

# MVE420: Nya teknologier, global risk och mänsklighetens framtid

<http://www.math.chalmers.se/Math/Grundutb/CTH/mve420/1819>

Etik och framtida generationer

2 april 2019

Olle Häggström

Ur kurshemsidans instruktioner för era projekt:

“Rapporten bör också ta upp hur vi (mänskligheten) kan agera för att ge bästa chanser till ett positivt utfall, och vilka **moralfilosofiska idéer** som (utifrån analysen till Deadline P2) ligger till grund för dessa överväganden. Detta kan betyda att lyfta några relevanta värderingsfrågor som aktualiseras av de nya teknologierna. Det bör genomgående framgå vilka påståenden och argument som kommer från litteraturen, och vilka som är studenternas egna.”

Men vad har etik och moralfilosofi med teknikutveckling att göra?

Men vad har etik och moralfilosofi med teknikutveckling att göra?

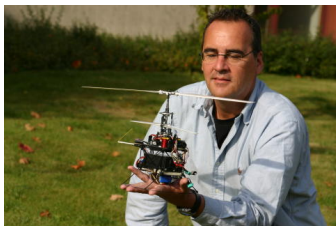
Jag börjar med ett konkret exempel.

# Opposition to autonomous warfare swells to 16,000 signatories

Artificial intelligence community comes together in unprecedented numbers to call for a ban on AI-controlled weaponry



**Ur brevet:** *"If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. It will only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and selectively killing a particular ethnic group. We therefore believe that a military AI arms race would not be beneficial for humanity."*



Patrick Doherty, professor i datavetenskap vid Linköpings universitet, i augusti 2015:

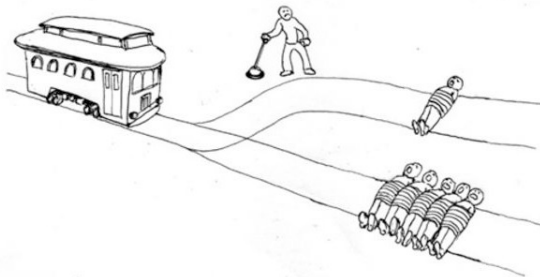
*"Jag har inte skrivit under det öppna brevet, ty det finns inga onda eller goda teknologier, det finns bara onda och goda användningar."*

Patrick Doherty och jag tycker olika i denna etiska fråga.

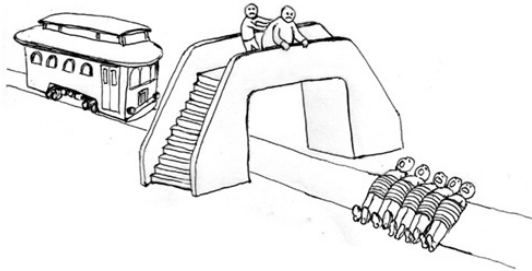
Det akademiska ämnet *moralfilosofi* (även kallat *etik*) finns till för att ge struktur åt etiska diskussioner och tillåta rationellt grundade resonemang kring dem, istället för att låta det hela förfalla till ett allmänt tyckande.



Populär gren av moralfilosofin på senare år: **trolleyologin**.



Känd variant på trolleyproblemet: **fat man**.





**Karim Jebari** @KarimJebari · 4h

När man undervisar i filosofi är det alltid någon student som säger något i stil med "men om jag sliter loss spaken och slänger ned den på spåret kanske jag kan stoppa tåget". Det är väldigt tacksamt, för då kan du förklara att det är att missa poängen med ett tankeexperiment.

[Translate Tweet](#)

1 1



**Karim Jebari** @KarimJebari · 4h

Syftet med experimentet är inte att diskutera de bästa sätten att rädda folk från tåg, utan att fundera på de etiska intuitionerna som får oss att tveka (eller inte).

[Translate Tweet](#)

1 1



**Karim Jebari** @KarimJebari · 4h

Med andra ord: tankeexperiment är teoretiska redskap för att resonera kring abstrakta principer. De är inte scenarion för att planera inför en viss situation eller träna för att lösa praktiska problem.

[Translate Tweet](#)

1 2



**Karim Jebari** @KarimJebari · 4h

Allt detta är enormt okontroversiellt. Det som dock är märkligt för mig, är att så fort filosofer (och andra) ska börja ägna sig åt tillämpad etik, så glömmer de distinktionen mellan ett tankeexperiment och ett scenario.

[Translate Tweet](#)

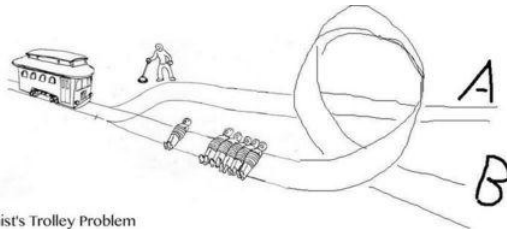
1 5



**Karim Jebari** @KarimJebari · 4h

Detta är särskilt påtagligt i diskussionen om självkörande bilar. Det har uppstått en hel litteratur som handlar om att postulera att trolley-problemet och andra tankeexperiment faktiskt handlar om verkligheten.

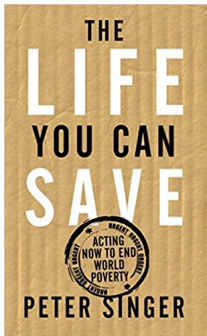
En mindre känd trolley-variant:

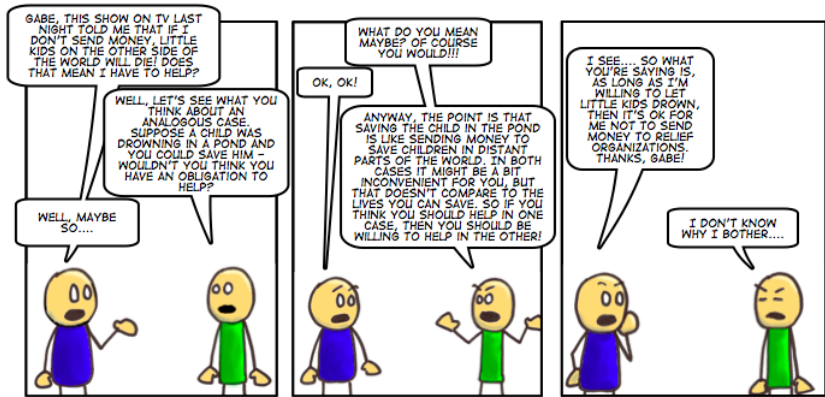


Hedonist's Trolley Problem

The track is heading towards B.

If you pull the lever, it will switch to A but it won't do the totally sick loop-da-loop.





Tre olika etiska teorier:

**1. Konsekventialism:** en handling bedöms efter dess *konsekvenser*.

**2. Deontologisk etik:** en handling bedöms efter dess *inneboende värde*.

**3. Dygdetik:** en handling bedöms utifrån *vad en dygdig person skulle ha gjort*.

Tre olika etiska teorier:

**1. Konsekventialism:** en handling bedöms efter dess *konsekvenser*.

**1a. Utilitarism** förordar de handlingar som leder till mest *nytta* i världen.

**2. Deontologisk etik:** en handling bedöms efter dess *inneboende värde*.

**3. Dygdetik:** en handling bedöms utifrån *vad en dygdig person skulle ha gjort*.



Tre olika etiska teorier:

**1. Konsekventialism:** en handling bedöms efter dess *konsekvenser*.

**1a. Utilitarism** förordar de handlingar som leder till mest *nytta* i världen, där nytta t.ex. kan likställas med lycka (hedonisk utilitarism) eller uppfyllda preferenser (preferensutilitarism).

**2. Deontologisk etik:** en handling bedöms efter dess *inneboende värde*.

**3. Dygdetik:** en handling bedöms utifrån *vad en dygdig person skulle ha gjort*.

Tre olika etiska teorier:

**1. Konsekventialism:** en handling bedöms efter dess *konsekvenser*.

**1a. Utilitarism** förordar de handlingar som leder till mest *nytta* i världen, där nytta t.ex. kan likställas med lycka (hedonisk utilitarism) eller uppfyllda preferenser (preferensutilitarism).

**2. Deontologisk etik:** en handling bedöms efter dess *inneboende värde*.

**2a. Pliktetik**

**2b. Rättighetsetik**

**3. Dygdetik:** en handling bedöms utifrån *vad en dygdig person skulle ha gjort*.

En annan indelning av utilitaristiska moralteorier: oavsett val av nyttobegrepp kan vi skilja mellan...

- ▶ **handlingsutilitarism**, och
- ▶ **regelutilitarism** (låt samhället styras av *de regler som leder till störst nytta*).

Oavsett val av etisk teori gör vi klokt i att hålla isär **faktapåståenden** och **värdepåståenden**.

Oavsett val av etisk teori gör vi klokt i att hålla isär **faktapåståenden** och **värdepåståenden**.



**Humes lag:** Det går inte att härleda värdepåståenden från enbart faktapåståenden.

Exempel:

**Premiss.** Arsenik är ett dödligt gift.

Exempel:

**Premiss.** Arsenik är ett dödligt gift.

**Slutsats.** Det är fel att lägga arsenik i sina middagsgästers mat.

Exempel:

**Premiss.** Arsenik är ett dödligt gift.

**Slutsats.** Det är fel att lägga arsenik i sina middagsgästers mat.

Om vi accepterar Humes lag är detta en felaktig slutledning.



Nytt försök:

**Premiss 1:** Arsenik är ett dödligt gift.

Nytt försök:

**Premiss 1:** Arsenik är ett dödligt gift.

**Premiss 2:** Du skall icke dräpa.

Nytt försök:

**Premiss 1:** Arsenik är ett dödligt gift.

**Premiss 2:** Du skall icke dräpa.

**Slutsats:** Det är fel att lägga arsenik i sina middagsgästers mat.

Vetenskap

# Maria Gunther: Låt vetenskapen gå före politiken

PUBLICERAD 2016-12-27

**KOMMENTAR.** 2016 har varit ett extremt år, både för klimatet och i världspolitiken. Men vi måste till varje pris komma ifrån att klimatet blivit en politisk fråga.



Maria Gunther >

+ Följ skribent

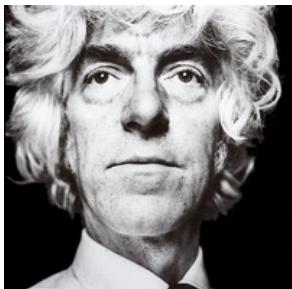




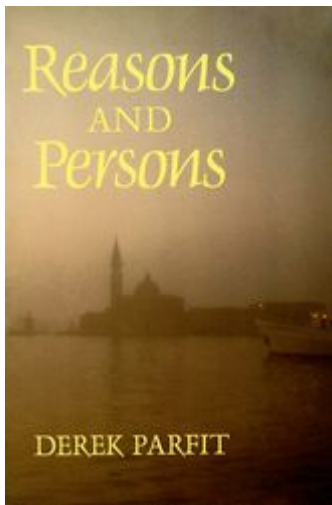
Christian Azar protesterar, i DN den 21 april 2017:

Det påstås att klimatfrågan blivit så viktig att vetenskapen bör gå före politiken. Det vore ett misstag. [...]

Vetenskaplig kunskap måste naturligtvis spela en mycket viktig roll när vi utformar vår politik om ett så komplext område som klimatet. Men om vi skulle överlämna klimatfrågan från politiken till vetenskapen, och låta forskare bestämma mål för klimatpolitiken, ja då skulle vi göra våld på vad som är vetenskapens uppgift och vi skulle kraftigt inskränka demokratin.



Derek Parfit (1942-2017)



## Parfit (1984) säger...

I believe that if we destroy mankind, as we now can, this outcome will be *much* worse than most people think. Compare three outcomes:

(1) Peace.

(2) A nuclear war that kills 99% of the world's existing population.

(3) A nuclear war that kills 100%.

(2) would be worse than (1), and (3) would be worse than (2). Which is the greater of these two differences?

## Parfit (1984) säger...

I believe that if we destroy mankind, as we now can, this outcome will be *much* worse than most people think. Compare three outcomes:

- (1) Peace.
- (2) A nuclear war that kills 99% of the world's existing population.
- (3) A nuclear war that kills 100%.

(2) would be worse than (1), and (3) would be worse than (2). Which is the greater of these two differences?

Most people believe that the greater difference is between (1) and (2). I believe that the difference between (2) and (3) is very much greater.



Parfits poäng är att om (3) inträffar går vi miste om inte bara de 7,6 miljarder människor som lever idag, utan även *alla framtida generationer*.

Parfits poäng är att om (3) inträffar går vi miste om inte bara de 7,6 miljarder människor som lever idag, utan även *alla framtida generationer*.

De kan bli många människor. Om vi antar att...

- ▶ mänskligheten överlever i en miljard år,
- ▶ en befolkning om 10 miljarder människor kan upprätthållas hållbart,
- ▶ en medellivslängd på 100 år,

så blir antalet mänskoliv som går förlorade

$$10^9 10^{10} 10^{-2} = 10^{17}.$$

Parfits poäng är att om (3) inträffar går vi miste om inte bara de 7,6 miljarder människor som lever idag, utan även *alla framtida generationer*.

De kan bli många människor. Om vi antar att...

- ▶ mänskligheten överlever i en miljard år,
- ▶ en befolkning om 10 miljarder människor kan upprätthållas hållbart,
- ▶ en medellivslängd på 100 år,

så blir antalet mänskoliv som går förlorade

$$10^9 10^{10} 10^{-2} = 10^{17}.$$

Nick Bostrom föreslår, i sin uppsats *Existential risk prevention as a global priority* från 2013, ännu mycket högre siffror baserat på rymdkolonisering och/eller uppladdning av mänskligt medvetande på datorer.

Dessa slags (utilitaristiskt anstrukna) överväganden dominerade de etiska diskussionerna vid 2017 års GoCAS-gästforskarprogram *Existential Risk to Humanity*, tills...

Dessa slags (utilitaristiskt anstrukna) överväganden dominerade de etiska diskussionerna vid 2017 års GoCAS-gästforskarprogram *Existential Risk to Humanity*, tills...

...Karin Kuhlemann dök upp och förordade en mer rättighetsbaserad etik, där rättigheter tillfaller *verkliga* människor och inte *hypotetiska*. Hon menade att det utilitaristiska perspektivet reducerar människor till "*containers for utility*".



Till **Deadline P2 (10 april)** behöver ni inkomma med en kort (en halv till en sida) preliminär idé om hur de etiska aspekterna på det valda ämnesområdet för ert projekt skall hanteras, strukturerat (i huvudsak) enligt något av följande två scheman.

## Praktisk metod för etisk analys: Hanssons förslag

1. Identifiera intressenter och deras roller, och beskriv situationen ur deras synvinklar.
2. Undersök noga vilka faktiska förhållanden som påverkar den etiska bedömningen, samt precisera vilken information som saknas.
3. Jämför med hur liknande fall skulle ha bedömts.
4. Undersök problemet med flera olika moralteoriers hjälp. Leder de olika teorierna till samma rekommendation?
5. Försök konstruera nya handlingsalternativ.
6. Undersök noga om förslagna lösningar är praktiskt genomförbara.

## Praktisk metod för etisk analys: van de Poel och Royackers förslag

1. Etisk problemformulering. Vilka värden kan komma i konflikt?
2. Problemanalys, med beaktande av (a) vilka intressenter som finns, (b) relevanta värden, och (c) relevanta fakta.
3. Identifiera handlingsalternativ.
4. Värdera de olika handlingsalternativen på basis av olika moralteorier.
5. Reflektera (*“differen ethical frameworks... do not necessarily lead to the same conclusion... intuitive judgement and common sense are also important”*) och gå tillbaka till steg 3 om det behövs.