Computer Based Decision-Making: Three Maxims

by Jeff Robbins

A lone hacker uses her computer to covertly access a mainframe which controls the assets of a banking conglomerate. Within minutes, she has illegally transferred millions of dollars into her checking account.

A hospital administrator decides to replace several physicians with a diagnostic computer which is 25 percent more accurate than an average doctor in matching symptoms to a disease.

What is the fundamental similarity between these two situations? And more importantly, what is the primary difference?

The focus of this discussion is computer ethics. I shall first give a brief description of three ethical perspectives to be used in our discourse. Then I shall define computer ethics and demonstrate that most of the ``computer ethics'' discussed are superfluous, as computers are just new tools used to accomplish old goals, and the ethical questions regarding computers are unchanged from more familiar scenarios. I will then demonstrate that the area in which computers have potential to raise new moral issues is in the field of computer decision-making. I shall point out some of the new ethical issues raised by automated reasoning and derive several moral principles regarding the use of computers as decision-makers.

Three Ethical Perspectives

We will view computer ethics from three different angles. I will attempt to introduce the reader to each of them; however, my treatment will necessarily be brief.

The first viewpoint is that of **Aristotle**, who postulated that the *telos*, or function, of humans is to reason. To be good, an entity must perform its function well; therefore, a good human is one who reasons well. Through reasoning, a human may attain a state of *eudaimonia*, or human flourishment. Followers of Aristotle focus on human character and virtue as the source of morality; others may choose a rule-base approach to make moral choices.

The second perspective is that of **Immanuel Kant**, who commands us to ``Act only according to that maxim by which you can at the same time will that it should become a universal law'' [1]. Suppose I hold the following maxim: ``Anytime I want a book owned by someone else, I will steal it.'' Now, to be universalizable, my maxim must pass two tests:

1. I would not mind if everyone follows my maxim.

2. If everyone follows my maxim, no logical contradictions would exist.

My maxim fails the first test because I would not want everyone going around stealing books (especially if they were *my* books). I could not will that everyone follow my maxim; therefore, it is faulty. Looking at the second condition, we see that a logical contradiction is created if my maxim becomes a universal rule. If everyone steals a book whenever they choose, the concept of owning a book becomes null and void. If everyone made it a habit to steal books owned by others, no one would be able to own any tomes because they would all get stolen on a regular basis.

Lastly, we will consider **preferential utilitarianism**. This school of thought holds that we should always strive to maximize the preferences of everyone involved in a given situation. For example, a preferential utilitarian fire fighter might be on his way to see a movie one Saturday when he notices that a nearby skyscraper is burning. Since the preference of many people to survive is greater than the preference of one person to go to the cinema, the fire fighter would immediately rush to the aid of the people in the towering inferno. Flaming buildings are much more dramatic than a Macintosh; however, utilitarianism is a useful perspective to apply to computer ethics.

"Computer Ethics" as Presently Defined

"Computer ethics" is an over-used term. Codes of ethics for computer users forbid actions that closely resemble immoral practices that existed before the advent of computers. The *Trinity University Code of Ethics for Computing*, for example, forbids practices such as damaging university computer equipment by "pounding, kicking, or moving it to another location" [2]. Damaging another's property is generally regarded as unethical, regardless of whether the item is a computer or not.

Another action forbidden by the *Trinity University Code of Ethics for Computing* is making "unauthorized copies of software that is copyrighted," a situation that presents a severe problem to a Kantian [3]. The maxim that the ``software pirate'' acts upon might be stated as follows: ``Whenever violating intellectual property rights suits my purpose, I will violate those rights."

If everyone were to illegally copy software, the intellectual "property" is no longer property because its creator retains no control over its use. If everyone were to violate intellectual property rights, intellectual property would not exist. This ethical dilemma did not originate with computer software; it has existed for some time in the area of print media.

Accessing computer systems when one is unauthorized to do so is also cited as a violation of computer ethics. However, it is difficult to see the ethical difference between unauthorized access to a computer system and unauthorized access to a building or a car. Donald Gotterbarn succinctly expresses the misuse of the term ``computer ethics'' by stating that using a computer to swindle money is no more a

violation of ``computer ethics'' than forging a signature for illegal gain is a violation of ``ball-point pen ethics'' [4]. The hacker arranging illegal money transfers via a modem is simply playing old tricks with new tools.

Computer Ethics in Regards to Decision-Making

Computer-based decision-making is the area in which a clear code of computer ethics is needed. Computers, unlike other non-human objects in the world, can make decisions upon which the welfare of humans depends. Computer decision-making, or automated reasoning, is one function of computers. According to Aristotle, something is good if it performs its function well. A knife's function is to cut; a good knife will cut well. Computers are programmed to make decisions; we hope they will make these decisions well.

If deciding is a form of reasoning (as I shall soon demonstrate), humans are no longer unique in being the only physical beings that have rational thought as their telos, or function. As computers advance, they will be able to make more complex decisions, reducing the need for humans to make certain choices. I might have a computer decide what I should eat for supper. I desire two hamburgers and french fries; however, reason tells me a plate of steamed broccoli and a glass of skim milk would be much better for my body. A computer allowed to make this choice for me would, on the basis of rational rules regarding health, choose the broccoli for me. This is one advantage of computer decision-making; the computer makes a decision on the basis of programmed (and hopefully rational) rules without regarding other factors. A computer in charge of allocating public funds would do so according to rational economic rules; a politician might forsake reason to appease a campaign contributor. However, a danger lies in allowing computers to make too many decisions.

A person who lets the computer make all decisions would no longer have a need for rational thought. Without the opportunity to reason, humans would not perform the function of reasoning and thus could not achieve eudaimonia. The rational facility of humans acts in much the same manner as the physical part: If one does not exercise the body, it soon weakens. In a similar manner, a person who relies on computers to make most decisions would soon have a weakened or non-existent rational capacity. Since rational thought is generally regarded as a good activity, one must clearly state how to allow humans to benefit from machine decision-making without losing the opportunity for rational thought.

In the cases where computers make decisions for humans, we want them to make a ``good decision.'' The problem lies in defining a good decision. Humans, in general, prefer to live in a ``good'' world, and ``bad'' decisions will not lead to a good world. A philosopher in the preferential utilitarian school would define a good decision as one that maximizes the satisfaction of people's preferences. Aristotle would postulate that a good decision is one that leads to eudaimonia. However, this definition poses a strange contradiction: a computer that makes decisions for humans is taking away their opportunity to reason and thus flourish. A Kantian decision would have to be universalizable; a world could exist in which everyone acted upon the maxim on which the decision was made. We could define the universalizable maxim behind a good decision as ``Always decide to act in a manner which will maximize the

satisfaction of people's preferences and which does not involve treating other people as only a means to an end." This definition of a ``good decision" combines Kantian and Utilitarian theories, and decisions made in regard to this maxim would lead to a good world.

Making a good decision is a form of reasoning. Haugeland defines reasoning as the manipulation of meaningful symbols according to rational rules and states that ``according to a central tradition in Western philosophy, thinking (intellection) essentially is rational manipulation of mental symbols" [5]. This view of reasoning says that thoughts can be moved around like physical objects. Computers are token manipulators of the most rational kind; they make decisions by using fixed rules to manipulate symbols. Although the process of automated reasoning is somewhat different than human thought, I regard machine decision-making as a form of reasoning.

Three Maxims For Computer Decision-Making

Maxim One

Because machines make more decisions as they become more advanced, it is necessary to establish a clear set of principles regarding what choices computers should make. James H. Moor proposes and then discounts three ``Dubious Maxims" regarding the use of computers in decision-making [6]. The first maxim discounted by Moor is ``Computers should never make any decisions which humans want to make." Moor refutes this by providing the example of a medical diagnosis program which could do a better job than human doctors; he states that:

> If the computer's diagnosis and suggestions for treatments would result in a significant savings of lives and reduction of suffering compared with human decision making, then there is a powerful moral argument for letting computers decide [7].

From the preferential utilitarian standpoint, one could argue that because the computer would do a better job than humans, more people would be cured as a result of better diagnosis and the satisfaction of their preferences (for wellness) would be increased. However, the physicians replaced by the computer might not be happy at the loss of their power and jobs. As only the physicians' desires are thwarted and the desires of many patients for health are satisfied, it appears that in this case, letting a more competent computer make the decisions instead of a less qualified person is the ethical choice.

Kant postulated that benevolent intention, or Good Will, is an essential factor in determining whether a certain action is good or not. If we take a Kantian perspective of the medical diagnosis example, we find that the concept of a Good Will is not present in the computer. A computer cannot have a Good Will in the sense that a human does. Computer programmers write programs which ``force" the computer to help the patient by providing the best possible diagnosis. Doctors who diagnose patients themselves are (hopefully) acting on the basis of a Good Will; the computer is not. However, the computer programmers who created the computer program could also be acting from a Good Will (albeit a Good

Will removed one level from the patient). Therefore, in both cases the potential for a Good Will exists.

Another example which refutes Dubious Maxim One is in the context of law. Suppose a computer is so competent in legal matters that it can determine guilt or innocence to a higher degree of accuracy than a traditional twelve-person jury. Kantians would prefer that the computer be used in place of a jury, as the right of innocent people to be free and the right of guilty people to be punished would be fulfilled more often if the computer were used. The satisfaction of people's preferences for justice and fewer free criminals would also be fulfilled, making such a computer good from the utilitarian standpoint. Therefore, we should use the computer to determine innocence or guilt even if there are humans who would like to perform such a function, as it would be unethical to utilize the more fallible human jury. I therefore define my Plausible Maxim One as follows: ``In any case where a computer is more likely to make a good decision than a human, a computer should make the decision."

Maxim Two

We will now discuss Moor's Dubious Maxim Two: ``Computers should never make any decisions which humans can make more competently." Moor refutes this because:

> Some activities, e.g. certain kinds of factory work or prolonged space travel may be so boring, time-consuming, or dangerous that it would be morally better to use computers, even if this involved sacrificing some competency in decision making, in order to spare humans from enduring such experiences [8].

Moor's refutation of this dubious maxim also appears satisfactory. Few people have the desire to perform dangerous and boring work, even if they can perform the task competently. Moor does not, however, raise the issue of decisions which are not dangerous or boring, yet which are undesirable to make. For example, consider a Roman magistrate given the opportunity to release one prisoner condemned to death. He could release an innocent prisoner or the guilty murderer that the mobs outside his gate prefer would go free. The magistrate might appease the crowd's desires by killing the innocent prisoner, or he could free the innocent man and face the wrath of the crowd. The magistrate's desire, in this case, is definitely not to make the decision. One might argue that he should decide; since he is the magistrate, he has the duty to make a decision. Yet, if he was forced to make the decision, he might abandon the rational choice of freeing the innocent due to his appetite for political popularity. How could he make a ``good" choice without giving in to his desire for voter approval? If a computer is appointed to make our previously defined ``good decision," it would first look for the opportunity to maximize the satisfaction of everyone's preferences. The crowd wishes for the death of the innocent man, and only the magistrate and possibly the innocent man himself wish otherwise. However, a sufficiently advanced computer would realize that executing the innocent would be satisfying the crowd's desires by treating a human as a means and would thus forbid the execution of the innocent man.

Therefore, humans should not make decisions that they do not wish to make. Humans do not desire to make some decisions because they are torn between making a good decision and one which would

satisfy their desires or appetites. Aristotle tells us that most men lead ``vulgar'' lives and identify good with pleasure [9]. Even ``vulgar'' people desire good; however, when faced with the choice of satisfying desires or doing good, they choose the most pleasurable path. A vulgar person who desires good does not wish to be faced with a decision involving the choice of good or personal pleasure and would be content to let a computer make the difficult choice. On the other hand, those who live the ``contemplative life'' would prefer to make the rational choice and might also desire to make the decision without help from a machine. I now define Plausible Maxim Two as: ``When a human desires not to make a choice, a computer should make that choice.''

Maxim Three

Moor defines Dubious Maxim Three as ``Computers should never make any decisions which humans cannot override.'' His refutation of this maxim is as follows:

> Further suppose in those cases when humans override computer driving decisions, the accident rate soars. Under such circumstances there is a pervasive moral and prudential argument to have computers do the decision making and not to allow humans to override their decisions [10].

Moor's refutation of the third maxim at first sounds convincing. We certainly would not want a drunken party-goer to suddenly decide that he can drive his computer-controlled car better than the damn piece of silicon which usually does the job. The plausible maxim which Moor is implying is ``Computers should make decisions which humans cannot override, if the overriding of the computers' decision might allow a human to make a choice which results in less good.'' However, this maxim presents a problem from the Aristotelian point of view. One might decide that computers are better at making all decisions than humans (and one could postulate that they will someday actually will reach this point), and humans are thus precluded from making any decision of importance. Humans would therefore have no reason to make decisions, and would have no decisions to reason over. Humans would never be able to reach a state of eudaimonia if they could not perform their function of reasoning.

The preferential utilitarian might also feel that humans should always have the authority to override computer decisions. Since most humans value free will, the loss of final autonomy to a computer could be seen as the thwarting of a major preference and thus be viewed as unfortunate. I view this loss of human autonomy as a greater tragedy than any disaster which the superior decision-making faculties of a computer could prevent. I therefore state my Plausible Maxim Three as: ``Humans should always have the authority to override a decision reached by a computer.''

Conclusions

The area in which computer ethics needs the most discussion is that of computer decision-making. Automated reasoning creates a new type of environment in which human beings are no longer the only objects which can make major decisions that affect humans. I have developed three maxims regarding

the use of computers as decision makers, which are as follows:

1. In any case where a computer is more likely to make a good decision than a human, a computer should make the decision.

2. Computers should make any decisions that humans have no desire to make.

3. Humans should always have the authority to override a decision reached by a computer.

By these maxims, computer will be used to make decisions in cases where they perform better than humans or where boredom or danger dictates the use of a machine. However, humans will always have the ability to override the decision of a computer, thus retaining the right to reason.

References

1. Brennan, Joseph G. *Foundations of Moral Obligation*. (Newport, Rhode Island: Naval War College Press), p. 91.

2. Trinity University Code of Ethics for Computing. San Antonio, Texas (Trinity University Computing Center), p. 2.

3. Trinity University Computing Center, p. 2.

4. What is Computer Ethics? [video recording] (New Haven, Connecticut Educational Media Resources, 1992, cassette).

5. Haugeland, John. *Artificial Intelligence: The Very Idea*. (Cambridge, Massachusetts: The MIT Press, 1985), p. 4.

6. Moor, James H. "Are There Decisions Computers Should Never Make?" In *Ethical Issues in the Use of Computers.* (Ed.) Deborah B. Johnson and John G. Snapper (Belmont, California: Wadsworth, 1985), p. 128.

7. Moor, p. 128.

8. Moor, p. 128.

9. Aristotle, Nicomachean Ethics

10. Moor, p. 128.