



On our best behaviour



Hector J. Levesque

Dept. of Computer Science, University of Toronto, Toronto, Ontario M5S 3A6, Canada

ARTICLE INFO

Article history:

Received 25 September 2013

Received in revised form 11 February 2014

Accepted 21 March 2014

Available online 28 March 2014

Keywords:

IJCAI Research Excellence

ABSTRACT

The science of AI is concerned with the study of intelligent forms of behaviour in computational terms. But what does it tell us when a good semblance of a behaviour can be achieved using cheap tricks that seem to have little to do with what we intuitively imagine intelligence to be? Are these intuitions wrong, and is intelligence really just a bag of tricks? Or are the philosophers right, and is a behavioural understanding of intelligence simply too weak? I think both of these are wrong. I suggest in the context of question-answering that what matters when it comes to the science of AI is not a good semblance of intelligent behaviour at all, but the behaviour itself, what it depends on, and how it can be achieved. I go on to discuss two major hurdles that I believe will need to be cleared.

© 2014 Elsevier B.V. All rights reserved.

1. Intelligent behaviour

This paper¹ is about the *science* of AI. Unfortunately, it is the *technology* of AI that gets all the attention. The general public could be forgiven for thinking that AI is just about all those whiz-bang applications, smart *this* and autonomous *that*. Those of us in the field know that for many applications, the term “intelligent” is no more than a buzzword (like the term “delicious” in “red delicious apples”). And along with the many possibly beneficial AI applications under consideration, we often have serious misgivings about the potential misuse of AI technology (in areas like weaponry).

But AI is more than just technology. Many of us are motivated not by any of the AI applications currently being considered, but by the scientific enterprise, the attempt to understand the world around us. Different sciences have different subject matters, and AI is the study of *intelligent behaviour* in computational terms. What could be more fascinating? The human brain is a remarkable thing, perhaps the single most complex object we know of in the universe. But even more remarkable is what a human brain is capable of *doing*. Our intelligent behaviour at its best goes well beyond what we have any right to expect to emerge out of purely physical matter. Indeed, the overarching question for the science of AI is:

How is it possible for something physical (like people, for instance) to actually do X ?

where X is one of the many instances of intelligent behaviour. This needs to be contrasted with a related question:

Can we engineer a computer system to do something that is vaguely X -ish?

about which we will have much more to say later.

E-mail address: hector@cs.toronto.edu.

¹ This paper is a written version of the Research Excellence Lecture presented in Beijing at the IJCAI-13 conference. Thanks to Vaishak Belle and Ernie Davis for helpful comments.

Note that the science of AI studies intelligent behaviour, not *who* or *what* is producing the behaviour. It studies natural language understanding, for instance, not natural language understanders. This is what makes AI quite different from the study of *people* (in neuroscience, psychology, cognitive science, evolutionary biology, and so on).

What sort of behaviour do we care about? Different researchers will quite naturally focus on different aspects. The behaviour may or may not depend on perceptual or motor skills. It may or may not include learning. It may or may not be grounded in emotional responses, or in social interactions. For some researchers, the main concern is intelligent behaviour seen in a variety of animals, like the ability to find a desired object in a room. For others, the focus is on behaviour seen in humans only, like the ability to play chess. (These two groups sometimes engage in methodological disputes, with the former arguing that we cannot expect to understand human behaviour until we understand its more basic forms, and the latter responding that this is not how science works at all. At this stage of the game, there is really no reason to take a doctrinaire position one way or another.)

1.1. Answering questions

In this paper, I intend to examine one basic form of intelligent behaviour: answering certain *ad-hoc* questions posed in English. Consider a question like the following:

Could a crocodile run a steeplechase?

Even if you know what crocodiles and steeplechases are,² you have never really thought about this question before, unless you happened to have read an early paper of mine [6]. Nor can you simply look up the correct answer somewhere. And yet, an answer does occur to you almost immediately. Here is another question from the same paper:

Should baseball players be allowed to glue small wings onto their caps?

Again, you have never thought of this before, but again an answer occurs to you. (In this case, you might even wonder if there is some sort of trick to the question that you may have missed. There is none.)

In this paper, I want to consider our ability to answer one-shot questions like these, and for four reasons:

1. This is behaviour that is clearly exhibited by people. We are indeed capable of answering questions like these without any special training or instructions.
2. This is behaviour that is difficult to crack. We have as yet no good idea about what people do to answer them. No existing computer program can duplicate our ability.
3. Our behaviour in answering questions like these appears to underly other more complex (and more ecologically significant) forms of behaviour.
4. Being clear and precise about the form of behaviour we care about even in this simple case will also help clarify what it means for the science of AI to be successful.

As we will see, however, there will be good reasons to move to answering questions of a more restricted form.

2. Behavioural tests

Given some form of intelligent behaviour, how do we know that the computational story told by AI researchers actually explains the behaviour? The answer, going all the way back to Turing, is this: a computational account is adequate if it is able to generate behaviour that cannot be distinguished over the long haul from the behaviour produced by people.

This, of course, harks back to the famous Turing Test [12]. We imagine an extended conversation over a teletype between an interrogator and two participants, a person and a computer. The conversation is natural, free-flowing, and about any topic whatsoever. The computer is said to *pass the Turing Test* if no matter how long the conversation, the interrogator cannot tell which of the two participants is the person.

Turing's point in all this, it seems to me, is this: Terms like "intelligent," "thinking," "understanding," and the like are much too vague and emotionally charged to be worth arguing about. If we insist on using them in a scientific context at all, we should be willing to say that a program that can pass a suitable behavioural test has the property in question as much as the person. Adapting the dictum of the movie character Forest Gump who said "*Stupid is as stupid does,*" we can imagine Turing saying "*Intelligent is as intelligent does.*" This is a very sensible position, it seems to me, and I have defended it elsewhere [7].

² For those who do not know, a steeplechase is a horse race, similar to the usual ones, but where the horses must jump over a number of hedges on the racetrack. So it is like hurdles for horses.

Download English Version:

<https://daneshyari.com/en/article/376899>

Download Persian Version:

<https://daneshyari.com/article/376899>

[Daneshyari.com](https://daneshyari.com)