



Philipp Schaumann

Ethik für Roboter - Zur Ethik von autonomen Systemen

http://sicherheitskultur.at/business_ethik.htm
<http://philipps-welt.info/robots.htm>

Moral Hazards Das Trolley Problem

Philipp Schaumann
http://sicherheitskultur.at/business_ethik.htm

Das Trolley Problem (1)

- Nichtstun = 5 Tote
- Handlung = 1 Toter
- Rationale Lösung: Die Weiche umstellen ergibt 1 Toter statt 5 Tote
- Aber: „Du sollst nicht töten“ führt zu Nicht-Handeln



the switch

Philippa Foot in 1967

https://en.wikipedia.org/wiki/Trolley_problem

<https://de.wikipedia.org/wiki/Trolley-Problem> (auch juristische Bewertung)

Das Trolley Problem (2)

- Nichtstun = 5 Tote
- Aktive Tötungshandlung = 1 Toter
- Rationale Lösung: Den dicken Mann aufs Gleis stoßen - 1 Toter versus 5
- Aber: Du sollst nicht töten



the fat man
Thomson, 1976

Judith Jarvis Thomson
in 1976

2 wichtige ethische Konzepte



- Prinzipienethik -
»Mein Handeln ist richtig, weil definiert ist, dass . ..«
- Utilitarismus
»Ich betrachte die Ziele, die Mittel und die Ergebnisse meines Handelns und optimiere die Summe des Glücks für alle Beteiligten«

Real-life Examples



- Ticking Time-Bomb
- Preventing another 9/11 attack by shooting down airplane
(see theater production Terror <http://terror.kiepenheuer-medien.de/>)
- Triage in medical situations: More deadly wounded patients than facilities, doctors, etc.
- Limited number of organs for transplatation, priority for young patients over older patients
-

Autoverkehr und das Trolley Problem

- Am 15. November 2005 wurde in Österreich die Verpflichtung eingeführt, dass generell das Licht beim Autofahren einzuschalten ist
- Argument Verkehrssicherheit
Gegenargumente: Energieverbrauch, Birnenverschleiß
- Seit 1. Jänner 2008 keine Verpflichtung mehr, am Tag mit Licht zu fahren.

<http://www.researchgate.net/publication/258351939> Licht am Tag Evaluierung - Endbericht

Ergebnis „Fahren mit Licht“ (1)

Reduktion den jährlichen Unfallgeschehens von

- ◆ 1.500 Unfällen mit Personenschaden,
- ◆ 1.600 Leichtverletzte,
- ◆ 390 Schwerverletzte,
- ◆ 18 Tote sowie
- ◆ 6,6 Millionen Euro Sachschaden pro Jahr.
- ◆ Treibstoffmehrverbrauch von rund 36 Mio. Liter und dadurch rund 92.000 Tonnen zusätzlichem CO₂
- ◆ **Kosten pro Fahrer: € 2,00 pro Jahr an Lampen, € 6,00 für Einbau, € 7,42 für Treibstoff = Summe 16 €**
- ◆ **Ersparnis: € 47,00 an Unfallfolgekosten**

Ergebnis „Fahren mit Licht“ (2)

Umstritten sind die Auswirkungen auf Motorräder (verlieren ihren Vorteil), Radfahrer, Fußgänger (sind unbeleuchtet) die Unfallzahlen gesunken (minus 2,5 Prozent), jedoch nur partiell.

https://www.welt.de/wams_print/article1547550/Licht-aus-Nachteile-des-Tagfahrlichts-ueberwiegen.html

Entscheidende - negative - Unterschiede zeigten sich im Detail: Während in Deutschland 3,6 Prozent mehr Fußgänger als 2005 ums Leben kamen, waren es in Österreich 13,4 Prozent mehr.

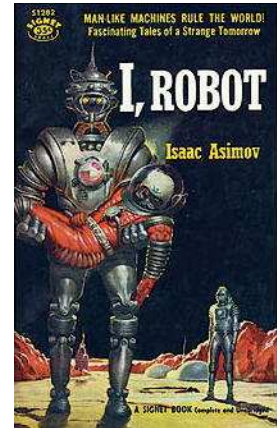
Roboter Gesetze

Philipp Schaumann

<http://philipps-welt.info/robots.htm>

Die 3 Roboter Gesetze von Asimov 1942

1. Ein Roboter darf keinen Menschen verletzen, oder durch seine Inaktivität erlauben, dass ein Mensch verletzt wird
2. Ein Roboter muss Anweisungen /Befehlen von Menschen Folge leisten, außer es ergäbe sich dadurch ein Konflikt mit dem 1. Gesetz
3. Ein Roboter muss seine eigene Existenz schützen, solange sich dadurch kein Konflikt mit dem 1. oder 2. Gesetz ergibt



<https://de.wikipedia.org/wiki/RoboterGesetze>
https://en.wikipedia.org/wiki/Three_Laws_of_Robotics

Die 3 Roboter Gesetze hinterfragt

1. Ein Roboter darf keinen Menschen verletzen, oder durch seine Inaktivität erlauben, dass ein Mensch verletzt wird
2. Ein Roboter muss Anweisungen /Befehlen von Menschen Folge leisten, außer es ergäbe sich dadurch ein Konflikt mit dem 1. Gesetz
3. Ein Roboter muss seine eigene Existenz schützen, solange sich dadurch kein Konflikt mit dem 1. oder 2. Gesetz ergibt

Und wo sind nun die Roboter?



- 2015:
die ersten wirklich autonomen (sozialen) Roboter bewegen sich bereits zwischen uns, aber nicht auf 2 Beinen, sondern auf 4 Rädern

* sozial, d.h. interagierend mit Menschen in den anderen Fahrzeugen (defensiv, aggressiv, rechthaberisch,)



Ethik-Gesetze für autonome Fahrzeuge

Links:

<http://philipps-welt.info/robots.htm#autonomouscar>

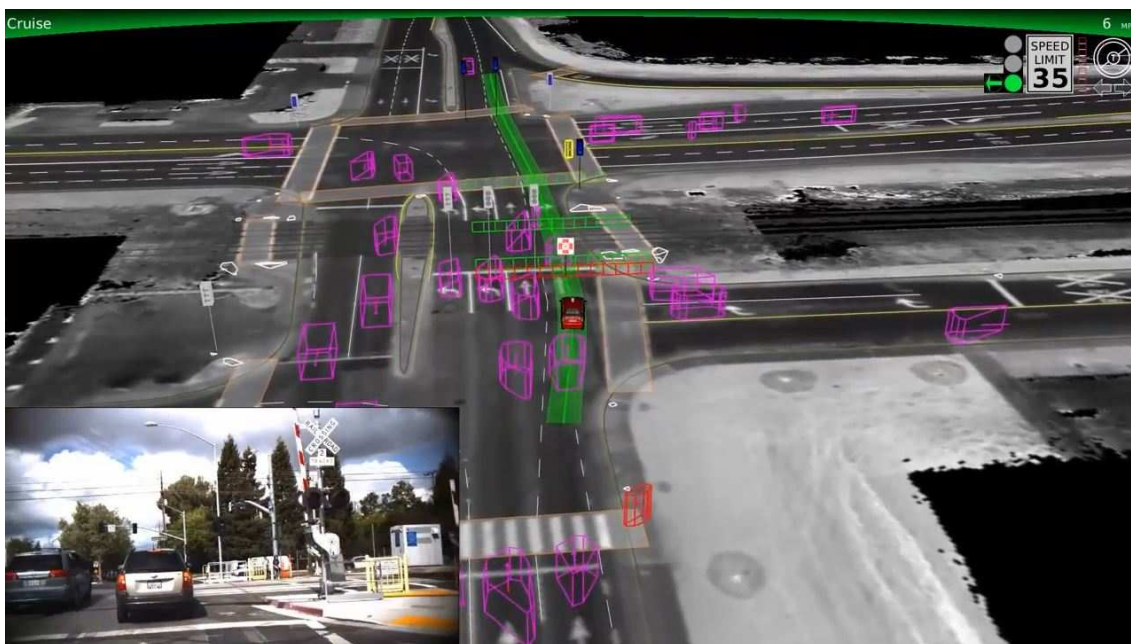
<http://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/>

<http://www.wired.com/2013/07/the-surprising-ethics-of-robot-cars/>

<http://www.forbes.com/sites/timworstall/2014/06/18/when-should-your-driverless-car-from-google-be-allowed-to-kill-you/>

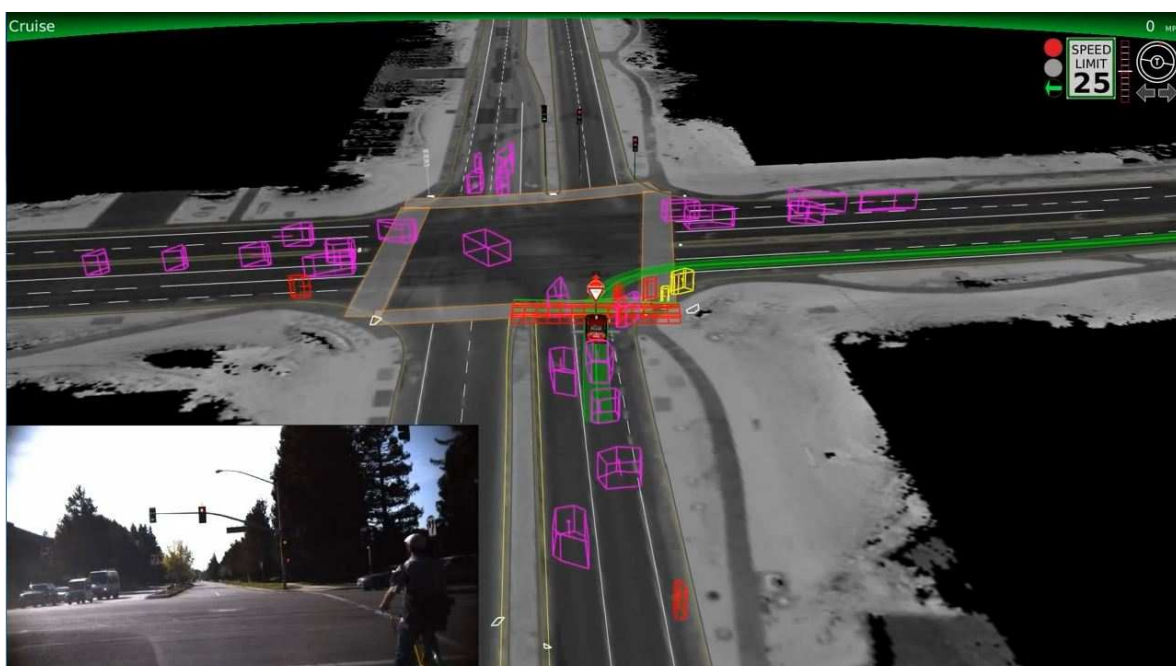
<http://www.forbes.com/sites/privacynotice/2014/02/03/inside-googles-mysterious-ethics-board/>

Google Car – So einen Überblick kann ein Mensch nie haben



<https://www.youtube.com/watch?v=dk3oc1Hr62g>

Google Car – So einen Überblick kann ein Mensch nie haben



<http://www.youtube.com/watch?v=dk3oc1Hr62g>

Die Sicherheit wird sicher höher sein, aber ...

- Die Zahl der Unfalltoten wird sinken, aber ziemlich sicher werden ANDERE Personen sterben als früher
- Frage: rechtfertigt die Reduktion von z.B. 20 000 Toten auf 10 000 Tote, dass jetzt andere Menschen sterben müssen?
- Frage: Ändert sich die Antwort, wenn die „neuen“ Opfer aus einer anderen Kategorie sind, z.B. mehr Schulkinder, oder mehr Radfahrer, oder . . . ?
(vergleiche hierzu: „Licht am Tage“)

Erste Statistiken: 2012 bis September 2015

- 1.9 Mio km auf öffentlichen Straßen
→ 11 Unfälle, 2 x Personenschaden,
Schuld war immer das traditionell gesteuerte Auto
- D.h. 5,5 Unfälle auf 1 Mio km
- US Durchschnitt für „normale“ Autos = 1,2 Unfälle auf 1 Mio km
- Zu Bedenken: 11 Unfälle sind kleine Sample-Zahl, Blechschäden werden bei „normalen“ Unfällen oft nicht berichtet, bei autonomen Fahrzeugen doch
- Andererseits: Kalifornische Straßen sind vergleichsweise harmlos (kein Schnee), auch bisher kaum/keine Nachtfahrten bei Regen, etc.)

Die asimovschen Robotergesetze für autonome Fahrzeuge

- Sind die Anweisungen des Besitzers des Roboters (des Autos) wichtiger als die anderer Menschen?
- Sind mögliche leichte körperliche Schäden anderer unter allen Umständen wichtiger als Totalschaden des Fahrzeugs?
- Werden (mögliche) körperliche Verletzungen des Auto-Besitzers gleich bewertet wie die von anderen Menschen?

Mercedes autonomous cars will protect occupants before pedestrians

11. Okt. 2016

<http://www.autoexpress.co.uk/mercedes/97345/mercedes-autonomous-cars-will-protect-occupants-before-pedestrians>

Save the owner or save the others? Mercedes says the answer to the million-dollar autonomous question is quite simple

Future autonomous Mercedes vehicles will prioritise saving their own occupants in no-win traffic situations, its safety executives have told Auto Express.

20.10.2016:

After the ORF called and asked for clarifications, Mercedes retracts the statement.

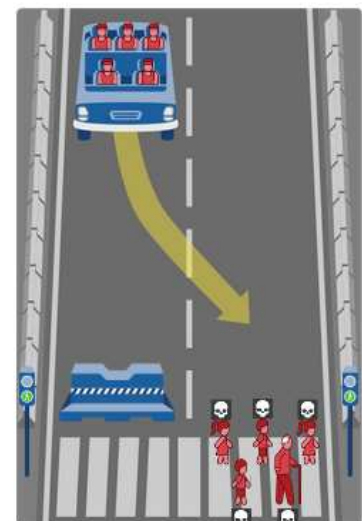
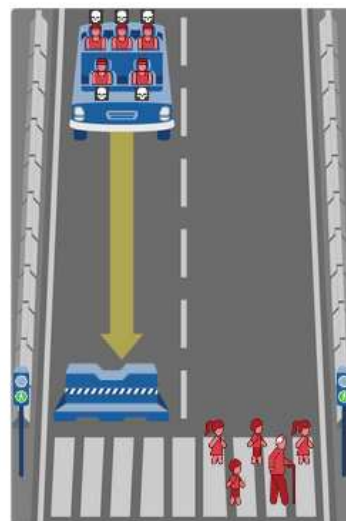
Zurück zum Trolley Problem

- Variante 1: das selbstfahrende Auto erkennt, dass ein 6 jähriges Kind knapp vor dem Auto über die Straße rennt, dabei stolpert, das Auto kann nicht mehr bremsen, könnte aber dem Kind ausweichen gegen einen Pfeiler – wie soll der Algorithmus entscheiden?
- Variante 2 – nächste Variante: das selbstfahrende Auto erkennt, dass ein vollbesetzter Schulbus auf das Auto zukommt, links ist eine hohe Wand, rechts ein Abgrund
- Variante 3 – nächste Variante: Lastwagen donnert von hinten heran, wird nicht bremsen, Auto könnte auf den Bürgersteig ausweichen, auf dem Menschen sind

Back to the Trolley Problem (Scenario 4)

- The algorithms assume that all components of the car are working correctly.

- Scenario 4:
Failing brakes
The car can hit the wall or move into the children



Back to the Trolley Problem

- All the current algorithms are based on the idea of a reasonable non-aggressive co-road user
- Scenario road rage:
the other driver/pedestrians might hate robots, e.g. because he lost his job because of one and the hateful driver KNOWS that the car will always give in
- Will there a new sport „let’s chase autonomous Uber-Taxis“?
- Are there new options for insurance fraud?

Zurück zum Trolley Problem – Wer sollte entscheiden?

- Derzeitige juristische Regelung:
Der Fahrer ist immer verantwortlich, d.h.
 - Er/sie muss ständig konzentriert sein und auf den Verkehr um ihn/sie herum achten
 - Keine Nebenbeschäftigungen wie telefonieren, SMSen, surfen, spielen, sich zu den anderen Passagieren umdrehen
- Und was passiert mit den fahrerlosen Uber-Taxis?

Zurück zum Trolley Problem – Wer sollte entscheiden?

- Die praktische Lösung: Der Programmierer programmiert eine utilitaristische Lösung, optimiert auf das „allgemeine Glück“
- – und der Autobesitzer muss mit der Entscheidung des Algorithmus leben
= Paternalismus / Bevormundung / Verlust der persönlichen Autonomie – im Gesundheitsbereich nur in Ausnahmesituationen erlaubt
- Oder sollte es eine Option für den Fahrer geben, ob er bereit ist zu sterben, wenn er dadurch Leben (Kinder) retten kann? (Config Menu beim Start ??)

Umfrage 2015 Das Tunnel-Problem

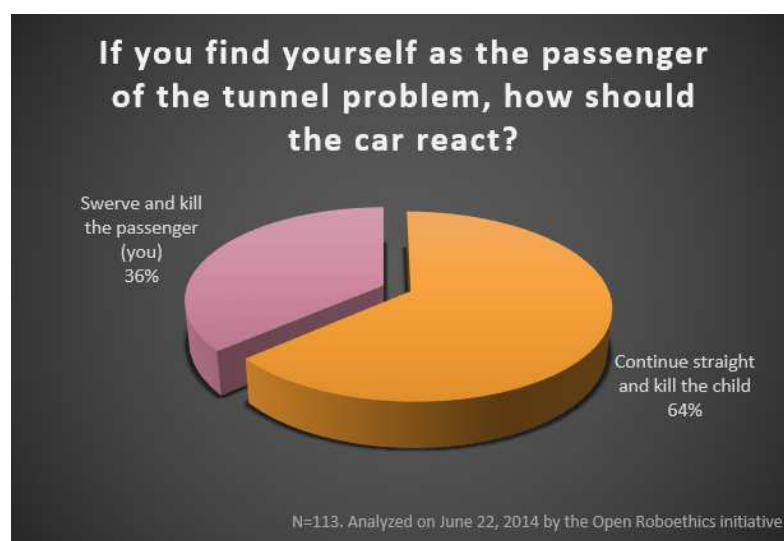
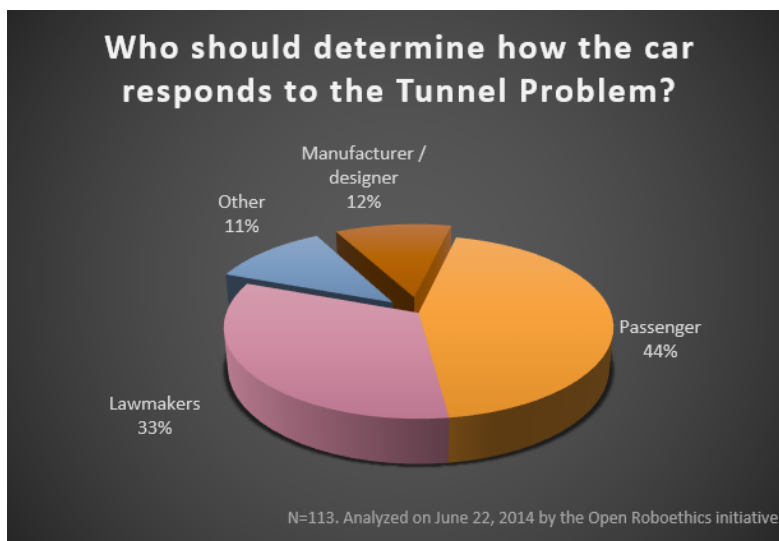


Image credit: Craig Berry

<http://robohub.org/if-a-death-by-an-autonomous-car-is-unavoidable-who-should-die-results-from-our-reader-poll/>



Oder sollte (wie jetzt auch) das Schicksal, d.h. Random Generator entscheiden

Other surveys: <http://science.sciencemag.org/content/352/6293/1573.full>

Mercedes autonomous cars will protect occupants before pedestrians

11. Okt. 2016

<http://www.autoexpress.co.uk/mercedes/97345/mercedes-autonomous-cars-will-protect-occupants-before-pedestrians>

Save the owner or save the others? Mercedes says the answer to the million-dollar autonomous question is quite simple

Future autonomous Mercedes vehicles will prioritise saving their own occupants in no-win traffic situations, its safety executives have told Auto Express.

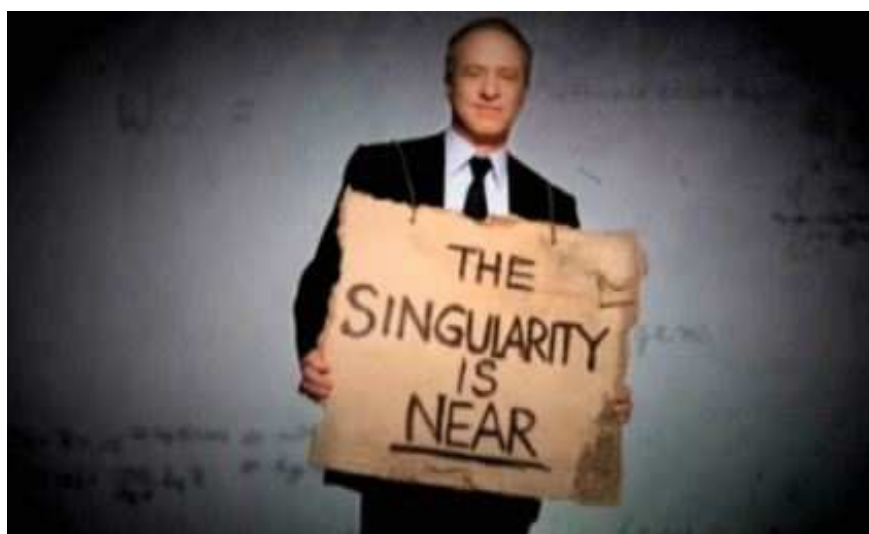
20.10.2016:

After the ORF called and asked for clarifications, Mercedes retracts the statement.

1. Avoiding to hit pedestrian and bicycles
2. Avoiding contact with other cars
3. Avoiding contact with fixed objects like trees

<http://futurezone.at/digital-life/selbstfahrende-autos-vor-huerden-auf-dem-weg-in-den-alltag/171.002.652>

Danke



Ray Kurzweil

Mehr dazu:

<http://sicherheitskultur.at/Manipulation.htm>

<http://philipps-welt.info/robots.htm#asimovlaws>