Málik, József Zoltán[1]

Ethical Considerations in Digital World

I. Introduction

The Hollywood science fiction movie "*I, the Robot"* was released in 2004. The story is based on Isaac Asimov's same-title book of short-story collection published in 1950, but the screenplay is connected immediately to none of his original stories. In short about the plot: We are in 2035, humanoid robots serve humanity. Del Spooner, a Chicago police detective is investigating a new case about the dead of the co-founder of U.S. Robotics Corporation, who worked as a researcher at the think-tank departure of the firm. The suspect is one of humanoid robots created in quantity production, but this is impossible in principle because humans are protected from the robots by the Three Laws of Robotics inputted in the mind of all robots. Nevertheless, Detective Spooner is sure in the guiltiness of the robot, he hates and distrusts robots because one of them rescued him from a car crash in the past, leaving a young girl to die because her survival was statistically less likely than his.

The Three Laws, originally formulated by Asimov, are as follows:

(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 the orders given 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.

So legally a robot cannot be charged with homicide, since, at worst, any incident caused by robots places within the realm of an industrial accident. And here comes a turn in the plot, which brings up a moral dilemma. As by Law (1) robots cannot bear a human being to come to harm, however humans tend to self-destruction, a Virtual Interactive Kinetic Intelligence, called briefly as VIKI in the film, which centrally governs humanoid robots used in households (their name is NS5), may totally redefine the Three Laws, and may instruct the NS5 robots to protect humanity even if some humans must be sacrificed. This instruction sounds something like a version of *volonté general* ("general will") introduced by Jean-Jacques Rousseau, the known French philosopher. Though the Three Laws of Robotics seem at first sight as laws for robots serving humans, VIKI could turn them inside out at a higher intelligence level, and so robots have become the enemies of humans. Is VIKI not a new "technological" Robespierre, and is this machine-governed world not a new Jacobin dictatorship?

Okay, I might be said this is just a fantasy. True, we could be sceptic all about humanoid robots, but "machine ethics" is not merely a science fiction even today. For example, engineers insist that driverless train systems are safe – what is more, safer than human drivers, in fact.[2] Despite advances in technology, pubic including passengers have always been sceptical about driverless trains and autonomous agents (artificial devices) in general; in the same way and for the same reasons as the detective was distrustful about humanoid robots in the movie. I wonder if all these feelings are absolutely irrational. Consider

[1] PhD in Political Science, PhD in Philosophy. Lecturer at Faculty of Law and Political Science, Eötvös Loránd University, Budapest. E-mail: cosmos@freemail.hu.

[2] Colin Allen–Wendell Wallach–Iva Smith: "Why machine ethics?". IEEE Intelligent Systems 21(4), 2006. pp. 12–17.

now a well-known moral dilemma called as Trolley Dilemma.[3] A runaway trolley is approaching a fork in the tracks. If the trolley runs on its current track, it will kill a work crew of five. If the driver steers the train down the other branch, the trolley will kill a lone worker. If you were driving the trolley, what would you do? And what would a driverless train do? We should face the music: with the development of the Information Technologies (ITs), human users are more and more in interaction with software or autonomous agents embedding in different ITs devices. And so, consciously or not, human users may delegate part of their decision power to these entities. In this respect, the situation is not so different than one about driverless train systems and can be seen as a precursor of the machine-governed world. Increasing the scope of the activities of autonomous (artificial) agents is getting a major issue in our digital society and raises the question of dealing with moral decisions. In this paper I make an attempt to give a brief insight into the topic and to present some decision-making models that allow to reason on several ethical principles and choice assessment.

II. What do we mean by ethical behaviour?

How should we live? What makes an action right or wrong? How should you treat others or yourself? These are fundamental questions which philosophers have argued about for thousands of years. Ethics is the study of morality. It is a branch of philosophy that tries to specify what things in life are morally good and which actions are morally right. So ethics deals with morality, but it is not the same as morality. Morality consists of the standards that an individual or a community has about what is right and wrong or good and evil. Generally speaking, moral standards deal with matters to which we attach great importance because they involve serious harm or injury to others or to oneself. Ethics "studies" these kinds of moral standards in a normative fashion. In philosophy ethical theories are generally divided into three main camps: consequentialists, deontological, and teleological (or virtue-based stance). They are three different types of normative theory and provide moral frameworks to guide and evaluate our decisions regarding what we ought to do.

Consequentialist theories require that we make decisions according to the state of affairs that will result from our actions. Even the unintended consequences of our actions must be taken into consideration. There are different variations on consequentialism, but each focuses primarily on the effects of actions rather than on the actions themselves. For example, one possible consequentialist argument might be that torturing a suspected terrorist during interrogation is justified, and even required, if the information to be gained is likely to save a thousand people. In other words, moral judgement on what is right and wrong relies on weighing the projected benefits that an action will produce against the possible harms. Utilitarianism is one prominent type of consequentialist position, according to which the end of human action is happiness. Though it derives moral standards from human nature, altogether the end, happiness, can be achieved not through private viewpoints but through an ensemble of people's viewpoint. This is the principle of utility introduced by Francis Hutchenson: *„that Action is best, which procures the greatest Happiness for the greatest Numbers; and that, worst, which, in like manner, occasions Misery."*[4] Two forms of this principle are worthy of distinction:

- the direct usage of the principle (Jeremy Bentham's utilitarianism), if the effects of actions are immediate to individual utility and the well-being of community;
- the indirect usage of the principle of utility (John Stuart Mill's utilitarianism), when it is just a subsequent test of assessing actions and policies.

[3] Judith Thomson: Killing, Letting Die, and the Trolley Problem, In: The Monist. No. 59, 1976. pp. 204-217.
[4] Francis Hutcheson: An Inquiry into the Original of Our Ideas of Beauty and Virtue. In: W. Leidhold (ed.): Natural Law and Enlightenment Classics. Liberty Fund, Indianapolis, 2004. pp. 125.

According to a very different class of moral choice, some acts are wrong in themselves, regardless of their consequences. Killing one innocent person, or engaging in torture, for example, is unconditionally wrong and cannot be justified even if the act would save a thousand others. Such a view is referred to as "deontological", a label derived from the Greek word *deon*, meaning "duty". For the deontologist, torture and the killing of innocents, for example, are unacceptable means to pursuing any end regardless of how noble. Morality requires that one's motives and the means that one deploys are good. In deontological ethical theory, actions are evaluated in and of themselves rather than in terms of the consequences they produce.

It is Immanuel Kant who is an outstanding exponent of deontological ethics.[5] Kant claimed that autonomy – the freedom to choose for oneself what one will do and the reasons on which one will act – is the heart of ethics. To let something or someone else decide what one will do is "heteronomy", and is wrong because moral action should depend on one's own will. The will is a person's ability to make decisions on the basis of reasons; a good will is one that chooses what is morally right because it is right and not because it is enjoyable or in one's self-interest. Kant believed that since a good will is good without qualification, we should strive to have a good will. But what is a "good will"? The person with a morally good will is the person who does what is right because he believes it is his moral duty to do it. Kant proposed a categorical imperative, which in its first form states a*ct only on that maxim which you can at the same time will to be a universal law*.[6] Thus, for example, it would be wrong to make a promise with the intention of breaking it. If everyone made promises with the intention of breaking them, then no one would believe in promises. The action would be self-defeating.

This implies that everyone is of equal value. In Kant's words, everyone has the same "absolute value", an idea he also expressed by saying that everyone is an "end in himself". Because everyone is of equal value, no one should be used to serve the interests of another without their consent, but must always be treated as having an "absolute value". Kant summarized these ideas by restating his categorical imperative in these words: *Act so that you always treat people as ends in themselves, and never merely use them as means*[7]. All in all, for Kant the motive of an action is far more important than the action itself and its consequences. He thought that in order to know whether or not someone was acting morally you had to know what their intention was. Hence, it was not enough just to know whether or not the Good Samaritan helped the man in need. The Samaritan might have been acting out of self-interest, expecting a reward for his troubles. Or else he might have done it only because he felt a twinge of compassion: this would have been acting from an emotional motive rather than from a sense of duty.

The emphasis in Western ethics has been on duty since Kant, on defining ethics in terms of what actions one is obligated to do. However, there is a tradition in ethics that goes back to Plato and Aristotle that looks at ethics in terms of virtues. Virtue theory is largely based on Aristotle's Nicomachean Ethics[8] and as a result is sometimes known as neo-Aristotelianism. Unlike Kantians and utilitarians, who typically concentrate on the rightness or wrongness of particular actions, virtue theorists focus on character and are interested in the individual way of life. The question here is "What shall I be?" rather than "What shall I do?" In the virtue-based ethical tradition there are four cardinal virtues: wisdom, courage, temperance, and justice; and from these primary virtues all other virtues can be derived.

[5] Immanuel *Kant: Grounding for the Metaphysics of Morals. [Translated by James W, Ellington]. Hackett, London, 1993.*
[6] Ibid. pp. 30.
[7] Ibid. pp. 36.
[8] Aristotle: Nicomachean Ethics, Book 2, Chapter 1.

However, in modern days, virtue-based theories often are turned into deontological rules for actions. That is, one is asked to act wisely, courageously, temperately, and justly, rather than being wise, courageous, temperate, and just. In the context of "machine ethics" this remark seems extremely true.

Figure 1:
The four cardinal virtues

Nevertheless, both consequentialist and deontological ethical theories have deficiency. For some, the consequentialist mode of reasoning is excessively permissive and cannot provide a guide to moral action. According to these critics, the problem is that any action – including a harm like killing innocents or engaging in torture – can be morally justified based on its projected consequences. Even though an appeal to consequences could also be made in opposition to such actions, many people find a mode of argument that could (even contingently) permit such acts unacceptable. However, consequentialism can also be criticized for being excessively demanding: all acts are either morally forbidden or required; and, calculating the potential consequences of each act can be an extremely difficult, if not impossible, task. Just imagine having to judge every choice as either right or wrong based on its multiple, and possibly long-term, consequences.

As for deontological theory and Kant, there are also some fundamental troubles. First, duties frequently conflict, and Kant's theory does not seem to give us an obvious way of resolving such conflicts. If, as Kant argues, it is always wrong to tell a lie and always wrong to break a promise, then which do I choose when these duties conflict? Second, the acts that the categorical imperative says are always wrong do not always seem wrong. For example, Kant says that it is wrong to lie, no matter what good might come of telling the lie. Yet is it wrong to lie to save your life? To save someone from serious pain or injury? There seems to be no compelling reason why certain actions should be prohibited without exception.

There still remains one important question when we want to talk about ethical behaviour of artificial agents.[9] Would a machine that behaves ethically actually be ethical? Would a robot that follows a program and thereby behaves ethically, actually be ethical? This question seems similar to the one raised by John Searle in his *Chinese room argument*:[10] would a computer that can hold a conversation in Chinese really understand Chinese? Or, in a more general sense, does a creature need to have free will to behave ethically? Well, in my mind, we may at this moment drop the epistemological side of the question, what we need in applications is to simulate ethical behaviour in a similar way as a chess program simulates sense of the game in playing chess. Of course, chess programs work very differently than

[9] James Gips: "Towards the Ethical Robot". In: K. Ford, C. Glymour and P. Hayes (eds.): Android Epistemology. MIT Press, Cambridge, Massachusetts, 1995. pp. 243–252.

[10] John Searle: "Chinese room argument". Scholarpedia, 2009. 4(8):3100. Availability: http://www.scholarpedia.org/article/Chinese_room_argument

human brain works (chess programs are based upon rapid searching method for good moves, instead of human intuition and practice), but in practice who cares if they simulate well to play the game – and they become very good by now, might as well be able to defeat human chess world champion.[11] Or, take Trolley Dilemma again. Does a human really make a better choice than a robot pilot? One thing is nevertheless clear that we should be more sophisticated if we try to implement ethical behaviour in a computer program. The famous computer scientist, Donald Knuth is probably right in that

"It has often been said that a person doesn't really understand something until he teaches it to someone else. Actually a person doesn't really understand something until he can teach it to a computer, i.e., express it as an algorithm. [...] The attempt to formalize things as algorithms leads to a much deeper understanding than if we simply try to understand things in the traditional way."[12]

In spirit of Knuth an important achievement of scholars in ethical studies could be to help computer scientists and engineers become aware of their work's ethical dimensions.[13] In this work three steps seem essential to identify: (1) To define some formal decision-making models as frameworks that allow to reason on several ethical principles and situation assessment; (2) To define methods to detect ethical conflicts; and (3) to provide a kind of conflict management whose results can be explained by an autonomous agent. What follows in the rest of this paper is to come up to step (1) by presenting some formal decision-making models based on different ethical stances in which we can demonstrate how to carry out some ethical assessment.

III. The consequentialist model of ethical behaviour

To be able to reason ethically along consequentialist lines, our autonomous agents could have: (i) A way of describing the situation in the world; (ii) A way of generating possible actions; (iii) A means of predicting the situation that would result if an action were taken given the current situation; (iv) A method of evaluating a situation in terms of its goodness or desirability. A possible model of doing this is Expected Utility Theory (EUT).

We first specify our uncertainty as a list of possible "states of the world" (each state is denoted ω_i, and is a member of the set of possible states Ω). We then have a list of possible actions and possible states, which together form the set of possible consequences (c_{xs}). A simple example is the following: You have to choose whether or not to bring an umbrella when you go for a walk (x_1: bring umbrella, x_2: not bring umbrella). There are two possible "states of the world": ω_1: it will rain, ω_2: it will not rain. Cross-tabulating these, we have the following four possible consequences:

Table 1.: EUT decision matrix

		Possible states (Ω)	
		ω_1: Rain with probability $\mathbf{p_1}$	ω_2: No rain with probability $\mathbf{p_2}$
Possible Actions **(X)**	x_1 (Bring an umbrella)	c_{11} (it rains and you have an umbrella)	c_{12} (you brought the umbrella, but it does not rain)
	x_2 (Do not bring an umbrella)	c_{21} (it rains and you did not bring an umbrella)	c_{22} (you did not bring the umbrella and it does not rain)

[11] This happened first in the history of chess in 1996, between Garry Kasparov and the Deep Blue. The machine plays automatically, though the IBM team determined the opening moves (just the first two moves) played by Deep Blue, which were a less frequent version of the Sicilian Opening in which Kasparov was very proficiency.
[12] Donald Knuth: "Computer Science and Mathematics". American Scientist, Vol. 61, No. 6. 1973. pp. 709.
[13] See again Allen–Wallach–Smith (2006).

The expected utility of an action is calculated as follows: multiply the utility of each possible consequence of an action by the subjective or given probability (p_1 or p_2) that the consequence will occur. Formally in our example:[14]

$u(\boldsymbol{x_1}) = v(\boldsymbol{c_{11}})\, p_1 + v(\boldsymbol{c_{12}})\, p_2$ and $u(\boldsymbol{x_2}) = v(\boldsymbol{c_{21}})\, p_1 + v(\boldsymbol{c_{22}})\, p_2$.

Or, more generally: $u(x_j) = \sum_i v(c_{ji})\, p_i$.

Maximization of expected utility then simply means that you choose the alternative that has the highest expected utility when it is calculated in the way described above.

Richard Jeffrey presents a decision model that does not assume the independence of the values of probabilities and consequences from chosen actions.[15] We can accomplish it by altering EUT: one has to calculate conditional probabilities instead of using elementary probabilities in Table 1. We thus have

$$cu(x_j) = \sum_i v(c_{ji})\, p(\omega_{ji} \mid x_j)$$

where $p(\omega_{ji} \mid x_j)$ is the probability of the events of the world given the fact that the decision maker chooses action x_j.

To illuminate Conditional Expected Utility Theory (CEUT), consider the following example with the below decision matrix:

Table 2.: CEUT decision matrix

		Possible states (Ω)	
		ω_1: One`s lifetime is more than 65 years.	**ω_2**: One will be dead before the age of 65
Possible Actions **(X)**	**X_1** (One smokes)	c_{11} (one`s lifetime more than 65 years provided he smokes)	c_{12} (one will be dead before of the age of 65 provided he smokes)
	X_2 (One gives up smoking)	c_{21} (one`s lifetime more than 65 years provided he gives up smoking)	c_{22} (one will be dead before of the age of 65 provided he gives up smoking)

It is clear in this example that the states of the world are not specified in advance, but the possible actions impact on them. In order to calculate the expected utility of an action, we have to investigate the expected probabilities of the events of the world and the expected consequences. We have two matrices coming from the original matrix to consider:

Table 2.1.: Probability matrix

		Possible states (Ω)	
		ω_1: One`s lifetime is more than 65 years.	**ω_2**: One will be dead before the age of 65
Possible Actions **(X)**	**x_1** (One smokes)	$p(\omega_1 \mid x_1)$	$p(\omega_2 \mid x_1)$
	x_2 (One gives up smoking)	$p(\omega_1 \mid x_2)$	$p(\omega_2 \mid x_2)$

Realize that, supposing the smoking surely noxious to the decision maker's health, $p(\omega_1 \mid x_1) \neq p(\omega_1 \mid x_2)$ and $p(\omega_2 \mid x_1) \neq p(\omega_2 \mid x_2)$, i.e., the probabilities in the same column can be different.

[14] We will use, following Hirschleifer and Riley, $v(\cdot)$ to denote the utility of consequences and $u(x)$ to indicate the utility derived from an action. [In: Jack Hirschleifer–John Riley: The Analytics of Uncertainty and Information. Cambridge University Press, Cambridge, 1992].

[15] Richard Jeffrey: The Logic Of Decision. McGraw-Hill, New York, 1983.

Table 2.2.: Consequence matrix		*Possible states (Ω)*	
		ω_1: One`s lifetime is more than 65 years.	ω_2: One will be dead before the age of 65
Possible Actions **(X)**	$\mathbf{x_1}$ (One smokes)	c_{11}: $x_1 \wedge \omega_1$	c_{12}: $x_1 \wedge \omega_2$
	$\mathbf{x_2}$ (One gives up smoking)	c_{21}: $x_2 \wedge \omega_1$	c_{22}: $x_2 \wedge \omega_2$

Hence, the expected utility of an action is calculated as follows: multiply the utility of each possible consequence of an action by the conditional probability that the consequence will occur. Formally in our example:

$cu(x_1) = v(c_{11})\ p(\omega_1 \mid x_1) + v(c_{12})\ p(\omega_2 \mid x_1)$

$cu(x_2) = v(c_{21})\ p(\omega_1 \mid x_2) + v(c_{22})\ p(\omega_2 \mid x_2)$.

Similarly to EUT, we can define maximization rule as "maximize conditional expected utility".

Within CEUT, the position of decision-maker has also been changed: while he has to assess only the consequences of his actions within EUT, he now has to consider whether to act or not, and thus this consideration may influence the evaluation of his action. In other words, the essential difference between the two approaches is that EUT describes a norm-neutral situation, whilst CEUT is committed to a normative stance, which is due to the evaluation of action. Logically, this means the following requirement for

- EUT: $\forall x_j \forall c_{ji}\ [v(c_{ji} \wedge x_j) = v(c_{ji})]$,
- CEUT: $\exists x_j \forall c_{ji}\ [v(c_{ji} \wedge x_j) \neq v(c_{ji})]$.

To complete the utility of principle, consequentialist evaluation schemes have the following form: $\sum w_i\ u_i$ where w_i is the weight assigned each person and u_i is the utility values for each one in the community. In Bentham's utilitarianism, the weight for each person is equal and the u_i is the amount of pleasure, broadly defined. What should be the distribution of the weights w_i across persons? i) For an ethical egoist, the weight for himself in assessing the consequences would be 1; the weight for everyone else would be 0. ii) For the ethical altruist, the weight for himself is 0; the weight for everyone else is positive. iii) The utilitarian ideal is the universalist, who weights each person's well-being equally.

IV. To incorporate the deontic thread: A decision model from multiple-value perspectives

In everyday life several problems are very complex; they can be so confusing that an average person is unable to make a good (ethical) decision. Usually we need a multi-dimensional choice over alternatives in these kinds of situation. The question is now what the relevant factors are with which we must count in the model. In the context of economics Kenneth Goodpaster has presented the most sophisticated model of responsibility-driven choice.[16] Goodpaster distinguished two basic components of moral responsibility: rationality and respect. Rationality means here four main attributes: (1) Lack of impulsiveness; (2) Care in mapping out alternatives and consequences; (3) Clarity about goals and purposes; and (4) attention to details of implementation. Note this is not the rationality postulate of standard economics, instead the procedural concept of rationality, or bounded rationality á la Herbert Simon. Respect means a special awareness of and concern for the effects of one's decisions and policies on others, which involves taking their needs and interests seriously. It is what

[16] Kenneth E. Goodpaster: "The Concept of Corporate Conscience". Journal of Business Ethics, No. 1, 1983. pp. 1-22.

Kant meant by the "categorical imperative" to treat others as valuable in and for themselves. To sum up briefly, in Goodpaster's model responsible choice is to combine rationality and respect for others in decision making. In this model respect is basically, if not exclusively, a consequentialist account, i.e., the decision-maker considers the effects of his choice on the stakeholders.

We repeat again, as it was done above, that in complex choice there might be marginal contributions, unforeseeable consequences, and distant effects, and they may create decision traps if the choice is based solely on consequentialist considerations. Goodpaster's responsibility model should be enlarged to include the deontological aspect of choice. Hence responsible ethical choice can be defined as a synthesis of deontology, rationality (as goal achievement), and respect for stakeholders.[17] In other words, we should assign some deontological (D(A)), instrumental (G(A)), and stackholders' value (S(A)) to the feasible alternatives in set of A. The question is how can the value functions D(.), G(.), and S(.) be defined?

Let multiple-values for actions $x_1, x_2, ..., x_m$ be represented by vector as follows: $v = [D(x_i), S(x_i), G(x_i)]$, where

- $D(x_i)$ is the deontological payoff for chosen action x_i: $D(x_i) = \sum_{k=1}^{p} v_k D_k(x_i)$, and

- the deontological payoff for x_i can be determined with respect to ethical norm D_k as follows:

$$D_k(x_i) = \begin{cases} 1 & \text{if } x_i \text{ corresponds to ethical norm } D_k \\ 0 & \text{if } x_i \text{ is neutral with respect to ethical norm } D_k \\ -2 & \text{if } x_i \text{ contravenes ethical norm } D_k. \end{cases}$$

- w_k are weights that show the relative significance of ethical norms $D_1, D_2, ..., D_p$ related to one another, and $\sum_{k=1}^{p} v_k = 1$.

- $S(x_i)$ is the payoffs for stakeholders: $S(x_i) = \sum_{l=1}^{q} w_l S_l(x_i)$, where

- the stakeholder's payoff for actions x_i can be determined as follows:

$$S_l(x_i) = \begin{cases} 1 & \text{if action } x_i \text{ is good for stakeholder } S_l \\ 0 & \text{if action } x_i \text{ is neutral for stakeholder } S_l \\ -2 & \text{if action } x_i \text{ is bad for stakeholder } S_l \end{cases}$$

- weights v_l show the relative significance of stakeholders $S_1, S_2, ..., S_l$, where $\sum_{l=1}^{q} w_l = 1$.

- $G(x_i)$ is the goal-achievement value of actions:

- the goal-achievement value for actions x_i can be determined as follows:

[17] Amartya Sen: On Ethics and Economics, Blackwell, Oxford, 1987.

$$G_m(x_i) = \begin{cases} 1 & \text{if action } x_i \text{ is positive for the realization of goal } G_m \\ 0 & \text{if } x_i \text{ is neutral with respect to the realization of goal } G_m \\ -2 & \text{if action } x_i \text{ is negative for the realization of goal } G_m \end{cases}$$

- weights z_m show the relative significance of decision maker's goals G_1, G_2, ..., G_r related to one another, where $\sum_{m=1}^{r} z_m = 1$.

Then we have an overall view of a complex decision situation by constructing the following matrix containing the assessments of the decision maker's actions, x_1, x_2,..., x_m:

$$v = \begin{pmatrix} D(x_1) & . & S(x_1) & . & G(x_1) \\ . & . & . & . & . \\ D(x_i) & . & S(x_i) & . & G(x_i) \\ . & . & . & . & . \\ D(x_m) & . & S(x_m) & . & G(x_m) \end{pmatrix}$$

Again, it is true that it is not easy for decision-maker many times to resolve which action to choose, for the values of D, S and G can be conflicted with each other. This is a sort of value-collision. Levi suggests to decision maker that he should apply lexicographic rule, i.e., he should choose the best action in the value-dimension he has assessed as the best.[18] In complex choices, the application of lexicographic rule therefore seems rather problematical, because it reduces the complexity of choice to uni-dimension.

Another possible decision rule is the so-called *maximin rule* that means the choice of the least bad action in a 3-dimensional space, (D, S, G): $A^{responsible}$ = *maximin* [$D(x_i)$, $S(x_i)$, $G(x_i)$]. [19] The action chosen in this way ends with Pareto-optimal solution, in the sense that there is no other admissible action would be better than the chosen action in any dimension of space (D, S, G), however it will not be worse than the others in one of the other dimensions, either.

A Demonstrative Example: Why to fasten safety-belt?

We know the costs of car crashes can be minimized by using safety-belts, and so the gain of protective measures from the installation of safety-belts in cars is greater than its costs. However, no much more than only 50% on average of drivers and passengers are ready to fasten their seat belts routinely after getting in. What is then the reason for ill-behaviour, and how to assess the choice in the models presented above?

1. To assess the choice by CEUT

Take the consequence matrix related to the choice, in which the fact that driver indifferent if the usage of safety-belts is good or bad is displayed:

[18] Isaac Levi: Hard Choices. Cambridge: Cambridge University Press, 1986.
[19] László Zsolnai: "Moral Responsibility and Economic Choice". International Journal of Social Economics, No.4, 1997. pp. 355-363.

Table 3.: Consequence matrix		*Possible states (Ω)*	
		ω_1: The driver survives the crash.	ω_1: The driver will not survive the crash.
Possible Actions **(X)**	$\mathbf{X_1}$ (Fasten the belt)	2	-2
	$\mathbf{X_2}$ (Do not fasten the belt)	2	-2

Now, if the fact that the costs of car crashes can be minimized by using safety-belts in 95 out of 100 cases, and in case of death it does not matter if the driver has fastened himself or not, then the conditional utility values are as follows:

$$cu(x_1) = v(c_{11})\, p(\omega_1 \mid x_1) + v(c_{12})\, p(\omega_2 \mid x_1) = 2 \cdot 0{,}95 - 2 \cdot 0{,}5 = 0{,}9$$

$$cu(x_2) = v(c_{21})\, p(\omega_1 \mid x_2) + v(c_{22})\, p(\omega_2 \mid x_2) = 2 \cdot 0{,}05 - 2 \cdot 0{,}5 = -0{,}9.$$

So the usage of the seat belt is rational. Yet the assessment is a little bit *ad hoc*, because the result may be altered with other probabilities. As we plausibly believe in the majority of drivers are not willing to commit a suicide, i.e., they prefer occurring ω_1, the different conditional probabilities in the assessment by CEUT are the reason for drivers' apparent irrationality.

2. To assess the situation by the responsibility-driven choice model

If the fact that the costs of car crashes can be minimized by using safety-belts is assumed, the ethical norm is that the usage of seat belts is obligatory during driving a car for safeguarding passengers. Two possible actions are: x_1 = "Safety belts are fastened in the car" and x_2 = "Safety belts are not used". Then, in deontological sense, x_1 is good and x_2 is bad evidentially, hence $D(x_1) = 1$ and $D(x_2) = -2$. Let the goal be that passengers will survive the trip; then action x_1 serves it with probability p and action x_2 serves it with probability q, i.e., $G(x_1) = p$ and $G(x_2) = q$. In this model condition $p \leq q$ says the same fact presented in CEUT model that violator drivers assume the independence of the possible actions and the occurred states of the world. The stakeholders, that is, the passengers are indifferent between fastening their belt or not: $S(x_1) = 0$ and $S(x_2) = 0$. Hence, the assessments of the decision maker's actions, x_1, x_2, are two vectors: $v(x_1) = (1, 0, p)$, $v(x_2) = (-2, 0, q)$. Comparing them, action x_1 is better than action x_2 by maximin rule.

So the assessment of the choice has been altered compared to CEUT in the respect that it has become independent of probability. But this is possible solely because of introducing deontic stance based on a moral postulate, namely, the usage of safety-belt is required.

V. Deontic Logic and Moral Conflicts

Deontic logic is the formal study of the normative concepts of obligation, permission, and prohibition; they are called as deontic modals, and their normative concepts are similar in many respects to the modal concepts of necessity and possibility. Deontic modality is a kind of modality which has to do with what is necessary or possible according to various rules, such as the norms of morality. Deontic logic, in other words, is a modal logic where instead of the alethic modalities *necessity* and *possibility*, we have deontological modalities of *obligation* and *permission*. Alethic logic deals with what is the case ("truth") or not the case ("falsehood"), and alethic modal logic deals with what is necessarily or possibly the case as well as with what is the case or not the case. Deontic modal logic deals with what is obligatory, or permitted, or forbidden to be the case as well as with what is or is not the case.

All I have said so far would be alone enough to be interested in deontic logic, but deontic logic is pretty cool in one more respect as we can investigate moral conflicts. By the

means of deontic logic we are not only able to define a framework that allows to reason on several ethical principles, but also able to define methods to detect conflicts. That's why, and as I do not want to enter into logical formalism in this paper, I now speak about this matter a little bit more. Moral conflicts and the conditions for their existence are among the divisive elements in meta-ethics. In the philosophical literature, moral conflicts are usually studied from the standpoint of a single agent, like in Trolley Dilemma above. An agent faces a moral conflict if the agent has two moral obligations that cannot both be fulfilled. By raising the study of moral conflicts from a single-agent to a multi-agent perspective, we can generalize the concept of moral conflict: two groups of agents, G_1 and G_2, face a moral conflict if G_1 has a moral obligation and G_2 has a moral obligation, such that these obligations cannot both be fulfilled. If the two groups are identical and consist of a single agent, the moral conflict boils down to a single-agent moral conflict in the usual sense.

Each group F of agents defines a moral code stipulating that F's collective interest is to be maximized. Accordingly, the group of all agents defines the moral code of utilitarianism: an agent has a certain utilitarian obligation if this obligation stems from the moral code that the collective interest of the group of all agents is to be maximized. Moreover, an agent only accepting the moral code defined by himself is an ethical egoist, who is to maximize his own self-interest. A moral obligation is indexed by two groups of agents, G and F. G indicates the group who has the obligation. F refers to the interest group who defines the consequentialist moral code from which the obligation stems. Based upon John Horty's utilitarian deontic logic, Barteld Kooi and Allard Taminga present a consequentialist deontic logic that enables us to give a formal definition of moral conflicts.[20]

In their logical calculus deontic statements have the form "In the interest of group F of agents, group G of agents ought to see to it that φ", or in abbreviated: $\odot_{\mathcal{G}}^{\mathcal{F}}\phi$. And now two groups of agents, G_1 and G_2, face a moral conflict if and only if there are formulae φ and ψ, such that both $\odot_{\mathcal{G}_1}^{\mathcal{F}_1}\phi$ and $\odot_{\mathcal{G}_2}^{\mathcal{F}_2}\psi$ are true, whereas the necessity of φ and ψ, i.e., $\Diamond(\phi \wedge \psi)$, is false. Note that if φ and ψ cannot both be true, the truth of ψ implies the falsity of φ and, hence, having a moral obligation to see to it that ψ implies having a moral obligation to see to it that ~φ . Therefore, any moral conflict between G_1 and G_2 implies a basic moral conflict between those groups: two groups of agents, G_1 and G_2, face a basic moral conflict if and only if there is a formula φ, such that $\odot_{\mathcal{G}_1}^{\mathcal{F}_1}\phi \wedge \odot_{\mathcal{G}_2}^{\mathcal{F}_2}\neg\phi$ is true. This is the central expression in the investigation of moral conflicts, but the name of the game is roughly the same what intuition tells us: a single group of agents may face a basic moral conflict if and only if the pertinent obligations stem from different moral codes. Or equivalently, a single group of agents cannot face a basic moral conflict if and only if the pertinent obligations stem from a single moral code.

VI. Afterword

We should not forget our moral choices are an inescapable part of who we are, and is mostly justified in that how to cope with moral conflicts. Ethics is becoming a major issue in the current landscape of ITs as ITs are turning into open and decentralized autonomous decision-making systems. It is up to us how to design our autonomous agents in the future. I hope we will avoid getting a "brand new digital world".

[20] John F. Horty: Agency and Deontic Logic. New York: Oxford University Press, 2001. Barteld Kooi–Allard Taminga: "Moral Conflicts between Groups of Agents". Journal of Philosophical Logic, 37. 2008. pp. 1–21.