Ethics and Autonomous Vehicles

CI453

1

Thoughts about autonomous vehicles

- The tram thought experiment or ‘trolley problem’
- The autonomous vehicle ‘tunnel problem’
- Autonomous vehicles in reality
- The tram thought experiment revisited
- Some suggestions

2

Photograph by Joe Thompson

A runaway tram…

- Philippa Foot, 1967, Oxford Review 5
- "…may rather be supposed that he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one on the other; anyone on the track he enters is bound to be killed." (p.6)

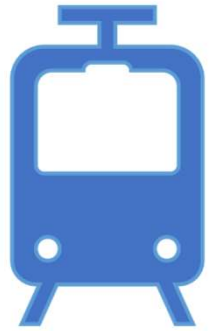
Original artwork by Fran Polanski

A runaway tram..

Foot's thought experiment was taken up many years later -

JJ Thompson 1985 'The Trolley Problem', The Yale Law Journal, Vol. 94, No. 6 (May, 1985), pp. 1395-1415

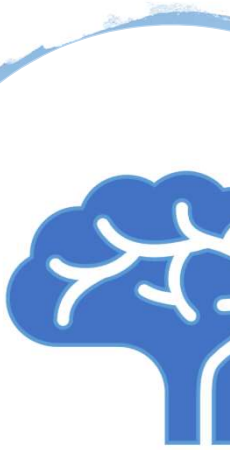
Thompson introduced 'the bystander problem' which is the version now more usually discussed.



A runaway tram..

- Now – 'Trolleyology' - and very important in both psychology and philosophy.

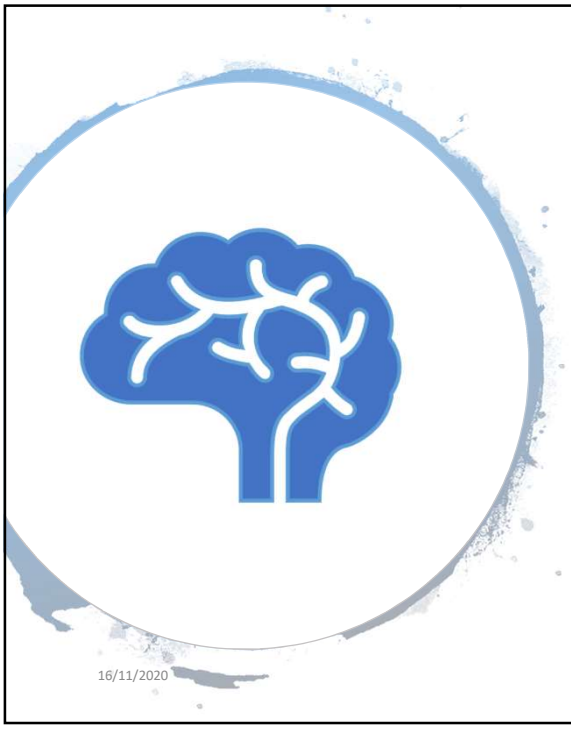
7



A runaway tram..

- Now – 'Trolleyology' - and important in both psychology and philosophy.
- Thompson's 'bystander variant' is the one most commonly discussed but there are at least 7 'standard variations' and countless 'fringe variations' .

8



A runaway tram..

- Now – 'Trolleyology' - and important in both psychology and philosophy.
- Thompson's 'bystander variant' is the one most commonly discussed but there are at least 7 'standard variations' and countless 'fringe variations' .
- I shall stick with Foot's original 'tram problem'

9

The 'tunnel problem'

A RECENT VARIATION INVOLVING AUTONOMOUS VEHICLES.

UNDERGOING THE SAME SORT OF GROWTH AND DEVELOPMENT AS THE ORIGINAL TRAM PROBLEM…

10

The 'tunnel problem'

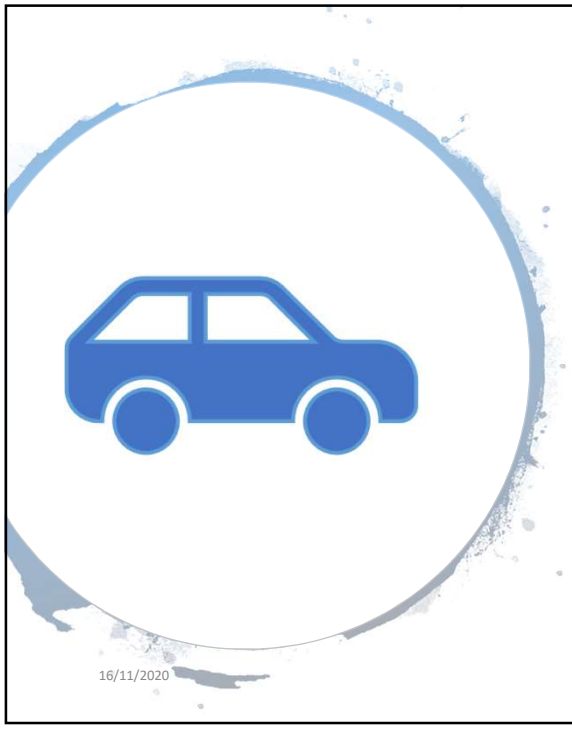
- Jason Millar, 2014, An ethical dilemma: When robot cars must kill, who should pick the victim?, Robohub
- "The car has but two options: hit and kill the child, or swerve into the wall on either side of the tunnel, thus killing you. How should the car react?"

Aspects of the 'tunnel problem'

WHAT IS THE BEST, OR LEAST WORST, OUTCOME? (MILLAR THINKS THERE IS 'NO RIGHT ANSWER')

WHO SHOULD DECIDE?

HOW?



The 'tunnel problem'

"Just because designers hold the technical abilities to engineer autonomous cars does not give them the authority to impose particular moral decisions on all users."

Millar J., 2014, op cit

Autonomous vehicles in the real world

- Please note: just as there is no horse in a horseless carriage, *so there is no human 'self'* in a self-driving car.

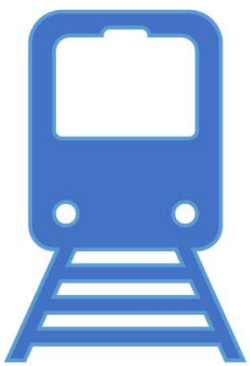
Lessons from the railways

- Autonomous vehicles have a long history.
- In many (maybe most) cases automation was introduced in order to reduce the number of accidents.
- Reduction > 90% has been achieved.
- 5 Grades of automation:

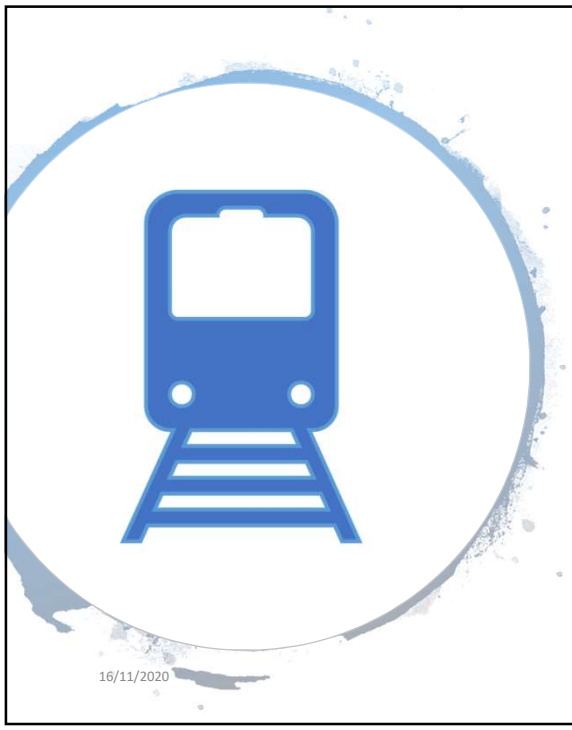
15

Some abreviations …..

- Automatic train operation (ATO)
- Grades of Automation (GoA)
- Automatic Train Control (ATC)
- Automatic Train Protection (ATP)
- Light Detection and Ranging (LIDAR)
- Artificial Neural Networks (ANN)



16



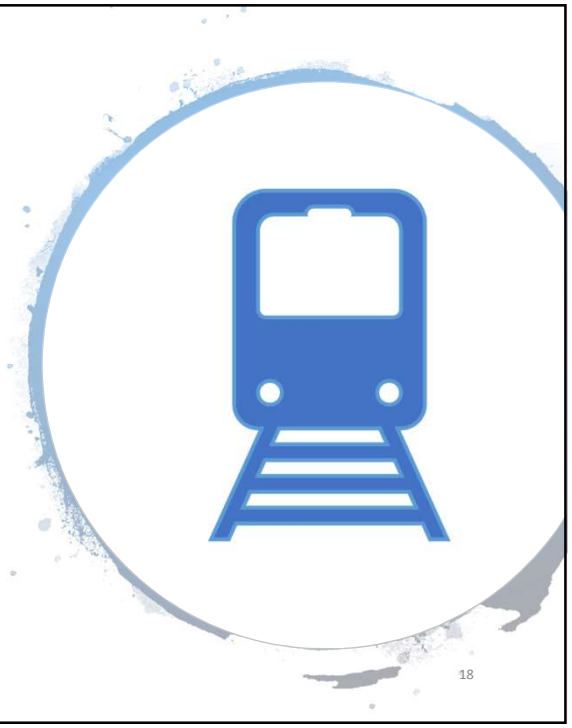
5 Grades of ATO

- GoA 0 is fully manual – visual avoidance of traffic
- GoA 1 is manual train operation with automated signaling (ATC/ATP)
- GoA 2 starting and stopping are automated but a driver in the cab starts the train
- GoA 3 is driverless train operation
- GoA 4 is unattended train operation

17

5 Grades of ATO

- There's no reason why the 5 GoAs in the railway case should not map on to cars.
- However some differences should be mentioned.
- Without a 'block system' automatic car stops are not so attractive.
- Existing and established technology does much of what an autonomous car would have to do



18

Existing and established technology

- For more than 10 years cars have been sold with various forms of autonomous cruise control – which, as well controlling speed, theoretically maintain a safe distance from vehicles in front.
- Many, many different proprietary systems with many different names.

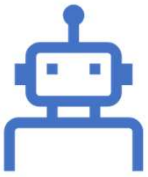
Existing and established technology

NEXT GENERATIONS OF ACCS WILL BE 'CO-OPERATIVE': TAKING DATA FROM ADJACENT VEHICLES.

THERE IS A CURRENT DEBATE ABOUT *HOW EXACTLY* TO TAKE THIS DATA FROM NEARBY VEHICLES

Industry classification of autonomous vehicles

1) 'Autonomous cars' – which take over from a human driver under certain circumstances



2) 'Self-driving cars' which don't require a human driver

21

Industry classification of autonomous vehicles

- The reasons for this classification – and the insistence on choosing their own terminology - are probably much more social, legal, and marketing-related than technical.
- Industrial players feel that governments and/or international bodies need to act now.

22

‘The Google Self-Driving Car’

Scheduled for general sale in 2017-20.

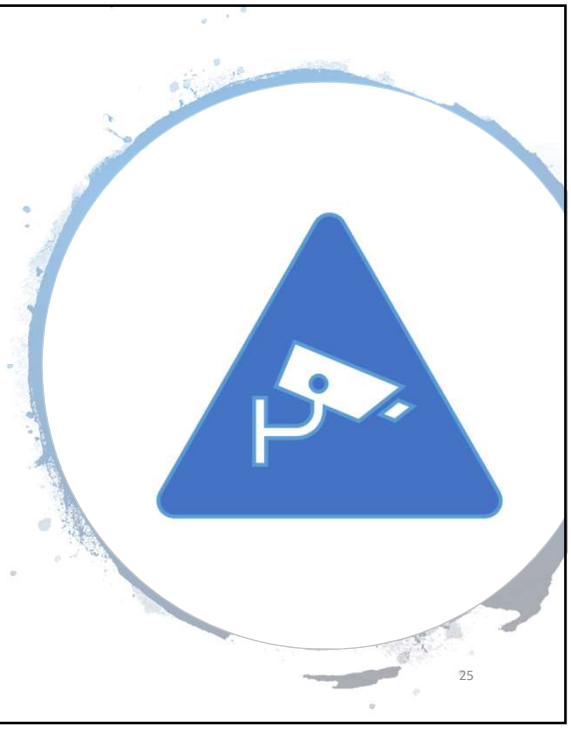
Uses an off the shelf Lidar and pre-programmed high accuracy maps.

‘The Google Self-Driving Car’

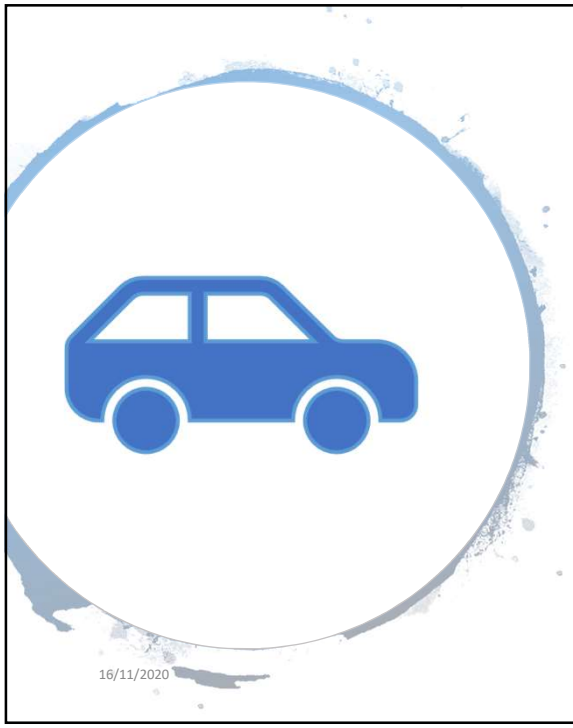
- Lidar is very poor in fog or mist.
- Can’t (at present) respond to temporary traffic lights or being ‘flagged down’ by the police.
- These deficiencies may be fixed before release but ‘failsafe’ driving is current technological practice.

Current technology and the 'tunnel problem'

- Sensors only detect *obstructions* – they cannot distinguish a child.
- There is no known, or foreseeable sensor technology which could determine the *relative moral worth* of any obstruction.
- Are humans any different?



25



Current technology and the 'tunnel problem'

- There is no current decision-making algorithm (or ANN) which would enable a driverless car to deal with this problem.
- There is some research – Tufts, The Bristol Robot Dilemma (tinael)
- Are humans any different?

26

My claim:

(In spite of these limitations…)

- We can and *should* reduce road casualties by introducing driverless cars as soon as practicable.

27

Some counter-arguments

1) The current level of road deaths is acceptable.

28

Some counter-arguments

- 1) The current level of road deaths is acceptable.

… If people wanted to reduce road deaths there would be more of an outcry.

29

Some counter-arguments

1) The current level of road deaths is acceptable.

1a) …It's Health and Safety gone mad as usual.

1b) …It's morally OK to kill 5 rather than 1

1c) …to find the answer to the question 'how many deaths are acceptable?' we should look at what people accept.

30

Some counter-arguments

2) Unlike humans, self-driving cars can be hacked.

Some counter-arguments

- 2) Unlike humans, self-driving cars can be hacked.

This is a serious risk.

[– the IT industry has a serious lack of professionalism and a truly lamentable record – I have said so repeatedly for 3 decades…]

Some counter-arguments

- 3) People expect much higher standards from AI than they typically get from humans.

33

Some counter-arguments

3) “Human drivers may be forgiven for making an instinctive but nonetheless bad split-second decision, such as swerving into incoming traffic rather than the other way into a field. But programmers and designers of automated cars don’t have that luxury, since they do have the time to get it right and therefore bear more responsibility for bad outcomes.”

Lin, P. 2103, The Ethics of Autonomous Cars, The Atlantic.

34

Some counter-arguments

- 3) People expect much higher standards from AI than they typically get from humans.

This seems to be psychologically true:

Malle et al, 2015, Sacrifice One For the Good of Many? People Apply Different Moral Norms to Human and Robot Agents, HRI '15 Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction

Pages 117-124

Some counter-arguments

- 3) People expect much higher standards from AI than they typically get from humans.

BUT:

Ethically, legally, and insurance-wise the best is the enemy of the good. People will die while we wait until systems are perfect enough.

Some counter-arguments

- 3) People expect much higher standards from AI than they typically get from humans.

BUT:

Ethically, legally, and insurance-wise the best is the enemy of the good. People will die while we wait until systems are perfect enough.

They only need to be better than humans.

37

Some counter-arguments

- 4) Who should decide?
- AI programmers and motor manufacturers do not have the moral authority to kill or let die.

38

Some counter-arguments

- 4) Who should decide?
- AI and motor manufacturers do not have the moral authority to kill or let die.
- The distinction between killing and letting die may be technologically important here.
- It is often held (by consquentialists in particular) not to be ethically important.

39

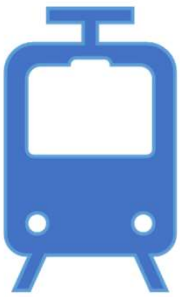
Some counter-arguments

- 4) Who should decide?
- AI and motor manufacturers do not have the moral authority to kill or let die.

[Oh yes they do - and it's high time they accepted responsibility and behaved as professionals.]

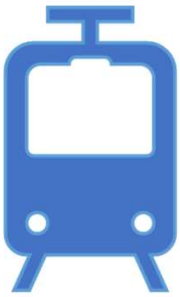
40

Back to the tram problem



- Philippa Foot, 1967
- "Imagine you are the driver of a runaway tram…"

Back to the tram problem



"Imagine you are the driver of a runaway tram…" and :-

"...self-driving vehicles could eliminate 90% of all auto accidents in the U.S., ..and save thousands of lives, according to a report by consulting firm McKinsey and co." Wall Street Journal, 5th March 2015

Time to choose…

24,580 Killed or seriously injured

2,500 Killed or seriously injured

Original artwork by Fran Polanski

Some suggestions:

- Legislation – preferably world-wide on a similar basis to civil aviation - is needed now.
- Product liability should be *relaxed* for driverless cars.
- Accidents should be investigated on a no-blame model and recommendations made to avoid any repetition.

Some suggestions:

IN EXCHANGE: ALL COMMERCIAL ALGORITHMS FOR THE SORT OF CONTROL DECISION DISCUSSED TODAY SHOULD BE MADE PUBLIC.

ALL DESIGN AND DEVELOPMENT OF DRIVERLESS SOFTWARE SHOULD BE LEGALLY REQUIRED TO HAVE OPEN ETHICAL SCRUTINY AT ALL STAGES.

References

- Whitby, B.R. (2105) Automating Medicine the Ethical Way in Machine Medical Ethics, Volume 74 of Intelligent Systems, Control and Automation: Science and Engineering pp 223-232
- Whitby, B.R. (2012) Do You Want a Robot Lover? in Robot Ethics: The Ethical and Social Implications of Robotics Lin, P., Bekey, G. , and Abney K., MIT Press.
- Whitby, B.R. (2012) On Computable Morality in Anderson S. and Anderson M, Machine Ethics, C.U.P.
- Whitby, B.R. (1996) Reflections on Artificial Intelligence: The Social, Legal and Moral Dimensions, Oxford: Intellect Books.
- Whitby, B.R. (1988) Artificial Intelligence: A Handbook of Professionalism, Chichester: Ellis Horwood.

Blay Whitby
blaywhitby.org
blayw@sussex.ac.uk

Originally delivered by