

Look at What Happens to Telltales and Buffaloes!

From: AAI Technical Report FS-97-02. Compilation copyright © 1997, AAI (www.aaai.org). All rights reserved.

Jean-Bernard Billeter

Laboratoire de microinformatique
Ecole Polytechnique Fédérale
1015 Lausanne / Switzerland
billeter@di.epfl.ch

Abstract

The introduction of autonomous robots into the public and semi-public domain (streets, parks, shops, etc.) may backfire if certain deep-rooted human reactions are not taken into consideration. I first present some potentially delicate situations in order to bring the problem into focus and then suggest a series of items to be included in the specifications of a robot/human interface which might make for easier coexistence.

A foreseeable but not planned irruption

The introduction of new technology seldom follows a well thought out social strategy. At most, it may be launched by a marketing campaign, following which its dynamics will depend on the conjunction of many factors, e.g. satisfied needs, aroused expectations and market logic. This process will be accompanied by effects, of which some may have been predicted, others not, although they might have been, and yet others which could not have been foreseen. The pattern is well known: it has been repeated since the dawn of humanity, since the discovery of fire and flint. As Jacques Neiryneck puts it (1990):

We pretend that our destiny is shaped by philosophical and political debate which is then purportedly given material existence by technical means. But in point of fact, our fate is governed more and more by an uncontrolled evolution of technology which we dress up, ex post facto, in ideological frippery.

This pattern is due to be enacted once more in the coming decades with the entry of robots into our daily life. We can indeed imagine that the existing animal species (including man) will soon share this Planet with a variegated crowd of mobile machines sufficiently autonomous to tend to their business without constantly expecting our instructions. Delivery robots, sanitation robots, day-and-night-watchman robots, pet robots... the list may be long. These mechanical creatures will be part of our environment or even part of our families, helping us in many ways, and, of course, creating new problems, some predictable, some not.

I thought it would be useful to look at some of the problems which, I am sure, are going to crop up. Useful for three reasons:

1. The fact that new technologies have, heretofore, seldom been introduced as a result of a well-considered social strategy does not mean that we should definitely abandon this option and give a free rein to commercial logic alone.

2. If we want ultimately to learn to inflect the course of things intelligently, it may be useful to keep a regular record of our reasoned predictions so as to eventually confront them with reality.

3. Anticipating the possible conflicts that future robots may provoke should help to pave the way for their better acceptance. A robot which flunks its entry on the market is a socially unintelligent, an uncouth robot!

The analysis I present is not exhaustive. In particular, the impact of robotics on unemployment (i.e. on the division of labor and thus on social life as a whole) will not be addressed because it is not specific to mobile robotics. In line with the point 3 above, I will content myself to anticipating two categories of foreseeable conflicts: *threats to human peace of mind*, and *threats to the safety of robot*; and I will present some simple ways to prevent them.

Threats to human peace of mind

Understanding what a machine is doing

Let's consider a washing machine. We know that it faithfully executes a pre-set program. No one coming into the laundry room will wonder about its "intentions". At best, you might press the STOP button should you notice that there are woollens in the drum while the temperature is set at 60° C.

The working of a computer is not more difficult to interpret than that. If it deviates from our instructions, or, rather, from the instructions we *thought* we gave it, we can be sure that the mistake is ours. The answer to the question "What in Heaven's name is it doing?" is: "Exactly what you told it to do"!

Automobiles, on the other hand, caused real anxieties: since they could move "by themselves", they at first appeared as autonomous agents, a sort of animal to be wary of. In (de Roulet 1998), a groom in the 20's panicked when the klaxon of his master's car, just parked in the garage, went off:

Sir, sir, don't go in there, the car is gone crazy, it might take the bit between its teeth and run you over.

This example shows that our instinctive reaction is to assimilate a moving machine to an animal. The problem with cars was soon resolved as people understood that the only will that moved a car was the will of its driver and that, in the absence of a driver, a car obeys physical laws that are simpler than for a horse-drawn carriage... especially if the horses had bolted.

With autonomous mobile robots, things will be more complicated. Let's look at the reason why.

Transparent intentions

The first robots will be given simple chores like, say, cleaning the streets. At first, they might arouse suspicion, or even irrational reactions. But soon, they will become familiar and trusted as we shall learn how to quickly identify them (the sanitation department will paint them to its colors), and understand what it is they do and how to adapt to them (for instance, by being careful not to get sprayed by an ill-directed stream of water). We might end up not noticing them at all, just as we frequently take no notice of men and women cleaning our offices.

The passage from mistrust to peaceful coexistence requires that, when encountering a robot, we should be able to instantly grasp:

1. its function (what is it doing? why is it here?)
2. its behavior (what might it do? in what respect do I have to be careful?)
3. its owner (to complain if things go wrong)
4. its PIN (to specify what robot went wrong).

A robot should identify itself like a soldier meeting his superior! We shall not go into the precise modalities of this presentation; let's just say that all traditional modes are allowed: colors, flags, symbols, auditory signals and messages, etc.; more modern means can be added, such as radio or internet messages, etc.

Note that this first list of information to be grasped instantly when meeting a robot applies equally to human contacts: when one's presence puzzles another person, it is customary to introduce oneself. The details which will be volunteered depend on the situation: "I'm the postman", or "My name is XYZ, I work for the Polling Institute; I'd like to ask you a few questions about your eating habits".

Non-transparent intentions

Besides cleaning, handling and exploration, the most frequently-mentioned task in the autonomous robotics literature is surveillance. It is not farfetched to imagine a team of robots patrolling a nuclear power-plant site. But this type of work will soon be extended to "plain clothes" surveillance: robots used to prevent shop-lifting will be made to look like merchandise-handling or shopping robots.

At this point, when robots will be made to conceal their true mission behind a faked activity, our coexistence will become more problematic. As soon as we realize that there

are dissemblers among robots, we will start mistrusting *all* robots. And justifiably so.

The distress caused by the difficulty of predicting an animal's or a human being's intentions is not an innocuous feeling. Understanding such intentions is often a matter of life or death. A significant part of our cerebral activity is habitually involved in this task, and a suspicious mobile robot will inevitably trigger these functions.

Compared with other machines, the novelty with robots is that we are entering into a human-type interaction system involving observation, deciphering of intentions, communication and behavioral reactions. Up to now, the uneasiness caused by the mistaken attribution of "human" characteristics to machines has been quickly dissipated:

- an automobile is again seen as just machine (even if body stylists play upon our automatic "animal" reading of the shape);
- back in the sixties, it took only some time for my grandfather to understand that the TV speaker could not see him, and that there was thus no need to straighten up the drawing room and to put on a jacket for the news!

The question about machines used to be "How does it work?" Since most people are not technically minded, this bothered only a small minority of users. On the contrary, the cerebral functions of behavior interpretation and intention deciphering are generally permanently activated in individual in their waking state. Which means that as soon as robots become just a little bit smarter, they will take the same place as living beings in our lives. They will become, if not real conspecies, at least "con-social", that is agents whose behavioral economy is linked to ours. If they bother us, we shall react the way we do towards an obnoxious animal or human being.

Elusive identity

In addition to the lack of transparency of a robot's intentions, there is also the problem of a lack of clarity about its identity. When we watch an animal, we see it as a complete, identifiable organism, say, a monkey, or a snake. With an unknown robot on the other hand, there is no obvious proof that the thing is all there is, that it is complete: could it be communicating with other agents? could it be but a part of some complex entity whose organs are distributed among a number of different machines and the brain located who knows where? In these circumstances, there is no hope of being able to decipher its intentions by observing its behavior. For us, humans, this is a nightmarish situation.

Response: A better interface, better laws

To neutralize such frightening feelings, the following information should be added to the four specifications mentioned above:

5. An exhaustive list of permitted activities
6. A list of areas where these activities are allowed

7. Time periods when these activities are allowed
8. Control mode (who is the brain).

For instance, for a cleaning robot:

5. Cleaning the streets, excluding any interaction with humans
6. Streets and public parks, City of Lausanne
7. Working days, from dawn to sun-down
8. Strictly autonomous while working, teleoperated while going back to base.

For an auxiliary of the police force:

5. Any measure necessary to maintain public order in an emergency
6. Public and semi-public domain, City of Lausanne
7. Day and night, throughout the year
8. Dual: autonomous or teleoperated; the choice is up to supervisor agent XYZ.

This information has to be believed if it is to have the desired reassuring effect. A necessary (but not sufficient) condition for it to be believed is that it is enforced by legislation. Of course, this law, like all laws, may be infringed. What proof have we that the robot's purported activity is really the one for which it is here? We might thus suspect a window-cleaner robot to be working hand in glove with burglars, or with the tax authorities; a shopping robot of having a secret deal with some laundry-soap manufacturer, etc. Thus, our trust in robots will ultimately depend on the prevailing social climate and on the confidence we place in our authorities! In other words, the launching of robots into the public and semi-public domain will *not* be frictionless. It will not be as smooth as the development of the Web.

Threats to the safety of robots

Robots suspected of serving as informers are exposed to acts of depredation and destruction, the way speed checking radars are. But a robot does not have to be suspected to be attacked: it acts as a trigger, in and of itself, to destructive behavior:

In (DeLillo 1971), a Vietnam veteran relates the horrors of the war to his brother:

You see, there's a primal joy in hitting a thing in motion. It's one of the oldest pleasures there is. Something moves, boo, you wing it. Beast, bird or human, the thing to do is knock it down. It's primal, Davy. It's basic to the origin of the species. I'm learning to live with it.

A case in point is the extermination of buffaloes, in the US, in the second half of the 19th. century. Their number plummeted from 30 million to just 25 heads! Some of the animals were shot for their coat or for their meat; but a big proportion were shot down simply for the fun of it, often from moving trains.

The recurrent dropping of stones from bridges over freeways shows that the instinct endures.

Mobile robots will be obvious targets, they will be sniped at, and vandalized. Naturally, their owners will object! Appealing to conscience and reason might help: this already prevents most of us from throwing stones at cars, temptation notwithstanding. Repressive measures might help too. But prevention, both passive and active, has to be built into the robot.

On the passive side, robots can be reinforced, armored. They can be made to carry a camera to continuously telerecord the surroundings so that, after an act of vandalism, the stored images might help to identify the wrong-doer. But this measure might be a double-edged sword: a spying camera may actually elicit aggressive reactions!

On the active side, we may play upon other atavistic features such as endowing robots with appeasing behaviors. In a piece of fiction, Miedaner (1977), quoted by (Hofstadter and Dennett 1981), writes about a little robot who scurries around the house and which, when hit, makes heart-breaking sounds! Of course, the pre-recorded sounds are not provoked by any feelings of pain, but, in the story they cause the same effect: the character who was meant to stop the machine by hitting it with a hammer found it very difficult to do this!

Should this type of behavior be incorporated in the robot/human interface of the street-cleaning robots to come? I would say *yes*!

Conclusion

Some roboticists and some ethologists think that a mutual cooperation could be fruitful. The roboticists want behavior recipes, ethologists need models and validations. But a new type of collaboration may develop; its object: to determine ways and means to accomplish a smooth and peaceful introduction of autonomous mobile robots into public life.

References

- DeLillo D. 1971. *Americana*. Houghton Mifflin, and Co.
- de Roulet D. (to be published in 1988) *Trois kilo trois*. Seuil, Paris.
- Hofstadter D.R. & D. Dennet. 1981. *The Mind's I*. Basic Books.
- Miedaner T. 1977. *The Soul of Anna Klane*. The Church of Physical Theology.
- Neiryneck J. 1990. *Le huitième jour de la création*. Second edition. Presses Polytechniques Romandes, Lausanne.