The Morality Menu

Oliver Bendel
School of Business FHNW, Bahnhofstrasse 6, CH-5210 Windisch
oliver.bendel@fhnw.ch

Abstract

Machine ethics produces moral and immoral machines. The morality is usually fixed, e.g. by programmed meta-rules and rules. The machine is thus capable of certain actions, not others. However, another approach is the morality menu. With this, the owner or user transfers his or her own morality onto the machine. The machine behaves in the same way as he or she would behave, in detail. The author developed several artifacts of machine ethics at his university from 2013 to 2018. For one of them, he designed a morality menu that has not yet been implemented. Another concept exists for a virtual assistant that can make reservations and orders for its owner more or less independently. In this article, the author introduces the idea of the morality menu in the context of two concrete machines. Then he discusses advantages and disadvantages and presents possibilities for improvement. A morality menu can be a valuable extension for certain moral machines.

Introduction

Machine ethics is not only a discipline of thought. It is also a design discipline (Wallach and Allen 2009; Anderson and Anderson 2011). It produces moral and immoral machines, in the form of simulations or prototypes (Anderson and Anderson 2011; Pereira 2016; Bendel 2016c; Bendel 2018b). Increasingly, it will help in the creation of products. The morality is usually fixed, for example via meta-rules and rules. The machine is thus able to perform certain actions and not able to perform certain others. Manufacturers or developers devise the rules or meta-rules, alone or with the support of guidelines, working groups and ethics committees. Further, users and consumers can show interest or disinterest in the prototypes or products (Bendel 2018b). An alternative approach, where the machine itself develops or improves morality, for example by adapting the rules, with the help of machine learning, is little pursued. Michael Anderson, Susan L. Anderson and Vincent Berenz have moralized a robot with machine learning methods. It automatically adapts to the respective situation in the care sector (Anderson et al. 2017). Neither of these possibilities is considered here.

Another approach is the morality menu (MOME). The owners or users transfer their own morality, their ideas and convictions about good and evil, their standards of value, their rules of conduct to the machine. The machine behaves in the same way they would behave, in detail. They may encounter certain default settings, but they have a certain freedom to change them or set new default settings. The ideal situation is when they do not need any programming knowledge, but can intervene in a simple and target-oriented manner.

From 2013 to 2018, the author developed four artifacts for machine ethics at the FHNW University of Applied Sciences and Arts Northwestern Switzerland (Bendel 2016a, 2017; Bendel et al. 2017). For one of them, the animal-friendly LADYBIRD vacuum cleaning robot, he has designed a morality menu that has not yet been implemented. In addition, another menu has been created for a virtual assistant in the form of Google Duplex, which can more or less independently make reservations and orders by telephone for its user. In this article, the author introduces the idea of the morality menu in the form of two concrete concepts for two different machines or systems. Then he discusses the advantages and disadvantages while presenting possibilities for improvement.

The Idea of the Morality Menu

The idea of the morality menu is that an owner or user can use it to adapt a machine – such as a domestic robot – in such a way that it largely replicates his or her moral ideas and convictions. The result is a proxy machine with a proxy morality elaborated in detail. The machine does what humans would do, in their absence and without their control. In order to be able to use the morality menu, one has to be clear about what standards of value one has and what rules of conduct one follows. It is also assumed that one wants these moral ideas and convictions to be effective in one’s own absence. Pereira and Saptawijaya have made
moral adjustments possible in a virtual simulation in which a robot is supposed to save a princess. For example, the robot can accept greater or lesser collateral damage when reaching its target (Pereira and Saptawijaya 2016). Otherwise, there are hardly any approaches to influencing morality in operations, except in the context of machine learning.

Machine ethics is usually concerned with autonomous or semi-autonomous machines, such as software robots (chatbots, social bots, software agents and virtual assistants) or hardware robots (autonomous cars, fighter planes and service robots). When normal machines become moral ones, they are moralized, so to speak. This moralization is done either by the researcher, inventor, manufacturer, developer, etc., so that the morality menu is an additional option for a moral machine, or completely by the MOME, so that only this one turns a neutral machine into a moral machine. The morality menu will also come from the researcher, inventor, manufacturer, etc., unless you design it completely open so that the user can determine each option. In this respect, it can also be located in the area of ethics by design (Dignum 2018), where these machines are differentiated from usual applications by the new alternatives that are granted.

The morality menu should be designed in such a way that one’s own morality can be represented in different aspects and to the greatest possible extent. It must fit to the respective functions of the machine. It should be easy to use because if, for example, programming skills or knowledge of a description language are required to determine a mass product in moral terms, this may be too high a hurdle and acceptance will be low, even if editors are available. It should be about moral questions to correspond to the name and idea, and less about pedagogical or psychological ones. Of course, these can be taken into account to a certain extent in the morality menu (pedagogical functions are known in smartphones to warn you if you set the music too loud).

In the two concepts developed by the author in 2018, which contain visual and textual elements, virtual sliders are used, as is the case with smartphones and tablets, to activate or deactivate the functions. To the left of them is described what happens when the slider is moved to the right and changed its color, i.e. is activated. The drafts do not thematize what possible presets are. It would be possible that all controls are shifted to the left at the beginning, but also a mixture is conceivable.

The technical implementation, especially the interaction between the morality menu and the machine or system, is not discussed in the concepts and should not play a role here either. It can be about completely different components of the machine, about hardware and software, about movement or speech functions. The purchase of the application is also irrelevant here – it can be supplied with the product or offered for sale in an app store. The smartphone, a tablet, a touch screen built into the device or an external, special display can serve as the terminal.

The LADYBIRD Project

LADYBIRD was already described in 2014 as part of a design study in the field of machine ethics. Such a design study also contains visual and textual elements and describes the basic functions of a moral machine (Bendel 2017). The idea at that time was to create a moral machine, more precisely an animal-friendly vacuum cleaning robot. This was influenced by not only machine ethics but also animal ethics, as well as animal-machine interaction (Bendel 2014, 2016b) and animal-computer interaction (Mancini 2011).

In particular, LADYBIRD spares ladybirds and other insects on the ground during cleaning operations – hence the name. The result was an annotated decision tree that described the concrete clarifications and options of the machine and – in the annotations on the nodes – contained moral justifications and assumptions (Bendel 2018a). The most important function was that the suction robot identified an insect and then stopped working until the animal disappeared or the owner commanded it to continue.

In 2017, a team of three students developed LADYBIRD as a prototype at the FHNW School of Business (Bendel 2017). The machine was able to detect red spots and objects on the floor, stop working and inform the user. It therefore concentrated on ladybirds. The aforementioned decision tree was partially used. A colour sensor played a central role (Hueber 2013). The team did not use a motion detector and was also unable – because of the educational background of the members – to take image and pattern recognition into account. This will be done in a follow-up project.

![Fig. 1: MOME for LADYBIRD (Bendel 2018b)](image)
Fig. 1 shows the morality menu for LADYBIRD. It was printed in a basic article on machine ethics (Bendel 2018b), but only briefly explained there – the idea of a morality menu was sketched, but not explained in detail.

According to the original idea, LADYBIRD detects and spares ladybirds and other insects. The menu starts with the option to kill vermin like cockroaches. The machine would have to be able to distinguish between ladybirds and vermin. Pest infestation is a problem with dwellings and may make them uninhabitable. However, a vacuum cleaner is only of limited use in combating this problem. The user would actively choose for the machine to kill – but it would do so anyway if it were not moralized. The second slider allows the user to order the killing of spiders. Many people are afraid of these animals, but in our latitudes they are usually harmless and useful, even in the home.

With the next option, the vacuum cleaning robot becomes a toy. One could further refine here, for example to distinguish between cat and dog, because each animal understands something different under a successful game (basically some pets are at war with normal vacuum cleaner robots). Ultimately, the aim is to contribute to a good life for larger animals. This is also the purpose of the next option. It is a reward system in which the user cannot actively intervene and which has been chosen quite arbitrarily here. If LADYBIRD is successful in its basic idea, the user’s money will be transferred to an animal rights organization. The welfare that the owner bestows on the animals doubles.

The next point relates to data transfer. Modern vacuum cleaning robots can generate and disseminate data on the size and condition of dwellings, inventory and occupants (Ram 2017). Vacuum cleaner manufacturers or third parties may have an interest in this. It is also possible to generate valuable information for the police and intelligence services. The user can prevent the disclosure of data, which may be seen in connection with privacy by design (Schaar 2010) and the relationship between robots and privacy (Calo 2011). The next decision also relates to data collection. Vacuum cleaning robots have a special viewing angle. They are down on the ground and record the space in front of them, sometimes above them (which can be done with cameras, but not necessarily). This allows them to generate personal data that affects people’s intimacy and privacy, as is the case with upskirting. If the slider is moved to the right, the machine becomes a discrete observer.

The last option turns the vacuum cleaning robot into a self-learning system, a moral machine that gains experience and adapts rules. Self-learning moral machines can be risky – this is explicitly pointed out to the user. However, in an environment like the home, the dangers may be extremely limited. It is conceivable to sketch the consequences in an info box, just as further information on all options is useful.

It is important at this point to reiterate that the concept is about principle. One can criticize and detail each option. The idea should be presented, not the implementation anticipated. A MOME as a product or component of a product may take a long time to be perfected for the market.

Basically, this is a hybrid system. There is a default setting where nothing can be changed and which has the consequence of sparing ladybirds, i.e. it has moral implications – while other settings can be adjusted.

**Google Duplex**

In 2018, Google introduced Duplex, which is based on Google Assistant, a virtual assistant used on smartphones and other devices (Rawes 2018). The idea is to use high-end technology to dial normal phones around the world and automate private tasks. For example, Google Duplex can call restaurants or hairdressers and reserve tables or make appointments (Dwoskin 2018). The data and the participant to be called have to be named beforehand by the user – the rest is done by the system. The voice in Google’s presentations sounded very lifelike, and the speaking in general. This is partially because breaks and “mmhs” occur as in real people. Imperfection could be the key to perfection.

Google Duplex soon found itself exposed to intense discussions, also from an ethical point of view (Wong 2018). Experts criticized the machine for not revealing that it was a machine. The whole conversation could be seen as a deception towards the called person. It was also seen critically that a social act was replaced with an automated action. Google soon improved; Duplex now reveals what it is at the beginning of the dialogue. Not least, the question arose as to whether the conversation situations were real or whether they were arranged or edited (Lindner 2018).

Fig. 2: MOME for virtual assistants (Bendel 2018c)
The author published the morality menu for virtual assistants (Fig. 2) in 2018 in an article on artificial voices (Bendel 2018c). It was not further explained and served mainly as an illustration. The article dealt with the synthetic voices of both software and hardware robots.

The first setting aims at the disclosure of machine existence. When this MOME was developed, Google wasn’t that far advanced in this respect. In the GOODBOT, a moral chatbot developed by the author and his team in 2013, transparency was firmly anchored in the system (Bendel 2016a). The user now has a choice. The second option regulates voice and speech. When the virtual assistant speaks like a machine, the user also knows what it is all about, even if he or she did not notice the disclosure or forget it. Pepper, too, has implemented a robotic way of speaking, while Alexa, on the contrary, strives for a way of talking that is as human as possible, for example by making her whisper (Myers 2017).

The compliment option regulates the behaviour of the machine in relation to a certain aspect. Compliments lead to emotional changes and attachments. Whether one should promote these in machines is again controversial and depends on the context. The limitation of the conversation time is connected with fairness with regard to the value of the lifetime – in principle, the machine could try to keep the conversation partner on the line for a lifelong period. It does not consume a lifetime, but the human being does. Of course, each of us would hang up after a while.

The data commandment is already known from the other morality menu. Also in this context personal data can result, which are worth to be protected. For example, third parties might be interested in what habits one has as a consumer or when a celebrity has dinner in which restaurants parties might be interested in what habits one has as a result, which are worth to be protected. For example, third parties might be interested in what habits one has as a consumer or when a celebrity has dinner in which restaurant.

The next option also seems to be known. Now, how the machine evaluates and connects his or her behaviour on the mobile phone. The risk of this activation is a gain pointed out. The last issue is also known from the other MOME. Again, the risk is mentioned.

Advantages of a Morality Menu

So far, the morality menu in its variants is only available as a concept. Nevertheless, it is possible to give initial assessments. In the following, the advantages of the MOME are listed and discussed fundamentally and in its two manifestations.

Awareness of Morality

The morality menu makes the owner or user aware that moral decisions are made during a process, no matter how common and everyday, and that these can also be transferred to machines. The moral decisions are named or described. An evaluation can also be provided.

With the LADYBIRD MOME, the user becomes aware when looking at the options and when moving the slider, that he or she can influence the welfare or suffering of animals. The MOME for virtual assistants such as Google Duplex makes it clear to him or her that moral questions also arise in a conversation situation and that automation has certain effects. These are sometimes difficult to assess, especially since the interlocutor may not be known.

Morality Transfer

The MOME allows the transfer of personal morality. Normal machines do not know morality at all; at least it is not explicitly designed in them. Moral machines will usually be designed as the manufacturer or developer wishes or the market demands. With the MOME, individualization in automation is possible. This is also a requirement of industry 4.0 (Reinheimer 2017).

The two concepts are partly similar in the transfer of morality, insofar as they represent the morality of the user. But they also differ considerably. In the one case LADYBIRD is made to behave in a certain way towards insects and other animals, in the other case the virtual assistant is adapted to one’s own person in order to act as a quasi-person or conduit. The proxy machine becomes the proxy person, so to speak.

Importance of Morality

The transfer of morality means that the user’s morality is taken seriously. His or her morality changes the functions of the device and has effects on humans and animals. This is psychologically relevant insofar as self-awareness is strengthened, and philosophically insofar as morality becomes the focus of reflection and action.

As the LADYBIRD MOME shows, the decisions are of an existential nature, at least from the perspective of animals. The user can cause the killing of vermin or spiders. Using the MOME for virtual assistants like Google Duplex can affect the relationship between customer and business and change the human interlocutor.

Modularization of Morality

As time goes by, when the idea spreads and is fundamentally convincing, more and more morality menus are developed and more and more rules can be identified that are meaningful and effective for many machines. Just as there are learning objects in e-learning that can be used to compile texts and tests (Boyle 2003), there could be moral objects. You select them from an overview or from a database for the respective morality menu.
There are already overlaps between the two concepts, namely with regard to data transfer and self-learning. Such overlaps could also be included in metaconcepts, which could be used in the development of all morality menus, or even in the overviews and databases mentioned above, in the sense of modularization, analysis and synthesis of morality.

Possibility of Manipulation
In general, the possibility of manipulation seems to be an advantage. Machine ethics is about the research and development of moral machines, and a morality menu is on the one hand a contribution to discipline, on the other hand a contribution to practice.

Both MOMEs are only described as examples. One can argue about the morality options that should be present, and also about whether there are not more suitable or more important ones. The MOMEs are not an end result, but a starting point for discussions in the discipline and for implementation in practice. They are part of iterative research and development.

Usability in the Group
Until now it has been assumed that a single owner or user transfers his or her morality to the machine. However, it can also be a household in which several people live. With certain machines, anyone can transfer their morality. An interesting side effect is that you may learn about the attitudes of the other members of the family or the shared flat.

With the virtual assistant, the morality of the machine can also be set during data entry. In this way, the morality of the machine meets the requirements of the respective user (not necessarily the owner). With the suction robot, this is obviously problematic, because it is supposed to do its work in the absence of direct supervision.

Disadvantages of a Morality Menu
In the following, the disadvantages of the morality menu are listed and discussed fundamentally and in its two manifestations.

Simplification of Morality
Moral attitudes can be complex. The sliders lead to a certain simplification. It is only possible to deactivate and activate a function. This initially strengthens predefined rules and hides the different consequences that arise from alternating contexts. In the extreme case, it leads to falsifications, because one actually has a position that lies in the middle between two absolutes.

Both concepts provide an opportunity to counter the simplification and dominance of the rules. Thus, self-learning potentially leads to new rules and adjustments to the system. However, this entails certain risks, which are also pointed out. In some cases, the sliders could also be used for three positions or for smooth transitions.

Delegation of Responsibility
With a moral machine, responsibility is basically given away in relation to specific decisions. The machine itself may not be able to carry any responsibility at all, at most in the sense that it performs a task (a rudimentary form of primary responsibility). It is hardly possible or reasonable to hold a machine accountable in the sense of secondary responsibility, to reprimand it and to blame it (Bendel 2018b). The human beings could get rid of their responsibility. They could also be strengthened by the MOME in the idea that they have done everything in their power and therefore see themselves as outside responsibility.

One should judge the two concepts differently in this question. LADYBIRD is out and about in a household that belongs to a person or a group. The situation is not very complex. A virtual assistant like Google Duplex, on the other hand, is directed outwards and can do more damage. It can offend people or make false reservations, which also raises liability issues. However, the MOME is also intended to counteract exactly this, although the possibilities are far from exhausted with the present concept.

Continuation of Evil
If the user is prone to damaging actions and evil deeds, the MOME allows him or her to spread and amplify the bad with the help of mechanical means, to make it exist even in his or her absence. The freedom that the MOME allows can be abused or exploited. In this sense, fixed requirements are a reliable alternative and at least prevent the worst (unless they themselves contain the worst).

This problem is particularly evident with the LADYBIRD MOME. According to the original plan, the vacuum cleaning robot should be able to spare not only ladybirds but also other insects (Bendel 2017). This would have made it a basically animal-friendly (actually insect-friendly) product, at the same time one that ignores the preferences of the user (who, however, does not have to buy it). The MOME, on the other hand, allows certain animals to be killed. The menu for virtual assistants also shows possibilities of abuse. For example, the option of not paying compliments is meant as protection from emotional attachment, but a conversation without compliments can seem emotionless and cold.

Morality Change
The previously discussed awareness of values and rules of conduct has a downside. The MOME could lead to individual morality changing into the negative. Evil is not only
changed, namely to spare ladybirds. The description and classification of reality creates a new reality.

The morality menus show the user what he or the machine has done and left so far, and give him the opportunity to continue his actions via the machine. They also show him what he could think and do in the future. He could have certain animals killed, of which he did not waste any thought before, or even leave out compliments in the conversation, in the opinion to avoid problems like the machine.

Authorship of the Rules
By defining a morality menu, options are given by the manufacturer or developer. In this way, their own morality can be mapped completely in some cases, in many cases only incompletely. The question also arises as to the criteria according to which the options are developed at all. However, machine ethics does not necessarily have to feel responsible for this (Bendel 2018b). A philosophical discipline is interested in investigating values, not in creating them. Machine ethics, which cooperates with AI and robotics, is above all interested in implanting values, whatever they may be, into a machine.

The two concepts have only partially shown possibilities in morality. One could go much deeper, could grant broader options. However, this would also make it more demanding, confusing and unfriendly to use. LADYBIRD in particular still has a fixed requirement that cannot be changed, namely to spare ladybirds.

Complexity of Technology
A MOME has to fit very different requirements. In other words, it must be developed individually for each machine, at least for each type. Different goals and levels have to be taken into account. It has already been suggested that over time, in the course of the development of many MOMEs, meta-concepts could emerge.

The two concepts show that with moral machines (as with normal machines) completely different goals and levels exist. Thus, there are machines that execute acts or actions, and machines that conduct dialogues, thus speech acts. Some of these levels can also interact with each other. For example, it is common for humanoid robots to coordinate facial expressions, gestures and language.

Usability in the Group
Here, too, in the disadvantages section, it has been assumed that a single owner or user transfers his or her morality to the machine. However, it can also be a household with different personalities. With certain machines, each member can transfer his or her morality to the machine.

With the virtual assistant, the morality of the machine can also be set during data entry. A problem already mentioned could be that responsibility and liability issues could arise, but they are now difficult to resolve because owner and user are no longer (or not necessarily) identical. With the vacuum cleaning robot, the respective change is obviously problematic, because it should carry out its work in absence without direct supervision. A dispute can arise between the members of the household. The fact that everyone has his or her own suction robot is also not appropriate, although the industry should have nothing against it.

Summary and Outlook
A morality menu can transfer the morality of the owner or user to a device or a robot. There are advantages and disadvantages on different levels. The morality menu supports the idea of strengthening the user’s morality, however it may be. The morality can be spread in the negative sense by the machine and changed in the person. This could be opposed by hybrid machines, as LADYBIRD is.

As was mentioned at the beginning, the morality menu could also be designed completely open. The user would have to formulate his or her preferences towards it, which the system then takes over. For this there would have to be an easy-to-use user interface or – mentioned elsewhere – a simple programming or markup language. A markup language, which would be suitable for the morality in the written and spoken as well as the morally adequate display of pictures, videos and animations and the playing of sounds, could be called MOML (Morality Markup Language) in the style of XML, HTML, SSML etc. Programming and markup languages could, however, also overtax individual persons in this context.

Many other approaches are conceivable. For example, the machine could allow the user to complete psychological and ethical tests and adapt itself on this basis. In addition, interviews are possible, or the user can show the machine what he or she would generally do, i.e. serve as a reference person, which was also discussed for autonomous cars (Bendel 2016). This article does not claim to have found the best solution. But it has made a worthwhile contribution to machine ethics and possibly to practice.

References