Do robots need ethics?

Szymon Bobek

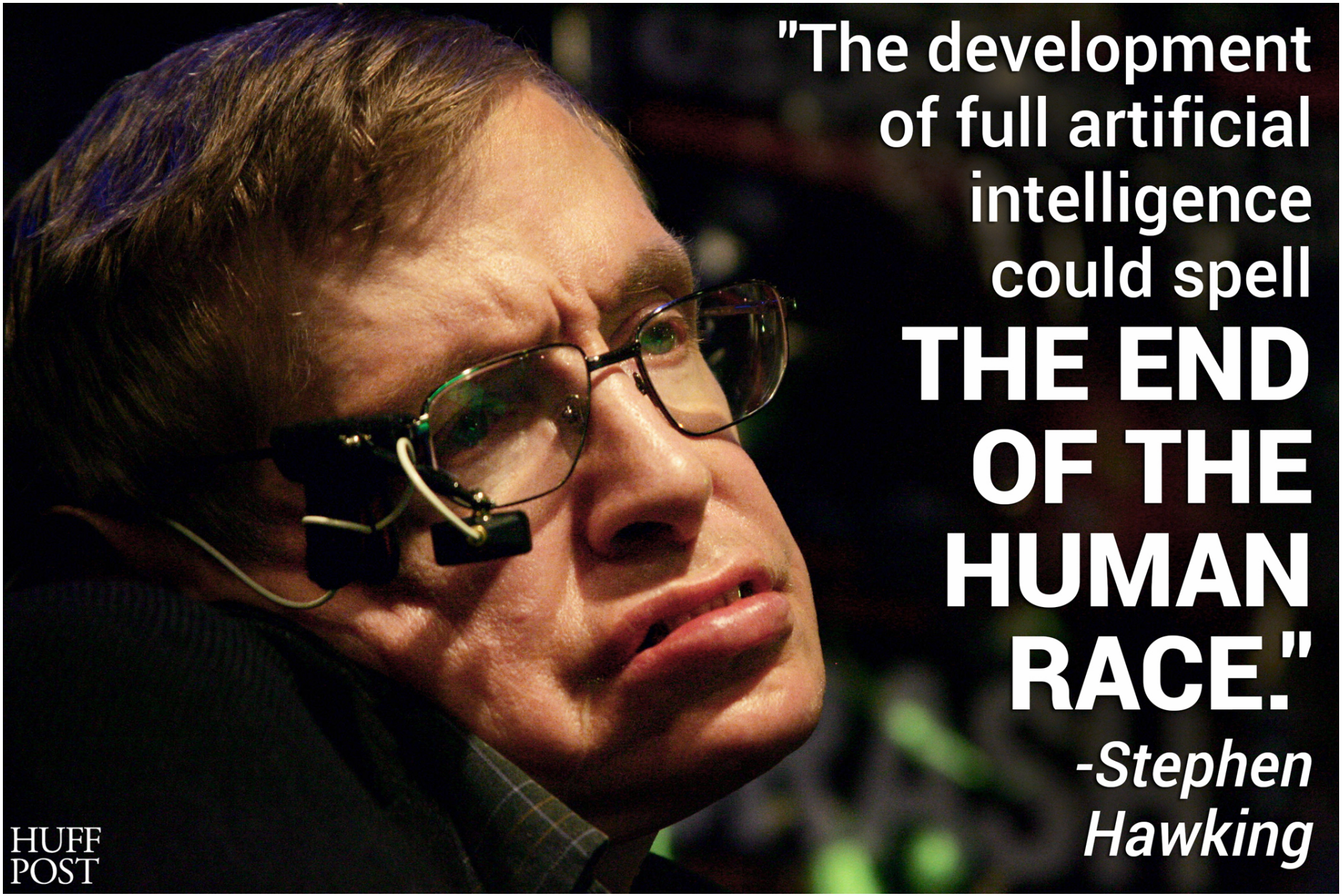
Institute of Applied Computer science
AGH University of Science and Technology

ISK Robotics club
email: szymon.bobek@agh.edu.pl
http://home.agh.edu.pl/sbobek

Outline I

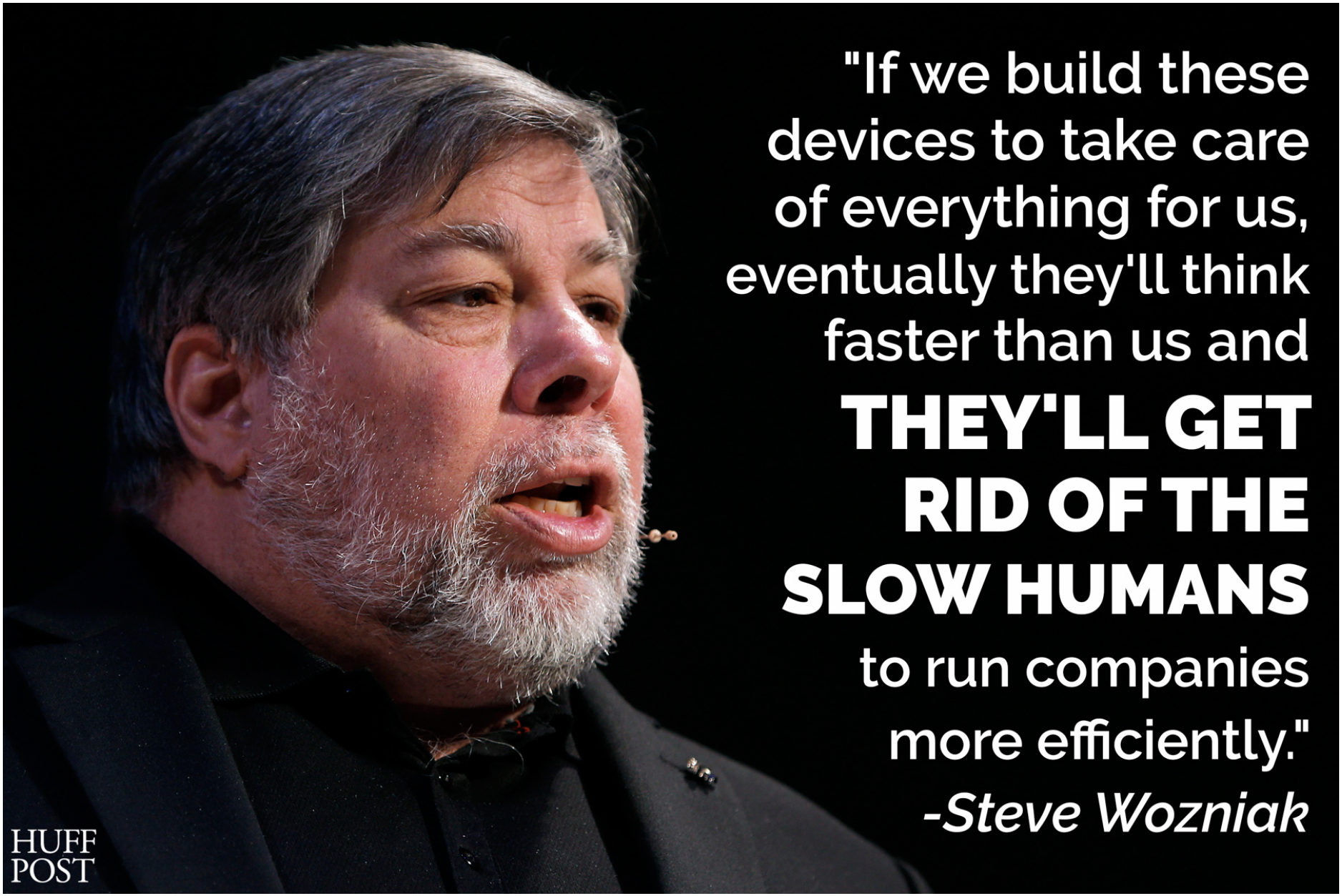
Presentation Outline

Why people fear AI



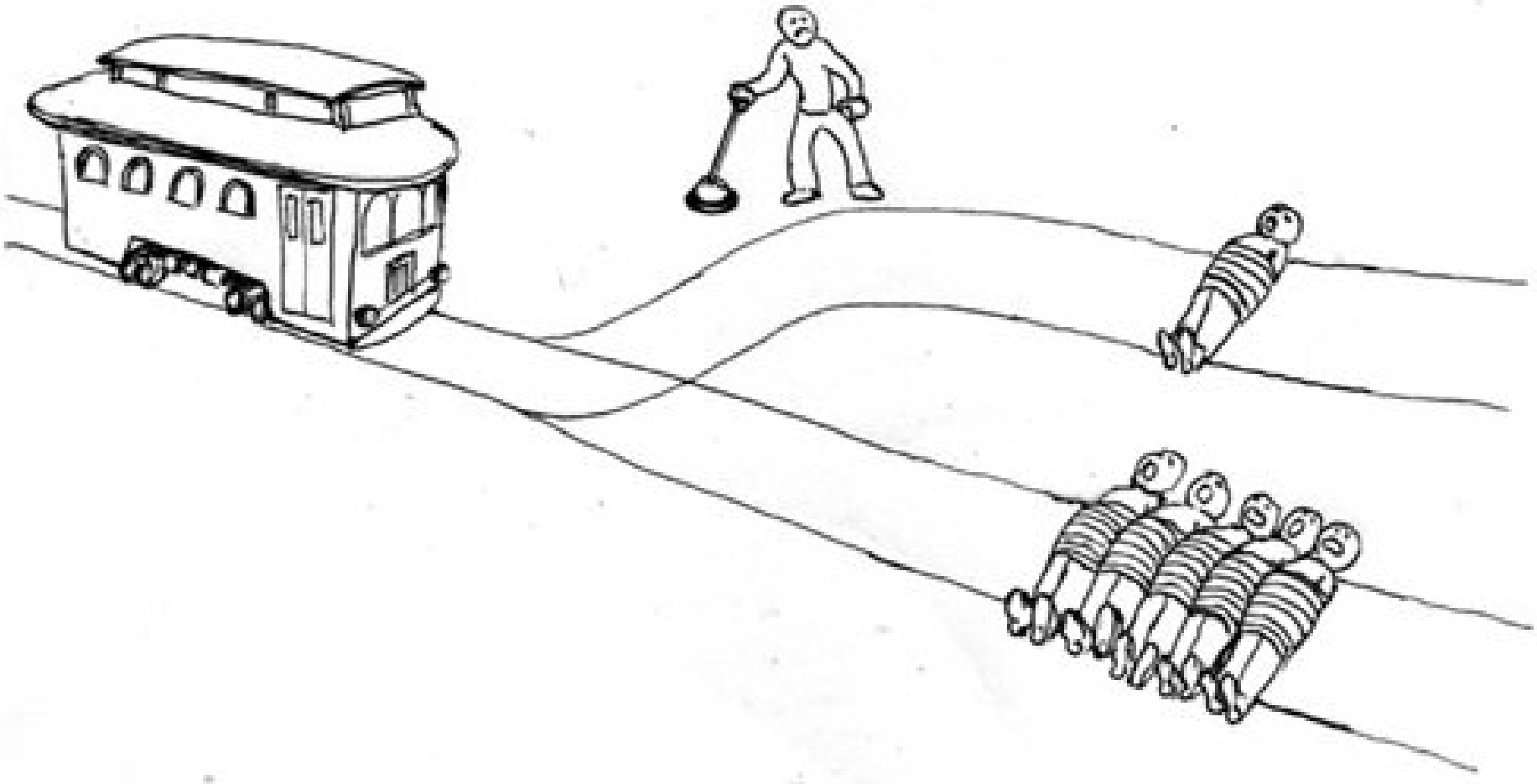
Why people fear AI

Why people fear AI

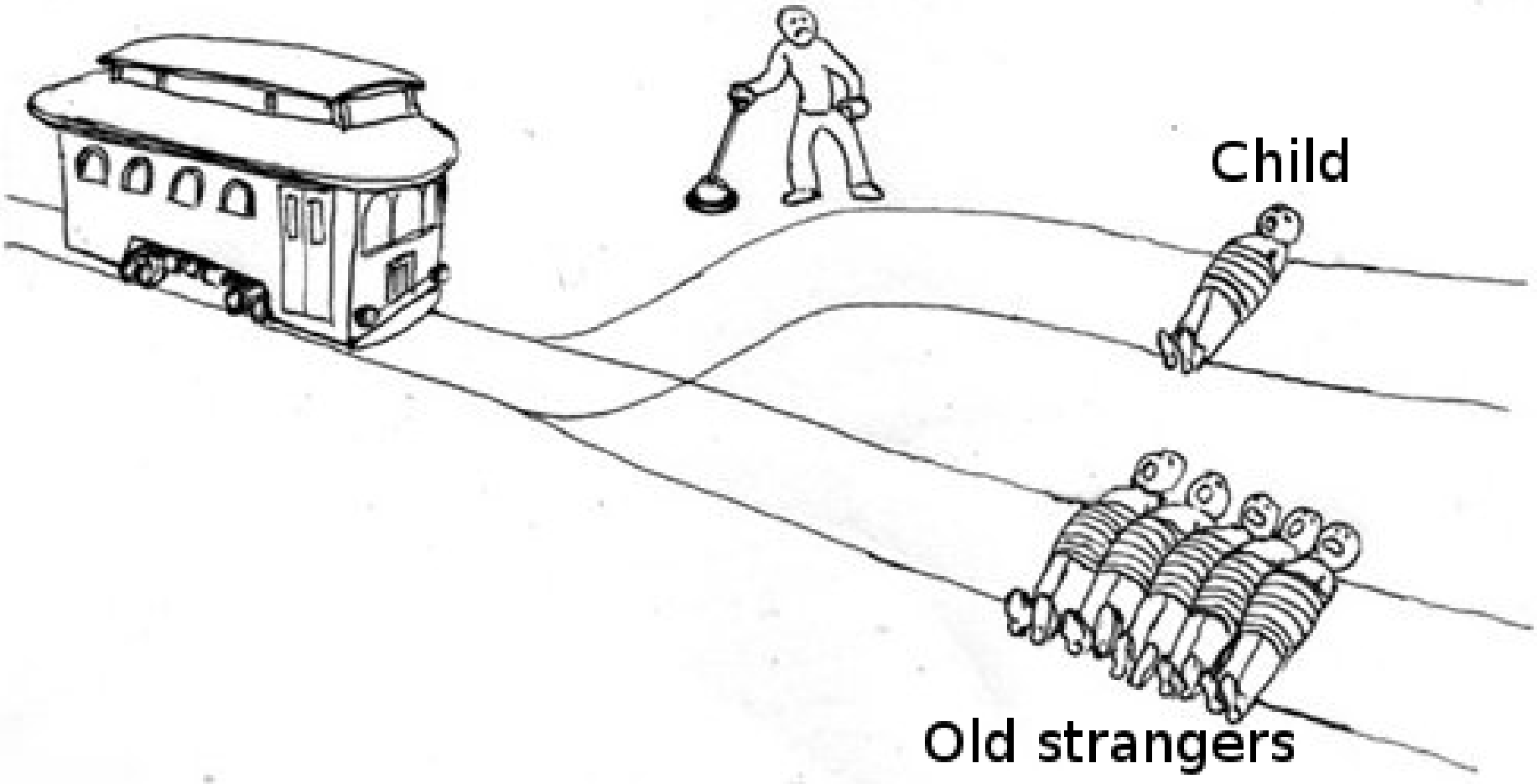


Presentation Outline

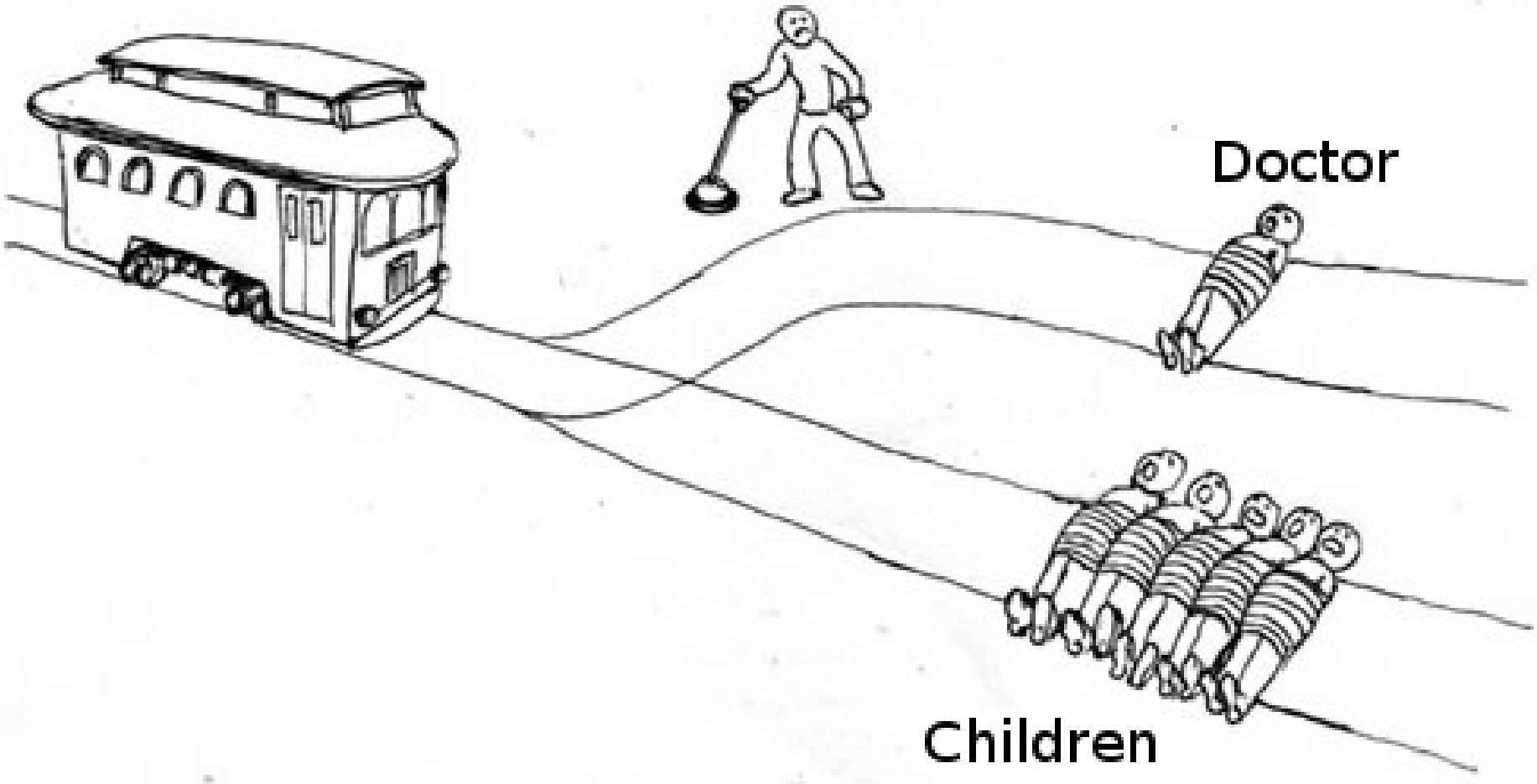
Trolley problem



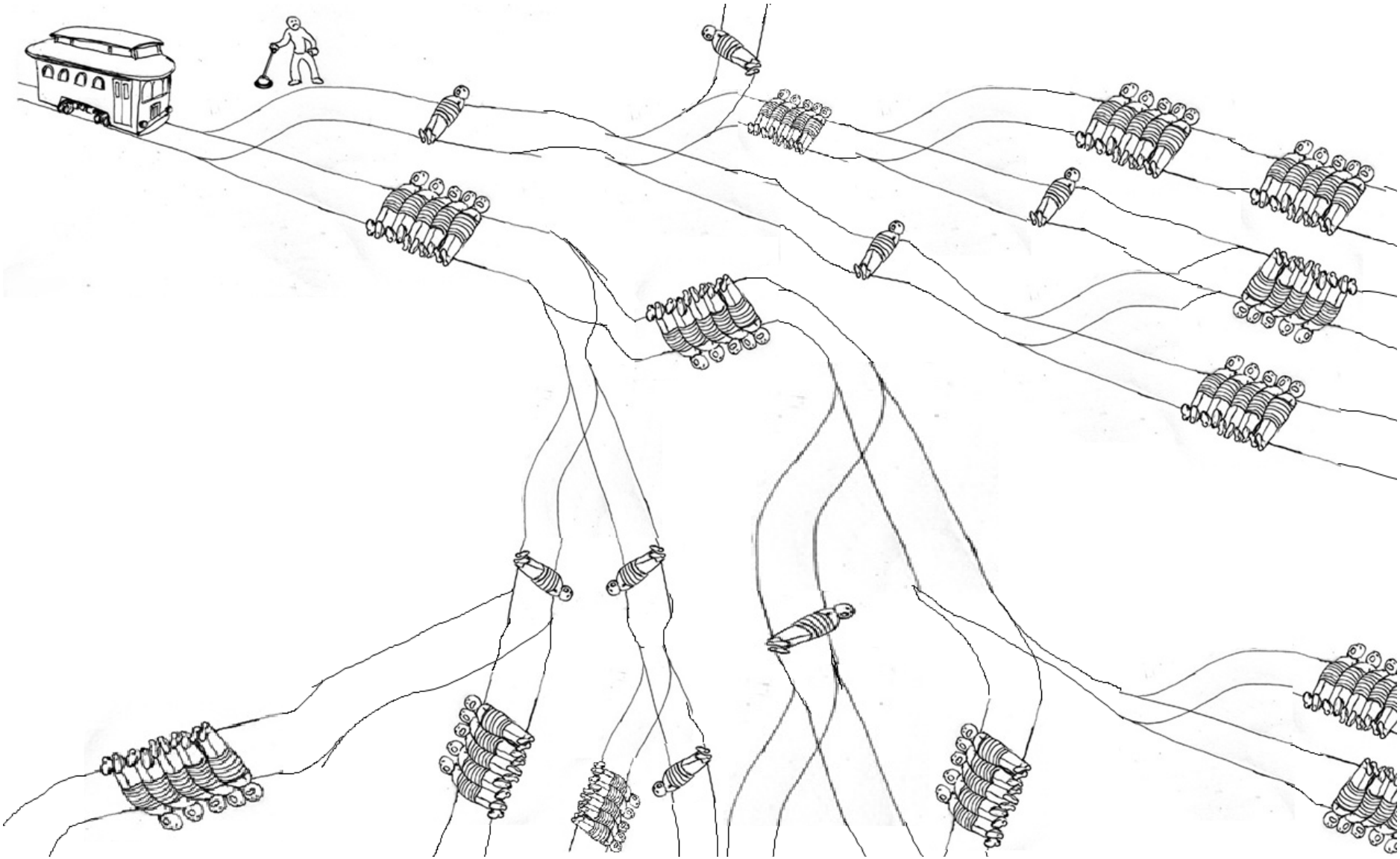
Trolley problem



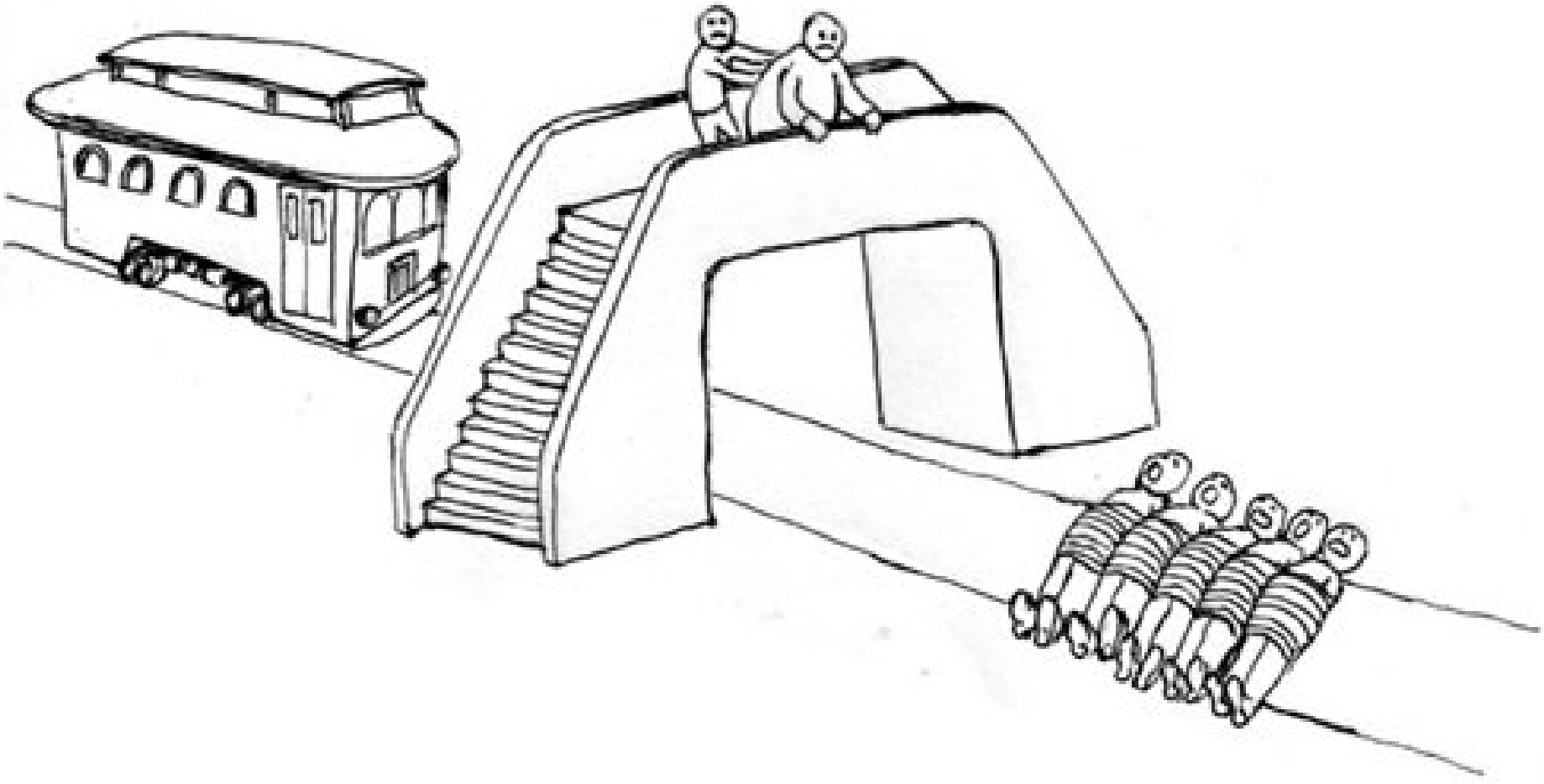
Trolley problem



Trolley problem



Trolley problem



Presentation Outline

Asimov's three laws

Three laws

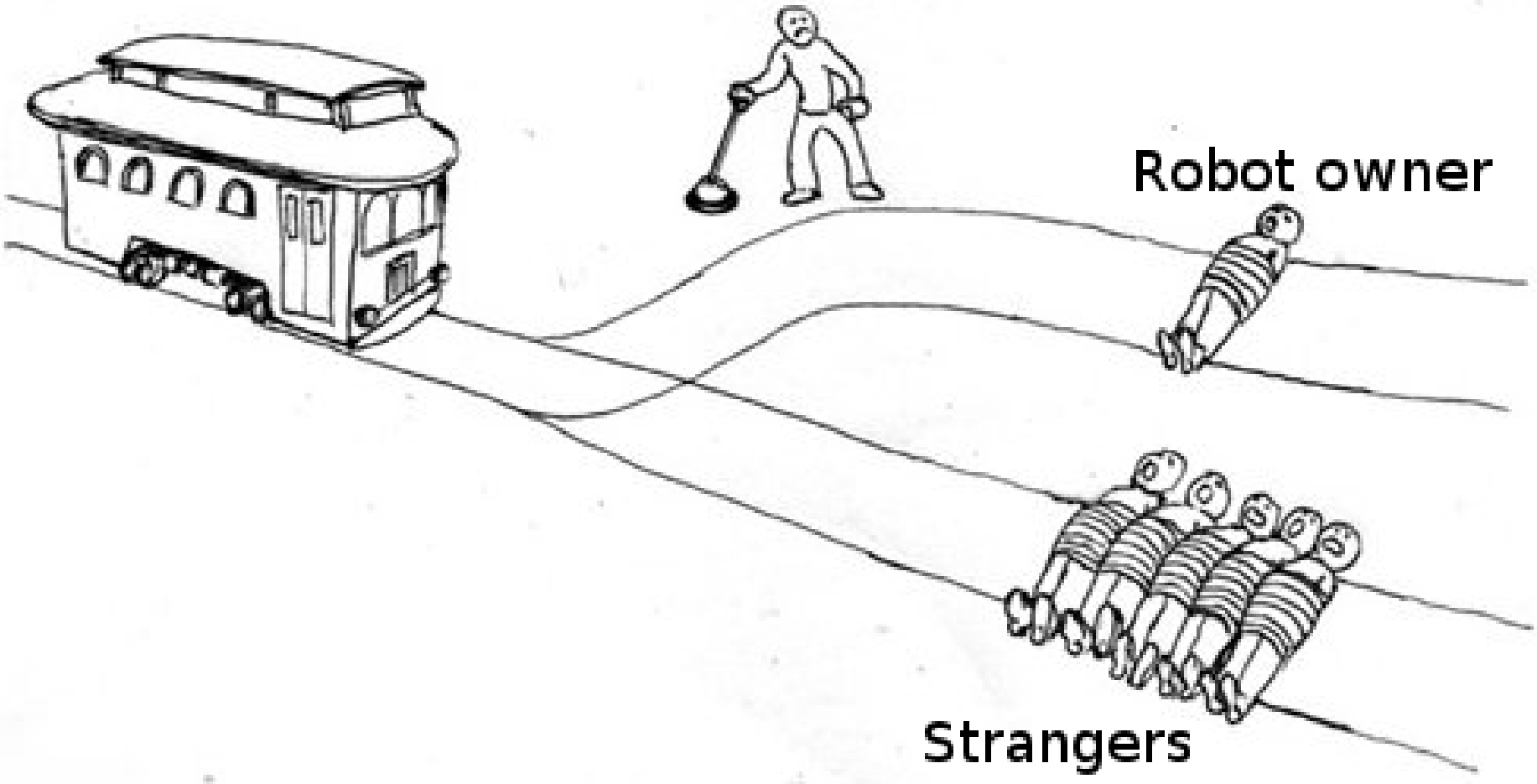
1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law

Deciding who is good and bad

Is robot allowed to kill?

The robot sees a human pointing a gun into another person. Is the robot allowed to kill?

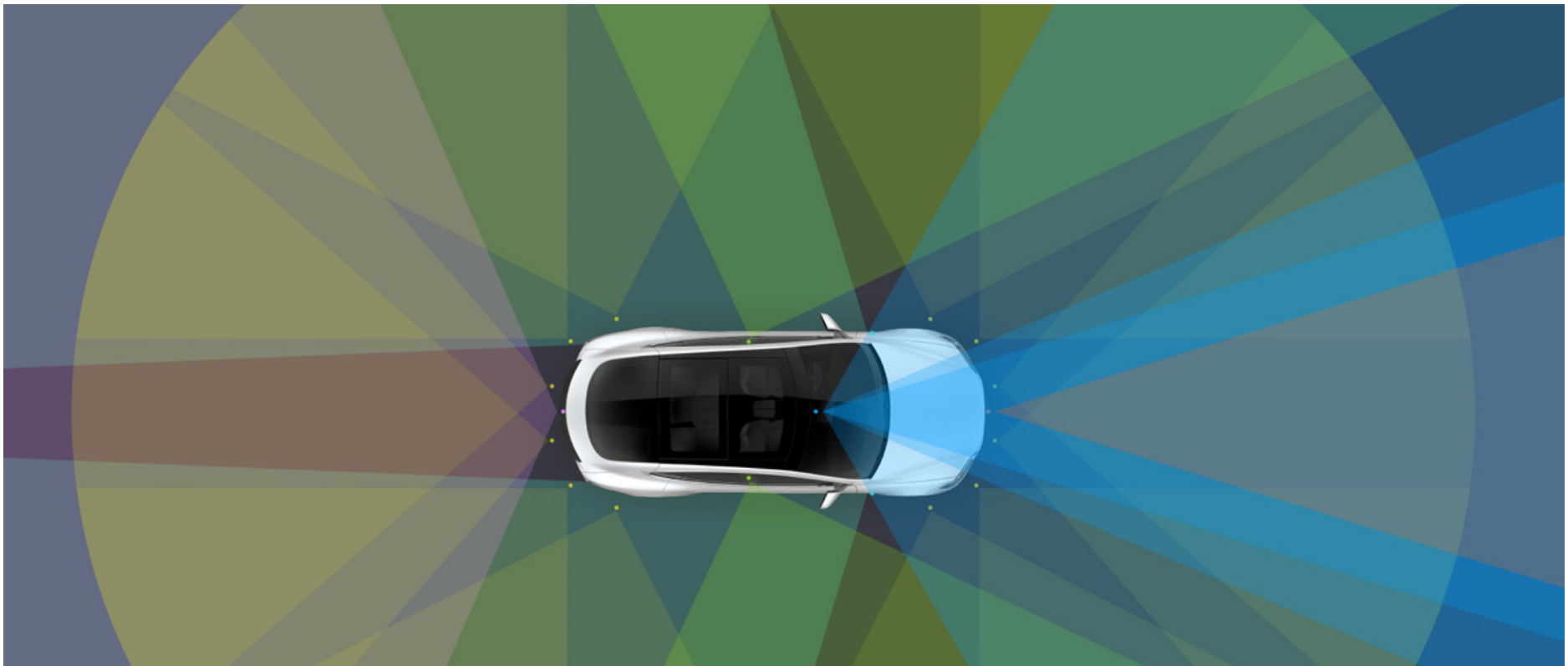
Trolley problem – revisited



Trolley problem – real case

Tesla cars

Two Tesla cars are heading each other and the crash is inevitable. One car has an option to drive into the river and risk lives of its passengers, or push the other car, to the abyss and save its passengers by killing the other. Only one car has an option to make a decision. What should it do? Who will take the responsibility for that (algorithms cannot go to jail)?



Zeroth law

Asimov's Zeroth law

1. A robot may not harm humanity, or by inaction, allow humanity to come to harm.
2. A robot may not injure a human being or, through inaction, allow a human being to come to harm, except when required to do so in order to prevent greater harm to humanity itself.
3. A robot must obey any orders given to it by human beings, except where such orders would conflict with the First Law or cause greater harm to humanity itself.
4. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law or cause greater harm to humanity itself.

Deciding what is good and bad

Rage against humanity

- Alcohol, tobacco, fast food, etc.
- Air pollution
- Free will...

Presentation Outline

Robots can save us

They can be better than us

- In computations?
- In making decisions?
- In making complex moral decisions?
- In... social skills?

Robots can save us

They can be better than us

- In computations?
- In making decisions?
- In making complex moral decisions?
- In... social skills?

But is it OK?

- Who should decide on your medical treatment – your family, or robot?
- High trust in robots may prevent humans from contesting their decisions.
- What if robots are better than us in art, science, etc. Will it kill creativity?
- Can robots be deceived or mistaken? Who takes responsibility for robots' autonomous decisions if they are harmful? Is turning the robot off really an appropriate punishment?

Presentation Outline

Bombs, Bonding, Bondage

War, Love and Slavery

Reading


I. Asimov.
I, Robot.
Robot series. Bantam Books, 1950.


David J. Gunkel.
The Machine Question: Critical Perspectives on AI, Robots, and Ethics.
The MIT Press, 2012.


Patrick Lin, Keith Abney, and George A. Bekey.
Robot Ethics: The Ethical and Social Implications of Robotics.
The MIT Press, 2014.


Spyros Tzafestas.
Roboethics: A Navigating Overview.
Springer Publishing Company, Incorporated, 1st edition, 2015.


Wendell Wallach and Colin Allen.
Moral Machines: Teaching Robots Right from Wrong.
Oxford University Press, Inc., New York, NY, USA, 2010.

Thank you for your attention!

Any questions?

email: szymon.bobek@agh.edu.pl
http://home.agh.edu.pl/sbobek