No psychological machines

Thilo Deussen

Department of Artificial Intelligence
University of Ulm, 89069 Ulm, Germany
mail@thilodeussen.de

1 Introduction

Ascribe psychological attributes to machines, particularly personal computers, and you are in trouble. At least, if you did not identify typical pitfalls that occur in doing so. After a certain, not yet reached, point it will be normal not only for a minority to lead daily conversations about practical things *with a machine*. It will be discussed if these machines should be called, then, intelligent or not.

2 Intelligent machines

When talking about "intelligent machines" one is applying psychological attributes to a machine as in "the computer is crazy" or "the PC understands me or doesn't understand me". "He, the machine, has a problem." Actually the user of those phrases might have a problem if not metaphorically used.

Behaviouristic view From the viewpoint of behaviourism one will find: something is intelligent, that *behaves* intelligently. The Turing-test (in [4]) or any of its variations offers a level of abstraction, that makes this behaviouristic view possible (see [2] for explanation of levels of abstraction). Indeed, there are many arguments against a Turing-test, finding that it by far not covers all aspects of whole human intelligence. But you don't get the Turing-test on his weak side, when arguing this way. A programme passing a Turing-test can easily be discredited as a non-understanding programme due to Searle's Chinese room argument (as explained in [3]).

From a functional point of view, it makes no difference for a human, chatting or communicating by other means, if there is a human or a machine on the other side. As long as you get your questions answered as expected for some bounded area of interest, there is no difference if the other side is $\mathsf{intelligent}_{\mathsf{human-full-intelligence}}$ or $\mathsf{intelligent}_{\mathsf{behaves-intelligent}}$. It simply does not matter.

Further, there are many applications where it is not necessary or even not wanted, that the other side behaves like a human, i.e. also error-prone and irreliable. If you intend to switch off your television, you don't want to find a remote control starting a discussion about it's bad mood answering "no, I will not do this now". A coffee machine, that can switch itself on and off and brew great coffee, is an often-used example for a intelligent device. Actually, the coffee machine should fulfill its function in the sense, that it can handle typical problems and

exceptions that occur during coffee machine-usage. To do this, it does not have to experience the smell of coffee and feel the great taste. It's sufficient if it has sensors that are accurate enough to comply with correspondent requirements to determine if the coffee is brewed well.

The demand for intelligent machines is not the call for (totally) autonomous machines, but rather the call for machines, that fulfill autonomously their function. If this is the case, the human user can do other things, because he can rely on the machine and doesn't have to care about these things.

3 The jowedoiness of things

Intelligence Today's talking about intelligence, especially intelligence of machines, may be wrongly led. As shown before, often we even don't want intelligent machines. In the following, a thing, that works well and does, what it was designed for, will be called a jowedo thing.

What's the help of not talking about "intelligent" things? If ten people are asked about a short definition of intelligence, one sure will get ten different answers, or even more, considering some of them being philosophers giving more than one answer.

Imagine replacing all usages of the words "intelligence" and "intelligent" with "jowedoiness" and "jowedo" respectively. What do we win doing this? The reader will not stumble over the original words and leave the discussion just because he has another preassumption on intelligence as the author might have. In contrast, he will ask "what is being jowedo?" and, hopefully, the author will introduce a well-defined meaning in the beginning. Inflationary use of the term "intelligence" is stopped and instead of discussing about a blurred definition of the former we have a operationalisable term "jowedoiness" which can be used to evaluate certain qualities of a machine.

Knowledge Another current exaltation is around compound words that include the term "knowledge". If one just wants to have a nice naming to impress, it might be acceptable talking about knowledge-storing, knowledge-creating and knowledge-based machines.

What it is to "know" something is strongly dependent on one's beliefs about knowledge. One knows Goethe, one knows how to drive a car, one knows a recipe, one knows how to cook a dish. The term "knowledge" is so colourful and diverse, that almost anybody can shape a correct sentence about knowing.

When we read "X knows Y" like "X shows certain behaviour that – if X were human – would lead to the conclusion of X knowing Y", surely we won't get into trouble, since knowing for humans is somehow defined, as humans are psychological entities. As mentioned earlier we want to ask ourselves, if talking about machines that know clarifies something.

Clean language As shown in [1] talking about a brain knowing is senseless. When Neuroscientists commit a fallacy mixing up predicates used for a human being with predicates not literally applicable for a human brain we see how important

it is to keep scientific language unambiguous and "clean". What is "clean"? If terms used have only one well-defined meaning then we can call a special purpose language clean. We shall not argue that simplifications in daily life are to be avoided. If an end user is talking annoyed about his personal computer not *wanting* to do something, there is nothing bad in it. Nobody wants to forbid neither "brain's food" nor "somebody's brain's thoughts" if used colloquially in everyday speech. But for science we should prefer stating "food which is favourable for a well-nutritioned brain" and "somebody's thoughts". Likewise we should state that "the computer was not in a configuration that would allow to execute a certain task" rather than that "he did not want to do it".

A possible approach to clean up language is to replace any phrase of the form "X is/does Y", where X is a machine and Y is a psychological predicate or behaviour with "it seems like as if X were being/doing Y". What happens? If the sentence still has its original meaning it is fine and – having in mind this metaphorical speech – can be restored. If the meaning changes, e.g. the statement becomes weaker or uninteresting, check the usage of those psychological attributes and if they are defined beforehand to not confuse the reader.

Strong/weak AI The discussion outlined above should not be confused with the arguments brought forward in strong or weak theses about artificial intelligence. Although there might be seen a parallel, it is not argued pro or contra either one of these positions but rather noticed that, for constructive and productive work, statements involving psychological attribution to machines should be handled with care.

Opening the discussion Investigation in computer science has an area called artificial intelligence. Researchers in this area in particular often have a natural science background and no intensive study of humanities. The more concepts like tractability, efficiency, algorithm or recursion are clarified and well-understood, the less concepts like intention, consciousness or intelligence are well-defined or unambiguous. It would be welcomed, if arguments could be presented that are in favour of using psychological attributes for machines and are not rebutted by analogous argumentation like in Bennett's and Hacker's work about the mereological fallacy in neuroscience [1, chapter 3], e.g. metaphoric use, metonymic use or homophonic use of language.

References

1. M. R. Bennett and P. M. S. Hacker. *Philosophical Foundations of Neuroscience*. Blackwell Publishing, Oxford etc., 2003.
2. Gian Maria Greco, Gianluca Paronitti, Matteo Turilli, and Luciano Floridi. How to do philosophy informationally. In *LNAI "Professional Knowledge Management"*. Springer, 2005 (to be published).
3. John R. Searle. Minds, brains, and programs. In *Behavioral and Brain Sciences 3*, pages 353–373, 1980.
4. Alan M. Turing. Computing machinery and intelligence. In *Mind 59*, pages 433–460, 1950.