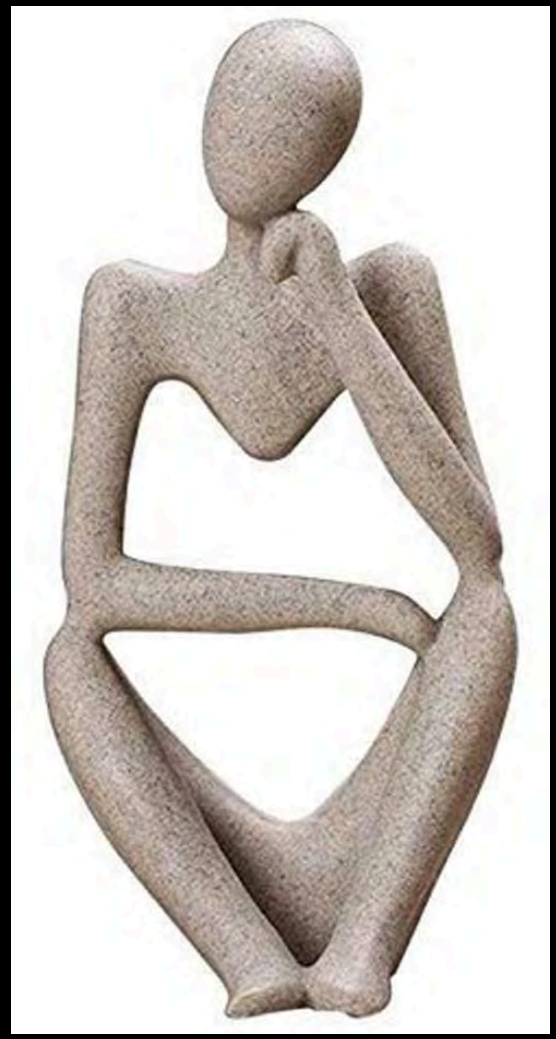
# Das ‚Ich' als abstraktes Selbstmodell, das im Gehirn entsteht



- Ontologie: Das ‚Selbst' ist keine eigene Substanz und keine autonome, selbsttragende Entität.
- Ein ‚Ich entsteht durch Verarbeitungsprozesse im Gehirn.
- Es ist ein „Selbstmodell der Subjektivität".
- Das ‚Ich' ist eine Art von ‚Trick' des Gehirns, um das Gesamtsystem (unseren Körper) besser zu steuern.
- Damit das Gehirn den menschlichen Körper durch die Welt navigieren kann, braucht es ein Modell der Welt, in der dieser als das ‚Selbst' enthalten.
- Das Gehirn schafft deshalb eine phänomenale Erste-Person-Perspektive, indem es ein repräsentationales Konstrukt phänomenaler Gegenwart erzeugt, das zu einem virtuellen, zugleich subjektiv realen im Jetzt anwesenden Subjekt wird.
- Sensorische und selbstrefenzielle Inputs führen zu Selbsterkenntnis über innere Zustände.
- Mit Sprache können Informationen zum Narrativ des eigenen Selbst als handelnde Person in der Zeit verbunden werden.
- Das »Ich« als erzählerisches Konstrukt und abstrakte Vereinfachung hoch komplexer neuronaler Prozesse und lingualer Interaktion.

# Die Umdeutung eines Abstraktums zur Person

- Das ‚Ich' wäre keine materielle Entität, sondern eine modellartige Abstraktion unseres Gehirns.
- Reduktion der Person auf das vom Gehirn konstruierte abstrakte Selbst, das dieses als Modell zur Steuerung von Handlungen entwirft.
- Dieses Modell der verkörperten Einheit einer Person wird zur eigentlichen Person umgedeutet.
- Ein solches Abstraktum soll als vollständiges Informationsmuster mit allen funktional relevanten Aspekten des Körpers in der Zukunft vollständig erfassbar und vom biologischen Substrat auf eine silizium-basierte Hardware übertragbar sein.



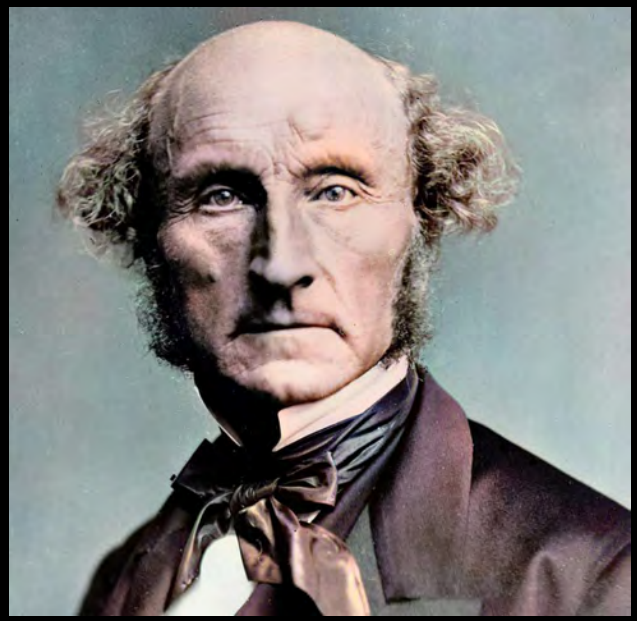
# Qualitativer Hedonismus - Mill

- ***Einfügen einer qualitativen Komponente***
  Kulturelle, intellektuelle und spirituelle Befriedigungen haben einen höheren qualitativen Wert als die rein körperliche Befriedigung.

- ***Das höherwertige Wesen erkennt den Unterschied***
  „Von zwei Freuden ist diejenige wünschenswerter, die von allen oder nahezu allen, die beide erfahren haben […] entschieden bevorzugt wird (Mill 1863).“

- ***Die Zufriedenheit der Schweine und Narren***
  „Es ist besser, ein unzufriedener Mensch zu sein als ein zufriedenes Schwein, besser ein unzufriedener Sokrates als ein zufriedener Narr.“

  „Und wenn der Narr oder das Schwein anderer Ansicht sind, dann deshalb, weil sie nur die eine Seite der Angelegenheit kennen. Die andere Partei hingegen kennt beide Seiten.“

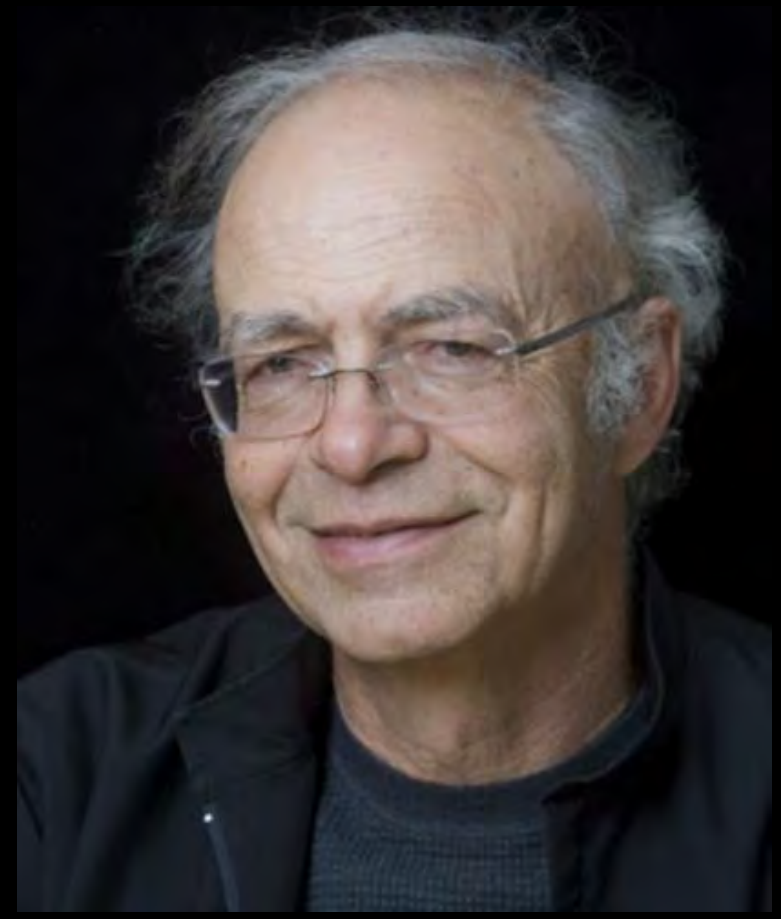


# Anti-Speziezismus - Singer

- ***Anti-Speziezismus***
  Absage an ein Denken, dass dem Menschen qua Menschsein einen besonderen, herausgehobenen moralischen Wert zuweist.

- ***Keine Diskriminierung aufgrund der Spezies***
  „Dem Leben eines Wesens bloß deshalb den Vorzug zu geben, weil das Lebewesen unserer Spezies angehört, würde uns in eine unangenehme Position bringen. Sie gleicht jener der Rassisten, die denen den Vorzug geben, die zu ihrer Rasse gehören (Singer 1979).“

- ***Keine Absage an qualitative Unterschiede***
  „Je höher entwickelt die mentalen Fähigkeiten eines Wesens, je größer der Grad von Selbstbewusstsein und Rationalität und je umfassender der Bereich möglicher Erfahrungen, umso mehr würde man diese Art des Lebens vorziehen, wenn man zwischen ihm und einem Wesen auf einer niedrigeren Bewusstseinsstufe zu wählen hätte.“



# Populationsethik - Parfit

- ***Populationsethik***

  Forschung nach Bedingungen, die zu einer besseren oder schlechteren Bewertung einer Population führen. Sie untersucht die ethischen Fragen, die sich ergeben, wenn wir die Konsequenzen unseres Handelns für die zukünftige Population in qualitativer und quantitativer Hinsicht in den Blick nehmen.

- ***Anwendung Konsequentialismus auf Populationen und langfristige Zukunft***

  Herausforderungen:
  - Axiologischen Frage, wie der Wert einer Population zu messen ist
  - Normativen Frage, was dann das richtige Handeln für die heute Lebenden ist

- ***Fundamentale Unterschiede nach Utilitarismus-Variante***

  (1) Totalansicht: In die zu maximierende Nutzensumme fließen die Zustände *aller* in der Gegenwart und in der Zukunft lebenden, empfindungsfähigen Wesen ein.
  (2) Vorherige-Existenz-Ansicht (personenbezogene Variante): In die zu maximierende Gesamtsumme fließen nur die Zustände jener Wesen ein, die unabhängig von der in Frage stehenden Handlung existieren oder existieren werden.
  Dieser Ansatz leugnet einen moralischen Wert, der darin bestünde, Freude dadurch zu vermehren, dass zusätzliche Wesen geschaffen werden.