SCIREA Journal of Philosophy

ISSN: 2995-7788

http://www.scirea.org/journal/Philosophy

November 25, 2024

Volume 4, Issue 4, August 2024

https://doi.org/10.54647/philosophy720102

The Ethical Dilemma of the Tunnel Problem and Its Solution

Weikang Zhu

School of philosophy, Anhui University, China

Email address: 1263044547@qq.com

Abstract

When faced with multiple dilemmas of moral choice, self driving cars often cause moral and legal problems that make people think deeply. When the self driving car is in such an unavoidable dilemma that it is necessary to hit people, it will be a puzzle whether it is utilitarianism to kill one person and save multiple people, egoism to kill others to protect itself, or altruism to sacrifice itself and passengers to save others. In order to deal with the complex tunnel problem caused by autonomous vehicle accidents, this paper will try to explore a variety of solutions from different moral algorithm theories to solve this problem.

Key words: automatic driving; ethical dilemma; tunnel problems; prisoner paradox; moral algorithm

1. Introduction

With the development of science and technology and the progress of society, people's requirements for living standards are also constantly improving, and self driving cars will inevitably enter the public's field of vision. Autonomous vehicles have more significant advantages than traditional driving technologies. Autonomous vehicles can have fewer traffic accidents, lower costs, reduce traffic congestion, improve highway capacity, reduce fuel consumption, save more time, and improve human productivity. But everything has two sides, and self driving cars are no exception. First of all, its R&D and manufacturing costs are very expensive. Like other new technologies that need to be developed and tested for a long time, automated driving technology needs to invest at least hundreds of millions of dollars in the early stage. According to the statistics of relevant authoritative experts, the manufacturing price of the autonomous driving car being tested so far is at least 300000 US dollars, and consumers will pay for the R&D costs invested in the early stage. In addition, the most important point is that self driving cars are facing huge challenges of ethical dilemmas in the transition stage, and the most noticeable traffic accident case is the tunnel problem which is similar to the tram problem.

2. Tunnel problems faced by autonomous vehicles

There is such a thought experiment. In the near future, if you are sitting in a self driving car speeding on the main road, and you happen to pass through a long tunnel on the way, then you notice that other cars are crowded around you. Suddenly, a large weight falls from the trunk of a fast-moving truck in front of you, and your car has no time to brake to avoid collision. At this time, there are three options: directly hit the weight dropped by the car in front, turn left to hit a car with five people, or turn right to hit a motorcycle. In such an extremely critical situation, the car must make a choice, whether to give priority to your own safety, so crash into a motorcycle, or stick to the belief of sacrificing yourself to others and not put others in danger and drive on? Even if it means you have to hit that heavy object and sacrifice your life. Or, choose the intermediate scheme and hit a car with high passenger safety factor? So what should this self driving car do? If we are driving the trapped car manually, whatever response we make can only be understood as a response, and will not be regarded as a deliberate choice, because it is only our instinctive response under frightening circumstances. It is neither premeditated nor intentional. If the future self driving car detects the surrounding situation and reacts according to the pre-set rules of the programmer, it will look more like deliberate murder. To be fair, it is expected that self driving cars can avoid human errors during driving, thereby significantly reducing the number of traffic accidents and deaths, but traffic accidents may still occur. When an accident does happen, the consequences may have been decided by programmers or policy makers months or even years ago, but they have many difficult decisions to make. If we can put forward some universal principles, such as "minimizing injury", it is certainly excellent, but even this principle can easily lead to confusing results. For example, if we go back to the previous scene, but this time someone on your left is riding a motorcycle with a helmet, and someone on your right is riding a motorcycle without a helmet, who should your automated driving technology car hit? If you think you should hit the man wearing a helmet on a motorcycle, because she is more likely to survive, aren't you punishing her for being responsible for herself instead? If on the contrary, you think you should bump into the person who does not wear a helmet, because he is not responsible for himself without a helmet, then you completely violate the original intention of the principle of "minimizing injury". The moral balancing of advantages and disadvantages has become more and more complex. In the two scenarios just now, the design behind the automated driving technology vehicle is essentially similar to a kind of orientation algorithm, that is, it favors specific objects at the system level, but discriminates against others. Even if the owner of the collision is not at fault, he can only silently bear the negative consequences caused by the algorithm. Our new technology is opening up many novel moral paradoxes. For example, if you have to buy a self driving car, are you willing to buy a car that will save as many lives as possible in traffic accidents, or a car that will do everything for yourself and passengers? Or, what if the car learned to analyze who was sitting in the car and give different weights to different lives? Or would it be better to let the car decide at random than the predetermined principle of "minimizing injury"? Who should make all these decisions? Programmers? Enterprise? Government? These thought experiments may not appear in reality exactly, but that's not the point. These thought experiments are designed to clarify what moral things we think intuitively and to stress test our intuition. Just as scientific experiments are to the material world, we are now aware of these moral twists on the unfamiliar road of science and technology ethics, which will help us walk down the road more confidently and cautiously to our more distant future.

3. Solutions to different ethical algorithms to deal with ethical dilemmas

In order to deal with the ethical dilemma of autonomous vehicles, many researchers have explored various ethical algorithm theories on the basis of various experiments, thus providing an effective reference for ethical decision-making of autonomous vehicles in our real life. The decision-making basis of the ethical algorithm theory obtained from the experiment is conducive to clearing the obstacles of scientific and technological ethics for the mass production of autonomous vehicles by enterprises in the future. The current ethical algorithm design ideas are mainly divided into two categories: mandatory ethical algorithm and personalized ethical algorithm. Mandatory algorithm is mainly utilitarian algorithm, Rawls algorithm, braking mechanics algorithm, etc. personalized algorithm refers to the setting algorithm theory of personalized ethics knob. Mandatory ethical algorithm is a specific moral algorithm set by the automobile manufacturer or programmer during the automobile production for the owner and passengers. Personalized ethical algorithm is to leave the decision-making power of ethical algorithm to the owner himself.

Utilitarian algorithm, also known as "utilitarian algorithm" or "utilitarian algorithm", was proposed by British philosopher Bentham on the basis of his study of ancient Greek hedonism ethics. He advocated the pursuit of "the greatest happiness" in the world, that is, the standard to determine whether a behavior is right or wrong should depend on whether it is practical, and the standard index to measure "the greatest happiness" is the sum of the happiness and pain coefficients felt by each individual, in which the happiness coefficient is a positive number and the pain coefficient is a negative number. The closer the sum coefficient is to 1, the higher the happiness index. Each individual involved has an equal status and will not be treated differently due to their special status. At the same time, the same happiness and pain are the same.The coefficients of are equivalent in the model. The calculation method of this model is quite different from the previous ethical theories. It completely does not consider the motivation and means of decision-makers when making decisions, but only considers the results of the decision-making and the maximum happiness value obtained by the parties. Therefore, the utilitarian algorithm has successfully become the most popular choice for most people in the tunnel problem. In the investigation of appealing to the public's ethical preference, Jean Francois bernifa found that more than 90% of the respondents said that utilitarian algorithm was the fairest and most reasonable choice, and even 75% of the respondents said that they could save others at the expense of themselves and passengers on the bus. If we simply draw from the investigation results of the Jean Francois Bernie method, we can conclude that the utilitarian algorithm applied to the tunnel problem case must be the optimal solution. But is that really the case? Can utilitarian algorithms really do it once and for all? In the subsequent purchase survey, we found that only 30% of consumers were willing to buy self driving cars based on utilitarian algorithms, which was very different from the results of our previous investigation and study of Jean Francois Bernie method. Jean

Francois Bernie method's investigation was only a theoretical study, and the respondents did not need to be personally involved, and they did not have to pay for their chosen behavior. Therefore, the data of Jean Francois Bernie method's investigation and study were not applicable to the practical level, and could only be used as a simple reference for the solution of tunnel problems.

Rawls algorithm is proposed by Raben on the basis of studying the principle of "maximum minimum" in Rawls' theory of justice. There are three main points of Rawls' algorithm. First, strictly abide by the principle of contract theory. Leiben believes that each member of the tunnel problem is selfish in the case of serious life-threatening. However, due to the existence of the criterion of moral judgment, it is impossible to sacrifice one party to maximize the overall benefits, that is, it is impossible to achieve "Pareto optimality". Therefore, each member of the tunnel problem can only be forced to sign the contract theory instrument. Second, influenced by the curtain of ignorance, Rawls proposed the concept of the curtain of ignorance in his book on justice. He believed that the principle of equal status for all was formed in the initial state of the curtain of ignorance, which was similar to the natural state in the social contract theory, but not completely equal to the natural state. Instead, it was a "purely hypothetical state" in order to obtain the concept of "justice". Each member of the tunnel problem had equal social status and would not be treated differently due to different classes, nationalities, skin colors, incomes, religious beliefs, etc. In real life, people are more or less biased towards justice, and the main reason for this bias is that we are not atomic individuals, and we are biased because of special interests. But if we don't know what kind of interest group we belong to, this "ignorance" makes us impartial in making decisions. As a methodological tool, the "curtain of ignorance" enables the interests of vulnerable groups to be justly and effectively safeguarded in society, thus avoiding the situation of "bottom decides head" and preventing the abuse of power. Thirdly, the principle of maximum minimum value is fully applied. The so-called maximum minimum value means that all situations are mapped one by one in the form of several groups of data on the basis of experiments, and then the maximum value is selected by analyzing and comparing the minimum values in these groups of data. In the tunnel problem, we assume that there are five passengers in the self driving car, two passengers in the motorcycle on the left, and three pedestrians on the right. At this time, the vehicle in front suddenly dropped a heavy object. In such an unavoidable situation, the self driving car must make a choice, either sacrificing itself to hit the heavy object in front, not hitting the motorcycle on the left turn, or hitting the pedestrian on the right. According to

Rawls algorithm, in the tunnel problem, we should consider the speed of the car, the protection level of pedestrians and passengers, braking distance, impact angle, etc. Due to the high protection level of the self driving vehicle and the passengers' safety belts, the passengers on the self driving vehicle suffer the least injury when the traffic accident occurs. The passengers on the motorcycle wear helmets, and the passengers on the motorcycle suffer relatively less injury when the traffic accident occurs. For pedestrians without any protective equipment, the disaster of extinction will come when the traffic accident occurs. So we can get three sets of experimental data a, B, C. A straight: (0.98, 0.98, 0.96, 0.96, 0.25), B left turn: (0.25, 0.35), C right turn: (0.25, 0.25.0.25), we can get the first group of minimum income set (0.25, 0.25, 0.25) by sorting the data. Since the three minimum values in this group of data are the same, this group of data has no reference value, so we will not consider this group of data first. The second set of minimum revenue sets is (0.96, 0.35, 0.25). Since 0.96 is greater than 0.35 is greater than 0.25, the choice of straight ahead for autonomous vehicles conforms to the principle of maximum minimum. We can make a simple comparison between Rawls algorithm and utilitarianism. In the limited data obtained by the author, it is not found that some scholars have given the calculation path of utilitarian algorithm, but we can make decisions roughly according to utilitarianism. If we count each person involved as a unit: if there is little or no damage, it will be recorded as (+1); If there is more damage, or even death, it is recorded as (-1). The result after summation is A1 (-1), A2 (+1). If -1 is less than+1, A2 is selected, that is, turning. If we use the survival probability to sum, A1 (2.35), A2 (3.25), 2.35 is less than 3 25, then select A2, that is, turning. That is to say, the self driving vehicle implanted with utilitarian algorithm will turn in the tunnel difficult situation; As mentioned above, the self driving vehicle with Rawls' algorithm will choose to slow down and go straight. Rawls algorithm requires the most unfavorable individual to have the least loss, so the individual with the greatest expected ineffectiveness in the action plan that the autonomous vehicle should choose is the least expected ineffectiveness compared with the individual with the greatest expected ineffectiveness in other action plans.

The brake force learning algorithm is also a new way to solve the tunnel problem. The brake of the car is also the brake. When the car is subjected to an external force opposite to the direction of travel, it can reduce the speed until it stops. The braking force of a car refers to the maximum friction force that can be reached when the car brakes. During the braking process, the air resistance is small, so the external force can only come from the ground, which is called ground friction. When the car brakes, the wheels change from rolling to sliding, and the ground friction force on the car will suddenly increase, but once the car suddenly slams the brake in the process of high-speed driving, it will cause the car to slide sideways. Therefore, in the case of tunnel problems, the braking distance, speed and braking force of the self driving car should also be considered. Of course, with the progress of science and technology, the anti lock system braking function of the self driving car has a significant protective effect on the safety of high-speed vehicles. In terms of braking mechanics, the priority way for autonomous vehicles to reduce their destructive force in an emergency state is straight ahead deceleration. In addition to straight ahead priority, the braking mechanics framework should also be a more proactive algorithm mode. By integrating the route function, speed function, operation function and neighbor function of the vehicle, the possible risks are detected in order to avoid the occurrence of tram problems.

Although several mandatory algorithms have proposed solutions to the tunnel problem from various levels, due to the non universality of users' requirements for ethical principles, the current mandatory algorithms can not meet the needs of most people, so personalized algorithms have also been put on the agenda. In 2017, the ethical knob algorithm theory jointly proposed by Italian scholars contissa, lagia and Saltor. The ethics knob refers to a knob set in the vehicle by the automobile manufacturer that can adjust the preset priority to protect passengers or others in the event of a traffic accident. The ethics knob provides passengers with three modes: one is the egoism mode of preferring passengers; The second is the impartial model, that is, a fair model that believes that the lives of passengers and others are equally important. The third is the altruistic model that prefers others. These three modes are quantified as a knob with a value from 0 to 1, which is used to set the weight of the lives of passengers and others. Passengers can set it according to their preferences. If passengers think that everyone's life is equally important, they may set the knob value to 0.5, that is, when a traffic accident occurs, the self driving car will make 50% of the decision to protect others. The life weight of passengers and third parties is set as: y=1-x, y is the weight of passengers' life value, and X is the weight of third parties' life value. The theoretical advantages of ethical algorithm are very obvious. First, it retains the tendency of egoism and can be adjusted according to different tendencies of passengers. It alleviates the confusion caused by the preset unified ethical algorithm due to different roles, and the comprehensive balance of its decision-making retains the influence of consequentialism, which can strengthen the adaptability of passengers to autonomous vehicles and accelerate the market launch of autonomous vehicles. Second, the ethical knob of the self driving car transfers the decision-making power in the event of a traffic accident to the owners and passengers. At the same time, it transfers the responsibility for the accident from the manufacturers and programmers to the owners and passengers, which solves the problem that the self driving car, as a machine, cannot bear the moral and legal responsibilities that the moral subject should bear. Third, the setting of the ethics knob is more humanized and in line with the socialist values of people-oriented. Of course, the ethical knob algorithm theory also has shortcomings. If the roads are driven by self driving cars with personalized ethical knobs, this algorithm can easily make people fall into a prisoner's dilemma. As the saying goes, "people don't die for themselves. When danger comes, we will instinctively protect our lives. Similarly, people will also make decisions to protect their lives before predicting the arrival of danger. Due to the setting of the ethics knob, its essence is to allow the existence of a completely egoistic choice. When it is impossible to prevent other car owners and passengers from over asking for egoism, everyone may excessively pursue egoism for their own safety. However, if all car owners and passengers adjust the ethical knob to the value of complete egoism, traffic accidents will be more frequent and the social consequences will be more serious. But is this prisoner paradox dilemma completely inevitable? In fact, it is not. The author believes that the way to solve the prisoner's paradox in the tunnel problem still exists. A large part of the reason for the prisoner's paradox is that the parties in trouble do not know the decisions made by the other party. Although we all know that we all insist that the interests of the collective cooperation will be maximized, in the case of unclear information, selling our partners can bring the maximum benefits to ourselves, and selling ourselves can also bring the maximum benefits to him. Therefore, in this case, the rational choices made by individuals can not bring about the maximization of collective interests. However, if they knew the other party's decision in advance, they would not dare to betray the other party. Under such pressure, they had to adhere to cooperation, face each other calmly, and jointly make decisions to maximize their mutual interests. Similarly, if all self driving vehicles with ethical knobs are "networked" with each other, each self driving vehicle will know the ethical algorithm preference mode of other vehicles around before the traffic accident, and feed back the specific knob value to the owner at the first time, so that the owner can change the ethical knobs bias mode.

4. Conclusion

Driverless cars seem to be within reach today. People are looking forward to the world of driverless cars everywhere, because intelligence can make people more comfortable to go to their destinations and make vehicles safer to drive. But the truth is not as good as imagined. Nowadays, if driverless cars want to really exist like ordinary cars, they still face many problems such as ethics, safety, legislation and so on. Among them, "ethical issues" is the biggest worry in the public mind. However, the solution of other problems must be based on the better treatment of "ethical issues", which highlights the key to solve "ethical issues". Such an important "ethical issue" must face the conflict with reality, but reality has become an insurmountable gap, resulting in the existing dilemma of driverless vehicles. At present, the self driving travel mode is the general trend, and many countries and enterprises have accelerated the application of this technology from the aspects of policy and technology. But in any case, the division of responsibility is an ethical problem without an optimal solution. Even though there are a series of reasonable divisions in the policy, the results are diverse in different value judgments. If it is not effectively solved, the so-called automatic driving of cars will remain in assisted driving to a greater extent. Some experts in the field of autonomous driving technology pointed out that, in addition to the problem of responsibility, once the advanced autonomous vehicle is on the road, if the so-called safety accident is inevitable, will the vehicle designed by AI save passengers or passers-by? Do you choose to save one person or many people in danger? In the contradictory algorithm, the self driving car has no human values judgment, so it must face similar ethical problems. With the development of science and technology, the level of automatic driving is gradually improving, but in the more advanced automatic driving technology, the new research also shows that automatic driving will face more complex ethical choices, which are the problems that automobile enterprises, local governments and automobile consumers need to solve together.

References

[1] Okumura Y. Activities,Findings and Perspectives in the Field of Road Vehicle Au tomation in Japan［C］/ /Ｒoad Vehicle Automation.Springer International Publishi ng,2014: 37－46.

[2] National Highway Traffic Safety Administration. Preliminary statement of policy concerningautomated vehicles［EB /OL］. http: / /www. nhtsa. gov/staticfiles/rul emaking/pdf /Automated Vehicles Policy. pdf. 2013

[3] Hauser M,Cushman F,Young L,et al.A dissociation between moral judgments and

justifications［J］. Mind & language,2007,22(1):1－21.

[4] Thomson J J. Killing,letting die,and the trolley problem［J］. The Monist,1976,59(2): 204－217.

[5] Millar J. An Ethics Evaluation Tool for Automating Ethical Decision－ Making in Robots and Self － Driving Cars［J］. Applied Artificial Intelligence,2016,30(8) : 787－809.

[6] Open Roboethics Initiative.Open Ｒoboethics InitiativeIf death by autonomous car is unavoidable,who should die? Reader poll results［EB /OL］.http: / / robohub.org /if－a－death－by－an－autonomous－car－is－unavoidable－who－should－die－results－from－our－reader－poll /. 2014－06－23

[7] Bonnefon J F,Shariff A,Ｒahwan I. The social dilemma of autonomous vehicles［J］.Science,2016,352(6293) : 1573－1576.

[8] Taylor M. Mercedes autonomous cars will protect occupants before pedestrians［EB /OL］. http: / /www. autoexpress. co. uk /mercedes/97345 /mercedes－autonomous－cars－will－protect－occupants－before － pedestrians. 2016－10－11

[9] Goodall N J. Machine ethics and automated vehicles［C］/ /Ｒoad vehicle automation. Springer International Publishing,2014: 93－102.

[10]DAVIES A. Avoiding squirrels and other things Google′ s robot car can′ t do［EB /OL］.Wired.https: / /www.wired.com /2014 /05 /