The Belief-filter Component

Antoni Diller*
School of Computer Science
University of Birmingham

A.R.Diller@cs.bham.ac.uk

April 28, 1999

Abstract

One of the goals of AI is to design and build an android who can pass undetected as he lives and interacts with human beings. Although this aim cannot be realised, it makes sense to design androids initially who have skills and abilities similar to those possessed by humans. Among such abilities is that of learning from other people. It is this that I consider in this paper. I propose a two-phase model of belief-acquisition. The first stage consists of a collection of easily applicable rules that are used to decide what to do with the many statements that a person hears and reads every day. The second stage comprises a critical methodology for the thorough investigation of a small number of selected theories. I argue that this methodology should be an anti-justificationist one. An android's belief-filter component should reflect this two-stage model if he is to interact effectively with human beings.

1 Introduction

1.1 Background

Some people working in AI see its ultimate goal as being that of designing and building an android that will be able to pass as a human being.[1] The idea is that this humanoid robot would be able to live in human society and go about his everyday business without people suspecting that he is an android. It would be impossible for the android to be detected from his outward behaviour and his interactions with human beings. If the android were to be taken apart, however, then it would be possible to see that he

*I am grateful to Debra Barton for reading a number of drafts of this paper and for making many helpful suggestions.

[1]For example, Charniak and McDermott write [9, p. 7], 'The ultimate goal of AI research (which we are very far from achieving) is to build a person, or, more humbly, an animal.' Some of the people working on the MIT Cog Project are more explicit. They write [6], 'Building an android, an autonomous robot with humanoid form and human-like abilities, has been both a recurring theme in science fiction and a "Holy Grail" for the Artificial Intelligence community.'

was not human. I must admit that I have grave reservations about the achievability of this goal, because I think that, even if all the engineering and mechanical problems could be overcome, it would still be possible to differentiate people from androids in terms of some aspects of their outward behaviour. (It would be out of place for me to elaborate on this point here, but I do say a little more about my reasons for believing this in the appendix to this paper.) At present, however, I cannot think of any reasons why it should not be possible in principle to design and build an android that looks like a human being and is capable of imitating some of the things that human beings do. Some of the robot's abilities may even considerably outstrip their human counterparts. Perhaps such an android would even be conscious.

Although I think that the goal of *replicating* a human being non-biologically is not achievable, the objective of building a human-like robot may well be possible and in this paper I look at some of the problems involved in designing one of the components of such an android's 'brain', namely the belief-filter component. I approach the issues quite abstractly and a great deal of work would have to be done before anything could be built. Such an initial high-level discussion of the issues involved, however, needs to be done before any low-level design work can even begin.

The goal of designing and building a human-like android needs to be made more specific. This is because human beings live in groups and in those groups there are people who have many different roles. Furthermore, people differ greatly in their skills, abilities and accomplishments. First, we need to select the kind of society in which the android is going to live and then we need to pick a role that the android is going to have in that society. We also need to select the level of expertise that we want the android to have. So, to begin this narrowing-down process I will say something about the kind of society I imagine the android living in. Not surprisingly, I think of the android as inhabiting the society that I live in. This is not because I think that this is the best kind of society to live in, but rather because it is the society that I know best. I live in England and so I imagine the android that I am writing about as also living here. Although there are differences between English culture and those of the other older Western democracies, I do not think that they are so great as to make my discussion irrelevant to the problem of designing an android to live in them. (When I talk about the older Western democracies I mean those countries that were referred to as the Western democracies before the fall of communism in Europe.)

People play many different roles in our society and these roles often require specific skills and abilities. So, we need to decide what kind of human being we want to imitate. Designing an android to pass as a human being who is brain-dead and existing in a persistent vegetative state raises different issues and problems from those raised by the attempt to design an android to pass as the Prime Minister. The sort of person that I am concerned with in this paper is the creative research worker and my interest is in the skills and abilities of such people. Designing androids to imitate other kinds of people would raise different issues and problems from those that are my main concern in this paper. Designing those sorts of android is a worthwhile task, but it is impossible to do everything in a single paper and so I have had to be selective in what I do here.

The researcher has many abilities and the two that I am particularly interested in

are his ability to learn from other people and his ability to creatively transform what he receives from other people. I am interested in the way such people use what they learn from other people. Although I am also interested in the creative process itself, in this paper I say nothing about that process. Here I am interested in, so to speak, the context in which research takes place, its antecedents and some of its consequences.

1.2 Specific Issues

I am interested in the growth of knowledge and I agree with Popper when he says that the 'advance of our knowledge consists in the modification and the correction of earlier knowledge' [29, p. 49]. The extent of our dependence on other people for our knowledge is increasingly being recognised by epistemologists and, although I too am interested in knowledge, people are more than a repository of beliefs and some of the other elements in a person's outlook have a strong influence on the content of that person's belief-system. By a person's *outlook* I mean something similar to what Kitcher calls an individual scientist's *practice* [19, p. 74]. For him this is a multidimensional entity consisting of the scientist's professional language, the problems belonging to his field that he regards as being significant, the theories in his field that he accepts, the collection of explanatory patterns that he recognises, the authorities that he recognises in his field, the experimental techniques that he accepts and his methodology. Epistemologists studying testimony are correct in pointing out that most of a person's belief-system is obtained from accepting other people's assertions, but they are wrong to ignore the equally true fact that we receive a great deal more from other people than just beliefs. Kitcher is correct to stress the fact that in addition to holding many theories a scientist also accepts and makes use of various other things, such as a methodology, but his restriction of a scientist's practice to what he makes use of in his professional life is an artificial abstraction. In addition to having all the components of a scientist's practice, the researcher also has all the other things that a person needs to be a fully-functioning member of human society. For example, he has a collection of moral principles by which he tries to live and also certain aesthetic sensibilities. Every adult member of human society has such an outlook and almost every part of that outlook has been obtained through interaction with other people or their artefacts. My main interest in this paper is with empirical knowledge and belief, but my approach differs from that of many other writers in that I think that in order to properly understand belief-acquisition and revision we need to realise that the people who acquire new beliefs and make changes to their belief-systems also have aesthetic values, methodological and heuristic rules, interests, desires, goals, ethical principles, etc., and that these have an influence on what beliefs they acquire and on how they go about modifying their belief-systems. Although in this paper I am interested in the design of the belief-filter component, such a component would need to interact with other components in the android's 'brain' that correspond to the other elements of a person's outlook.

Coady [10, p. 6] has urged that the answer 'Jones told me' should be as acceptable as the answers 'I saw it' or 'I remember it' or 'It follows from this' or 'It usually happens like that' to the question 'How do you know that?' I think that this is

only partially correct and in this paper I put forward a two-phase model of belief-acquisition which brings out the limitations of Coady's suggestion. The first phase of belief-acquisition deals with the criteria that we use on-line to assess the assertions that we hear, whereas the second phase relates to a more considered evaluation of the credentials of some claims that are particularly important to us. The answer 'Jones told me' is appropriate as an answer to the question 'How do you know that?' only in the realm of the first phase of belief-acquisition.

Distinguishing these two phases also helps us to better understand the ways in which knowledge is transmitted. Most of our knowledge comes from other people, but people make mistakes or they may deliberately misinform us or they may be sincere in what they tell us but were themselves misled by others. It is inevitable that most of our knowledge comes from other people, but error and incorrect information can be transmitted in this way just as easily as correct information. The first phase of belief-acquisition can only account for the transmission of information with practically no modification of what is transmitted. The second phase allows for the possibility of quite large changes in what is passed on to others.

Following a suggestion of Price's I elaborate a view in which the first phase of belief-acquisition consists of a collection of methodological rules. These encourage us to accept most of the assertions that we hear or read. An assertion has only to pass a few crude tests in order to be admitted. Although this idea is due to Price, his single rule is very vague. I isolate several factors that we take into account when assessing other people's assertions on-line. The rules that we use to evaluate the assertions that we encounter must, therefore, contain elements corresponding to these factors. Any android that we design would be severely hampered in achieving his goals if he did not have a similar collection of rules. The idea that the first phase of belief-acquisition consists of such a collection of rules has many advantages over the idea, used by Quine and Ullian [32, p. 14], that we can get by without any rules and just make do with meta-beliefs.

The second phase of belief-acquisition is not so easy to characterise. It involves a more rigorous examination of some of the knowledge that a person has. I look at the issues involved in very general terms and I distinguish between two types of empiricism, namely justificationist and anti-justificationist. To put the difference between these very simply, the justificationist is interested in the origins of beliefs, whereas the anti-justificationist is interested in the consequences of our knowledge. These two types of empiricism are not meant to exhaust the range of possible epistemologies, but they are, in my opinion, the two main dominant approaches. I argue that anti-justificationist empiricism is better than its justificationist cousin. One consideration involves the fact that some scientists are justificationists and some are anti-justificationists. Every scientist lives in a scientific community that contains scientists of both these persuasions (and probably also scientists who hold still other methodologies). A single methodology is not used by all scientists. Although this presents a problem for the justificationist, the anti-justificationist can take it in his stride. I argue that the best methodology for a scientist, who belongs to a community of scientists using a variety of methodologies, to adopt is an anti-justificationist one. If our goal in designing a hu-

manoid robot is to build an efficient and effective one, then we should equip him with an anti-justificationist methodology to employ in the second phase of belief-acquisition.

1.3 Organisation

In the next section I look at some of the components of what I have been calling a person's *outlook* and I indicate some of the ways in which a person's acquisition of empirical knowledge is affected by things which are not themselves empirical beliefs. I also give some examples of the power of the institution of learning from others and I briefly mention three different kinds of argument that have been put forward to suggest that the greater part of anybody's belief-system is correct. Sections 3 and 4 are more substantial and they deal, respectively, with our ways of doing on-line assessment and off-line checking of information. The conclusion, contained in section 5, summarises what has been achieved and considers what is left to be done. In the appendix I briefly state why I think it is impossible to design and build an undetectable android.

2 Knowledge in Context

2.1 Outlook Components

In order to live in a human society a person needs to be in possession of a great deal of information. Most of this knowledge comes from other people. It has been acquired by interacting with other people, listening to them and by reading what they have written. This fact is increasingly being realised by epistemologists and its ramifications are being explored.[2] Unfortunately, much of the recent work has concentrated on empirical knowledge to the exclusion of the other components of a person's outlook.[3] This is unfortunate because the empirical information we possess forms only one element of the mental equipment that we need in order to live in human society and, even if our interest is primarily in belief-acquisition and in the properties of a person's belief-system, the other components of a person's outlook influence those belief-related issues. In designing the 'brain' of an android we have to ensure that the belief-filter component can communicate with the other components in that 'brain' that correspond to the other elements of a human being's outlook. Thus, we need to ensure that the component in the android's 'brain' dealing with ethical issues, for example, can interact with the belief-filter component.

I do not claim that the following list is exhaustive, but it gives some ides of the components of a person's outlook. Rather than attempting to produce a historical account of the order in which a person acquires different parts of his outlook, this list is constructed from the vantage point of a mature adult. I think that this is the best starting-point from which to understand a person's outlook. Although I stress

[2]In saying this I am thinking especially of the recent resurgence of interest in the topic of testimony. See, for example, the following books and articles: [23], [10], [1], [2] and [22].

[3]See, for example, the influential book *Testimony* [10] by Coady. He does briefly discuss the transmission of mathematical knowledge, but virtually the entire book is about the transmission of empirical knowledge.

that a person's outlook is obtained through an extended period of socialisation, the current state of that outlook is easier to describe and analyse than some hypothetical reconstruction of how it was formed. A view about how part of that outlook was obtained is a theory (which may itself be part of someone's outlook). Theories about the origins of a person's outlook do not have a special or privileged position. For most people they are part of what we receive from other people and should be treated in the same way as the other theories that we acquire from other people. I do, however, give some indication of when the component was likely to have been acquired.

(1) Mature adults know one or more natural languages. The person who knows a human language has the ability to produce grammatical sentences belonging to that language which contain words that he knows the meaning of. People acquire their mother tongue (or tongues) through a lengthy process in early childhood. Learning a language involves the implicit absorption of a collection of grammatical rules and the acquisition of a certain vocabulary, but in addition we acquire the ability to extend our vocabularies in various ways and to correct our spoken and written language. There are vast individual differences in this respect and some people are more concerned about the correctness of their language than others. In addition, many people are taught how to use dictionaries and other works of reference in order to be able to communicate effectively, but not all of them use these resources to the same extent. It is reasonable to assume that most people living in the same society will have a large common vocabulary, but many people will, in addition, have specialised vocabularies relating to their work, interests, hobbies, etc. Whereas the largely common vocabulary is acquired in early infancy, specialised vocabularies are often learnt later on in life. Much, if not all, of a person's empirical knowledge is expressible in linguistic form and a person's ability to acquire new beliefs depends to a considerable extent on the vocabulary that he has mastered. Thus, someone who does not understand the expressions 'oceanic lithosphere' and 'subduction zone', for example, or their equivalents in English or some other language, cannot acquire the belief that oceanic lithosphere is consumed in a subduction zone. To summarise, adults living in a human society know a natural language and many of them have a number of auxiliary skills and abilities as well. These include the ability to consult appropriate reference works to improve their grammar and to correct or extend their vocabulary. A person's vocabulary affects his ability to acquire empirical knowledge.

(2) In the Western, liberal democracies religious tolerance has been accepted for quite some time and thus in those cultures we find people of various faiths and also many atheists. For many people their religious faith is an important part of their outlook and it influences many of the things that they do. A person's religion, if he has one, is usually acquired early in life, but does not have to be. Thus, some people grow up in a Christian environment and take this faith on board while others are brought up in an atheistic household and grow up as atheists. These convictions are not entirely determined by the enculturation process and some

people do undergo various conversion experiences later on in life. A person's religious faith or its absence may well influence the content of his belief-system. Someone brought up in a religious environment, say, a Christian one, is likely to have much information relating to the faith in which he grew up. He will not just have certain religious beliefs, but also many factual beliefs relating to that faith. For example, he may know the biographies of several important figures in his church.

(3) An important part of any person's outlook consists of the ethical principles that he accepts. Although in a liberal, Western democracy there is wide consensus about many important ethical issues, there are also many disagreements. As with the acquisition of a religious faith, people usually absorb their ethical principles quite early on in life. Often these ethical principles are linked to a religious faith but they need not be. A person's ethical principles may well affect what empirical beliefs he acquires. For example, someone who believes that it is wrong to read another's diary without that person's permission and who sticks to this principle will not acquire the information contained in a diary even if he has the opportunity to read it, whereas someone without this ethical principle may well gain empirical knowledge in this way.

(4) An important component of many people's outlooks are various aesthetic values and sensibilities that they have. There are many individual differences amongst people concerning their aesthetic tastes. Mozart is not everybody's cup of tea, but then neither are the Spice Girls. A person's aesthetic preference may well influence the content of his belief-system. Someone who does not care for Mozart will be unlikely to spend much time reading and finding out about his life, but someone who loves his music may well want to deepen his appreciation by learning more about the man, his contemporaries and his time.

(5) Most people in the Western, liberal democracies have a definite social role or even several such roles and associated with a social role there is a certain amount of information that is needed in order to fulfill that role.

(6) Most people have a certain amount of knowledge about some of the institutions of the society in which they live. For example, they have at least a rudimentary knowledge of the economic, political and legal systems of those societies. The amount of more detailed information that a person has depends on many factors, such as his interests, his intelligence, the problems he is working on and so on.

(7) Most people have a certain amount of historical and geographical information. They also know various ways in which they can extend their knowledge of these subjects. There are, of course, personal differences in the amount of information that anyone has.

(8) Most people have a certain amount of information about psychology, biology, chemistry and physics.

(9) People have a certain amount of mathematical knowledge. A large part of this is acquired in formal teaching in school or in a university. More sophisticated people also have some meta-mathematical knowledge and a methodology.

(10) Many people have a set of problems that they are keen to solve. My interest in this paper is especially in the knowledge possessed by researchers and such people usually do have a collection of open problems that they are working on.

(11) Many people have a collection of goals or aims that they strive to achieve. Although some people appear to live aimless lives, the kind of person that I am concerned about in this paper does have definite goals.

There is a vast amount of information available in the liberal, Western democracies. You just have to wander through a library or spend a few hours surfing on the Internet to realise this. It is impossible for any one person to possess all this information. Every individual has a selection of the available information and continually makes decisions about how to extend and revise his knowledge. This is one of the reasons why the idea of a person's outlook is so important. The reasons why a person chooses to acquire knowledge about a certain topic are many and various and often involve elements from that person's outlook other than the components consisting of collections of theories or beliefs.

2.2 The Power of Testimony

We have a very strong tendency to believe what other people tell us and what we read. This has been known for some time. For example, Thomas Reid formulated a principle of credulity in section 24 of part VI of his book *An Inquiry into the Human Mind on the Principles of Common Sense* (1764).[4] This 'is a disposition to confide in the veracity of others, and to believe what they tell us.' Everyday life would be impossible without some such principle. It enables us to make use of a great deal of information that we could not acquire for ourselves in any other way. That is the advantage of having some such principle, but its disadvantage is that we also acquire some false beliefs as a result of it. There are many examples from the history of science of people holding beliefs that we today regard as false, but I will give two non-scientific examples. The first of these concerns the number of words for snow in the language of the Eskimos of the Arctic. In some circles it is believed that there are dozens of words for snow in this language. This is apparently false. Someone who investigated this supposed fact found that there are in fact only three different words for snow in this language [5, p. 102]. My second example of a view that although false is widely accepted concerns John Osborne's play *Look Back in Anger*. Many reference books state that the original title of this play was *On the Pier at Morecambe*, but this is not the case [25, p. 35].

Accepting that it is highly likely that everyone has some false beliefs, some philosophers have wondered if there are any limits to the number of false beliefs that a person

[4] Reid's principle of credulity is quoted from Alston's article on it in *A Companion to Epistemology* [11, p. 366].

might have. At least three different kinds of argument have been used to support the proposition that the vast majority of our beliefs must be true. One of these involves evolutionary considerations, one makes use of the predictive strategy that Dennett has called 'the intentional stance' and one involves considerations relating to our ability to communicate with one another.

Dennett is one of those who makes use of evolutionary ideas in order to argue that only a small proportion of anyone's beliefs can be false. As he puts it [13, p. 96], 'if an organism is the product of natural selection, we can assume that most of its beliefs will be true and most of its belief-forming strategies will be rational.' (Others have used similar arguments. Many of these are mentioned, and their arguments evaluated, in chapter 3 of Stich's book *The Fragmentation of Reason* [35].)

The intentional stance is a strategy used to predict an agent's behaviour. You assume that the agent is rational. Knowing the agent's location and its purposes you work out what beliefs and desires it ought to have and putting all those considerations together you predict that the agent will try to achieve its aims knowing what it does. Dennett adds [13, p. 19], 'An implication of the intentional strategy ... is that true believers mainly believe truths.'

Davidson has argued that in order for communication to be possible at all between people those people must have a large number of beliefs in common and most of those beliefs have to be correct. In his own words [12, p. 295], 'Successful communication proves the existence of a shared, and largely true, view of the world.'

It would be out of place for me to examine any of these arguments thoroughly here. My concerns in this paper are different and I am interested in other issues. I am interested, for example, in the reasons why we do in fact have quite a few false beliefs, in what can be done to decrease the number of false beliefs that we acquire and in the ways in which people go about trying to find out which of the theories that they hold are false. I find it more useful to think about these issues rather than speculating about how much error there is.

3 On-line Assessment

3.1 Introduction

It is highly likely that some of my beliefs are false and I regard this to be an undesirable state of affairs. I cannot change the past, but is it possible for me to adopt some strategy in the present that will make it impossible for me to acquire any further false beliefs? If I stop listening to other people and reading, then I will manage to avoid acquiring any new false beliefs from those sources. But even then I may acquire new false beliefs as a result of my memory playing tricks or because I misperceive something or because I make an incorrect inference. If I continue reading and interacting with other people, however, then it is almost inevitable that I will acquire quite a few additional false beliefs. Although we cannot completely avoid acquiring new false beliefs, it is possible to take steps to cut down the number of false beliefs that we acquire. I will return to this possibility after I examine the on-line criteria that we use

to assess incoming assertions.

In order to understand human belief acquisition I think that at least two phases or stages need to be distinguished. In much of everyday life we are concerned with many complex tasks and difficulties and we are engaged in lots of different activities. We do not have the time or inclination to carefully consider each piece of information that we come in contact with. Furthermore, the amount of information we receive is so large that we are simply unable to check everything. Sometimes the source of our information, like the radio or television, is not within our control, so even if we would like to check an individual piece of information we cannot do so there and then because we cannot ask for the flow of information to stop. The radio and television have an on-off switch, but no pause control. Even when we are reading and we could stop in order to more fully investigate some statement we have come across, we do not often do so there and then, because too many interruptions impede good comprehension of a text. We tend to flag the statement as one to come back to later. We do sometimes, however, challenge people who say something to us which we find implausible to provide reasons for what they have said, but we could not challenge everything they say to us in this way. Thus, most of the time in all these circumstance we tend to use easily applicable criteria for deciding whether or not to accept what we read or are told. For example, if we ask a stranger for street directions or the time, we usually simply accept what we are told without giving the matter a second thought. Our goal is to get to a certain destination and if we decided to test what we were told that would force us to alter our current plans. If we had to arrive at our destination at a specific time, maybe we had arranged to meet someone there, then we may well miss our appointment if we got involved in an investigation of the correctness of the directions we were given. However, especially if we are involved in any sort of research, we do, at our leisure and often in the quiet of a study, thoroughly check a few of our beliefs or conjectures in order to ascertain their truth or falsity. There are, of course, several different ways of checking information and the method employed may also depend on the kind of information involved, for example, whether it is historical, mathematical, ethical or whatever. For convenience I will call the two phases of belief-acquisition that I have mentioned *on-line assessment* and *off-line checking*. I discuss on-line assessment in this section and off-line checking in the next.

My conjecture is that the criteria we use to assess information on-line are a set of rules that we apply to the things that we hear and read in order to decide what to do with them. For example, whether to accept or reject them or whether to flag them as worthy of further consideration. In this section I look at the factors that we take into account when assessing what we hear and read on-line and I also look at the form that the rules we use take. Working out what rules a particular person uses is more complicated than may at first sight appear. To devise a set of rules that could be programmed into an android would constitute a substantial research project.

3.2 Methodological Rules

We assess the locutions that we hear and read in at least two different ways. We have some criteria that we use on-line and we have a methodology that we use off-line. We

must have some easily applicable criteria because in everyday life we have to make decisions in real time about what to believe and what to reject (or suspend judgement about). I think that it is best to conceive of these on-line criteria as a collection of rules. The idea of construing these criteria in this way is based on some remarks made by the philosopher Price. Most of his discussion of how we learn things from other people in lecture 5 of his book *Belief* [31] centres around the awkwardly phrased principle, 'What there is said to be (or have been) there is (or was) more often than not' [31, p. 116], but at one point he says [31, p. 124]:

> There is however another way of interpreting the principle we are discussing. Perhaps it is not itself a proposition which we believe, still less a proposition believed with complete conviction. Instead, it may be more like a maxim or a methodological rule. In that case, it is better formulated in the imperative than the indicative mood. We might put it this way: 'Believe what you are told by others unless or until you have reasons for doubting it.'

Price's methodological rule is too general. It lumps together and treats in the same way every assertion that we encounter. It does not, for example, distinguish between different sources of information and the contexts in which we may encounter them. It applies to the answer we receive when we ask a stranger the time and also to the assertions made in an article published in a learned journal. It encompasses the gossip we hear from a friendly neighbour and also the information presented at a lecture at a prestigious conference. It covers the case when we are told that someone close to us has died and also the situation in which we hear that it is snowing in Scotland. We do treat the information we receive in different ways depending on its content, the source of the information, the context in which it was received, etc. Rather than having a single rule like Price's, I suggest that we have several rules that take these various factors into account. In the next three subsections I consider some of the factors that must be taken into account in the formulation of the rules that we use. These rules and the information they apply to could be classified in several different ways. I am going to discuss rules that apply to information obtained from different kinds of source separately. Some of the information we receive we obtain from other people face-to-face, some of it we get from the media, some from the Internet, some by reading articles and books and so on. I am not going to consider every conceivable source of information. In order to illustrate the issues involved I will look at three areas, namely information received in the course of a conversation, that received from the media and that received from a written text.

3.3 Assessing Personal Communications

Some of the information that we acquire we receive from other people face-to-face. Hume was aware of some of the factors involved in assessing the reliability of information conveyed in this way. In section X of his *An Enquiry Concerning Human Understanding* (1748) he mentions various factors that we take into account when

assessing the truth or otherwise of what other people tell us. He says that we consider the character of the person involved. If he is of doubtful character, then we do not necessarily accept his testimony. We consider whether or not the person has an interest in what he tells us. We also take into account the manner of the person's delivery. If he either hesitates or presents his testimony with 'too violent asseverations', then this may arouse our suspicions. Hume also says that we take into account the number of witnesses involved and whether or not they contradict themselves. Hume's observations are as relevant today as when he first made them and they are a good starting-point, but they need to be supplemented by some further considerations. In this subsection I consider the situation in which we receive information from one other person. We take at least the following considerations into account when assessing the truth of what we are told:

- The context in which the speaker makes his assertions is relevant to the way in which we receive them. Thus, something that is asserted in the course of an informal conversation with a friend is treated differently from the way in which we would consider a witness's statement in a court of law.

- We take the person's character into account if we have some knowledge about it. To take an extreme case, we are unlikely to believe what a person we know to be a habitual liar tells us.

- We take into account the social or political role of the person making the assertion. (People can, of course, have several different roles in society.) We treat differently, for example, a judge's statements in a court of law from those of the prosecuting barrister.

- We take into account the person's experience and expertise. Thus, we take more seriously what a lawyer who has been practising law for twenty-five years says, for example, about the law than what a newly-qualified solicitor might tell us.

- We take into account the manner in which the person gives us the information and his body language. Thus, we might suspect that someone is lying if he makes more eye contact with us than is usual or if he makes a lot of meaningless hand movements.

- If we think that the person has a good reason to deceive or mislead us, then we take that into account when considering the likely truth of what he tells us.

- We consider the content of the message and its importance and relevance to us. One consideration concerns the coherence of the message and its internal consistency. Furthermore, we tend to be less critical of assertions that are not particularly important to us than those that are really important to us. For example, if a person has little or no interest in rugby football and he hears by accident that Moseley have won their latest game, he is likely to simply accept this. However, if he has published articles about BSE in which he contends that it is caused by a virus and someone tells him that it is really caused by a prion, he is unlikely to accept this there and then.

- Whether or not we accept another's assertion may be influenced by our pre-existing knowledge. For example, if the assertion is straightforwardly inconsistent with what we already know and we are confident of the truth of the statements that it is inconsistent with, then we are unlikely to accept the assertion outright. We may, though, flag it as something we should investigate more fully later. This consideration includes the case when our knowledge is that of what this speaker has already told us either earlier on this occasion or an a previous one.

- We consider the obvious consequences and repercussions of accepting or rejecting the message. For example, if the message is such that accepting it would have a profound effect on my current plans, my life-style or my belief-system, then I am unlikely to accept it outright, even if it comes from a reliable source who shows no signs that he is lying. In such a case I would probably flag the assertion as one that I need to consider thoroughly at some later time.

My conjecture is that people have a collection of rules that they use to decide what to do with the assertions that they hear from another person. The main options available are to accept what is heard, to reject it or to flag it as something worth investigating more thoroughly later. Whatever we decide to do with the assertion, we usually remember who made that particular assertion. Thus, if Jones tells us that it is safe to eat beef, then, whether we decide to accept what he says or reject it or flag it as worthy of further study, we will tend to remember that Jones told us this and we will usually also add the belief that Jones told us this to our belief-system with some indication of when he said this. Such information may come in useful later in helping us to decide whether or not to treat Jones as a reliable source of information.

In order to give an example of the kind of rule that occurs in a person's first-phase rule set, consider the case of talking to a person you have not met before and about whom you know very little. In this case we cannot take the person's character into account nor his social role nor his level of expertise. As the conversation progresses, we may learn something about these things and start taking them into account. We can, however, from the start take the person's body language into account, the content of his assertions, our pre-existing knowledge and the consequences of accepting his assertions. The rule we use initially in these circumstances will, therefore, look something like this:

> Believe what you are told unless:
>
> (1) the person's body language and manner of delivery suggest that he is lying;
>
> (2) the content of what he says is the sort of thing that people usually lie about;
>
> (3) the assertion is incoherent or is obviously inconsistent with something the speaker said earlier;
>
> (4) the assertion is straightforwardly inconsistent with your current belief-system; and

(5) the content of the assertion is of vital importance to you and your acceptance of it would entail making major changes to your current plans, your life-style or your belief-system.

In addition we make use of some auxiliary rules like these:

> If you decide not to believe what you hear, but you consider it worth thinking about, then flag it as such and come back to it later.
>
> Whatever you decide to do with this assertion, add the belief that it was asserted by this particular person on this occasion to your belief-system.

3.4 Assessing Information Received from the Media

Most people in the older Western democracies receive a great deal of information from the media and the information obtained in this way is not restricted to the current cultural, economic and political situation. The following is a list of some of the factors that we take into account in order to evaluate the information that is produced by the various media.

- We take the medium into account. For example, we treat newspapers, the radio and television differently. Furthermore, we do not treat all radio stations, for example, in the same way. Radio stations differ in the care with which they vet the information they put out.

- We take the owner of the newspaper, radio station, television station or whatever into account. This is especially true of newspapers where proprietors have historically tended to exercise a lot of influence over the way in which stories are presented in their papers.

- We take the country in which the organisation is based into account. This is because there is, for example, a different amount of government control over the media in different countries. Thus, an English person is more likely to believe what he hears in a news bulletin broadcast on BBC Radio 4 than what he hears broadcast on Radio Baghdad. That is not to say that such a person accepts uncritically everything asserted in a news bulletin on Radio 4, but he does not treat everything as being of dubious authenticity.

- In the case of the radio and television, we take the kind of programme during which the assertions were made into account. Not all programmes are factual. How we assess a piece of information depends on whether it was put out during a news programme or a documentary or a short story or a play or whatever.

- In the case of the radio and television, we take into account what we know about the person who asserts the piece of information. In the case of the print media, we take into account any knowledge that we might have about the author of the piece in question. The considerations involved are similar to those that were discussed above in connection with our assessment of personal communications.

- We take the content of the message into account. Russell briefly touches on this point in *The Problems of Philosophy* [33, p. 77]:

  > If the newspapers announce the death of the King, we are fairly well justified in believing that the King is dead, since this is the sort of announcement which would not be made if it were false.

  The Problems of Philosophy was originally published in 1912 and Russell was writing in England. In some other countries and in different times announcements about the health of the monarch may well be attempts at disinformation. Having said that I do think that Russell's point is largely correct. The more significant an assertion is the more thoroughly it would have been checked before being included in a newspaper article.

- Our pre-existing knowledge may be relevant. This is the same consideration that was mentioned in relation to the ways in which we assess assertions made in the course of a conversation.

- The consequences of our accepting the message may be important. This is the same consideration that was mentioned in relation to the ways in which we assess assertions made in the course of a conversation.

3.5 Assessing Written Sources

As mentioned in the introduction, the kind of person that I am concerned with in this paper is the research worker. Such a person receives a great deal of the information that he uses from written sources. There are very many different kinds of written material available in the older Western democracies and we take a number of things into account when assessing this material. Some of the factors that are relevant are the following:

- We take the kind of document involved into account. We treat information obtained from an article in a learned journal, for example, differently from that which we get from reading a cereal packet.

- If the information comes from a document with a named author (or authors) and we know something about this person, then we take this into account when considering the quality of the information involved. If the document is an article in a journal, say, then many people take the author's affiliation into account. An article written by somebody who is a full professor at Harvard tends to be treated more seriously, for example, than one produced by a junior lecturer at the University of Bournemouth.

- If the document is a book or a journal or something similar that is published in a definite country and at a specific time, then we take that into account in assessing the information it contains.

- We take the publisher of the item into account in many cases. This consideration is involved in the following comment from Quine and Ullian [32, p. 14]:

  > We all hold ... that those [beliefs] gained from respected encyclopedias and almanacs are more to be relied on than those gained from television commercials.

  There are, however, great differences between people in how they evaluate reference works. I doubt that Quine and Ullian would regard *Harper's Encyclopedia of Mystical and Paranormal Experience* [16] or *The Encyclopaedia of Occult, Paranormal and Magick Practices* [20] as respected encyclopedias, but there are people who would rate them very highly as authorities.

- We take the content of the assertion into account.

- Our pre-existing knowledge may be relevant.

- The consequences of accepting the assertion may be relevant.

The last three factors on this list are the same as the last three considered above in section 3.3.

3.6 The Two-phase Model

In this paper I am arguing for a two-phase model of belief-acquisition. The first phase consists in the employment of a set of easily applicable rules whereas in the second stage a sophisticated critical methodology is used. In order to illustrate the way in which the first-stage rules are employed and to contrast this with the utilisation of the second-stage methodology I will look at part of Watson's personal account of the discovery of the structure of DNA.

During the time leading up to the discovery of the double helix Watson tried out various ideas. One of these was the like-with-like idea in which adenine coupled with adenine, cytosine with cytosine, guanine with guanine and thymine with thymine. The possibility of such couplings was a consequence of information contained in respected textbooks. Watson was excited by the model he was able to build on these assumptions and mentioned it to Jerry Donohue who said the idea would not work. Watson recounts his reaction as follows [38, p. 192]:

> Though my immediate reaction was to hope that Jerry was blowing hot air, I did not dismiss his criticism. Next to Linus [Pauling] himself, Jerry knew more about hydrogen bonds than anyone in the world. Since for many years he had worked at Cal Tech on the crystal structures of small organic molecules, I couldn't kid myself that he did not grasp our problem. During the six months that he occupied a desk in our office, I had never heard him shooting off his mouth on subjects about which he knew nothing.
>
> Thoroughly worried, I went back to my desk hoping that some gimmick might emerge to salvage the like-with-like idea.

The considerations that Watson adduces are all what I would call first-stage ones. However, they are not the sorts of consideration that would convince the scientific community or even part of that community. They are the kinds of reasons that all of us consider and take seriously in our day-to-day research work, but they are not the stuff of which properly scientific argumentation consists. The reasoning contained in Watson and Crick's famous article "A Structure for Deoxyribose Nucleic Acid" [37] is completely different from that contained in the above passage. This is to be expected. In that paper they employ arguments which conform to the critical methodology that they accept. The employment of such a critical methodology is something that occurs in the second phase of belief-acquisition and I look in detail at this stage in the next section.

It is not only our belief-system that undergoes modification each day, our collection of rules also undergoes frequent modification because of the interplay between the rules and our beliefs. The rules are the gatekeepers that let beliefs in or keep them out or tag them as worthy of later investigation, but the rules themselves make use of our current belief-system. For example, one of the ways in which we assess assertions is by considering the character of the person making the assertion, but our views about his character are themselves beliefs that may undergo change as a result of information that we receive.

The second stage of belief-acquisition may also interact with a person's collection of first-stage rules. For example, we may have trusted Jones in the past and treated him as a reliable source of information, but Green, who we also consider to be reliable, has told us that Jones has a tendency to over-estimate the distance between places. This may be of sufficient importance to us that we investigate it using our second-stage methodology and conclude that it is correct. We then incorporate this information into our belief-system and as a result of that it also forces a change in the rule or rules that we use to assess Jones's assertions on-line.

At the beginning of this section I said that it is possible to do things to decrease the number of new false beliefs that we acquire. One of the things that can be done is to alter our first-phase rules in the light of new information that we receive. That is to say, we change the criteria that we use in order to allow beliefs into our belief-system. We do not have much control over the set of rules that we acquire during our enculturation into society in childhood, but when our critical faculties develop we begin to exercise more control over the rules that we use. Such a set of rules is something that forms part of a person's outlook.

4 Off-line Information Checking

4.1 Introduction

I have stressed the fact that most of the knowledge that a person has has been obtained from other people and that every person's belief-system is modified on a daily basis as new information is taken on board. Most of this information is accepted without being thoroughly checked out and some of it is bound to be wrong. It is, therefore,

safe to assume that everyone has some false beliefs. Some philosophers have tried to show that most of our communal beliefs are correct, but even if their arguments are valid they still acknowledge that people do have some false beliefs. Given that some of my beliefs are bound to be false and that I do not know which, it is profitable to ask, 'How do I decide which of my beliefs to check?' and 'How should I go about checking those beliefs?' I will consider the first of these questions at the end of this section, but to begin with I will look at how the information we have should be checked. It could be argued at this point that such an investigation would have to look at each type of knowledge in turn, because different kinds of empirical information are checked in different ways. I think that there is an element of truth in this objection and it is true that historical methodologies are very different from those employed in physics, but I think that some very general methodological and epistemological principles are shared by different disciplines and it is these that I want to look at here. I consider two kinds of empiricism, namely justificationist and anti-justificationist. I do not claim that these two methodologies are the only possible ones, but I do think that they are the two most important ones. The dominant methodology in the sciences is still what I call *justificationist empiricism* and although parts of this are being given up it still exerts a powerful influence over many scientists. I shall argue, however, that anti-justificationist empiricism is the better methodology to adopt. If our aim is to build an effective and efficient android, then we should equip him with this kind of methodology as it is the best one that is currently available.

4.2 Rationalism

In order to place the contrast between justificationist and anti-justificationist empiricism in context I will first say something about the rationalist tradition and then I will say how the two types of empiricism fit into this tradition.[5] One strand in that tradition is called *uncritical* or *comprehensive rationalism* by Popper [26, p. 230] and *panrationalism* by Bartley [3, p. 87]. The two main components of comprehensive rationalism are, according to Bartley [3, p. 87], the following:

> (1) A rationalist accepts any position that can be justified or established by appeal to the rational criteria or authorities; and (2) he accepts *only* those positions that can be so justified.

Several different rational authorities have been put forward over the years. Bartley talks of the 'depressingly similar pattern' of philosophical revolutions [3, p. 84]. For example, the church was replaced by intellectual intuition, which in turn was replaced by sense experience and that was replaced by a particular language system [3, p. 84].

Comprehensive rationalism fails for a number of reasons. In his book *Unfathomed Knowledge, Unmeasured Wealth* [4, pp. 230–232] Bartley cites four of these. The most important consideration relates to the fact that the two requirements of comprehensive

[5]Unfortunately, the word 'rationalism' is used in several different ways. There is a philosophical movement known as *Rationalism*, comprising such thinkers as Descartes, Spinoza and Leibniz, but I do not use the word in this narrow sense. In this paper a rationalist is simply someone who takes rational arguments seriously. In this sense rationalism contrasts with irrationalism rather than empiricism.

rationalism are mutually incompatible. That means that it is impossible to justify everything rationally and yet that is exactly what comprehensive rationalism demands. Thus, it is inconsistent.

Rationalists who appreciate the limitations of comprehensive rationalism have tended to embrace one or other form of limited or critical rationalism. Here the requirement to justify everything is abandoned. The limited rationalist realises that not everything can be justified. Those principles and statements that he cannot justify, but which he needs, he accepts irrationally. Clearly, something has gone wrong for the rationalist if he has to accept some things irrationally and Bartley spells out some of the difficulties in *The Retreat to Commitment* [3, pp. 96–107]. Although there are difficulties in limited rationalism, it does at least have the merit of being consistent.

What both comprehensive and limited rationalism have in common is their acceptance of justificationism. A consistent rationalism, without the difficulties of limited rationalism, is possible if one gives up justificationism. My strategy in this section is not to discuss justificationism in general, but rather to look at one particular form of justificationism, namely justificationist empiricism, that still exerts a powerful influence over many scientists, including many working in AI.[6] This is a form of limited rationalism and so it suffers all the defects of that kind of rationalism, but I will not here repeat Bartley's account of these defects. Rather, after characterising justificationist and anti-justificationist empiricism and comparing some of their components, I will present an argument which shows that anti-justificationist empiricism is the better methodology to adopt in a community containing people using both kinds of empiricist methodology.

4.3 Justificationist Empiricism

Empiricism has taken many forms over the years and it has been in and out of favour many times. Justificationist empiricism still exerts a powerful influence over the thinking of many scientists. It has the following main components:

- Knowledge is defined as being justified true belief. It is, therefore, something subjective.
- Justificationist empiricists are often engaged in, what Dewey called, the quest for certainty.
- The justificationist empiricist holds that scientific error can always be avoided and, therefore, that someone who makes a scientific mistake should be blamed for having made it.
- Sense experience or observation is singled out as being the sole rational authority or authoritative source of knowledge. Observation statements are validated by the rational authority and they form the foundation of the scientific edifice. Everything else is built on top of them.

[6] Galliers [14, p. 231], for example, after discussing foundationalist and coherence theories of rational belief change, goes so far as to say, 'All sides agree that justifications as reasons for beliefs are important'.

- A collection of inferential procedures are identified and these are used to infer things from the observation statements validated by the rational authority. Amongst these inferential procedures there is some form of induction which allows the justificationist empiricist to infer universal statements from a collection of singular premises.

- Criticism and justification are fused. In practice, this means that the way in which a theory is criticised is by seeing whether or not it can be derived from the statements sanctioned by the rational authority using acceptable methods of inference. If the theory can be so justified, it is retained. However, if it cannot be so justified, it is jettisoned. To avoid inconsistency the justificationist empiricist has to accept that there are a collection of statements that cannot themselves be justified. These are beyond criticism.

- Scientific knowledge is seen as growing by means of a non-evolutionary process of accumulation.

In the following passage, from Carnap's *The Logical Structure of the World* [8, p. xvii], he is urging philosophers to copy the way in which scientists work. The interest of the passage, however, is in the fact that it presents scientists as using a justificationist methodology and thus collaborates some aspects of the above summary account of this methodology.

> If we allot to the individual in philosophical work as in the special sciences only a partial task, then we can look with more confidence into the future: in slow careful construction insight after insight will be won. Each collaborator contributes only what he can endorse and justify before the whole body of his co-workers. Thus stone will be carefully added to stone and a safe building will be erected at which each following generation can continue to work.

Nowadays, few people would accept Carnap's philosophy in every detail, but his picture of the way in which scientists work would be one that many contemporary scientists share.

4.4 Anti-justificationist Empiricism

The anti-justificationist is not interested in questions of justification. Instead of asking a question like, 'How can we justify our theories or beliefs?', the anti-justificationist asks questions like, 'How does our knowledge grow?', 'How can we get rid of false theories?', 'What are the consequences of our theories and are they acceptable?', 'What are good ways of criticising theories?' and 'How can we decide which out of a group of competing theories is the best?' The following are the main components of anti-justificationist empiricism:

- Knowledge is seen as being fallible, conjectural, hypothetical and subject to revision. It is also thought of as being objective.

- Anti-justificationists are engaged in a quest for truth and not certainty.
- The anti-justificationist holds that scientific error cannot always be avoided and that a scientist who makes a mistake is not always blameworthy.
- Anti-justificationist empiricism is anti-authoritarian. No source of knowledge is treated as being above criticism. There are no secure foundations. In particular, observation statements are not seen as forming the foundation of the scientific edifice. They do, however, have an important role to play. They are used in the criticism of theories rather than in their justification.
- The only inferential procedures sanctioned by anti-justificationist empiricism are deductive ones. At best, induction is seen to be a useful heuristic device.
- Criticism is distinguished from justification and many different types of criticism are employed. No principles are deemed to be beyond the reach of criticism.
- Anti-justificationism is an evolutionary epistemology.

4.5 Comparison and Further Elaboration

In this subsection I look in more detail at several of the ways in which justificationist and anti-justificationist empiricism differ from one another.

4.5.1 Conception of Knowledge

Justificationists can hardly deny that people get a lot of their knowledge from listening to others and reading what they have written, but they regard the knowledge contained in books and other documents as being dependent on subjective knowledge. Anti-justificationists do not deny that subjective knowledge exists, but they see objective knowledge as being more important and subjective knowledge as being dependent on it.

Justificationism is characterised in various ways by different people. Lakatos, for example, thinks that the definition of knowledge is crucial. He [39, p. 54] sees the key feature of justificationism as being 'the identification of knowledge with proven, certain knowledge.' Jarvie and Laor [18, p. xv], however, see the view that 'all error is avoidable and thus culpable' as constituting justificationism, whereas Bartley [4, p. 230] singles out the assumption that 'rational opinion and action must be *justified* or given a foundation.' I think that there is value in all these observations, but I prefer to present a set of features that characterise justificationism without singling out one as being *the* defining attribute. The definition of knowledge as justified true belief is a significant feature of justificationism, but it is not the only important property that it possesses.

4.5.2 Observation Statements

The justificationist empiricist can agree with me that most of the knowledge that I have has been obtained from other people, but he will contend that if you track the

movement of this information through the various transmission chains which end with me you will find that the starting-points of all these chains are observation reports (and maybe also mathematical and logical principles). It is assumed that information is transmitted almost exactly as it is received or that some of the people along the way have passed on information that they have synthesised from what they have received using some acceptable inferential procedures. Such a picture has many faults, but here I want to look at the starting-points of the whole process. I grant that some of the starting-points of the transmission chains that end with me are observation reports. However, I do not accept that observation reports are beyond criticism. In addition, the observation reports that are actually made in any situation are only a tiny fraction of the possible reports that could be made. Which ones are actually made depends on factors other than those that occur in the perceptual environment of the person making the report. Furthermore, there are many statements that are acceptable starting-points which are not observation reports. I will look at these points in greater detail in turn.

We do not accept every observation report that we hear or read uncritically. Encountering an observation report there are several different responses that we can make to it even when we are only assessing the information that we are receiving. We can simply accept it, but we do not have to. If it is inconsistent with some other observation report that we have previously accepted, then we have to think through our response to the new observation statement. Similarly, if it is inconsistent with some theory that we accept, then we need to think things through carefully. Observation reports that are sufficiently important to us can be subjected to various types of criticism. A witness, for example, who says that he saw the accused in the "Goose and Granite" pub on the 3rd of April 1999, for example, can be discredited in a number of ways. We will reject his observation report if we can establish that he was not in the "Goose and Granite" on that occasion or that the accused was not there. If the witness is short-sighted and did not have his glasses with him and claims to have seen the accused in the far corner of the pub, then we are not likely to believe him. And so on.

It is true that people have the ability to make judgements about their perceptual environment, but there are indefinitely many judgements that can be made in any situation. For example, looking around me I can think of lots of true observation reports that I could make. Some of these are, 'The birds are singing outside', 'It is warm in here', 'It is getting dark outside', 'It is raining', 'There are books on the table', etc. Which of these, if any, I decide to make depends upon several factors such as my interests, the problems that I am working on, the point I hope to make by means of my utterance, etc. (In the case of the above reports, they were made in order to illustrate the philosophical point that I was making.) These considerations create problems for the justificationist, because he needs observation reports to occur as the starting-points of transmission chains, but observation reports themselves are made in a context containing features partly drawn from the outlook of the person making them.

It is reasonable to suppose that observation reports should not contain any theoretical terms, but it has proved surprisingly difficult to distinguish adequately between

observational and theoretical terms. It would be too much of a digression for me to discuss this fully here, but there is a good discussion of the difficulties and the untenability of the distinction in Newton-Smith's *The Rationality of Science* [24, chapter II]. Even if the distinction was tenable some starting-points would be statements containing theoretical terms.

I do not wish to endorse everything that Searle has said about institutional facts [34], but he has drawn attention to an interesting feature of our language. We may take 'Jones gave his wife a five-pound note' or 'There is a plain-clothes policeman standing near the telephone box' to be acceptable starting-points, but they are far from being observation reports. They can only be understood and asserted by people who have a shared understanding of the institutions of money, marriage and law-enforcement.

The anti-justificationist also allows universal empirical theories to occur as the starting-points of certain transmission chains. This is because the anti-justificationist sees theories as guesses whose origins are irrelevant to their value. Thus, a transmission chain can begin with a universal theory. The reason, however, why the theory continues to be propagated as a live option is because it has been subjected to various sorts of tests which it has survived. It is not only the theory that is transmitted from person to person, but also the critical discussion that has accumulated around the theory. What Popper [28, p. 59] says concerning our acceptance of Copernicus's theory is true more generally:

> To tell *the story of the observational evidence* which has accumulated for many centuries, starting with the Egyptians and Babylonians, would do no justice to the powerful reasons we have for believing in the truthlikeness of the model. ...
>
> Our real reasons for believing in the truthlikeness of the Copernican model are much stronger. They consist in *the story of the critical discussion*, including the critical evaluation of observations ... . It was not so much the accumulation of observations by Tycho as the critical rejection of many conjectures by Kepler, Descartes, and others, culminating in Newton's mechanics and its subsequent critical examination, which ultimately persuaded everybody that a great step had been made towards the truth.

4.5.3 Inferential Procedures

These days many different things are called *induction*. For example, Holland, Holyoak, Nisbett and Thagard [17, p. 1] take induction 'to encompass all inferential processes that expand knowledge in the face of uncertainty.' As examples they give the following:

> The mother of a four-year-old boy, observing that he has been unusually cranky and obdurate for several days, decides that he has entered a "phase". A laboratory rat, busily pressing a lever to obtain food, hears a distinctive tone, which is followed by an electric shock. The very next time the animal hears the tone, it hesitates in its lever-pressing activity, waiting, one is tempted to say, for the other shoe to drop. A nineteenth-century scientist

> observes the behavior of light under several types of controlled conditions and decides that, like sound, it travels in waves through a medium.

It is unhelpful to think of all these various things as examples of induction. In this paper I use the word 'induction' (1) to refer to a supposed rule of inference from singular premises to a universal conclusion and (2) to refer to an alleged rule of inference from a less general universal statement to a more general one. People sometimes talk of the *inductive method*. This makes use of inferential procedures that are inductive. The idea is that scientists begin by making observations, from these they infer universal laws and then they verify those laws. Such a supposed method is an integral part of justificationist empiricism.

Justificationist empiricists usually include some inductive principles among the inferential procedures that they use in order to justify the scientific theories that they hold from the observation statements that they take to be foundational. I cannot go into a full discussion of induction here, but I will point out what I consider to be the strongest argument against it. The argument is due to Popper. He points out that many scientific theories are inconsistent with their predecessors. For example, Newton's dynamics is inconsistent with both Galileo's terrestrial physics and Kepler's celestial physics [27, p. 198]. I will only mention a few of the incompatibilities. According to Galileo's theory a projectile fired on the earth would follow a parabolic path, whereas according to Newton's theory it would follow an elliptic path. Additionally, according to Galileo a free-falling body, say one dropped from a great height, would fall with constant acceleration, whereas according to Newton's theory as it fell its acceleration would slightly increase as it got nearer to the earth's centre [27, pp. 199–200]. In the case of celestial physics, Kepler's dynamics is inconsistent with Newton's. Thus, according to Newton's dynamics $a^3/T^2 = m_0 + m_1$, where a is the mean distance between two bodies of mass m_0 and m_1 and T is the time of a complete revolution of one around the other, whereas according to Kepler's third law $a^3/T^2 = k$, where k is a constant independent of the mass of the bodies involved [27, pp. 200–201]. As Newton's theory is inconsistent with both Galileo's and Kepler's theories it cannot be logically inferred from them (assuming that they are both consistent theories).

4.5.4 Conception of Criticism

Justificationists and anti-justificationists have very different ideas about criticism. The way in which a justificationist empiricist criticises a theory is by showing that it does not follow from observation statements. Anti-justificationists, by contrast, have a wider conception of criticism and use various strategies in order to criticise theories. The following is a selection of some of the ways in which a theory can be criticised.[7] As well as illustrating the anti-justificationist conception of criticism, this list also represents the sorts of thing that an android researcher would have to be able to do.

[7]The ways of criticising a theory contained in the following list are based on the four checks that Bartley identifies [3, p. 127], namely the checks of logic, sense observation, scientific theory and the problem, and some of Laudan's work on conceptual problems [21, chapter 2]. I do not claim that Laudan is an anti-justificationist, but some of his views on the ways that theories can be criticised can profitably be incorporated into an anti-justificationist framework.

- Is the theory consistent? If we discover that a theory is inconsistent, then that inconsistency needs to be removed. Although people sometimes work with inconsistent theories, this is just a stop-gap measure until the source of the inconsistency can be located and a more acceptable solution found. A theory can be inconsistent whether it is mathematical, empirical or metaphysical.

- What problem is the theory intended to solve? According to Popper's anti-justificationist epistemology, theories are put forward in order to solve problems and one way to criticise a theory is to show that it does not solve a genuine problem.

- Does the theory successfully solve the problem it was put forward to solve? Even if a theory is put forward in order to solve a genuine problem, it may be that it does not solve it very well. This is another way of criticising a theory. In this case, it may be that the theory is a good solution of some other problem.

The questions mentioned so far can be asked of any type of theory, but there are differences between the way in which empirical, mathematical and metaphysical theories can be criticised. The following questions are some of those that can be asked of empirical theories:

- Is this theory consistent with observed facts? One way of criticising an empirical theory is by showing that that theory, together with some initial conditions, entails a prediction which is contradicted by an observation report. This is only a criticism of the empirical theory if we have good reasons to think that neither the initial conditions nor the observation report is at fault.

- Is this theory better than its rivals? Even if a group of two or more empirical theories are all consistent, have all adequately solved the same problems and none of them have been falsified, it may still be possible to think that one of the theories is better than its rivals. We may decide, for example, to pick the simplest theory.

- Is the theory in conflict with some other scientific theory that has survived a lot of criticism? If there is a conflict of some sort, then either we have to give up or modify the proposed theory or else we have to give up the other theory. In order to decide what to do we would need to subject both theories to further criticism.

- Is the theory in conflict with the methodology of its parent discipline? If there is a conflict, then either the theory or the methodology has to go, but we would need to submit both to further criticism in order to decide which it is.

- Is the theory in conflict with some elements of the dominant cultural worldview? Popper does not think that such a conflict amounts to a rational criticism of the theory involved. As he says [30, p. 17], 'the historical and sociological fact that the theories of both Copernicus and Darwin clashed with religion is completely irrelevant for the rational evaluation of the scientific theories proposed by them.'

> I think that Laudan [21, pp. 46–47] is correct to disagree with this. If there is such a conflict, then either the theory or the element of the worldview involved has to give way. To decide which we would have to submit both to further criticism.

4.5.5 Accumulation or Revolution?

One of the ways in which justificationist and anti-justificationist empiricism differ from one another is that in the former scientific knowledge is seen as growing by means of a non-evolutionary process of accumulation, whereas in the latter knowledge is seen to grow by means of an evolutionary process. This difference needs to be clarified because several approaches to the growth of knowledge go by the name 'evolutionary epistemology'. (Campbell's article "Evolutionary Epistemology" [7] discusses quite a few of these.) The form of evolutionary methodology that I have in mind is that due to Popper. This uses terminology drawn from evolutionary theory metaphorically to describe the growth of knowledge. This process can be seen as having three stages. (1) One or more blind or unjustified variations are put forward as solutions to some problem or other. (Such variations are often called *random*, but they are not random in a statistical or probabilistic sense.) They can be either new combinations of existing elements or novel mutations. (2) The variations are subjected to a consistent and systematic selection or weeding-out process. (3) Those variations that have survived are preserved and propagated.

Unlike some other evolutionary epistemologists Popper does not draw conclusions for epistemology from any considerations relating to the way in which human beings, including their intellectual abilities, have evolved. He does not solder his epistemology to any scientific theory of evolution as, for example, Campbell does [7, p. 47]. Popper insists that his methodology is not an empirical theory [28, p. xxv]. It is, rather, a metaphysical theory that may well contain normative elements.

4.6 Prospering in the Scientific Marketplace

It is not the case that there is a single methodology that all scientists accept. Some scientists accept justificationism and some accept anti-justificationism. (There are probably others who accept other methodologies.) Whereas this fact poses a real problem for the justificationist, the anti-justificationist can take it in his stride. The considerations that I am about to present constitute an *ad hominem* argument, that is to say, I draw out unacceptable consequences for the justificationist on the assumption that his views about knowledge and methodology are correct. In connection with the other arguments that have been put forward against justificationism (some of which have been briefly mentioned or alluded to above) I think that this constitutes a refutation of justificationist empiricism.

Although individual scientists have their own goals and aims, the aim of science or the scientific community is the production of true explanatory and predictive theories. In doing their work, as has been stressed in this paper, scientists have to make use of information that was obtained from other people. On the whole the justificationist

empiricist sees scientific knowledge as growing through a non-evolutionary process of accumulation. What is important to him is that the knowledge was obtained using a fairly reliable method. I am not suggesting that all justificationists are reliabilists or that they all accept a reliability theory of knowledge.[8] What I am saying is that they all accept some form of the inductive method and they think that knowledge that is obtained by means of this method is generally reliable. For them the pedigree of a piece of information is of crucial importance.

The anti-justificationist empiricist sees things very differently. For him the way in which theories are produced is irrelevant from an epistemological or methodological point of view, though it may be interesting from a psychological perspective. What is important for him is how those theories are criticised using methods that do not involve the attempt to justify them. The anti-justificationist disregards what he knows about the origins of theories when he is involved in checking them off-line. If he comes across a theory that is claimed by its author to have been produced by using an inductive method, he disregards that when he is criticising that theory. He is prepared to criticise a theory allegedly obtained by using the inductive method just as much as one obtained in any other way. Hence, the existence of justificationists in the scientific community presents no epistemological problems. Their presence slows down scientific progress and acceptance of their methodology creates the illusion that scientific knowledge is especially reliable, but the anti-justificationist can accommodate himself to these things. He is willing to consider the theories propounded by justificationists on their merits irrespective of their origins. He considers the time the justificationist spent arguing for his theory inductively as having been wasted. He would have taken the theory seriously, if it had intrinsic merit, even if there were no such argumentation present.

The justificationist, however, has real difficulties accommodating the existence of anti-justificationists in the scientific community. These people propound theories and then they try to falsify them. They spend much of their time criticising theories rather than trying to conclusively establish them. They are not concerned with the origins of their theories nor do they claim they have a pure pedigree. Yet sometimes their theories are generally accepted and become, for a time at least, part of the fabric of knowledge. The way in which they work ensures that the knowledge they produce does not have a pure pedigree. The justificationist has to make use of knowledge produced by other people, but he cannot ascertain the pedigree of every piece of information that he uses. His goal of adding a few stones to the growing edifice of scientific knowledge is undermined by the fact that anti-justificationists build in the air. The presence of anti-justificationists spoils the pedigree, for the justificationist, of a great deal of scientific knowledge and, from his point of view, he cannot always know which pieces of information have been infected. Thus, in a community of scientists using different methodologies, the anti-justificationist has a definite advantage.

[8] Goldman [15, p. 2] summarises the reliability theory of knowledge as follows, 'to be a case of knowing, a true belief must be formed by a cognitive process or method that is generally reliable, i.e., one that generally produces true beliefs.'

4.7 What to Check?

It should be clear from what I have said above that I consider the application of an anti-justificationist methodology to be the best way in which to test the information that we receive. However, this still leaves unanswered the question of deciding which pieces of information to check. We make use of so much information that it would be impossible to check all of it and yet it is likely that some of the information that we use is incorrect, so it makes a great deal of sense to check some of it in order to ascertain whether or not we should continue to make use of it.

Sometimes we realise that one of our beliefs or one of the theories that we hold is false by chance, but it is not a good idea to leave the discovery of error entirely to chance. It does not make sense to adopt a strategy of checking our beliefs one by one in some arbitrary order which ensures that each belief gets checked eventually. It would make more sense to adopt a more systematic approach to the task which takes into account the fact that some of our beliefs are more important than others. It is not possible to devise a foolproof strategy for detecting error, but anti-justificationist methodology suggests a way of working that can help in the detection of error.

One of the most important considerations in the checking of information is the recognition of the fact that nothing is beyond criticism. The anti-justificationist should not protect any belief or principle from criticism, no matter how well they have stood up to criticism in the past. Another important consideration is the realisation of the big part that problems play in the growth of knowledge. The best way for the researcher to work is by trying to solve problems that are explicit and have been clearly articulated. The theories that the research worker should criticise are those that he needs in order to solve the problems that he is working on. These include his own theories and also those belonging to current received wisdom. The more important a theory to the solution he currently favours the more thoroughly it should be checked and criticised. Clearly, the importance of a theory or belief is not something that can be decided independently of the problem under consideration. Importance in this context is not an absolute notion, but one that is relative to a definite problem.

It is often difficult to criticise a theory if there are no alternatives to replace it with. So, if a researcher is aware of only one theory that solves the problem that he is interested in, then a preliminary step that he would often have to make, before he could begin to adequately criticise that theory, is to devise one or more alternatives to it. Since several of the ways in which theories are criticised presuppose the existence of alternatives, the proliferation of rival theories will help the theory in question to be criticised. For example, one way of criticising a theory is by showing that it is inconsistent with an observation report that has itself withstood criticism. If we are only aware of the theory in question, then it might be difficult to think of possible falsifying observations, but with two rival theories the choice of relevant observations is easier. It makes sense, for example, to choose to make an observation that would falsify at most one of the theories involved.

5 Conclusion

There are very many problems to overcome if we are ever going to build a humanoid robot with intellectual abilities analogous to those possessed by human beings. Although it is impossible to design and build an android whose abilities replicate those possessed by human beings, it is sensible to design androids, at least initially, whose abilities are similar to human intellectual ones. In designing an android it makes sense to design one that human beings can interact with. If the android was very different from us, then this would not be possible. In this paper I have concentrated on some of the problems that arise from the fact that human beings need a great deal of information in order to be fully-functioning members of any human society and the academic community in particular. It is impossible for them to generate all this knowledge for themselves. Most of this knowledge comes from other people. Any android that we design and build would be in the same position and in this paper I have looked at some of the issues involved in designing the android's belief-filter component.

Although most of our knowledge has been obtained from other people, rather than looking at the issues involved in a developmental way, I have started from the perspective of a mature adult. It is adults who bring up and educate children and what the formal education process and the informal enculturation process are trying to bring about are adults who can fit into society. The processes of education and socialisation are best understood by looking at what they produce. It is more profitable to study the process having studied the product rather than the other way around.

Although my main interest in this paper has been with empirical knowledge, I have stressed the fact that people are much more than large databases of beliefs. As well as holding many theories, people also have a variety of intellectual skills and they also accept various principles, norms, rules, values and so on. Furthermore, they have various goals that they want to achieve and various problems that they want to solve. This is especially true of the sort of person that I have concentrated on in this paper, namely the creative research worker.

I have presented a two-stage model of belief-acquisition. The first stage consists of a collection of methodological rules that we apply on-line in order to assess the information that we are receiving. On the whole we tend to believe what we are told and what we read. We need to have a reason not to believe something that we hear or read. If we decide not to accept an assertion, but we consider the matter sufficiently interesting, then we can decide to flag the assertion as something that we should investigate more thoroughly later on. The second stage is that in which this more thorough checking takes place. I have argued that the best methodology to use in doing this is an anti-justificationist one.

There is a lot of interaction between a person's belief-system, his collection of first-stage methodological rules and his second-stage critical methodology. The following is a list of the three main ways in which these elements interact:

(1) Beliefs are added to a person's belief-system if they pass through the set of rules that he accepts.

(2) The collection of rules itself makes use of various beliefs and so can change as the

person's belief-system changes. For example, the person may rate a particular reference work very highly until he discovers a number of errors in it. Form that time on information gleaned from that source will be treated differently from the way it was treated before the person's opinion was revised. Furthermore, such a change in evaluation may entail a revision of beliefs obtained from that source in the past.

(3) Beliefs are also added to a person's belief-system and removed from it as a result of the operation of that person's critical faculties.

When we come to accept a new statement or theory, this may have a knock-on effect on our pre-existing knowledge. We may have to revise some of the things that we previously believed. This is an issue that I have not looked at in this paper, because my interest was focused elsewhere.

It is impossible to check everything and so, even when a person is thoroughly checking out a selected position, he has to take on trust a vast amount of information that he could never himself investigate properly. Such knowledge was largely obtained from other people and it had only to pass through the person's set of first-stage rules.

A great deal of work still needs to be done if we are ever going to get close to the goal of building an android with human-like abilities, but I have given some indication in this paper of what we are aiming at in the case of belief-acquisition. I would like to single out two particularly fruitful research areas. The first would involve formulating in sufficient detail for them to be implemented a set of methodological rules for assessing incoming assertions on-line. The second would involve a theoretical investigation of the very many different ways in which people criticise theories in an anti-justificationist manner. In this task a great deal of theoretical work still needs to be done before we could even make a start at implementing anything.

A Why Computers Will Never Have *Human* Emotions

Some people regard the ultimate aim of AI to be that of designing and building an artificial person who will be able to go unnoticed as he lives in human society. I do not believe this goal to be achievable and in this appendix I want to briefly state my reasons for this.

The humanoid robot envisaged in the statement of the ultimate goal of AI is one that is made up out of mechanical rather than organic parts. It would be a more sophisticated version of the robots that are currently being designed and built at MIT, for instance, as part of the Cog Project.[9] The issues involved in producing an organic replicant are, on the whole, different from those involved in making a mechanical android and will not be discussed here.

In order to design any complex mechanism the designer has to make use of a large number of theories. This is as true of the design of a bridge, say, as it is of the design of an android. These theories will concern many different things. In the case when the

[9]For more information about the MIT Cog Project visit `http://www.ai.mit.edu/projects/cog`.

design is of an android some of these theories will relate to the engineering aspects of the task, whereas others will be about the mental life of the android and the outward manifestations of his internal mental states that the android will need in order to live in human society. I am here concerned about the latter group of theories. To make the argument easier to follow I will consider, for illustrative purposes, the problem of designing the android's emotion sub-system, but similar considerations would apply to every sub-system in the android's construction. Clearly, in order to be able to live in human society and pass unnoticed the android must be able to display and recognise human emotions. One of the problems facing the designer is that there are very many different theories of emotion to choose from. For example, Strongman in chapter 2 of his book *The Psychology of Emotion* [36] discusses about thirty different theories, but he does not claim that he has mentioned every theory of emotion that there is. He admits [36, p. 13], 'To describe all theories of emotion would necessitate a book in itself, so the number has been restricted.' Although this is a problem facing the designer, my argument does not depend on the fact that there are many different theories of emotion. It is not the number of theories that is important, but rather some of the properties that any conceivable theory must have. The considerations that I am going to present would apply even if there was only one universally accepted theory of emotion. Just because a theory is universally accepted by the scientific community, does not mean that it will continue to be accepted by that community forever nor does it mean that the theory in question is true. General theories cannot be either conclusively verified or conclusively falsified. They remain forever conjectural. Thus, the designer has to make use of a conjectural theory that he cannot know is correct and which cannot be proved to be true. It could be objected that, although the theory in question cannot be known to be true, it might by accident just happen to be true. I do not know of any conclusive argument which establishes that this could not happen, but there are several considerations that strongly suggest it could not happen. A believer in induction could at this point make use of what Newton-Smith calls *the pessimistic induction* [24, p. 183]:

> Past theories have turned out to be false, and since there is no good reason to make an exception in favour of our currently most cherished theories, we ought to conclude that all theories which have been or will be propounded are strictly speaking false.

He even goes so far as to say [24, p. 14], 'Indeed the evidence might even be held to support the conclusion that no theory that will ever be discovered by the human race is strictly speaking true.' Being an anti-justificationist I do not accept that there are inductive arguments, but that does not matter because the conclusion of Newton-Smith's argument is wholly in accord with anti-justificationist principles as expounded in section 4.4 above.

Returning to the question of designing the emotion sub-system of an android, let us say that the designer uses the theory of emotion T. Theory T is a model of human emotion, which we can assume is produced by some part P of the human organism. If T is used in the design and manufacture of the emotion sub-system of an android, then

that android will display emotions by means of some component C that encorporates theory T in its design. Thus, whereas human emotions are produced by P, android emotions are produced by C which was designed using a theory T which is a model of P. In these circumstances it would be possible for human behaviour and emotional response to falsify theory T, but it would not be possible for android behaviour and emotional response to falsify T. As a theory of human emotion T would be a falsifiable empirical theory, whereas as a theory of android emotion produced by component C it would be an unfalsifiable, non-empirical theory. It would be possible for theory T to be falsified by experiments involving human beings, but it would be impossible for android experience to falsify the theory. It is reasonable to assume that human knowledge will continue to grow in a revolutionary, non-incremental fashion and therefore that any theory of emotion that takes the place of T in the above argument will be at least a good approximation to the truth, but false nonetheless. This means that the difference between human emotional response and android emotional response will be established by any human behaviour that falsifies theory T and thus there will be a way of detecting the presence of such androids in human society.

Perhaps an analogy will make my reasoning clearer. Consider the problem of designing and building a physical model of the solar system. Such an orrery could be designed to illustrate one of several theories. For example, it could be designed to produce behaviour satisfying Ptolemy's theory or Copernicus's or Kepler's or Newton's. If the miniature solar system was designed according to Kepler's theory and then built, its behaviour could not deviate from that theory. No matter how many observations we took of the miniature solar system they would always be in conformity with that theory. It would be impossible for any observations of that model to falsify Kepler's theory. If we compared the orrery to the real solar system, then we would be able to discover discrepancies. In this analogy, the orrery corresponds to an android built according to some theory of emotion. Kepler's theory corresponds to that theory of emotion and the real solar system corresponds to a real human being. What the argument and the analogy show is that there will always be detectable differences between human emotional behaviour and that displayed by an android. The above argument concerns itself with the emotion sub-system of an android, but similar considerations apply to every sub-system in the android that affects the android's mental processes and their outward manifestations.

The above argument is a non-existence proof. It shows that humanoid androids would not be able to live undetected in human society. This conclusion would not bother someone who was trying to design and build androids as an engineering task; it would only bother someone who had other reasons for trying to make a human-like android. For example, someone who had a reductionist or mechanistic view of human beings. The above argument, being a non-existence proof, does not draw limits to what humanoid androids would be able to do. All it shows is that they could be detected from their behaviour. An interesting question remains and that concerns the drawing of such limits, but that is a problem that I will consider on another occasion.

References

[1] Jonathan E. Adler. Transmitting knowledge. *Noûs*, 30(1):99–111, March 1996.

[2] Robert Audi. The place of testimony in the fabric of knowledge and justification. *American Philosophical Quarterly*, 34(4):405–422, October 1997.

[3] William Warren Bartley, III. *The Retreat to Commitment*. Open Court Publishing Company, La Salle and London, second edition, 1984.

[4] William Warren Bartley, III. *Unfathomed Knowledge, Unmeasured Wealth: On Universities and the Wealth of Nations*. Open Court, La Salle (Illinois), 1990.

[5] Wayne C. Booth, Gregory G. Colomb, and Joseph M. Williams. *The Craft of Research*. Chicago Guides to Writing, Editing, and Publishing. The University of Chicago Press, Chicago and London, 1995.

[6] Rodney A. Brooks, Cynthia Breazeal, Matthew Marjanović, Brian Scassellati, and Matthew M. Williamson. The Cog project: Building a humanoid robot. This article is going to appear in a volume of the Springer-Verlag Lecture Notes in Computer Science series of books. It is also currently available on the Internet: `http://www.ai.mit.edu/projects/cog/publications.html`.

[7] Donald T. Campbell. Evolutionary epistemology. In Gerard Radnitzky and William Warren Bartley, III, editors, *Evolutionary Epistemology, Rationality, and the Sociology of Knowledge*, chapter II, pages 47–89. Open Court, La Salle (Illinois), 1987.

[8] Rudolf Carnap. *The Logical Structure of the World* and *Pseudoproblems in Philosophy*. Routledge & Kegan Paul, London, 1967. This volume includes a translation, by Rolf A. George, of the second edition of Carnap's *Der Logische Aufbau der Welt*.

[9] Eugene Charniak and Drew McDermott. *An Introduction to Artificial Intelligence*. Addison-Wesley Series in Computer Science (consulting editor: Michael A. Harrison). Addison-Wesley, Reading (Massachusetts), 1985.

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

[11] Jonathan Dancy and Ernest Sosa, editors. *A Companion to Epistemology*. Blackwell Companions to Philosophy. Basil Blackwell, Oxford, paperback edition, 1993.

[12] Donald Davidson. The method of truth in metaphysics. In Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, editors, *Contemporary Perspectives in the Philosophy of Language*, pages 294–304. University of Minnesota Press, Minneapolis, 1979.

[13] Daniel C. Dennett. *The Intentional Stance*. MIT Press, Cambridge (MA) and London (England), 1987.

[14] Julia Rose Galliers. Autonomous belief revision and communication. In Peter Gärdenfors, editor, *Belief Revision*, pages 220–246. Cambridge University Press, Cambridge, 1992.

[15] Alvin I. Goldman. *Philosophical Applications of Cognitive Science*. Westview Press, Boulder, 1993.

[16] Rosemary Ellen Guiley. *Harper's Encyclopedia of Mystical and Paranormal Experience*. HarperCollins, New York, 1991.

[17] John H. Holland, Keith J. Holyoak, Richard E. Nisbett, and Paul R. Thagard. *Induction: Processes of Inference, Learning, and Discovery*. The MIT Press, Cambridge (MA), 1986.

[18] I. C. Jarvie and Nathaniel Laor. Introduction to volume II. In I. C. Jarvie and Nathaniel Laor, editors, *Critical Rationalism, the Social Sciences and the Humanities: Essays for Joseph Agassi, volume II*, volume 162 of *Boston Studies in the Philosophy of Science (editor: Robert S. Cohen)*, pages ix–xvii. Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[19] Philip Kitcher. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions*. Oxford University Press, Oxford, 1993.

[20] Brian Lane. *The Encyclopaedia of Occult, Paranormal and Magick Practices*. Warner Books, London, 1996.

[21] Larry Laudan. *Progress and its Problems: Towards a Theory of Scientific Growth*. Routledge & Kegan Paul, London, 1977.

[22] Jack Lyons. Testimony, induction and folk psychology. *Australasian Journal of Philosophy*, 75(2):163–178, June 1997.

[23] Bimal Krishna Matilal and Arindam Chakrabarti, editors. *Knowing from Words: Western and Indian Philosophical Analysis of Understanding and Testimony*, volume 230 of *Synthese Library (managing editor: Jaakko Hintikka)*. Kluwer Academic Publishers, Dordrecht/Boston/London, 1994.

[24] W. H. Newton-Smith. *The Rationality of Science*. International Library of Philosophy, editor: Ted Honderich. Routledge & Kegan Paul, Boston, London and Henley, 1981.

[25] John Osborne. *Almost a Gentleman: An Autobiography, Volume II, 1955–1966*. Faber and Faber, London, 1991.

[26] Karl Raimund Popper. *The Open Society and its Enemies: The High Tide of Prophecy: Hegel, Marx, and the Aftermath*, volume 2. Routledge & Kegan Paul, London, fifth edition, 1966. Originally published in 1945.

[27] Karl Raimund Popper. *Objective Knowledge: An Evolutionary Approach*. Oxford University Press, London, 1975. Originally published in 1972.

[28] Karl Raimund Popper. *Realism and the Aim of Science*. Routledge, London and New York, 1983. From the *Postscript to the Logic of Scientific Discovery*. Edited by W.W. Bartley, III.

[29] Karl Raimund Popper. *In Search of a Better World: Lectures and Essays from Thirty Years*. Routledge, London and New York, 1992. Translated by Laura J. Bennett, with additional material by Melitta Mew. Translation revised by Sir Karl Popper and Melitta Mew.

[30] Karl Raimund Popper. *The Myth of the Framework: In Defence of Science and Rationality*. Routledge, London, 1994. Edited by M. A. Notturno.

[31] Henry H. Price. *Belief: The Gifford Lectures, 1960*. George Allen and Unwin, London, 1969.

[32] W. V. Quine and J. S. Ullian. *The Web of Belief*. McGraw-Hill, New York, second edition, 1978.

[33] Bertrand Russell. *The Problems of Philosophy*, volume 18 of *Oxford Paperbacks University Series*. Oxford University Press, Oxford, paperback edition, 1967. Originally published in 1912.

[34] John R. Searle. *The Construction of Social Reality*. Penguin, London, paperback edition, 1996.

[35] Stephen Stich. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. MIT Press, Cambridge (MA) and London (England), 1990.

[36] K. T. Strongman. *The Psychology of Emotion*. John Wiley & Sons, Chichester, second edition, 1978.

[37] J. D. Watson and F. H. C. Crick. A structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 25 April 1953.

[38] James D. Watson. *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. Weidenfeld and Nicolson, London, 1968.

[39] Wolfgang Yourgrau and Allen D. Breck, editors. *Physics, Logic, and History: Based on the First International Colloquium held at the University of Denver, May 16–20, 1966*. Plenum Press, New York and London, 1970.