Frank Kelly Freas

United Nations Ethics Office

Study Guide

Singularity Model of United Nations
SMUN2030

Table of Contents

Welcome Letter

First of all, we would like to thank you for your participation in this Model of United Nations SMUN2030. This will not only be a great boost for your academic curriculum but also a great opportunity for you to make new friends who share the same interests with you, and therefore expand your social links all across the globe.

Nowadays the spread of technology is present everywhere. From food to transportation, we are all living around new kinds of automation that evolve in an everyday basis. This edition of SMUN is based on the fact that science is evolving faster than never, and consequently our lives are doing so as well. Because of that, this simulation will take place in 2030, making the topics proposed for every committee a challenge for you, the delegates. You will have to place yourselves in 10 years, and think how the solution for the topics would be if we were in 2030.

Welcome to the future, and welcome to S'MUN2030!

Chair Biography

Chair Director:

My name is Adrià Gambús Crespo. I'm 17 years old and I live in Barcelona, Spain. Although I lived here for almost my whole life, I lived in Athens for two years. It was a great experience that made me open my multicultural barriers and offered me a great opportunity to meet and make great friends from throughout the world, that also allowed me to learn and practice my English. Now I'm studying 1r de Batxillerat, the Spanish equivalent to 11th Grade at Pare Manyanet Les Corts. I would like to study Physics as a career major, because I love science and how it can explain everything about everything, and as I loved math and physics since I was a kid I am planning to study it outside Spain.

In my free time I like listening to music, playing guitar and reading. I love collecting albums and records, playing and going to concerts.

In 2019 I took part in MUN as a delegate, and it was a great experience. The simulation provides you a great opportunity to realize how a day of an UN ambassador is helps you explore the insights of International Relations. I was in the Security Council, and my Chair Director was great.

I hope I can be as good as expected!

Chair Asssitent:

Hello delegates, my name is Martí Serra Figarola and I am an eighteen-year old student of Philosophy, Politics and Economics in Barcelona. Although I usually spend all my day in this wonderful city, I actually live in a little village by the sea called Vilassar de Mar. Because of this my favorite sports are those that imply water, even though I also love football. Furthermore, I also enjoy listening to music, going to the cinema or reading. Regarding my intellectual interests I am fascinated by all the major fields in social sciences as well as by moral philosophy and international relations. As I am sure it happens to lots of you, I do not know what do I want to work as in the future. Despite this fact, I am convinced that whatever it is that we will be working as, we will have to put our everyday efforts in constructing a better society. It is also because of this objective that I have previously participated several times in Parliamentary simulations and in MUN as a delegate.

But this time it is your turn. During these three days you will have to work in teams, dialogue and reach consensus. Along with all these things you will also be able to make new friends and maybe discover new interests. Finally, you will write, I am sure, a great resolution that will contribute to achieve a better society in the future.

Good luck to all of you.

Introduction to Committee

The UN Ethics Office is the organization in charge of the regulation and promotion of an ethical environment based on UN's core values of integrity, professionalism and respect for diversity. This office also takes part on the values outlined in the Code of Ethics for UN Personnel which include independence, loyalty, impartiality, integrity, accountability and respect for human rights. The Ethics Office assists the Secretary-General ensuring that all the UN staff members perform their functions according to the standards of integrity, as it is required by the Charter of the United Nations. This Office makes sure a secure confidential environment is provided to the staff, so topics regarding ethical issues can be consulted, and to proceed to the search of protection against retaliation for reporting misconduct.

The main points of the UN Ethics Office are:

- Its independence from management and all other UN offices;
- Its impartiality on how it treats individuals;
- The maintenance of the confidentiality of the information entrusted to it;
- The professionalism dealing with stakeholders.

The Office provides all the necessary resources for those who seek advice before engaging in an activity, in order to avoid and manage different types of conflicts (interest-based). The office helps staff carry out their jobs in a professional and fair way by providing clear and action-oriented advice. This Office also tries to help the UN workers to manage their private lives in a way that does not interfere with their official and professional duties. All staff are expected to cooperate, and act following the Ethics Office code of action and provide access to records and documents if requested.

The Ethics Office, stablished in 2006, was stablished to secure the highest standards of integrity of the UN staff members, according to the Article 101, paragraph 3, of the Charter of the United Nations, also taking into considerations paragraph 161 of the 2005 World Summit Outcome and pursuant to General Assembly resolution 60/248.

The Office carries out its work following five strategic functions:

1. Advice;
2. Protection against retaliation;
3. Financial disclosure;
4. Ethics training;
5. Coherence of ethical standards.

Topic: The Trolley problem and its regulation

History of the topic

The Trolley Problem, proposed in 1967 by Philippe Foot, defines the moral of the person trying to solve it. Because of that, it can be applied to the programming of automatic systems such as driverless cars. The measuring on the levels of morality in a case in which a human life has to be taken by force is extremely necessary in the process of creating a car such as an DC, because the real-life conflicts described in the Trolley Problem may appear during the use of an DC, and therefore they must be covered before the commercialization of an autonomous system.

This does not only apply to cars, there are no boundaries on Artificial Intelligence Ethics. As the main goal of Artificial Intelligence is to mimic the human behavior (and, finally, surpass the performance of a human), the ethical methodology of a human brain must be introduced in AI robotics. As you will see further into this Study Guide, the situation of AI Ethics is currently in a birth-state.

Countries have different opinions, laws and legislations over Artificial Intelligence. Your main goal is to investigate, research and work on them in order to merge it with the Trolley Problem.

Definition of Key Terms

Trolley Problem   This moral and hypothetical problem defines the patterns of how we divide right from wrong. It was introduced in 1967 by Philippa Foot, and it helps to illuminate the landscape of moral intuitions. The most common version of the trolley problem is the following one:

A runaway trolley is heading down the tracks toward five workers who will all be killed if the trolley proceeds on its present course. Adam is standing next to a large switch that can divert the trolley onto a different track. The only way to save the lives of the five workers is to divert the trolley onto another track that only has one worker on it. If Adam diverts the trolley onto the other track, this one worker will die, but the other five workers will be saved. (Switch case)
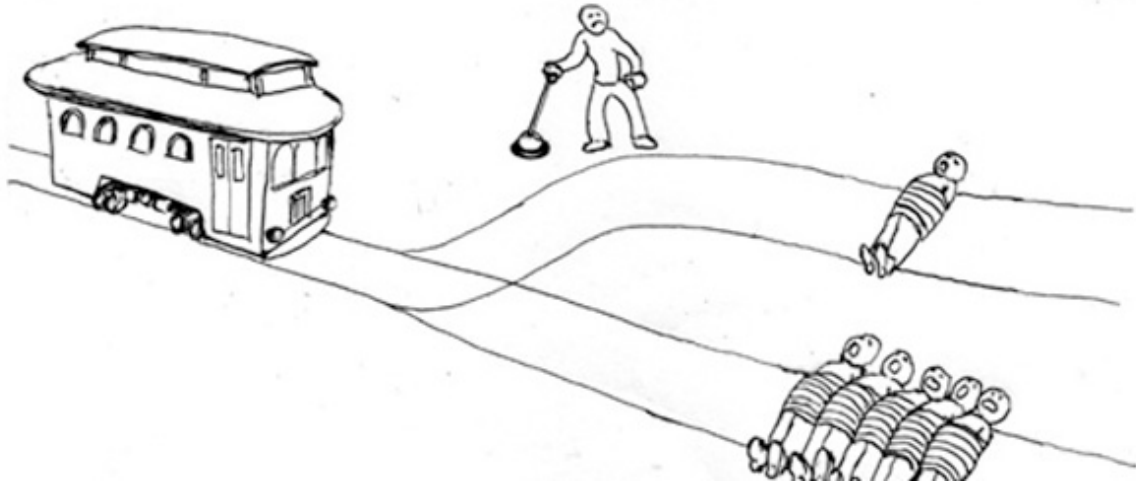
Autonomous car   An autonomous car is a vehicle that can guide itself without human conduction. This type of vehicle has become a concrete reality of a futuristic system where computers learn the art of driving, and eventually will take over standard cars.

Partially Observed Markov Decision Process (POMDP) → is a combination of an MDP to model system dynamics with a hidden Markov model that connects unobservant system states to observations. The agent can perform actions which affect the system (i.e., may cause the system state to change) with the goal to maximize a reward that depends on the sequence of system state and the agent's actions. However, the agent cannot directly observe the system state, but at each discrete point in time, the agent makes observations that depend on the state. The agent uses these observations to form a belief of in what state the system currently is. This belief is called a belief state and is expressed as a probability distribution over the states. The solution of the POMDP is a policy prescribing which action is optimal for each belief state.

Markov Decision Process (MDP) → is a discrete time stochastic control process. It provides a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker. MDPs are useful for studying optimization problems solved via dynamic programming and reinforcement learning.

Luddite   A person opposed to new technology or ways of working.

Timeline of Conventions, Declarations and Treaties

During the previous years, ethics and transparency have been two issues that have been developed by the UN organs. Because of that, some of them will be highlighted:

- Annex 2 of Resolution 70/305 on the Revitalization of the work of the General Assembly contains the Code of Ethics for the President of the General Assembly, will strengthen the capacity of the President of the General Assembly to exercise his or her duties and responsibilities whilst enhancing her moral authority, integrity and credibility,
- Members of the ACT (Accountability, Coherence, Transparency) Group circulate a non-paper on elements for consideration as part of a Code of Conduct regarding Security Council action against genocide, crimes against humanity and war crimes. ACT expanded on the work of the Small Five (S5) initiative which aimed to improve the transparency of the UNSC by suggesting that the P5 should explain why the veto has been employed in each situation.

Apart from this, we should mention the General Assembly resolution 60/248, where the Ethics Office was created. This regards the resolution of the bodies of the United Nations tackling ethical issues.

Current situation

- **Utilitarianism perspective vs Deontological perspective.**

This other version of the trolley problem should be considered:

A runaway trolley is heading down the tracks toward five workers who will all be killed if the trolley proceeds on its present course. Adam is on a footbridge over the tracks, in between the approaching trolley and the five workers. Next to him on this footbridge is a stranger who happens to be very large. The only way to save the lives of the five workers is to push this stranger off the footbridge and onto the tracks below where his large body will stop the trolley. The stranger will die if Adam does this, but the five workers will be saved. (Footbridge)
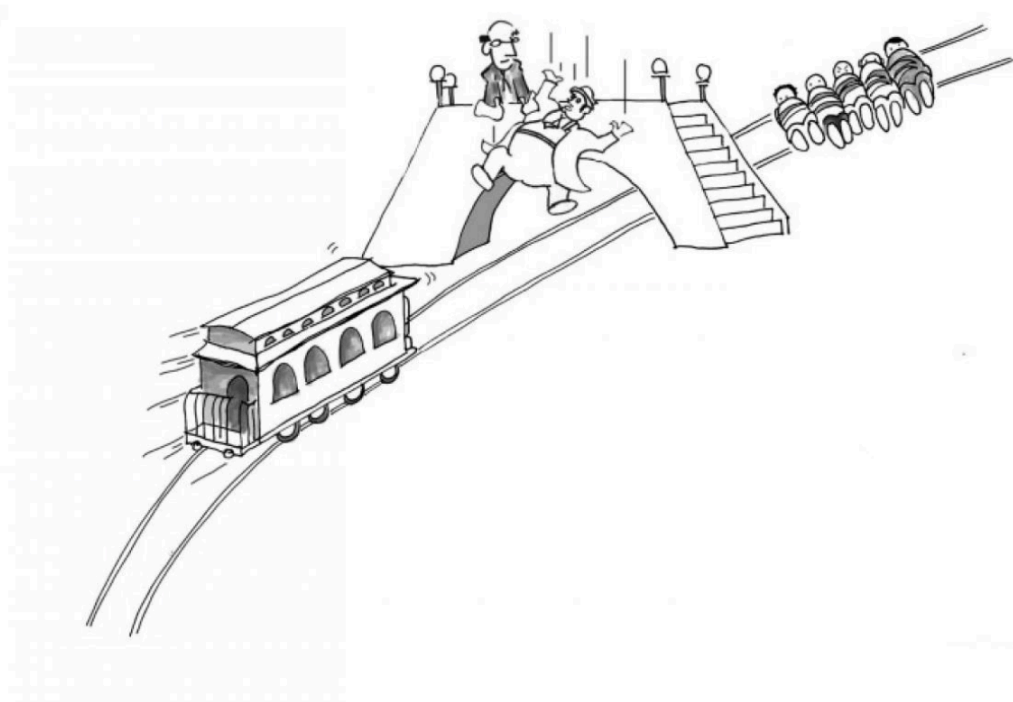
If the two versions of the same problem (the one defined on the Key Terms and the one exposed here), it is very possible that the solutions would be different in each case. The Trolley Problem helps to explain and highlight the tension between two branches of moral thought. The utilitarian perspective dictates that most appropriate action is the one that achieves the greatest good for the greatest number. Meanwhile, the deontological perspective asserts that certain actions – like killing an innocent person – are just wrong, even if they have good consequences. In both versions of the trolley problem above, utilitarians say you should sacrifice one to save five, while deontologists say you should not.

Psychological research shows that in the first version of the problem, the utilitarian version is the one mostly agreed, and then morally accepting to use the switch and kill own instead of five lives. But then in the second version of the problem, the deontological way of thinking is the one leaned, and therefore it's not acceptable to push a stranger to death without any consent, and then killing five instead of one.

Scientists think that, generally, the western moral intuitions evolved to make us good social partners. Because we learn from a very young age that violence towards others is typically punished, our moral intuitions tell us it's wrong to take actions that physically harm others. So, in versions of the trolley problem that involve physical contact, like the footbridge case above, harming one to save many is generally less acceptable than in versions that do not involve such contact, like the switch case.

- **The Thomson solution**.

Judith Jarvis Thomson is the scientist that has done more over the past forty years to draw the attention to the Trolley Problem. Even though the problem was created by Philippa Foot, Thomson developed the best-known solution.

First solution:
   o The bystander may flip the switch, because he makes what was threatening to five come to threaten only to one. Also, he does not by any means do so in order to constitute the infringement of any right of the one's.

The First Solution reflects a reality that many rights-based moral theories followers sympathize with, because it is based on minimizing the harm wherever possible. However, people's rights prevent them from being able to sacrifice a human life to do so.

Because of that, the First Solution had to be rejected. Thomson then offers a Second Solution: the bystander may turn the trolley because all six workmen belong to a group that at some point in the past the individual interest of the workers wanted the bystander to turn the trolley in order to save himself.

- **Artificial Intelligence and the Trolley Problem.**

The Trolley Problem seems like a perfect fit to be applied to autonomous cars. The Problem can help us develop the working and thinking of autonomous cars. Moreover, it can help in the measuring of the human influence over their decision-making processes, and the real ethical issues that face those advocating the advancement and deployment of autonomous cars in cities and towns all over the globe. Basing on the modern pathways that are being defined every day, we should try to focus on technology-based problems instead of runaway trolleys, bystanders and killing workers. Once we have that, then we are able to see how new ethical questions arise. This, being more complex and nuanced than their antecessors, provides an aid into the creation of an ethical and moral based future.

Autonomous systems like autonomous cars work using algorithms. The most commonly-used algorithm is a Partially Observed Markov Decision Process (POMDP). This is a variant of the Markov Decision Process (MDP). The MDP is a mathematical model used in control and planning problems proposed in the fields of engineering and computing. For example, the MDP is fairly used in an environment described as *fully observable*, has *discrete time intervals* and *few choices of action* in various conditions. This kind of models are useful when describing or programing a game of chess or tic-tac-toe. The algorithm knows the environment that it is going to be working on (the board, pieces and rules) and waits for

the opponent's move. Once that move is made, the algorithm calculates all the potential moves and then taking the "best" or "optimal" decision or counter.

The thing is that a real-world environment is not a tic-tac-toe or a chess game. A robotic system cannot have the complete knowledge of its environment, even if the system has a huge number of sensors connected to it. This is due to limitations in the range and fidelity of the sensors, latency between the reading and displaying of information (by the time that the sensor sends the information, the environment may have changed). Moreover, a robot in this kind of situation makes a decision based on the current observations, as well as a history of previous actions and observations. In a more precise way, a system is measuring everything it can at a specific state (s), and we define the finite set of states as

$S = \{s1, ..., sn\}$ in the environment. When a system observes itself in s, and takes an action named a, it moves to a new state, s', and then is able to take action $a2$ (s', $a2$). This shows. that the set of possible actions is $A = \{a1, ..., ak\}$. Then, at any given point, a system is deciding which action to take based on the present state it is at the moment of decision, its prior state (if there is one) and the expected future based on the transition between the future and the previous one.

Packing it all up, the main difference between a POMPD and an MDP is that in a POMPD, the system is in an environment where it has incomplete knowledge, therefore it works through probabilities. An autonomous vehicle does not work from an MDP, so it is more likely to be working with a POMPD.

- **AI Machines and the General Ethical Aspect.**

Nowadays AI machines surround us in our everyday life. As AI's impacts permeate our societies, the power from its transformation must be put at the service of people and the planet. At the same time, AI is also fueling new ethical concerns, never thought of before. These questions concern the trustworthiness of AI systems, the dangers of codifying and reinforcing existing biases (gender, race, etc.), infringing human rights and values, such as privacy. The concerns regarding AI grow exacerbating inequality, climate change, market concentration and the digital division. No single country or anyone has all the answers to these challenges. Therefore, an international co-operation response to guide the development and use of AI for the wider good is needed.

- **The ethics of automated jobs.**

Although the Luddite movement ended up a long time ago, some people still have a sense of fear when it comes to technology and the automation of jobs. With the development of AI systems, there is currently a general debate based on the possible "steal" of job from machines. There are reports that show why people have this fear. According to the data

from a report, up to 800 million jobs (20% of the global workforce) could be lost due to automation by 2030. For the first time, humans will start competing with machines on a cognitive level. Because of the fact that AI machines are able to compile ad learn way faster than humans, many economists are concerned that the society will not be able to adapt itself to this AI revolution, and ultimately left behind.

The ethical question that arises here is why we do not try to provide individuals that might lose their jobs in the future reachable alternatives, rather than just completely forgetting about them.

Past UN Actions

As this is a very recent issue and the use of AI in a daily-basis has just arrived to the society, the Unit Nations has not declared any state or has not made any action or declaration. It is expected that you make the first resolution regarding the Ethics in AI, taking in mind the time-situation (this MUN edition takes place in the future, 2030) you will have to write on.

Evaluation of current policies

In 2017, Patrick Lin defended the use of the Trolley Problem as an "intuition pump" to make us think about what sorts of principles we ought to be programming into AVs (autonomous vehicles). Thought-based experiments like this "isolates and stress-tests a couple of assumptions about how driverless cars should handle unavoidable crashes, as rare as they might be. It teases out the questions of whether numbers matter and whether the killing is worse than letting die." He also notes that due to the fact that driverless cars are a human creation over time, meaning that it is still not finished, "programmers and designers of automated cars […] do have the time to get it right and therefore bear more responsibility for bad outcomes," developing a resolution on whether there was enough intentionality for the act of being judged as morally right or wrong.

It is accepted to agree with Lin's affirmation, stating that, in some cases, philosophical situations cannot be applied to real-life situations, but in order to isolate and press upon our intuitions this does not mean that they are well-suited for all kind of purposes and scenarios. As Peter Singer notes, reducing "philosophy…to the level of solving the chess puzzle" is rather unhelpful, for "there are things that are more important". We need to take special care to see the asymmetries between cases like the Trolley Problem and algorithms that are not moral agents but make morally important decisions.

The easiest way of seeing this is to acknowledge that an AV utilizing a POMDP-based system in a changing environment is not making a decision at one point in time, but

making a series of sequential decisions based one on another. An AV makes a choice based on a distribution of probabilities about what act or choice will give the situation the highest reward function (or minimize the most cost) based upon prior knowledge of the environment, present observations and probable future states. This differs from the Trolley Cases, where there is one decision to make at one point in time, and therefore this is a different operation in comparison on how autonomous cars operate.

If the thinking that the Trolley Problem offers little guidance on the wider social issues at hand is agreed, then the acknowledge of a wide-ranging issue that society faces and will face with ACs can be acquired. As Kate Crawford and Ryan Calo explain, "autonomous systems are [already] changing workplaces, streets and schools. We need to ensure that those changes are beneficial, before they are built further into the infrastructure of everyday life." In a short term, the identification of the values that are needed to be actualized through the engineering, design and deployment of AI technologies, such as self-driving cars need to be identified. Because of all this, there is a double-sided work here: it is obvious that the software run in these systems will be trying to maximize their value functions, but it is also needed the verification that they are maximizing the value function of society too.

Guiding questions

- What are your country's policies about AI?
- How is the Ethics department in your country structured?
- What are the past actions that your country has done regarding ethics and, more specifically, AI?
- What is your country's policy about autonomous systems?
- What is the main goal of the Trolley Problem?
- How can the Trolley Problem be applied to Ethics in robotics?
- What is the next step into the appliance of the Trolley Problem to technology?
- Is there any expert from your country that has done a statement about the Trolley Problem?

Bibliography

- "United Nations Ethics Office." United Nations, United Nations, vwww.un.org/en/ethics
- Crockett, Molly. "The Trolley Problem: Would You Kill One Person to Save Many
- Others?" The Guardian, Guardian News and Media, 12 Dec. 2016, www.theguardian.com/science/head-quarters/2016/dec/12/the-trolley-problem-would-you-kill-one-person-to-save-many-others
- Peter A. Graham. Journal of Ethics and Social Philosophy. November 2017. Thomson's Trolley Problem. https://doi.org/10.26556/jesp.v12i2.227 Vol. 12, No. 2
- Crockett, Molly. "The Trolley Problem: Would You Kill One Person to Save Many Others?" The Guardian, Guardian News and Media, 12 Dec. 2016, www.theguardian.com/science/head-quarters/2016/dec/12/the-trolley-problem-would-you-kill-one-person-to-save-many-others
- OECD (2019), Artificial Intelligence in Society, OECD Publishing, Paris, https://doi.org/10.1787/eedfee77-en
- Bărbuță, Ionela. "The Trolley Dilemma in AI." Strongbytes, Strongbytes, 11 Feb. 2019, strongbytes.ai/the-trolley-dilemma-in-ai/
- Bărbuță, Ionela. "The Trolley Dilemma in AI." Strongbytes, Strongbytes, 11 Feb. 2019, strongbytes.ai/the-trolley-dilemma-in-ai/