HOW EMPIRICISM DISTORTS AI AND ROBOTICS

Dr Antoni Diller
School of Computer Science
University of Birmingham
Birmingham, B15 2TT, UK
email: A.R.Diller@cs.bham.ac.uk

Abstract

The main goal of AI and Robotics is that of creating a humanoid robot that can interact with human beings. Such an android would have to have the ability to acquire knowledge about the world it inhabits. Currently, the gaining of beliefs through testimony is hardly investigated in AI because researchers have uncritically accepted empiricism. Great emphasis is placed on perception as a source of knowledge. It is important to understand perception, but even more important is an understanding of testimony. A sketch of a theory of testimony is presented and an appeal for more research is made.

Keywords

Knowledge acquisition, testimony, evaluating information, world knowledge.

Introduction

Especially since the creation of the science-fiction genre, the idea of creating artificial beings in human form has gripped many people's imagination. For example, Fritz Lang's 1927 film *Metropolis* features a metallic robot that is transformed in a memorable scene into the likeness of Maria. The real Maria is a saint who comforts the oppressed workers whereas the android is an *agent provocateur* who incites a riot which provides a pretext for John Frederson, the ruler of Metropolis, to crush the rebellious spirit of the workers. The intricacies of the plot of *Metropolis* need not concern us here, but what I would like to single out is the fact that the android is able to pass undetected as Maria. It comes as a real shock to the watching crowd when the android changes back into its original metallic form. Much more could be said about the exploits of fictional robots, including Rachel in *Blade Runner*, the killing android in *Terminator*, Lieutenant Commander Data in *Star Trek: The Next Generation* and David in Spielberg's film *AI*, but what of the reality?

There are a number of important research projects in the USA whose goal is the manufacture of a humanoid robot. One of the best-known of these projects is headed by Rodney Brooks of the Massachusetts Institute of Technology [11, pp. 58–65]. His robot is called 'Cog'. Currently, Cog has no legs, but it does have robotic arms and a head with video cameras for eyes. It is one of the most impressive robots around as it can recognise various physical objects, distinguish living from non-living things and imitate what it sees people doing. Cynthia Breazeal worked with Cog when she was one of Brooks's graduate students. She said that after she became accustomed to Cog's strange appearance interacting with it felt just like playing with a baby and it was easy to imagine that Cog was alive.

There are several major research projects in Japan. A common rationale given by scientists engaged in these is the need to provide for Japan's ageing population. They say their robots will eventually act as carers for those unable to look after themselves. The robot carers would look after the physical needs of those in their care and also, presumably, they would be companions to their elderly or disabled charges. Having robot carers would free up humans to do more economically profitable work.

One of the most advanced Japanese machines is the Honda P3. This looks like a man wearing a spacesuit, but it is made entirely of electronic and mechanical components. It is capable of walking, going up and down stairs and opening doors. The team that designed and built the P3 was led by Masato Hirose and the development took place at Honda's Wako Research and Development Laboratory [11, pp. 42–45].

Unfortunately, despite their technological sophistication none of these projects will succeed in producing androids that behave like their fictional counterparts. I do not want to belittle the great scientific and technological progress that has already been made. Researchers have had to solve many difficult problems in order to build humanoid robots that can walk on two legs, climb stairs, grasp objects without breaking them, recognise different kinds of physical object, imitate the human behaviour that they see and so on. However, because of their shared assumptions about knowledge, none of these high-profile projects will ever result in an android that can interact meaningfully with a human being and engage in intelligent conversation with that person. The fundamental human ability that virtually all researchers have overlooked is that of learning from others by believing what they say and by accepting what they have written. These scientists have ignored the role of testimony in human belief-acquisition. There are still many outstanding problems to be solved before artificial natural-

language processing is as good as the human ability to understand language, but I am not concerned with semantic issues in this paper. My interests are epistemological. I am concerned with how humans and machines acquire knowledge about the world in which they live. One of my main points is that researchers in AI and Robotics have focused almost exclusively on perception as our main source of information about the world and, consequently, have neglected the importance of testimony as a source of knowledge, although most of the knowledge a person has about the world has been acquired through testimony, that is to say, by accepting what other people say and by believing what they have written. The reason for this neglect is that they uncritically use an empiricist theory of knowledge, but before explaining this the importance of testimony needs to be demonstrated.

The Importance of Testimony

In order to live in any human society, especially a complex, technological one, a person needs a great deal of information and an android would need a similar amount of knowledge. Some of this information could be programmed into the robot before it entered society, but not all of it. One of the reasons for this is that the android would need to have knowledge about things that change over time. This is especially true of those things that people talk about with each other on a daily basis and an android employed as a carer for an elderly person, say, would have to be able to engage in such conversation. For example, people talk about the weather, what their neighbours and friends are getting up to, the current political situation, how their football team is doing, recent advances in science and so on. Information about these topics is obtained from various sources including the radio, television, newspapers and other people.

The android would also have to know about those activities which most people engage in at some time or other, namely such things as looking after a home, travelling, shopping, going on holiday, looking after money, eating in a restaurant and so on. In order to do any of these you need to have a considerable amount of information. For example, in order to have a meal in a restaurant you need to have some knowledge about various restaurants and the sort of food they serve, you also need to know how to get to the restaurant you have chosen, how to read the menu once you get there, whether or not to leave a tip and how much to leave if it is appropriate to do so and so on. Such information changes less frequently than, say, information about the weather, but the android's knowledge about such things would still have to be kept up to date.

The Neglect of Testimony

It is a pity that the ability to learn from other people's assertions has been overlooked by scientists building androids, but it is easy to see why this has happened. Though some people may deny the fact it is nonetheless true that science and philosophy are inextricable linked. This is especially true of that branch of philosophy, known as *epistemology*, which deals with how people acquire knowledge. Einstein, for one, was aware of this connection. He wrote [9, pp. 683–684], 'Epistemology without contact with science becomes an empty scheme. Science without epistemology is—insofar as it is thinkable at all—primitive and muddled.' It is also true that an incorrect epistemology is just as bad as none at all. The tradition in which scientists building androids work has been profoundly influenced by that epistemology known as *empiricism* and two principles in particular have prevented those scientists from appreciating the importance of learning from testimony. The first of these is the principle that statements should not be accepted unless evidence is provided to support them and the second states that this evidence should be ultimately grounded in sense experience or perception.

The main problem with the idea that we should not accept assertions without checking their credentials is that we simply do not have the time to do this. To rationally evaluate any statement we read or hear is a very time-consuming process. Thus, the number of statements that we can check in this way is very small. For example, a physicist wanting to make use of Planck's constant or the value of the speed of light in a vacuum would look up their values in a standard reference book trusting that they had been correctly worked out. Their values are taken on trust and this saves the physicist wanting to use those values a great deal of time. On a small number of occasions a few people recalculate the values of these constants, but this does not happen very often. I am not suggesting that we should never check any of our beliefs. Just because we cannot rationally evaluate all our beliefs it does not follow that we should not check any of them. In fact, there are occasions when it is important for us to critically examine some of our views in order to weed out error and falsehood. However, the principle that a statement should not be accepted unless evidence is provided in its support is not one that anybody could live by.

The second empiricist principle states that we should only accept a statement if there is perceptual or observational evidence in its favour. In fact, empiricism is often taken to be the theory that all knowledge is derived from sense experience. Pollock is completely explicit about his acceptance of empiricism and he speaks on behalf of most people working in AI and Robotics when he writes [12, p. 52], 'The starting point for belief formation is perception. Perception is a causal process that produces beliefs about an agent's surroundings.'

The theory that knowledge is derived from perception cannot mean that all human knowledge is derived from one particular individual's sense experience. But accepting the position that knowledge is derived from the sense experience of many people completely undermines empiricism. Any statement that I utter based on my sense experience, no matter how certain it is to me, is testimony to other people which they may well reject. Furthermore,

any statement that I hear someone utter, even if to them it is based on sense experience and they have no doubt regarding its truth, is simply testimony to me which I may well reject. For example, I may not know how reliable a witness that person is or their statement may contradict the report made by another person who saw the same event and which I have already accepted. Many psychological experiments have shown that observers of the same event often contradict each other [3, ch. 15]. Furthermore, the greater the number of intermediaries the information has passed through the more opportunities there are for doubting its truth.

People's ability to reject eye-witness testimony is well illustrated by the case of the early deep-sea explorer William Beebe [1]. In 1934 Beebe made the deepest dive that had been made up to that time. His primitive bathysphere dived to a depth of half a mile. He carefully described in his diary the strange creatures that he observed, but the life-forms he wrote about were thought so outrageous by the scientific community that his observations were discounted. Only in recent years, when more people have seen the same creatures, has his reputation been restored.

Understanding Testimony

Before we can incorporate the ability to learn from others in an android, we first need to understand it in its human form. The goal of my research is to devise a model of this human ability that is sufficiently detailed to allow a computer program to be written to simulate it [8, 7]. My current proposal is that our acceptance or rejection of the assertions we encounter is governed by the defeasible rule to believe them. Many people on first encountering this *acquisition* rule think that it is either superfluous or simplistic. Before responding to these points, I will explain what is meant by defeasibility.

Several notions of defeasibility are used in AI. The one I use is best introduced by means of a legal example. Imagine a country that has had a law against murder for a long time and which has decided to permit euthanasia under certain circumstances. The law against murder is well understood by lawyers and citizens and so there is little call for it to be repealed. Fortunately, there is a mechanism in the legal system which allows one law to supplant another one which appears to apply to the same event. Thus, a law allowing euthanasia when certain conditions are met can be enacted without repealing the law forbidding murder. Then, when someone is killed by a doctor following the guidelines for euthanasia, the law against murder will not apply because it has been overridded by the law permitting euthanasia. So, the doctor who carried out the mercy killing will not be guilty of murder. In such a case the law prohibiting murder is said to be *defeasible*. One of the advantages of putting things this way is that it avoids the need for laws with many qualifications such as, 'Killing is murder except when the death was an accident or it was carried out under the euthanasia guidelines or it took place during a just war or . . . '

A recent trend in AI has been the tendency to see the human mind as being hierarchically organised with higher-level abilities emerging from lower-level ones without being explicitly designed. This idea is based on the anti-reductionist doctrine of emergence which states that entities with a sufficient level of organisation have properties that could not be predicted from lower-level properties. For example, the sweetness of sugar cannot be predicted from its chemical structure, thus it is an emergent property. In AI this is known as the *bottom-up approach* and one of its champions is Steve Grand whose robot Lucy, by the way, rivals Cog in its abilities [10]. From this perspective it may be argued that having an acquisition rule is unnecessary since our response to others' assertions will emerge once we have given an android the ability to understand its perceptual environment and react appropriately to it. The problem with this is that it fails to appreciate that perception and conception are radically different processes. Understanding other people's assertions and deciding whether to believe them or not, are completely different things from, say, the ability to pick out an apple form one's perceptual field. I am not suggesting that research designed to give androids the ability to recognise physical objects and to understand their perceptual environment is worthless. Far from it. It is absolutely essential for androids to have these abilities, but in addition they need to be capable of acquiring information from testimony. It is not a matter of either learning from perception or from other's assertions; it is rather a matter of being able to learn in both these ways (and no doubt others as well).

The acquisition rule may appear trivial, but in fact it is exceedingly fruitful because it forces us to consider those situations in which it is overridden. Understanding the factors which may cause us not to believe something that we hear or read is much more complicated than you may think [5, 6, 4]. Recall that my proposal is that our response to the assertions that we come across is governed by means of the defeasible rule, 'Believe what you hear or read'. Most of the time when we read a non-fiction book, say, there is no reason for us to override this rule. For example, in the *Macmillan Dictionary of the History of Science* I read that the Royal Society of London was founded in 1660 [2, p. 377]. Having read that I now believe it. There is no reason for me to doubt the veracity of this fact. For example, the Macmillan Press is a well-known publishing house with a reputation for producing reliable and authoritative reference books. It may not be as highly regarded as the University Presses of either Oxford or Cambridge, but it still needs to maintain its reputation. Therefore, I am confident that the editors of this volume and the various contributors have taken great care to ensure that the information it contains is correct. Of course, I also know that people are fallible and mistakes are made, so if the date of the foundation of the Royal Society is particularly important to me I will check it in other sources.

There are times, however, when I do not accept what I read. For example, in *Chariots of the Gods?* von Däniken writes about the Nazca lines on the plains of Peru and says that they are giant runways for space-craft [13]. Although I have read this, I do not believe that the Nazca lines are the markings of giant runways. This is because I have overridden the acquisition rule in this case. I have learned from other sources that von Däniken is often unreliable. Furthermore, I know that the Nazca lines are drawn on the pebbly surface of the desert and would be destroyed if an aircraft tried to land on them.

When reading a book, therefore, there are many factors that may cause us to override the acquisition rule [6]. These factors relate to such things as what we know about the author or editor of the book, the publisher, the edition, the date and place of publication and so on. For example, we would be wary of accepting what we read about the Stalin purges in an encyclopaedia if that encyclopaedia was published in Moscow before the fall of communism. Other factors that may cause us to override the acquisition rule relate to the reader and include such things as her pre-existing knowledge, her interests, her credulity, her intelligence, her maturity and so on.

We obtain information from other sources in addition to books. We acquire beliefs from other people in the course of listening to them, from journal articles, from the radio and television, from the newspapers and from the Internet. In each case our response is governed by the acquisition rule, but the collection of factors that may cause us to override this varies depending on the source of the information. In the case of the media, for example, the location of a news agency is important: we are more likely to believe a news item broadcast on BBC Radio 4 in England than one from Radio Baghdad in Iraq during the reign of Saddam Hussain. However, this consideration does not apply to other sources of information. Isolating all the potential overriding factors, how they operate and how they interact with one another is a complicated task. Although I have made a start on identifying many of the factors [5, 6, 4], much research still needs to be done on these problems.

It should be noted that the acquisition rule is not infallible. This means that, when using it to assess assertions, an agent sometimes accepts assertions that are in fact false and sometimes rejects assertions that are in fact true. Therefore, in my model of how people learn from testimony I have found it necessary to include a second stage of belief-evaluation [8]. Assessing assertions by means of the acquisition rule takes place in the first stage, but sometimes people end up with false beliefs and sometimes they fail to learn from true assertions. In the second stage a variety of methods are used to weed out false beliefs and to reconsider some previously rejected assertions to see whether or not they were correctly rejected. It is not possible to characterise the second stage of belief-evaluation as succinctly as the first, because different methodologies have to be employed in order to assess different assertions. For example, a historian investigating the death of General Sikorski in order to decide whether he died in an accident or was murdered uses a very different methodology from that employed by a physicist trying to calculate the value of the speed of light in a vacuum.

Conclusion

Research in AI and Robotics has made great progress in recent years. However, because of their shared assumptions about knowledge, none of the major projects currently being undertaken will succeed in producing an android which can interact meaningfully with a human being and talk about the topics that people like to talk about. To live and work in human society and to converse with people an android would need a vast amount of information which would have to be kept up to date. The ability to learn from others is not optional. Although fundamental to human beings it is but poorly understood. In order to program this ability into an android we first need to understand it in its human form. The purpose of my research is to construct a model of this ability. If this can be achieved, then we will have made significant progress in producing artificial companions with whom we can enjoy intelligent conversation.

References

[1] William Beebe. *Half Mile Down*. Harcourt Brace, New York, 1934.

[2] W. F. Bynum, E. J. Browne, and Roy Porter, editors. *Macmillan Dictionary of the History of Science*. Macmillan Press, London, 1981.

[3] C. A. J. Coady. *Testimony: A Philosophical Study*. Oxford University Press, Oxford, 1992.

[4] Antoni Diller. Evaluating information found in journal articles. In Ángel Nepomuceno, José F. Quesada, and Francisco J. Salguero, editors, *Logic, Language and Information: Proceedings of the First Workshop on Logic and Language: Instituto de Lógica, Lenguaje e Información, Universidad de Sevilla, 29, 30 de noviembre y 1 de diciembre de 2000*, pages 71–78. Kronos, Sevilla, 2000.

[5] Antoni Diller. Everyday belief-acquisition. In Gabriela P. Henning, editor, *Argentine Symposium on Artificial Intelligence (ASAI2000) Proceedings: Tandil, September 5–7, 2000*, pages 221–232. Sociedad Argentina de Informática e Investigación Operativa (SADIO), Buenos Aires, 2000.

[6] Antoni Diller. Acquiring information from books. In Max Bramer, Alun Preece, and Frans Coenen, editors, *Research and Development in Intelligent Systems XVII: Proceedings of ES2000, the Twentieth SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence, Cambridge,*

December 2000, pages 337–348. Springer, London, 2001.

[7] Antoni Diller. Designing androids. *Philosophy Now*, (42):28–31, July/August 2003.

[8] Antoni Diller. Modelling assertion evaluation. *AISB Quarterly*, (114):4, Autumn 2003.

[9] Albert Einstein. Remarks to the essays appearing in this collective volume. In Paul Arthur Schilpp, editor, *Albert Einstein: Philosopher-scientist*, pages 665–688. The Library of Living Philosophers, Inc., Evanston (Illinois), 1949.

[10] Steve Grand. *Creation: Life and how to make it*. Phoenix, London, 2001.

[11] Peter Menzel and Faith D'Aluisio. *Robo Sapiens*. MIT Press, Cambridge (MA) and London (England), 2000.

[12] John L. Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT Press, Cambridge (MA) and London (England), 1995.

[13] Erich von Däniken. *Chariots of the Gods? Unsolved Mysteries of the Past*. Souvenir Press, London, 1969.