Föreläsning på MVE420: Nya teknologier, global risk och mänskligheten framtid

Existentiell risk och effektiv altruism

16 april 2021

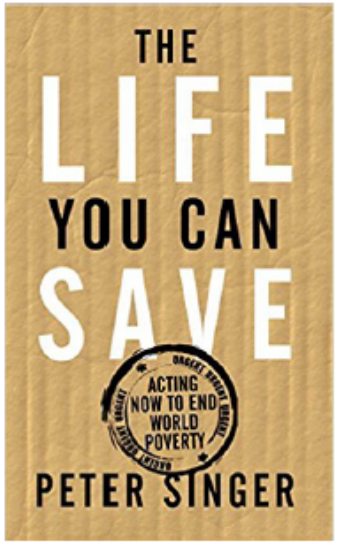
Olle Häggström

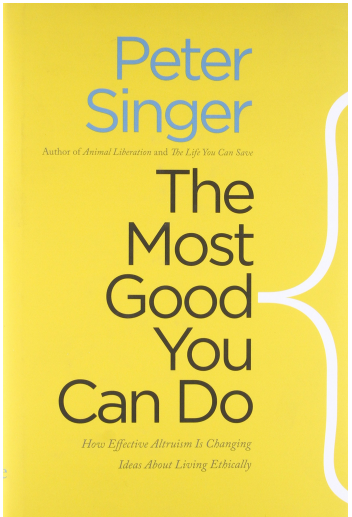
https://www.chalmers.se/en/Staff/Pages/olle-haggstrom.aspx

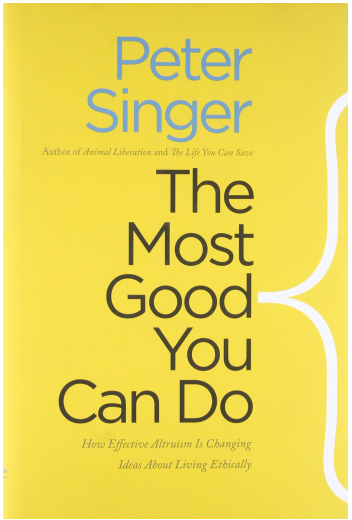
http://haggstrom.blogspot.com/

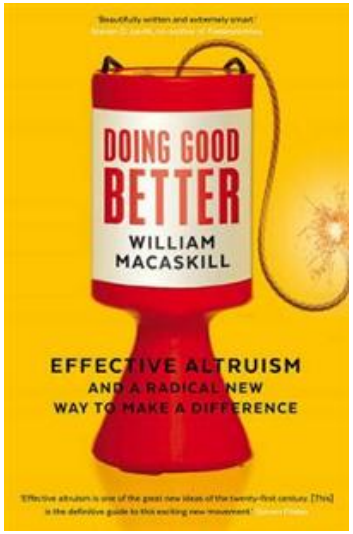
Ni minns väl Peter Singer?

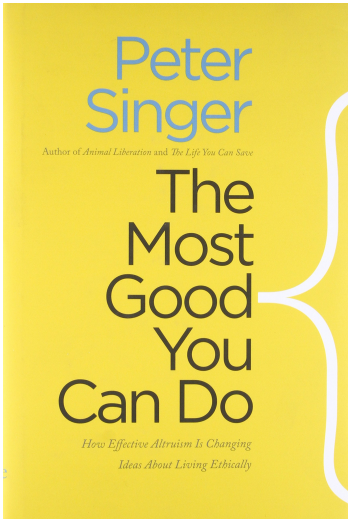
Ni minns väl Peter Singer?

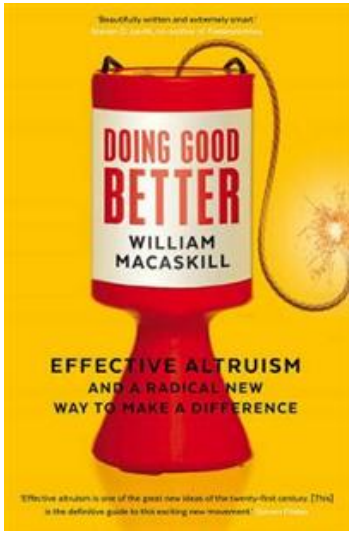



Peter
Singer
Author of Animal Liberation and The Life You Can Save
The
Most
Good
You
Can Do
How Effective Altruism Is Changing
Ideas About Living Ethically


Peter
Singer
Author of Animal Liberation and The Life You Can Save
The
Most
Good
You
Can Do
How Effective Altruism Is Changing
Ideas About Living Ethically

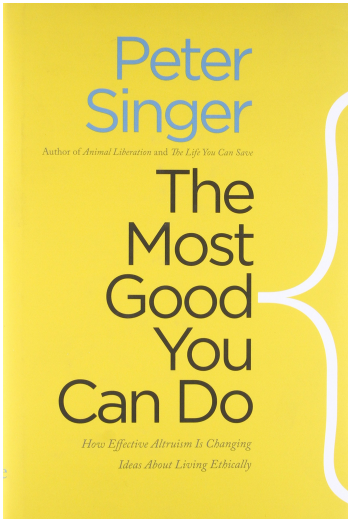

DOING GOOD
BETTER
WILLIAM
MACASKILL
EFFECTIVE ALTRUISM
AND A RADICAL NEW
WAY TO MAKE A DIFFERENCE

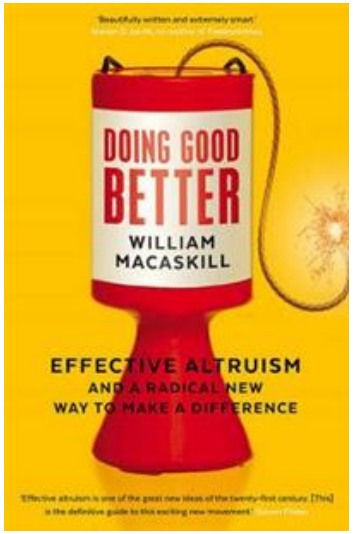




Effektiv altruism

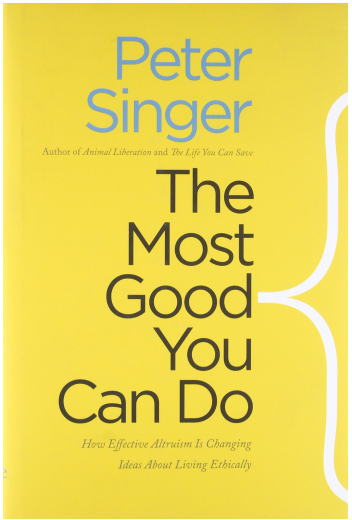
En filosofi och en social rörelse med grundtanken att göra världen bättre,

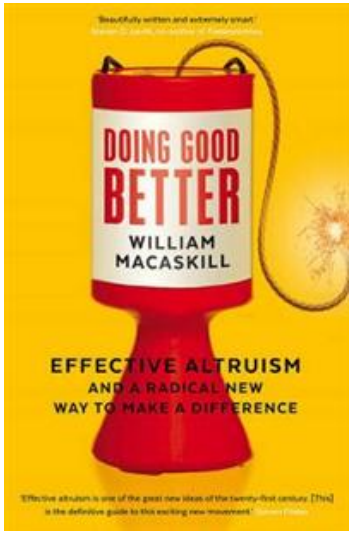




Effektiv altruism

En filosofi och en social rörelse med grundtanken att göra världen bättre, **så effektivt som möjligt**.

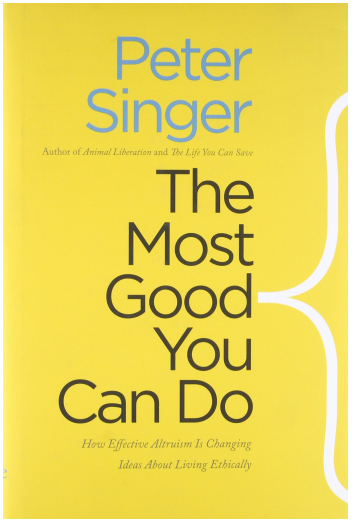


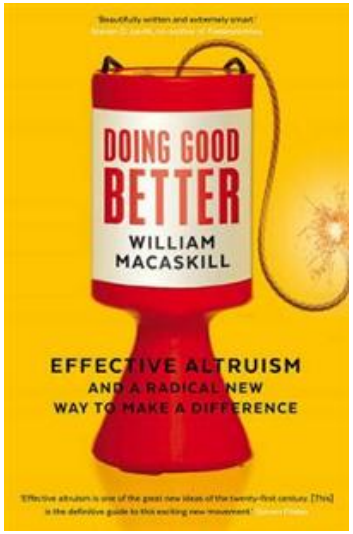


Effektiv altruism

En filosofi och en social rörelse med grundtanken att göra världen bättre, **så effektivt som möjligt**.

Prioriterade områden:



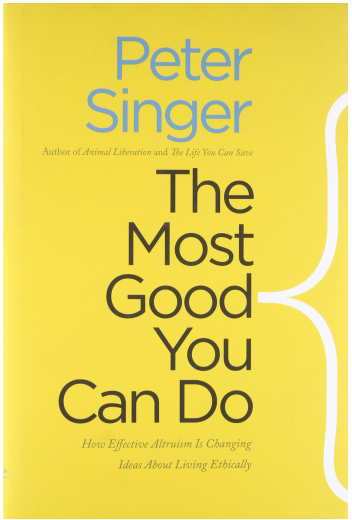


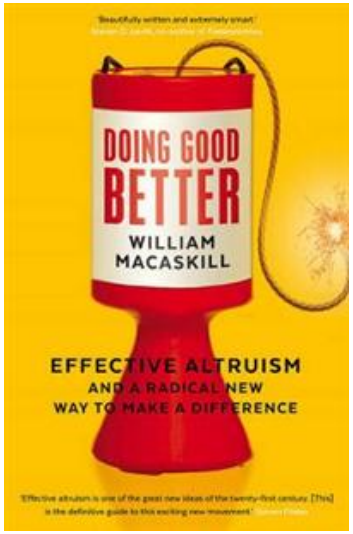
Effektiv altruism

En filosofi och en social rörelse med grundtanken att göra världen bättre, **så effektivt som möjligt**.

Prioriterade områden:

- Motverka fattigdom, svält och sjukdom i tredje världen



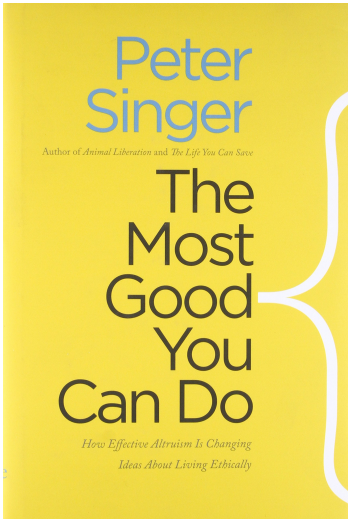


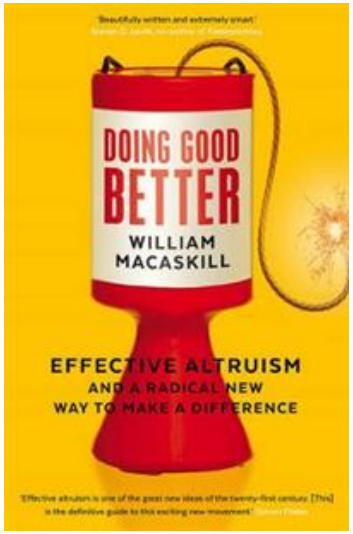
Effektiv altruism

En filosofi och en social rörelse med grundtanken att göra världen bättre, **så effektivt som möjligt**.

Prioriterade områden:

- Motverka fattigdom, svält och sjukdom i tredje världen
- Hejda djurplågeriet





Effektiv altruism

En filosofi och en social rörelse med grundtanken att göra världen bättre, **så effektivt som möjligt**.

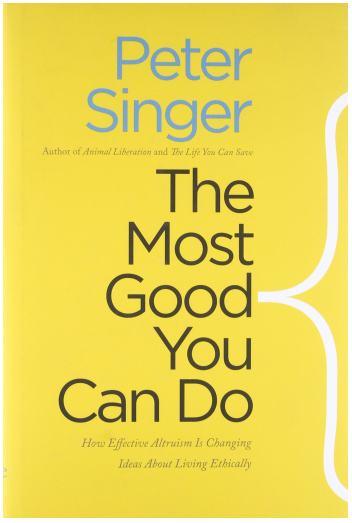
Prioriterade områden:

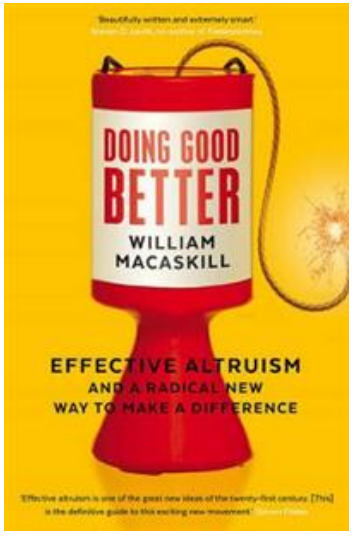
- Motverka fattigdom, svält och sjukdom i tredje världen
- Hejda djurplågeriet
- Avvärj existentiell risk

Nick Bostrom, i sin artikel *Existential risk prevention as global priority* **från 2013:**

Nick Bostrom, i sin artikel *Existential risk prevention as global priority* **från 2013:**

> *An existential risk is one that threatens the premature extinction of Earth-originating intelligent life or the permanent and drastic destruction of its potential for desirable future development.*



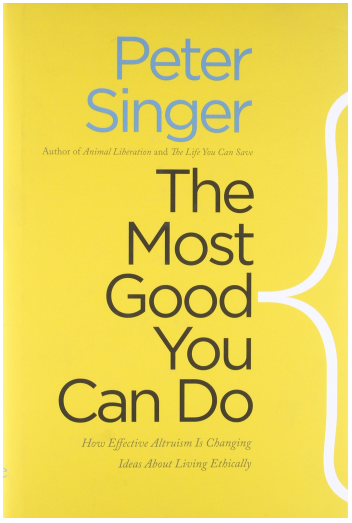


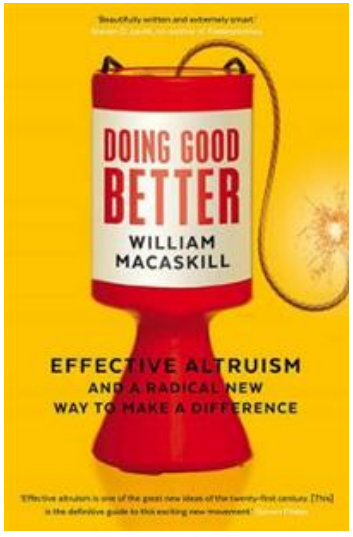
Effektiv altruism

En filosofi och en social rörelse med grundtanken att göra världen bättre, **så effektivt som möjligt**.

Prioriterade områden:

- Motverka fattigdom, svält och sjukdom i tredje världen
- Hejda djurplågeriet
- Avvärj existentiell risk



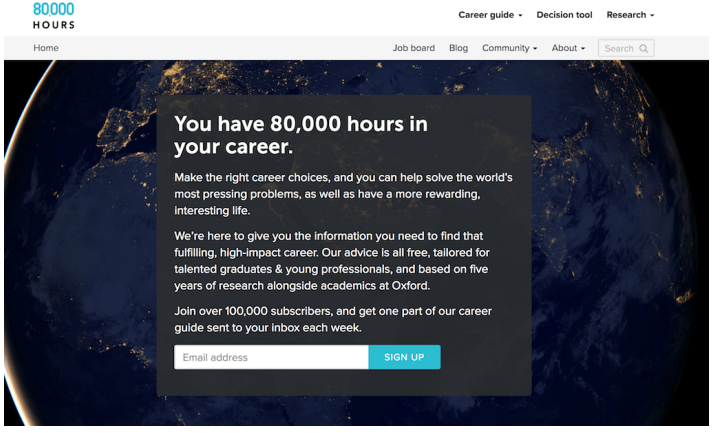


Effektiv altruism

En filosofi och en social rörelse med grundtanken att göra världen bättre, **så effektivt som möjligt**.

Prioriterade områden:

- Motverka fattigdom, svält och sjukdom i tredje världen
- Hejda djurplågeriet
- Avvärj existentiell risk



Max Tegmark, 2018: *It's now, for the first time in the 4.5 billion years history of this planet, that we are at this fork in the road. It's probably going to be within our lifetimes that we're either going to self-destruct or get our act together.*

Ni minns väl Nick Bostroms överslagskalkyler om hur mycket som står på spel?

Ni minns väl Nick Bostroms överslagskalkyler om hur mycket som står på spel?

I samma artikel från 2013 där han definierar existentiell risk gör han några överslagskalkyler. Om vi spelar våra kort rätt, och...

- mänskligheten överlever i 10^9 år,
- en befolkning på 10^9 kan upprätthållas,
- medellivslängden är 10^2 år,

så blir antalet framtida människoliv

$$10^9 10^9 10^{-2} = 10^{16}$$

Ni minns väl Nick Bostroms överslagskalkyler om hur mycket som står på spel?

I samma artikel från 2013 där han definierar existentiell risk gör han några överslagskalkyler. Om vi spelar våra kort rätt, och...

- mänskligheten överlever i 10^9 år,
- en befolkning på 10^9 kan upprätthållas,
- medellivslängden är 10^2 år,

så blir antalet framtida människoliv

$$10^9 10^9 10^{-2} = 10^{16}$$

Han föreslår även betydligt högre tal baserade på interstellär rymdkolonisering (10^{34}) och/eller uppladdning (10^{54}).

Senaste nytt från Toby Ord

Senaste nytt från Toby Ord

The Edges of Our Universe

Toby Ord*

This paper explores the fundamental causal limits on how much of the universe we can observe or affect. It distinguishes four principal regions: the *affectable* universe, the *observable* universe, the *eventually observable* universe, and the *ultimately observable* universe. It then shows how these (and other) causal limits set physical bounds on what spacefaring civilisations could achieve over the longterm future.

Introduction

How large is the universe? What exactly is the observable universe? Will we ever be able to detect things that are outside it? If so, what are the ultimate limits of observability? Are there fundamental limits on how far we could travel through space? How far away does something have to be such that it is completely causally separate from us? How do these relate to each other? And how do they change with time?

These are key questions for understanding the large scale picture of the universe, its spatial limits, and its causal structure. They are also key questions when attempting to understand how much a spacefaring civilisation might ever be able to achieve — including fundamental physical limits on our own civilisation. Yet very few people understand these limits and how they relate to each other. Many discussions conflate all the different causal limits together and assume that they are all set by the observable universe. Even astrophysicists and cosmologists make mistakes about these limits, including in their academic articles and textbooks.[1] Some of the most

Utöver den kvantitativa aspekten om hur **stort** och **långvarigt** det mänskliga samhället kan bli, finns också en mer kvalitativ fråga: hur **bra** kan våra liv bil?

Utöver den kvantitativa aspekten om hur **stort** och **långvarigt** det mänskliga samhället kan bli, finns också en mer kvalitativ fråga: hur **bra** kan våra liv bil?

Amanda Askell @AmandaAskell · Jan 19

I think my life on a grad student stipend was better than the life of a medieval king. Presumably this would be surprising to medieval people. I'm hoping that, at some point in the future, most people will think their lives are better than the lives of today's billionaires.

23 28 444

Jason Crawford @jasoncrawford · Jan 19

Louis XIV had smallpox, measles, gonorrhea, gout, and a couple of fevers that might have been typhoid/malaria... and then died of gangrene. Presumably you missed at least one of those in grad school

4 4 47

Amanda Askell @AmandaAskell · Jan 19

He also didn't have flush toilets, a shower, modern medicine and science, the internet, access to the world's music / food / art at the drop of a hat. All of which I did have during grad school.

1 22

Fisher (2021):

Imagine that one day you jumped in your time-travelling machine, and went back millions of years for a conversation with a pre-human hominin.

You: Hello there, pre-human!

Fisher (2021):

Imagine that one day you jumped in your time-travelling machine, and went back millions of years for a conversation with a pre-human hominin.

You: Hello there, pre-human!

Hominin: Hello.

Fisher (2021):

Imagine that one day you jumped in your time-travelling machine, and went back millions of years for a conversation with a pre-human hominin.

You: Hello there, pre-human!

Hominin: Hello.

You: Wow, you can talk! I didn't know you had language.

Fisher (2021):

Imagine that one day you jumped in your time-travelling machine, and went back millions of years for a conversation with a pre-human hominin.

You: Hello there, pre-human!

Hominin: Hello.

You: Wow, you can talk! I didn't know you had language.

Hominin: This is a thought experiment. And you can time travel.

Fisher (2021):

Imagine that one day you jumped in your time-travelling machine, and went back millions of years for a conversation with a pre-human hominin.

You: Hello there, pre-human!

Hominin: Hello.

You: Wow, you can talk! I didn't know you had language.

Hominin: This is a thought experiment. And you can time travel.

You: Good point.

Fisher (2021):

Imagine that one day you jumped in your time-travelling machine, and went back millions of years for a conversation with a pre-human hominin.

You: Hello there, pre-human!

Hominin: Hello.

You: Wow, you can talk! I didn't know you had language.

Hominin: This is a thought experiment. And you can time travel.

You: Good point.

Hominin: What can I do for you?

Fisher (2021):

Imagine that one day you jumped in your time-travelling machine, and went back millions of years for a conversation with a pre-human hominin.

You: Hello there, pre-human!

Hominin: Hello.

You: Wow, you can talk! I didn't know you had language.

Hominin: This is a thought experiment. And you can time travel.

You: Good point.

Hominin: What can I do for you?

You: Gosh, I have so many questions... but today I'll ask just this: what do you expect in the far future? When you evolve into *Homo sapiens*, what do you imagine the lives of your descendants will be like?

Hominin: You're asking what they can look forward to? Well, I guess they'd have a life of unlimited bananas! Wouldn't that be grand?

Hominin: You're asking what they can look forward to? Well, I guess they'd have a life of unlimited bananas! Wouldn't that be grand?

You: ... wait... bananas? But there's so much more to humanity's fut...

Hominin: You're asking what they can look forward to? Well, I guess they'd have a life of unlimited bananas! Wouldn't that be grand?

You: ... wait... bananas? But there's so much more to humanity's fut...

Hominin: No, that's what I want for my children. Now if you don't mind, I have lunch to forage.

Vad kan orsaka mänskligheten undergång?

Vad kan orsaka mänskligheten undergång?

- Kärnvapen

Valentin Savitsky

Valentin Savitsky

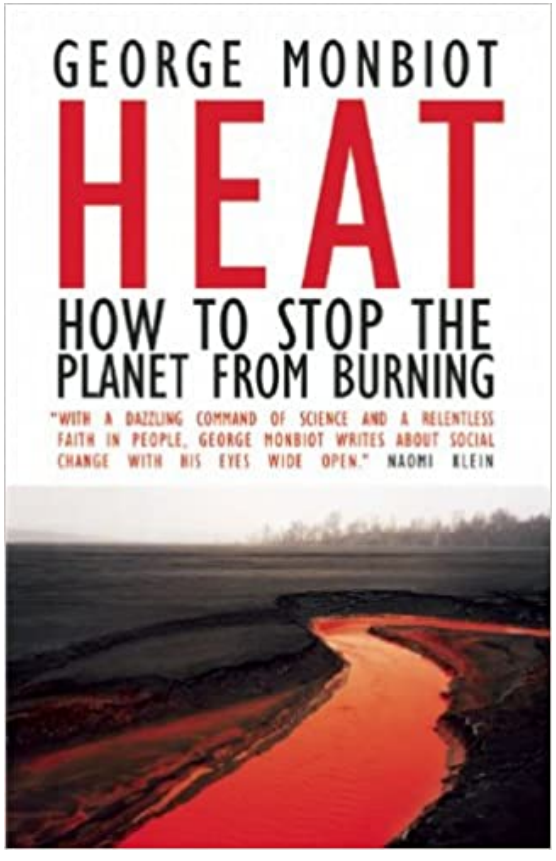
Vasili Arkhipov (1926-1998)

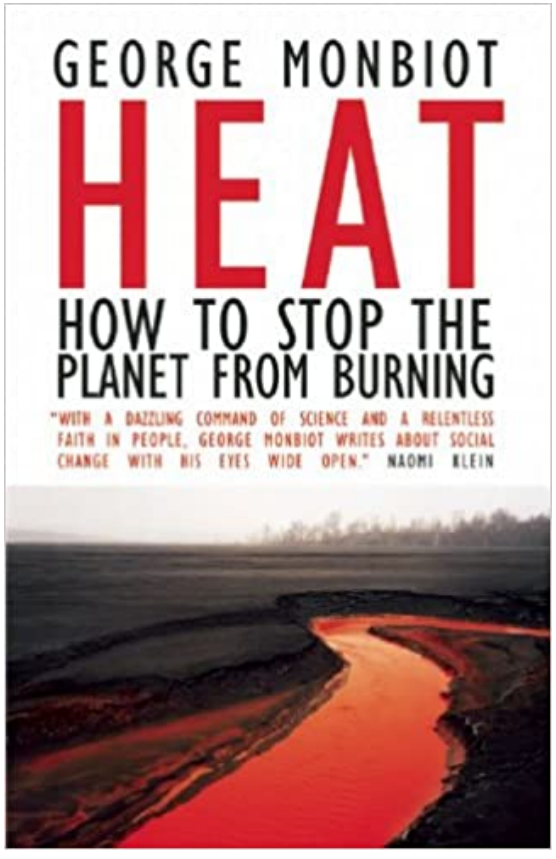
Vad kan orsaka mänskligheten undergång?

- Kärnvapen

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar

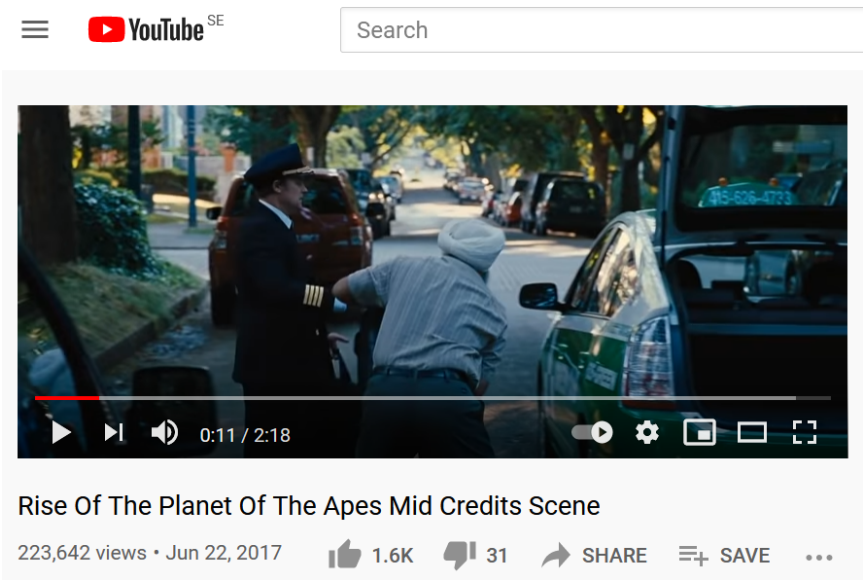

GEORGE MONBIOT
HEAT
HOW TO STOP THE PLANET FROM BURNING
"WITH A DAZZLING COMMAND OF SCIENCE AND A RELENTLESS FAITH IN PEOPLE, GEORGE MONBIOT WRITES ABOUT SOCIAL CHANGE WITH HIS EYES WIDE OPEN." NAOMI KLEIN


GEORGE MONBIOT
HEAT
HOW TO STOP THE PLANET FROM BURNING
"WITH A DAZZLING COMMAND OF SCIENCE AND A RELENTLESS FAITH IN PEOPLE, GEORGE MONBIOT WRITES ABOUT SOCIAL CHANGE WITH HIS EYES WIDE OPEN." NAOMI KLEIN


F4 L1 Y4 G2 S1 K5 A1 M3


RISE
OF THE PLANET OF THE APES


RISE
OF THE PLANET OF THE APES


YouTube SE
Search
0:11 / 2:18
Rise Of The Planet Of The Apes Mid Credits Scene
223,642 views • Jun 22, 2017
1.6K
31
SHARE
SAVE

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier

Vad kan orsaka mänskligheten undergång?

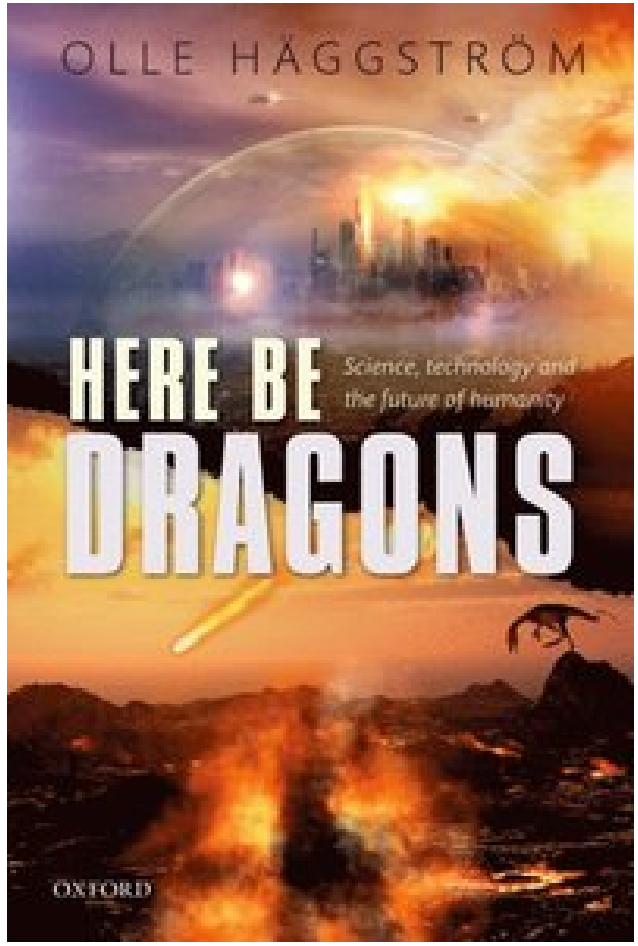
- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag

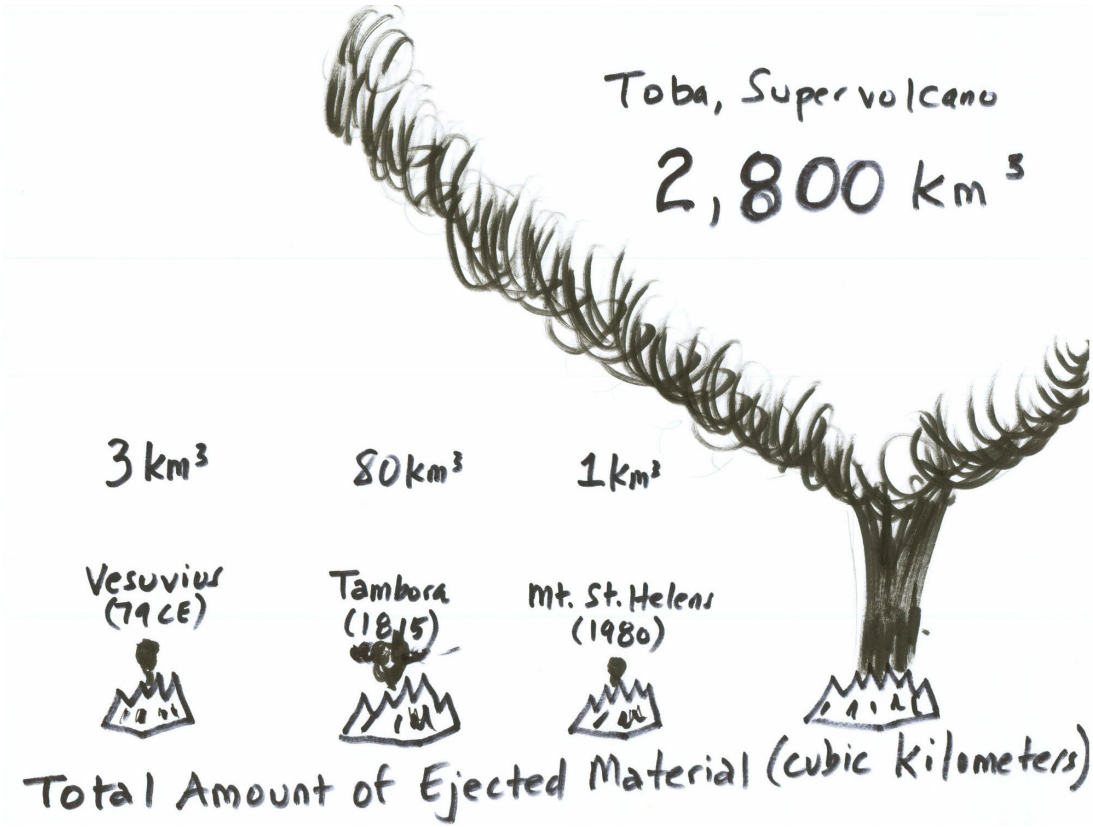

OLLE HÄGGSTRÖM
HERE BE DRAGONS
Science, technology and the future of humanity
OXFORD

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag
- Supervulkaner


Toba, Supervolcano
2,800 km³
3 km³
80 km³
1 km³
Vesuvius (79 CE)
Tambora (1815)
Mt. St. Helens (1980)
Total Amount of Ejected Material (cubic kilometers)

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag
- Supervulkaner

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag
- Supervulkaner
- Supernovor

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag
- Supervulkaner
- Supernovor
- Utomjordingar

Shocking life expectancy figures in Britain's poorest areas

Meghan Markle baby: Why Meghan and Harry had to pay more for...

Jesus Christ revelation: Major breakthrough as 'Goliath's skull...

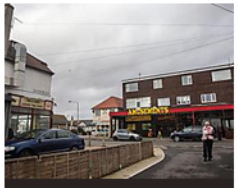
Kate with Paki

Alien alert: Scientist warns humans should NEVER make contact with ET - 'Risk too great'

HUMANS should avoid making contact with aliens, a scientist has warned, due to fears extraterrestrial life might wish us harm.

By SEAN MARTIN

PUBLISHED: 14:32, Wed, Sep 4, 2019 | UPDATED: 14:42, Wed, Sep 4, 2019



Shocking life expectancy figures in Britain's poorest areas

Meghan Markle baby: Why Meghan and Harry had to pay more for...

Jesus Christ revelation: Major breakthrough as 'Goliath's skull...

Kate with Paki

Alien alert: Scientist warns humans should NEVER make contact with ET - 'Risk too great'

HUMANS should avoid making contact with aliens, a scientist has warned, due to fears extraterrestrial life might wish us harm.

By SEAN MARTIN

PUBLISHED: 14:32, Wed, Sep 4, 2019 | UPDATED: 14:42, Wed, Sep 4, 2019

Aliens with the technology to reach Earth from the far reaches of the cosmos would likely have to potential to destroy us. Olle Häggström, a professor of mathematical statistics at Chalmers University and author of the existential risk book Here Be Dragons, said an advanced extraterrestrial civilisation could see humanity as a threat and destroy us. Mr Häggström told journalist Bryan Walsh, author existential risk book End Times: "There are optimists who say that good things can come out of establishing communications

Vad kan orsaka mänskligheten undergång?

- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag
- Supervulkaner
- Supernovor
- Utomjordingar

Vad kan orsaka mänskligheten undergång?

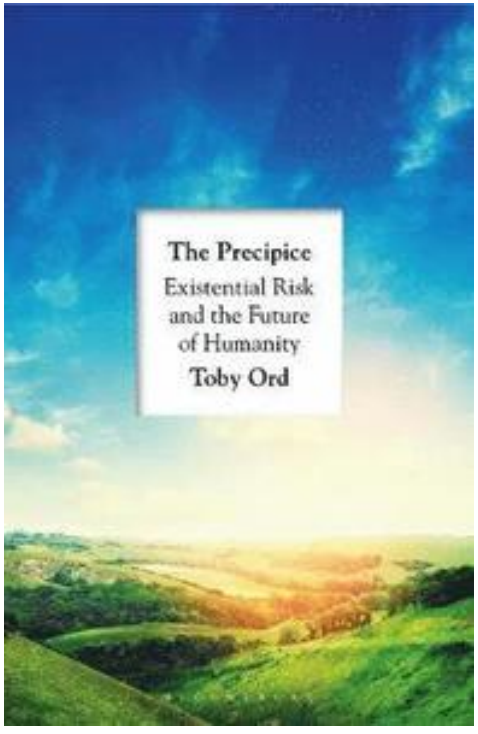
- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag
- Supervulkaner
- Supernovor
- Utomjordingar
- AI-katastrof

Vad kan orsaka mänskligheten undergång?

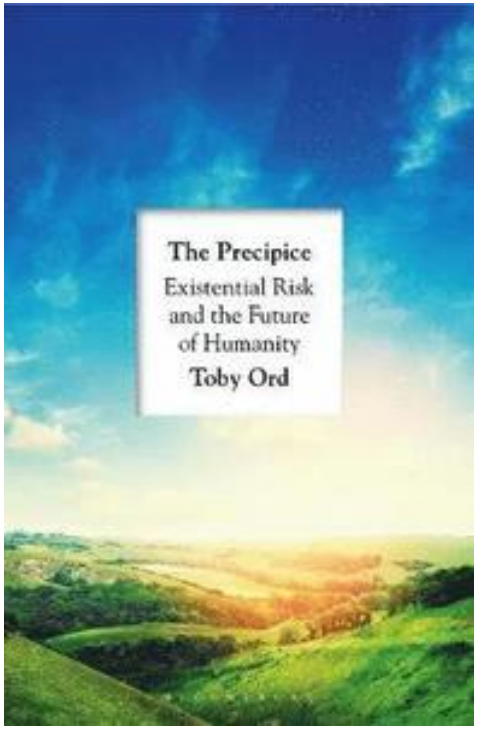
- Kärnvapen
- Klimatförändringar
- Pandemier
  - Av naturligt ursprung
  - Laboratorieskapade
- Asteroidnedslag
- Supervulkaner
- Supernovor
- Utomjordingar
- AI-katastrof
- ...

Ur Toby Ords *The Precipice*

Ur Toby Ords *The Precipice*



Ur Toby Ords *The Precipice*

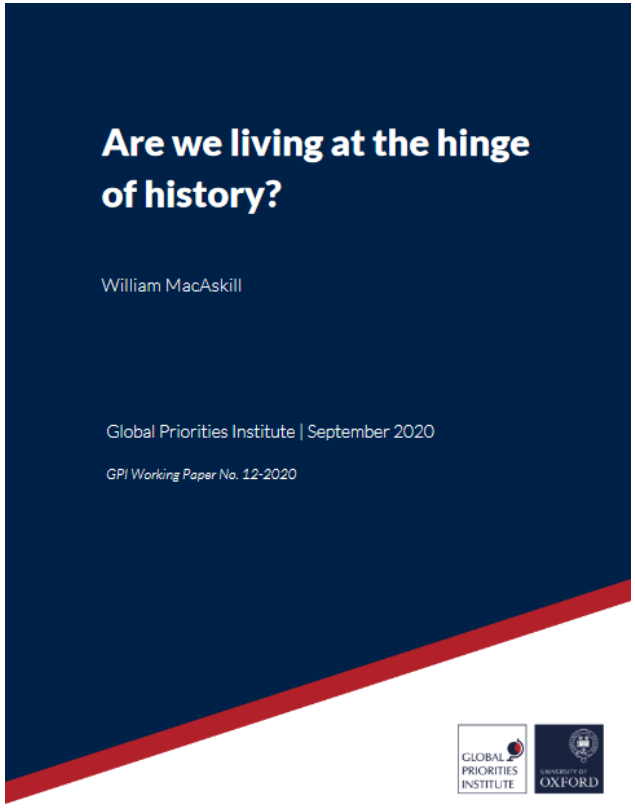


Existential catastrophe via	*Chance within next 100 years*
Asteroid or comet impact	~ 1 in 1,000,000
Supervolcanic eruption	~ 1 in 10,000
Stellar explosion	~ 1 in 1,000,000,000
Total natural risk	**~ 1 in 10,000**
Nuclear war	~ 1 in 1,000
Climate change	~ 1 in 1,000
Other environmental damage	~ 1 in 1,000
'Naturally' arising pandemics	~ 1 in 10,000
Engineered pandemics	~ 1 in 30
Unaligned artificial intelligence	~ 1 in 10
Unforeseen anthropogenic risks	~ 1 in 30
Other anthropogenic risks	~ 1 in 50
Total anthropogenic risk	**~ 1 in 6**
Total existential risk	**~ 1 in 6**

TABLE 6.1 My best estimates for the chance of an existential catastrophe from each of these sources occurring at some point in the next 100 years (when the catastrophe has delayed effects, like climate change, I'm talking about the point of no return coming within 100 years). There is significant uncertainty remaining in these estimates and they should be treated as representing the right order of magnitude—each could easily be a factor of 3 higher or lower. Note that the numbers don't quite add up: both because doing so would create a false feeling of precision and for subtle reasons covered in the section on 'Combining Risks'.

It's now, for the first time in the 4.5 billion years history of this planet, that we are at this fork in the road. It's probably going to be within our lifetimes that we're either going to self-destruct or get our act together.

Derek Parfit, 2011: *We live during the hinge of history. Given the scientific and technological discoveries of the last two centuries, the world has never changed as fast. We shall soon have even greater powers to transform, not only our surroundings, but ourselves and our successors. If we act wisely in the next few centuries, humanity will survive its most dangerous and decisive period. Our descendants could, if necessary, go elsewhere, spreading through this galaxy.*



Will MacAskill (2020): *In this article I try to make the hinge of history claim more precise, give arguments in favour and against, and assess whether it is true. Ultimately, I argue that the claim* [...] *is quite unlikely to be true, and that this fact can serve as part of an argument for the conclusion that impartial altruists should generally be investing their resources, rather than trying to do good immediately.*

Definitioner:

Definitioner:

- MacAskill definierar en tidpunkts *inflytelserikedom* som den största förväntade nytta som går att göra genom att spendera en given mängd resurser vid denna tidpunkt.

Definitioner:

- MacAskill definierar en tidpunkts *inflytelserikedom* som den största förväntade nytta som går att göra genom att spendera en given mängd resurser vid denna tidpunkt.
- *Hinge of History (HoH)*-påståendet är utsagan att den mest inflytelserika tiden någonsin är nu.

Definitioner:

- MacAskill definierar en tidpunkts *inflytelserikedom* som den största förväntade nytta som går att göra genom att spendera en given mängd resurser vid denna tidpunkt.
- *Hinge of History (HoH)*-påståendet är utsagan att den mest inflytelserika tiden någonsin är nu.

Om HoH-påståendet stämmer så har vi ypperliga möjligheter att göra gott genom att skrida till handling *nu*. Om HoH-tidpunkten istället ligger i framtiden, så går det att argumentera för att istället spara resurser för att använda då – så kallad **tålmodig longtermism**. Skälen kan vara dubbla:

Definitioner:

- MacAskill definierar en tidpunkts *inflytelserikedom* som den största förväntade nytta som går att göra genom att spendera en given mängd resurser vid denna tidpunkt.
- *Hinge of History (HoH)*-påståendet är utsagan att den mest inflytelserika tiden någonsin är nu.

Om HoH-påståendet stämmer så har vi ypperliga möjligheter att göra gott genom att skrida till handling *nu*. Om HoH-tidpunkten istället ligger i framtiden, så går det att argumentera för att istället spara resurser för att använda då – så kallad **tålmodig longtermism**. Skälen kan vara dubbla:

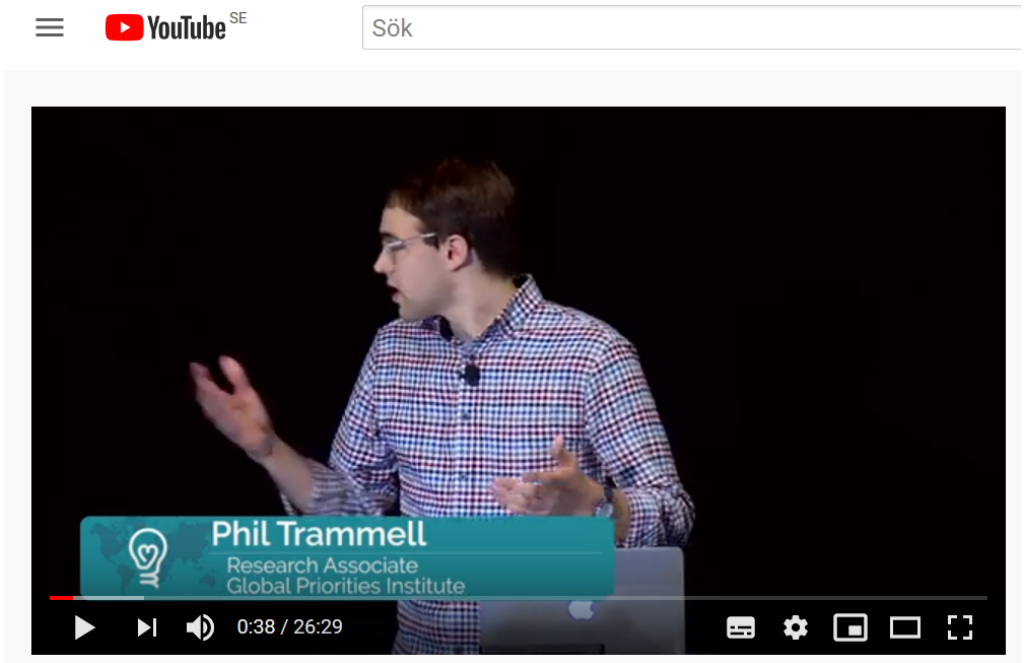
- Den större nytta resurserna kan göra vid den verkliga HoH-tidpunkten.

Definitioner:

- MacAskill definierar en tidpunkts *inflytelserikedom* som den största förväntade nytta som går att göra genom att spendera en given mängd resurser vid denna tidpunkt.
- *Hinge of History (HoH)*-påståendet är utsagan att den mest inflytelserika tiden någonsin är nu.

Om HoH-påståendet stämmer så har vi ypperliga möjligheter att göra gott genom att skrida till handling *nu*. Om HoH-tidpunkten istället ligger i framtiden, så går det att argumentera för att istället spara resurser för att använda då – så kallad **tålmodig longtermism**. Skälen kan vara dubbla:

- Den större nytta resurserna kan göra vid den verkliga HoH-tidpunkten.
- Om vi investerar förståndigt kan resurserna då ha hunnit växa.



Philanthropy timing and the Hinge of History

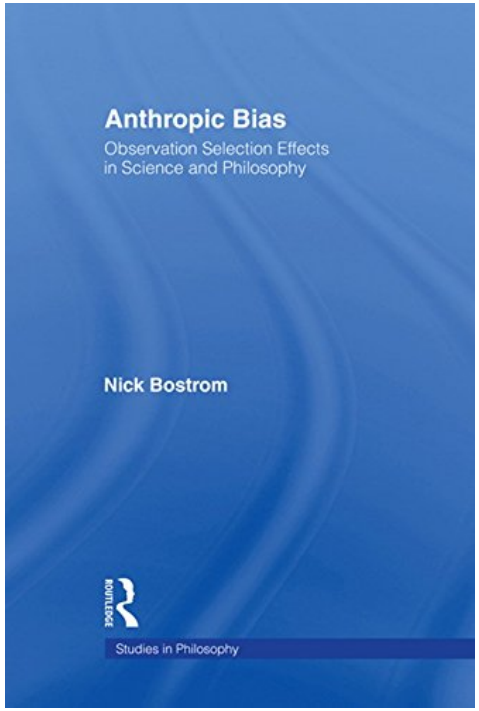
539 visningar • 29 aug. 2019

19 0 DELA SPARA

MacAskill godtar att det finns evidens för att HoH är nu. Men som goda bayesianer behöver vi använda denna evidens för att uppdatera a priori-sannolikheten att HoH är nu.

MacAskill godtar att det finns evidens för att HoH är nu. Men som goda bayesianer behöver vi använda denna evidens för att uppdatera a priori-sannolikheten att HoH är nu.

Denna a priori-sannolikhet är mycket låg, hävdar MacAskill med hänvisning till Nick Bostroms *Self-Sampling Assumption (SSA)*: Vår position i rum-tiden skall betraktas som vald på måfå bland alla observatörer – i det förflutna, nu och i framtiden.



MacAskill: *This principle seems compelling as a way of setting priors. The a priori probability that I am in the top 100 funniest people in Scotland today is 100 out of 5.4 million; the a priori probability that I am in the top 1000 strongest people in the UK today is 1000 out of 66.4 million.*

MacAskill: *This principle seems compelling as a way of setting priors. The a priori probability that I am in the top 100 funniest people in Scotland today is 100 out of 5.4 million; the a priori probability that I am in the top 1000 strongest people in the UK today is 1000 out of 66.4 million.*

[...]

For the purposes of my argument, what matters is not these precise numbers, but that any of them are astronomical. If there are a trillion trillion people to come, then the a priori probability that we are among the million most influential people ever is one in a million trillion. This is about the same probability as dealing a Royal Flush from a well-shuffled pack of cards three times in a row. But even if we assume that there are only a hundred trillion people to come, the a priori probability of being among the million most influential people ever is still one in a hundred million.

Idén är alltså att om framtiden är väldigt stor, som med Bostroms föreslagna 10^{16} människoliv, så blir a priori-sannolikheten att vi lever under HoH *nu* ytterst liten: cirka $10^{10}/10^{16} = 10^{-6}$.

Idén är alltså att om framtiden är väldigt stor, som med Bostroms föreslagna 10^{16} människoliv, så blir a priori-sannolikheten att vi lever under HoH *nu* ytterst liten: cirka $10^{10}/10^{16} = 10^{-6}$.

Med rymdkolonisering blir sannolikheten $10^{10}/10^{34} = 10^{-24}$, och med uppladdning $10^{10}/10^{54} = 10^{-44}$.

Idén är alltså att om framtiden är väldigt stor, som med Bostroms föreslagna 10^{16} människoliv, så blir a priori-sannolikheten att vi lever under HoH *nu* ytterst liten: cirka $10^{10}/10^{16} = 10^{-6}$.

Med rymdkolonisering blir sannolikheten $10^{10}/10^{34} = 10^{-24}$, och med uppladdning $10^{10}/10^{54} = 10^{-44}$.

Därför menar MacAskill att om vi vill hävda att HoH är nu, så gäller principen *extraordinära påståenden kräver extraordinära belägg*...

Idén är alltså att om framtiden är väldigt stor, som med Bostroms föreslagna 10^{16} människoliv, så blir a priori-sannolikheten att vi lever under HoH *nu* ytterst liten: cirka $10^{10}/10^{16} = 10^{-6}$.

Med rymdkolonisering blir sannolikheten $10^{10}/10^{34} = 10^{-24}$, och med uppladdning $10^{10}/10^{54} = 10^{-44}$.

Därför menar MacAskill att om vi vill hävda att HoH är nu, så gäller principen *extraordinära påståenden kräver extraordinära belägg*...

...och hans poäng är att beläggen för HoH inte är tillräckligt extraordinära.

Hans tillämpning av Bostroms SSA-antagande är diskutabel, av minst två skäl:

Hans tillämpning av Bostroms SSA-antagande är diskutabel, av minst två skäl:

- Inte alla tidpunkter i mänsklighetens historia behöver vara lika sannolika att vara HoH, eftersom det är ytterst rimligt att ansätta substantiell sannolikhet till att HoH inträffar mycket tidigt i mänsklighetens historia. I så fall fyller inte längre de enorma bostromska befolkningstalen 10^{16} och 10^{34} och 10^{54} den funktion MacAskill behöver för sitt argument.

Hans tillämpning av Bostroms SSA-antagande är diskutabel, av minst två skäl:

- Inte alla tidpunkter i mänsklighetens historia behöver vara lika sannolika att vara HoH, eftersom det är ytterst rimligt att ansätta substantiell sannolikhet till att HoH inträffar mycket tidigt i mänsklighetens historia. I så fall fyller inte längre de enorma bostromska befolkningstalen 10^{16} och 10^{34} och 10^{54} den funktion MacAskill behöver för sitt argument.
- Antropiska argument är fortfarande ett ganska outrett område, och varje slutsats av dem bör betraktas med skepsis, i synnerhet extrema slutsatser om att god evidens övertrumfas av en extrem a priori-fördelning.

Ytterligare tvivel rörande tålmodig longtermism:

Ytterligare tvivel rörande tålmodig longtermism:

- SSA-argumentet för låg apriori-sannolikhet för att leva under HoH kommer att vara tillämpligt även i framtiden. Om dagens evidens för HoH inte anses stark nog att övertrumfa denna låga prior är det tveksamt om tillräcklig evidens *någonsin* kommer att uppnås.

Ytterligare tvivel rörande tålmodig longtermism:

- SSA-argumentet för låg apriori-sannolikhet för att leva under HoH kommer att vara tillämpligt även i framtiden. Om dagens evidens för HoH inte anses stark nog att övertrumfa denna låga prior är det tveksamt om tillräcklig evidens *någonsin* kommer att uppnås.
- Är samhället tillräckligt stabilt för att investeringar på tusentals eller miljontals års sikt skall vara meningsfullt?

Ytterligare tvivel rörande tålmodig longtermism:

- SSA-argumentet för låg apriori-sannolikhet för att leva under HoH kommer att vara tillämpligt även i framtiden. Om dagens evidens för HoH inte anses stark nog att övertrumfa denna låga prior är det tveksamt om tillräcklig evidens *någonsin* kommer att uppnås.
- Är samhället tillräckligt stabilt för att investeringar på tusentals eller miljontals års sikt skall vara meningsfullt?
  - Vi har inte haft kapitalism mer än några hundra år.

Ytterligare tvivel rörande tålmodig longtermism:

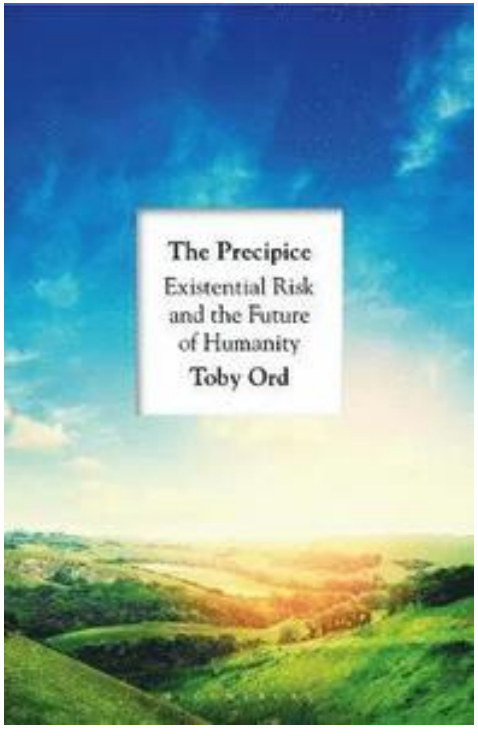
- SSA-argumentet för låg apriori-sannolikhet för att leva under HoH kommer att vara tillämpligt även i framtiden. Om dagens evidens för HoH inte anses stark nog att övertrumfa denna låga prior är det tveksamt om tillräcklig evidens *någonsin* kommer att uppnås.
- Är samhället tillräckligt stabilt för att investeringar på tusentals eller miljontals års sikt skall vara meningsfullt?
  - Vi har inte haft kapitalism mer än några hundra år.
  - Precis som etiken har förbättras sedan den barbariska tid då vi godtog slavhandel och enbart tillskrev vita män rättigheter, så kan den komma att förbättras i framtiden, kanske till en nivå där det blir överflödigt med medel i longtermistiska fonder.

Ytterligare tvivel rörande tålmodig longtermism:

- SSA-argumentet för låg apriori-sannolikhet för att leva under HoH kommer att vara tillämpligt även i framtiden. Om dagens evidens för HoH inte anses stark nog att övertrumfa denna låga prior är det tveksamt om tillräcklig evidens *någonsin* kommer att uppnås.
- Är samhället tillräckligt stabilt för att investeringar på tusentals eller miljontals års sikt skall vara meningsfullt?
  - Vi har inte haft kapitalism mer än några hundra år.
  - Precis som etiken har förbättras sedan den barbariska tid då vi godtog slavhandel och enbart tillskrev vita män rättigheter, så kan den komma att förbättras i framtiden, kanske till en nivå där det blir överflödigt med medel i longtermistiska fonder.
- Avtagande nytta. On HoH inträffar 1 000 000 år från idag, är det verkligen bättre att börja spara nu jämfört med att fixa nuvarande århundrades problem och sedan spara i 999 900 år?

Slutligen tycks det mig ofrånkomligt att HoH-frågan i hög grad kokar ned till om vi delar Toby Ords och andras bedömning att vi lever i en tid med mycket hög existentiell risk.

Slutligen tycks det mig ofrånkomligt att HoH-frågan i hög grad kokar ned till om vi delar Toby Ords och andras bedömning att vi lever i en tid med mycket hög existentiell risk.



Existential catastrophe via	*Chance within next 100 years*
Asteroid or comet impact	~ 1 in 1,000,000
Supervolcanic eruption	~ 1 in 10,000
Stellar explosion	~ 1 in 1,000,000,000
Total natural risk	**~ 1 in 10,000**
Nuclear war	~ 1 in 1,000
Climate change	~ 1 in 1,000
Other environmental damage	~ 1 in 1,000
'Naturally' arising pandemics	~ 1 in 10,000
Engineered pandemics	~ 1 in 30
Unaligned artificial intelligence	~ 1 in 10
Unforeseen anthropogenic risks	~ 1 in 30
Other anthropogenic risks	~ 1 in 50
Total anthropogenic risk	**~ 1 in 6**
Total existential risk	**~ 1 in 6**

TABLE 6.1 My best estimates for the chance of an existential catastrophe from each of these sources occurring at some point in the next 100 years (when the catastrophe has delayed effects, like climate change, I'm talking about the point of no return coming within 100 years). There is significant uncertainty remaining in these estimates and they should be treated as representing the right order of magnitude—each could easily be a factor of 3 higher or lower. Note that the numbers don't quite add up: both because doing so would create a false feeling of precision and for subtle reasons covered in the section on 'Combining Risks'.