

Can Computers Decide what is Right and Wrong?

By Jacob Palme, e-mail jpalme@dsv.su.se
CMC (Computer Mediated Communication)
K2Lab laboratory (DSV university department)

Introducing the Issue by an Example

Example: Suppose you equip a motorcar with a computer-controlled device (a breathalyzer), which makes it impossible for you to drive when your blood alcohol level is above a certain level. And suppose there is an exceptional case. John has a heart attack, and the only person available to drive to the hospital, Mary, has drunk a little too much. The computer stops Mary from driving the car, and John dies. They do not have a telephone in their remotely located summer cottage, and their mobile phone is out of reach for the network it is using.

My argument: This is an example of a computer making a decision. The computer decides that Mary is not allowed to drive the car, and in this particular case, the decision made by the computer might be ethically wrong. This is no easy issue; the best may still be to have such a computer device in the car. But the example illustrates the problem and the danger of programming computers to make decisions about right and wrong. And some (not all) such breathalyzer actually has an override button, even if the user of course has to report and explain why s/he used the override button, or lose their driver's license.

Counter-argument: No, the computer did not make any decision. The decision was made by the humans who programmed the computer. They may have weighed pros and cons, and decided that the advantage with such a drunk-driving-protection device is worth the risk that in some exceptional cases the outcome may be wrong.

My counter-counter-argument: This is becoming a discussion of the meaning of words. You do not accept that the computer made a *decision*. OK, let us then say that the computer made a *ruling*, or whatever word you prefer to apply to the case where a computer prohibits you from doing something. You are avoiding the ethical issues: In what way should we program computers to control human beings.

I am not a fanatical liberal who is against all laws and rules. I am quite willing to accept that in some cases it may be ethically right to program computers to prohibit you from drunken driving or stop children from downloading bomb-making recipes from the Internet. But I am advocating that in those cases you are programming the computer into making *decisions*, or *rulings* or whatever word you prefer to use. And this can be dangerous and you should be aware of the risks.

Solution: In this special case, a solution might be to allow the driver of the car to communicate with SOS Alarm, and allow them to send a code to override the breathalyzer lock in some very exceptional cases. By doing this, we are moving the final decision from computers to humans.

Explanation of Problem

One mode of human communication is the setting of rules. Some human beings make a list of rules. The rules may be a law, a local ordinance, ethical rules of a professional organization, company rules for employees or published in other ways. The human beings may also introduce ways of enforcing the rules, such as courts of law, committees on ethical conduct, etc. This is sometimes (not always) necessary even though the rulings made are sometimes wrong, like convicting innocent people. But we accept that this risk must be taken because without law and order society would not work.

An example: Even if it is forbidden for a pedestrian to step into the street against a red light, there are special cases where this rule does not apply. Suppose a child runs out on the street, and the only way to stop the child from getting run over by a car is to run out and catch the child. Such special cases are easily handled by humans. No human court would sentence a person for running against a red traffic light in such a case. But it is not so easy to teach a computer to understand such exceptions.

The danger is that people do not always understand, that putting such rules or laws into a computer, and programming the computer to enforce the rules, is something very different from having humans implement the rules by human decisions [[Grip 1975](#), [Palme 1975](#)]. Humans can understand the special conditions of special circumstances. A human might decide, in the example above, that in this special case the importance of getting John to a hospital is higher than the risk of Mary driving while intoxicated.

Counter-argument: A counter-argument made at [the ACM meeting](#) was that in this case it was humans who made the decisions, by programming the car computer, but their decisions were wrong. They did not take all circumstances into account. They should have made a more advanced program, which could take into account the special circumstance of the heart attack situation.

Counter-counter-argument: What you are doing, with this kind of argument, is to make the computer program more advanced and complicated, to reduce the risk that the decisions/rulings made by the computer are wrong. The path you are treading may make things worse instead of better. More complex and advanced computer rulings may increase the risk of wrong decisions instead of reducing them. The right solution may sometimes instead be to accept that the computer is not perfect, and thus that all rules do not have to be enforced by computers alone.

Italian Strike Example

A well-known method for strikes is to continue to work, but to adhere 100% to all rules while working. This rapidly causes many businesses to a complete standstill, or at least makes them work much slower and less efficiently before. The reason for this

is that even good and benevolent rules can have disastrous effects if adhered to 100%, the way a computer would do if programmed to enforce them.

Filling in a Form Example

If you fill in a form manually, you have the option of adding an accompanying sheet of paper explaining why you filled in the form in a particular way. In most forms, where there is a "yes-no" question, it is quite possible to omit the checkboxes and write a more nuanced explanation below or above. The human who receives the form will read and understand and interpret this. When the form becomes computerized, this freedom to not follow the prescribed way of filling in the form often disappears.

Society Evolves by Many People Doing Things in Better Ways

An important way in which human society is evolving is that many people make small and large decisions to try out new and better ways of doing things. If the computer program stops them from doing things in other ways than those specifically allowed, this will prevent people from finding better ways of doing things, and thus stop improvements [[Hoare 1975](#), [Palme 1997](#)].

Sometimes this may be necessary. For example, there is a human tendency to stop performing actions which are necessary only to avoid seldom occurring risks. Example: A pilot forgets an item on the pre-flight check list, or a night watchman forgets to go to a normally empty part of the building. In such cases, it may be necessary to use technical means to ensure that the human follows the rules, for example the night watchman must turn a key to show that he has passed that part of the building. But this does not forbid the night watchman from disobeying the rules in special cases, for example skip the empty corridor if there is a thief in another part of the building. The danger is when the computer does not allow you to do things in other ways than those foreseen when programming it.

People are in fact very clever in circumventing restrictions in order to do what has to be done.

[...]

First version: 4 June 1997. Last revision: 31 October, 2006
<http://people.dsv.su.se/~jpalme/reports/right-wrong.html>

TASKS

1. What is the main gist of the text? What's the general pattern the author used in this article (full Internet version)? Give reasons.
2. What if we let computers decide what is right or wrong? Give reasons
3. What is a feasibility/viability study? What is it used for? What are the main parts to be included?
4. Imagine you were to present the above topic ("*Can Computers Decide what is Right and Wrong?*") before a suitable audience. What's the general lay out you would use for your transparencies? Is

131051120 / 131061122 INGLÉS TÉCNICO

there any web-site devoted to public presentations and laying out slides/transparencies? Give reasons, as well as web-addresses (Please, note well: your suggestions should be different from YouTube video «5 Quick Tips to Effective Public Speaking», already included below).

5. Watch the following You Tube video clips, and then answer the questions:

5.1 «Seven-Language Interpreter»

(<http://youtube.com/watch?v=GBG47Enx79I&feature=related>)

5.1.1 What's the reason why the CEO thinks he's living a nightmare?

5.1.2 Do you think the temporary interpreter's behaviour is honest? What other solutions would you have offered?

5.2 «5 Quick Tips to Effective Public Speaking»

(<http://www.youtube.com/watch?v=xcMOzkne8ko&feature=related>)

5.2.1 How does the speaker rate herself as a public speaker?

5.2.2 Name the five tips she mentions, together with, if possible, a piece of advice she may give.

6. Look for a web-site tutorial, and, if possible, a video-tutorial (YouTube, DailyMotion, etc.) on public presentations/speech deliveries.

7. Your presentation group should already have started working. Decide on the topic, and possible contents. Try to fit them in into an outline. Assign time slots and tasks to each group member (all of you should present a part of the topic, during an average of 5-7 minutes. The spoken language should be English).

8. Your glossary should be enlarged by 50 words. By this week, you should have around 300-350 entries.

