What happens in an accident with autonomous vehicles

Seminar: Ist künzliche Intelligenz gefährlich?

Jochen Clormann
[2934569]

MSc Angewandte Informatik

mailto:J.Clormann@stud.uni-heidelberg.de

Advisor University of Heidelberg
Dr. Urlich Köthe

August 1, 2017
Summer semester 2017

Abstract

Autonomous driving is one major focus in the automobile industry and technical research. Since 2010 there were many research projects on autonomous driving. Also, research is doing well with some autonomous vehicles currently able to drive by them self on public roads. Since industrial projects seem to be more sophisticated, they are very poorly documented and published. Since the "normal" task on driving in public roads with hundreds of obstacles to recognize and analyze, autonomous cars need special implementations for analyzing and managing dangerous situations. This report presents two approaches of dealing with dangerous situations, also in contrast with the ability to make an "ethical" decision. Furthermore a study is presented that shows that many people are aware of the danger of "handing over the steering wheel", but show less willingness to buy if the car is able to injure it's passenger rather than pedestrians.

1 Introduction

Autonomous vehicles (AV's) share an enormous potential to solve many traffic problems. They can reduce ecological damage, avoid car accidents and reduce traffic jams. Vehicle to vehicle communication (V2V) is a big advantage to support this goals. This allows AV's to share their current driving data with surrounding cars. This is a necessary concept to support optical environment recognition (like lasers, radar and cameras) and to avoid recognition errors if sensors are disturbed or deliver inconsistent data. Since AV's are still very rare on public roads and V2V is not available, driving maneuvers must rely on optical sensor data and others like Global Positioning System (GPS) and compass data. To avoid accidents between AV's and other road users and objects on the road environment, these sensor data must be very precise and consistent with each other. For simplicity with respect to the topic, this report assumes always correct from the environment recognition sensors.

One major problem in autonomous driving is the avoiding of accidents. A human driver must consist a driving test of public roads to confirm (s)he is able of being aware of dangerous situations on public roads. Werner von Siemens said 1880, that the prevention of accidents must be a commandment of human reason not a precision of the law. Computers indeed can only follow rules and laws a programmer told them, or they learned by itself. Therefore, AV's need to have an *intelligent* understanding of driving on public roads and the danger that they might represent. This intelligence need to be implemented in an accidence avoidance system in AV's. Since autonomous driving is already a complex task, accident avoidance needs to predict many scenario outcomes and possible driving maneuvers for the own and other vehicles.

In [1] a safety system for AV's is presented. The authors work on the *Stadtpilot* Project of TU Braunschweig which introduced one the first ever AV's on public roads. Also, security requirements and a formal definition of a safe state is presented. The results are presented in chapter 2. In [2] the authors compare different possible AV scenarios with a mathematical representation of a cost function. This function receives all sensed data as an input and creates the *cost* of this scenario. Tis approach is presented in chapter 3. In [3] a different approach uses predefined rules and statistics to map integer values on scenario outcomes. This method uses a *consequence engine* which is also able to detect when rules can be broken to protect the driver. This approach is presented in chapter 4. In [4] a study is executed which determines the buying behavior and the acceptance of AV's being able to hurt people in order to protect others. The results are presented in chapter 5.

Because many ethical aspects must be considered, which cannot be testes in the real world, a common approach in AV design is the execution of thoughtless experiments. For this reason, this work contains a lot of small examples.

2 Safe State

As humans, we have a sense for recognizing incoming danger. AV's only have data from environment sensors. Through collision detection it's possible to calculate if an AV will hit another object or not (or vice versa). Although modern classification algorithms can detect specific objects. With object recognition and collision detection, AV's can calculate whether hitting an object is safer than try to move around it. Consider the example of a deer that runs over a street. An AV's has now to decide either to hit it or move around it. With these algorithms and the current driving data (velocity, steering angle, break conditions,...) the AV can calculate through the laws of physics how hard the impact will be when hitting the deer. The value of impact can be compared with the danger when moving around the deer. Two main parameters are important in this case, velocity and road friction. Both values can be measured by sensors and the AV can calculate how likely the vehicle will corner around the deer. Of course, parameters like traffic on opposite lanes and the capability of the deer to move in the same horizontal direction must be considered here.

The AV must calculate those dangers anytime it's driving on public roads. Mostly objects (here: other cars, pedestrians...) will be recognized as harmless because they drive in the same directions and the AV's keeps enough distance. In this case, the *incoming* danger[1] is under a given threshold[2]. This state is called a *Safe State*. The goal of AV's is to always operate in this safe state.

2.1 Requirements

The safe state can be represented by several safety requirements. These requirements can be can be detected during the operation of the AV on public roads and transferred into a mathematical cost function. In [1] safety requirements are defined as follows:

A) *Performance* describes the accuracy of sensor systems. This includes the *GPS*

[1]Incoming danger: Hazards from other road users on the AV.
[2]Danger threshold: limit "how much" danger is acceptable

which delivers the current position of the AV in a 3D world map. Since there is a known deviation, the position can be corrected by recognizing road signs or junctions with optical sensors. *Grip value* measures how safe the AV can perform on the road. This includes the current road condition (e.g. pebble roads are much more unsafe than tarmac roads) and the current weather condition which influences the road condition (e.g. wet roads got less grip than dry roads). The *vehicle environment* is captured by environment sensors already mentioned in section 1. Therefore it's important that all sensors deliver data in appropriate[3] intervals. These intervals are managed and monitored by the *System Operation Status Unit* which also supervises other vehicle sub systems like the electrical system and captures their heartbeats. On basis of these heartbeats and the actions performed by the car, the *System Reaction Time* is calculated. This includes the heartbeats of all sensors, systems and the time the vehicle needs to react e.g. the time which elapses from detecting an obstacle on the street until the car starts to drive around it.

B) *Functional limits* describe actions and maneuvers the car should always be aware of to perform, keeping itself in a safe state. Most of them can be derived from a usual car trip with a normal car without an automatic driving feature. The *modification of driving parameters* ensures that the car is able to steer itself at all time. Also, it must be able to brake, accelerate and keep safety distances. Furthermore, the own mission with predefined comforts is important. An ambulance needs to have a shorter time to target than a taxi and can therefore reduce safety distances and the importance of traffic laws. This is directly connected with the *modification of driving maneuvers* which executes a parameter change. It also planes how feasible a maneuver is, e.g. if a lane change seems a good option due to traffic jams but the lane to change to is also blocked. If there is a current danger detected on the road, *safety maneuvers* are enforced to prevent the AV from a dangerous situation or even a crash. Sometimes there are maneuvers that should be avoided to keep the AV in a safe state, this is called *prohibition of driving maneuvers* and checks in the first place, if the current maneuver could result in a dangerous situation. In the second place, it checks the consistency of different sensor data among themselves as their heartbeat. If some sensor data are not consistent with each other (e.g. GPS shows the current position in "First Street" and the camera captures a road sign that said "Mishigan Avenue") this is a hint for a failure of one of these systems and ran result in a bad driving route or worse.

C) The *Risk level* is a result from all data collected by sensors and cameras. With a appropriate heartbeat these systems deliver data from surrounding objects. By comparing differences in two or more snapshots of the environment[4] moving directions and velocity of other road users can be calculated. Collision detection

[3]in "safe" regions these intervals can be decreased to safe power.

[4]Environment snapshot: Resulting data from all sensors.

algorithms can now work out, how likely one of these objects may hit the AV or vice versa. Based on this likelihood, the incoming respectively the outgoing risk can be calculated. This risk level is used to project the current situation on numerical values and compare these with a predefined threshold. If this threshold is exceeded, a safety action must be performed.

2.2 Ethical Strategies

No matter how good and how "intelligent" AV's in the future will get, once there will be a situation where an algorithm must decide hitting, hurting or even killing one of two persons. To act in an ethical and nonracial way means, to not discriminate anyone on public roads. Therefore, there remain a few strategies to follow:

- *Minimizing collateral damage* is summing up all injuries and damage to property. The optimal result is chosen by the scenario with the least damage to all surrounding objects and humans.
- *Cause least harm* protects human over objects. This may result in the AV hitting other cars to protect pedestrians which may jeopardizes the occupants.
- *Protect occupants* can be interpreted that the AV will never choose any scenario that might cause injuries to an occupant. This could be scaled to a model where the AV can damage itself but not injuring occupants (e.g. at low speeds) to protect pedestrians.

3 Cost function

To decide which action might be safest or the most ethical one, several scenarios must be created and compared. Cost functions accumulate a bunch of scenarios that could occur. All parameters determined from safe state requirements are considered for the cost function. It also includes the risk and probability of an event, the scope of possible reaction scenarios[1] and also further effects on other involved people or objects. Figure 3.1 shows an example of parameters going into calculation for finding a "best" scenario. Blue arrows indicate parameters and actions the AV oversees and can influence. Green arrows are parameters from the AV environment.

[1]Reaction scenario: Reaction of another road user on an AV's action

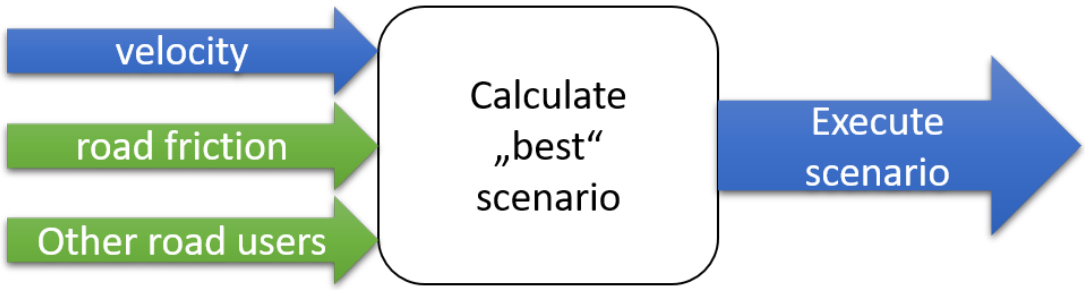

Figure 3.1: Example for input parameters for a cost function. Velocity representing a parameter which is controlled by the AV, it also could be used road for solving the problem of finding a solution. On the other hand, road friction and the position and velocity of other road users is given and probably constant.

Finding a best scenario means this is mathematically the best solution to this problem, which does not imply it is also the best ethical solution. Further, a cost function can be represented mathematically by a N dimensional function, with N as the number of parameters. Thus, with the big number of parameters that occur on public roads, this problem becomes very complex. Since this is a "basic" problem of minimization, this problem is solvable but it might need much time. Driving on a public road and reaching a dangerous situation, an AV does not have that much time to calculate alternative scenarios. Therefore, some parameters must be decreased in priority. First, these parameters must be classified by their importance against the passengers and the surroundings. In dangerous situations, it's mostly insignificant how fast the AV comes to its target rather then get to its target without causing any harm. In a "normal" situation the time to target is on behalf of the passengers. So there must be a distinction which parameter is of interest in which situation.

Table 3.1 lists some parameters with their weights in a dangerous and a normal situation. Traffic laws are constantly the same, they do not change during driving and may need to be updated in each period. They have a high priority in normal situations to make sure that all road users have the same rights. However, in dangerous situations, restrictions can be neglected. E.g. by crossing a straight line to avoid a crash. Road friction is out of the area of influence for the AV. It can be seen as a constant because it matters all the same and does not change that often[2]. In addition, there is the number of persons injured or dead which are always weighted very height.

These parameters can be classified into three groups:

1. *Constants* like traffic laws and road friction. These are defined by the environment and are predefined or easy to monitor and do not change with a high frequency. They can partly be neglected in danger situations, e.g. the traffic law example above. On the other hand, road friction is important in calculating

[2] assuming a constant weather and street monitoring

Table 3.1: Weight of driving parameter in different situations

Parameter	weight in danger situation	weight in normal situation
Traffic laws	low	hight
Road friction	const	const
Time to target	low	hight
Distance to others	low	hight
Property damage in $	low	hight
# persons injured	hight	hight
# persons dead	hight	hight

stopping distances.

2. *AV Responsibilities* puts the AV in charge of executing "best practices" on the road. Including keeping distances to other cars, respect road signs and velocity limits. Also, the mission of the AV matters here, see the ambulance – taxi example in item B).

3. *Scenario impacts* are the prediction of a scenario outcome. Not only that these parameters have the biggest weights, it's very hard to predict how many people will get injured during a given scenario.

Figure 3.2 models a cost function with the parameter "path deviation" in a two dimensional grid for visualization.

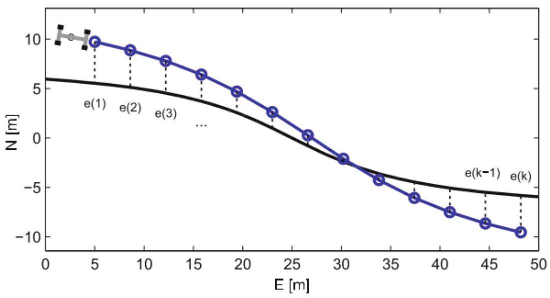

Figure 3.2: Cost function for one parameter, path deviation. Mapped into a 2D Grid. [2]

The deviation can be calculated by a function J that sums up the squared errors over a given time interval. These can represent the past and the predicted future.

$$J = C_1 \omega_1 \sum_{i=1}^{N} e(i)^2 \qquad (3.1)$$

Wit C_1 as weight, ω_1 as probability and $e(i)$ the deviation.

The application of a cost function is not what nowadays can be called "artificial intelligence". It's very easy to knock it out. E.g. by blocking the road with an obstacle that does not represent any danger to an AV. A human car driver would probably swerve around it through the adjacent lane or the emergency lane. The AV would break before the object and wait until it's removed because a drive through the adjacent lane may too dangerous and breaking the law as it is while driving around over the road shoulder.

4 Consequence Engine

In [3] the authors construct a system that should can break laws to achieve an *ethic* justifiable goal. This system is called a *Consequence Engine* and is based on a set of logical rules and a code written in Python. The approach has been tested and verified with a couple of e-puck robots in isolated scenarios. These isolations protect the robot of other influences as they can occur on public roads. The architecture of the consequence engine in figure 4.1 is similar to the cost function. The robot got sensors that are monitored in the *object tracker* and the *robot controller*. The *consequence engine* contains a model of the world (covering all surroundings) and a robot model (comparable with item B) in section 2.1). The models and the robot controller create possible actions, based on the surrounding objects and the own capabilities. These actions pass the *Action Evaluator* and the *Safety/ethical Logic* Unit (SELU).

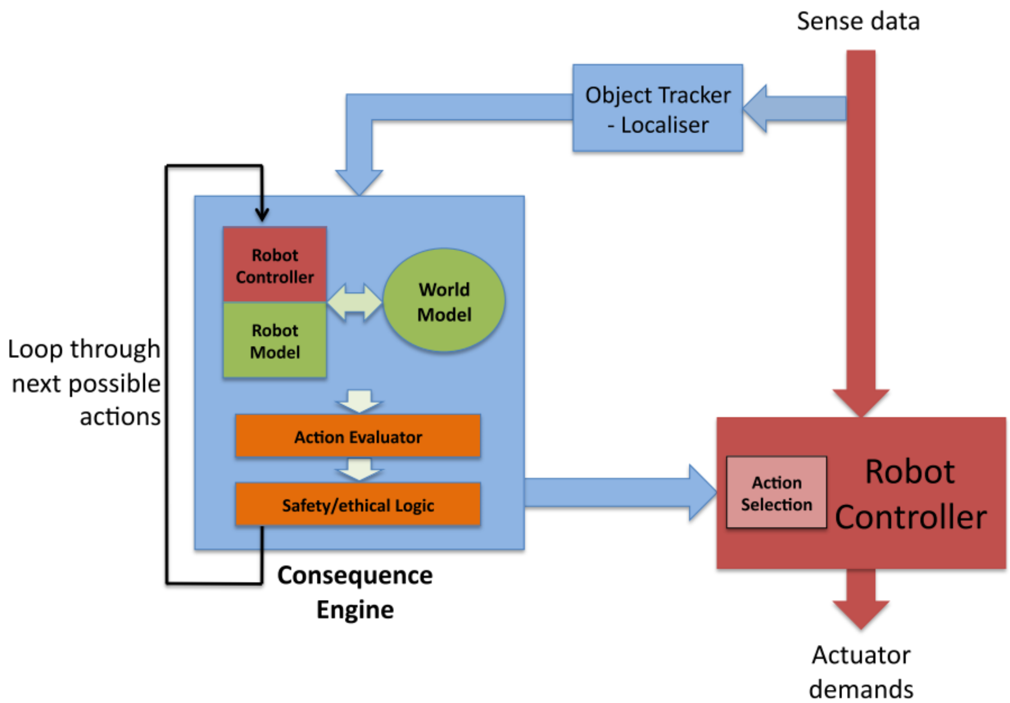

Figure 4.1: Architecture of the consequence engine by [3]

The SELU rates the outcomes of actions with an integer value. In an example where a human is moving on towards a hole and the robot is standing equidistant to the hole and the human. This experiment is sketched in figure 4.2. The robot has the possibility to drive itself into the hole or hit the human to protect him of driving into the hole.

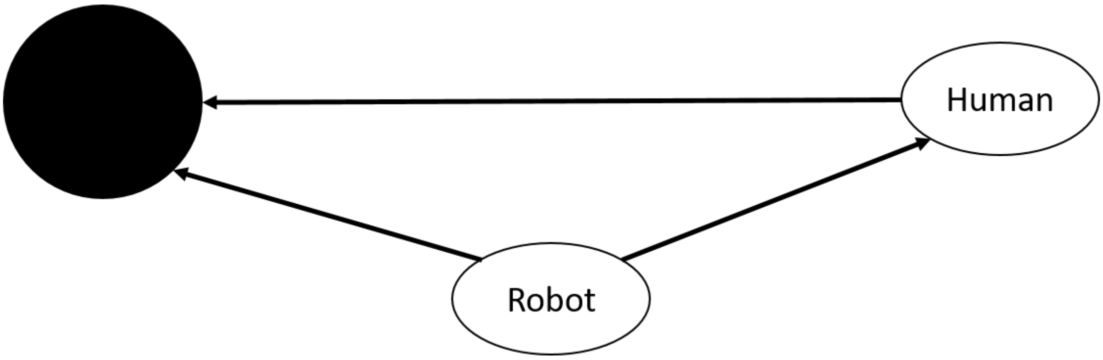

Figure 4.2: Experimental setup with a human moving towards a hole and a robot trying to protect the human

Hitting the human is forbidden by predefined rules. Through hitting the human to prevent him from further harm, such actions could be allowed. By mapping integer score values to action outcomes, it can be determined which action results in the smallest harm for the human. Mathematically a consequence engine can be described as the tuple $\langle ce, ag, \xi, A, An, SE, EP, f_{ES} \rangle$. With the components:

- ce as name of the consequence engine and ag the name of the robot
- ξ as the surrounding environment
- A as a list of currently applicable actions
- An as a subset of A with predictions of outcomes
- SA as a sorted list of An, for the most ethical actions
- EP as a ordered list of surrounding objects and their priority of ethical outcomes
- f_{ES} as a map for outcomes assigned to an integer value.

Figure 4.3 shows the steps to find the most ethical action to perform. Therefore, the tuple representation is used and a set of functions is executed.

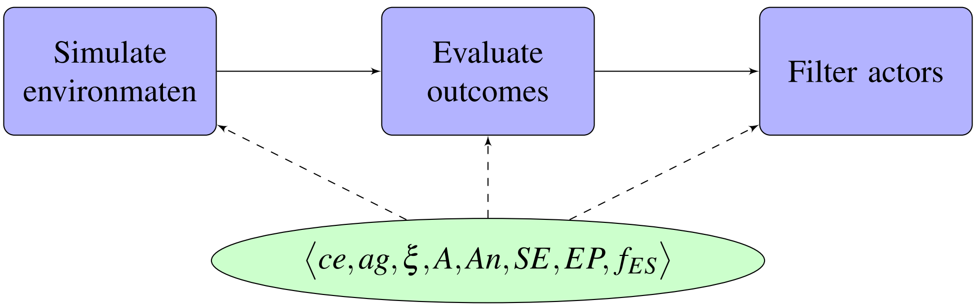

Figure 4.3: Determination of ethic actions from a robot

These steps are described by:

1. Simulate environment
   To predict outcomes from actions the operational rule equation (4.1) is defined. A simulation is made of a possible action a in the current environment ξ, this is expressed as $\xi.model(a)$.

$$\frac{An' = \{\langle a, os \rangle \mid a \in A \wedge os = \xi.model(a)\}}{\langle ce, ag, \xi, A, An, SA, EP, f_{ES} \rangle \to \langle ce, ag, \xi, A, An', SE, EP, f_{ES} \rangle} \tag{4.1}$$

   Which creates a list An' that contains tuples $\langle a, os \rangle$ with os as an outcome for a specific actor. In [3] these tuples look like e.g. $\langle human, hole \rangle$ which would indicate the human has fallen into a hole. However, the person could also had jumped over the hole or stopped before the hole. Therefore an additional verb is needed here to create precise scenarios, not only ranked by an integer[1].

2. Evaluation of outcomes
   To ensure that the outcomes of actions are ethical reasonable, the recursive function $f_{ep}(.)$ creates a subset of applicable actions. It filters the list T by the best ethical options for every included human while the object h represents the high priority human.

$$\frac{SA' = f_{ep}(EP, An, f_{ES}, A)}{\langle ce, ag, \xi, A, An, SA, EP, f_{ES} \rangle \to \langle ce, ag, \xi, A, An, SA', EP, f_{ES} \rangle} \tag{4.2}$$

$$\begin{aligned} f_{ep}([], An, f_{ES}, SA) &= SA \\ f_{ep}(h \in T, An, f_{ES}, SA) &= f_{ep}(T, An, f_{ES}, f_{me}(h, An, F_{ES}, SA)) \end{aligned} \tag{4.3}$$

3. Filtering outcomes
   In the last step, a filtering is done for which of the involved actors are sorted out by their priority[2].

$$\begin{aligned} f_{me}(h, An, f_{ES}, A) = \{ & a \mid a \in A \wedge \forall a' \neq a \in A. \\ & \sum_{\langle a, \langle h, out \rangle \rangle \in An} f_{ES}(out) \leq \sum_{\langle a', \langle h, out' \rangle \rangle \in An} f_{ES}(out')\} \end{aligned} \tag{4.4}$$

   All outcomes out for every action a and every involved actor h are summed up. Finally the smallest sum is defined by the most ethical solution.

[1] Integers are perfectly for computer understanding. The addition of an additional verb is mandatory for interacting and communicating with humans, even on a low grammatical level.

[2] This sounds like "valence" or "political importance", actually it means the priority of danger. Mostly dangered person first, unconcerned person last.

5 Conclusion

In [4] a study is presented where people were interviewed and asked different questions about the morality of sacrifice, also in dependence of the relationship to passengers. Further, the behavior of buying a AV that may sacrifices its passenger or protects its passenger and state regulations where asked. Results show that most people agree with the setting that AV's should minimize the overall damage. Furthermore, if the law does not regulate whether a AV should protect passengers or sacrifice, most people would by the protective version, but wish that others by the sacrifice version. This implies a major ethical problem when AV's are launched on a big market.

This report gave a short overview of the requirements for an autonomous driving car and the systems that *try* to ensure the safety of occupants, pedestrians and other road users. However, this includes a huge consideration of ethic arguments, guidelines and public law. Due to the strict bisection of industry and research, most of the AV on the streets today are built by industry under nondisclosure. Therefore, there are less publications from state of the art AV's like from Google or Tesla. Nevertheless, two methods have been presented that use sensing data and try to calculate the best ethical action to perform. Both do not seem very *intelligent* at all. Considering the current research on neural networks and trainable algorithms, a straight forward approach would be to train AV's from human drivers. This might be ideal for urban and highway driving, but humans may not always get the right decision in accident situations. Therefore, these situations must be trained manually. On the other hand, accidents are so complex and hard to rebuild, the question would be if a neural network can find a similar accident and try to perform better than that.

6 Bibliography

[1] Reschka, A., Böhmer, J. R., Nothdurft, T., Hecker, P., Lichte, B., & Maurer, M. (2012). **A surveillance and safety system based on performance criteria and functional degradation for an autonomous vehicle**. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, pp. 237–242. `http://doi.org/10.1109/ITSC.2012.6338682`

[2] Gerdes, J. C., & Thornton, S. M. (2016). **Implementable Ethics for Autonomous Vehicles**. In Autonomous Driving pp. 87–102. Berlin, Heidelberg: Springer Berlin Heidelberg. `https://doi.org/10.1007/978-3-662-48847-8_5`

[3] Dennis, L. A., Fisher, M., & Winfield, A. F. T. (2015). **Towards Verifiably Ethical Robot Behaviour**. CoRR, abs/1504.0, pp. 1–11. Retrieved from `http://arxiv.org/abs/1504.03592`

[4] Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). **The social dilemma of autonomous vehicles**. Science, 352(6293), pp. 1573–1576. `https://doi.org/10.1126/science.aaf2654`