

Testimony

Antoni Diller

1 December 2011

1 AI's Holy Grail

Many people working in AI see the construction of an android with human-like abilities as the main goal of AI. For example, Charniak and McDermott write [4, 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.

Those working on the MIT Cog Project are more explicit [3]:

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.

In recent years considerable progress has been made towards achieving this goal. Here are some ongoing projects:

- robotic kitten (Starbrain Project, Starlab, Belgium)
- Lucy the robotic orangutan (Cyberlife Research, England)
- Cog the humanoid robot platform (Cog Project, AI Lab, MIT)
- humanoid robot (Kawato Dynamic Brain Project, ATR Labs, Kyoto, and the University of Southern California in LA)
- P3 the humanoid robot (Honda Motor Company and Carnegie Mellon University)
- Elvis the humanoid robot (Chalmers University, Gothenberg, Sweden)

More information about these projects can be found on the ArtificialBrains.com website where we also find the following statements:

Many artificial intelligence scientists now believe that true artificial intelligence will only emerge through the evolutionary development of autonomous mobile robots. And in order to create human-like intelligence, the robotic body too needs to be humanoid.

2 An Overlooked Human Ability

Recall Brooks's claim that AI's Holy Grail is to construct an android with human-like abilities. Much progress has been made to emulate human perception and locomotion, the human ability to grasp physical objects and similar abilities. However all the people and projects mentioned have overlooked one fundamental human ability and that is our ability to learn from others, not by mimicing their behaviour, but by accepting and believing what they say and what they have written.

3 Why is Testimony Neglected?

Why do many people, including philosophers, cognitive scientists and AI researchers, downplay the importance of tradition and testimony or simply ignore it? (Note that 'testimony' refers to much more than eyewitness testimony. It refers to any information received from any source in linguistic form. It can be about anything, including logic, maths, history, geography, science, philosophy, metaphysics, theology, etc.)

The study of testimony is marginalised because of the overwhelming acceptance of empiricism by scientists. Empiricism gained popularity through the success of modern science which was created during the Scientific Revolution that started around 1550. The Middle Ages were dominated by books, especially the Bible and the works of Aristotle. C.S.Lewis stresses 'the overwhelmingly bookish or clerkly character of medieval culture.' He continues (*The Discarded Image* (1964), p. 5):

When we speak of the Middle Ages as the ages of authority we are usually thinking about the authority of the Church. But they were the age not only of her authority, but of authorities. If their culture is regarded as a response to environment, then the elements in that environment to which it responded most rigorously were manuscripts. Every writer, if he possibly can, bases himself on an earlier writer, follows an *auctour*: preferably a Latin one. This is one of the things that differentiate the period almost equally from savagery and from our modern civilisation. In a savage community you absorb your culture, in part unconsciously, from participation in the immemorial pattern of behaviour, and in part by word of mouth, from the old men of the tribe. In our own society most knowledge depends, in the last resort, on observation. But the Middle Ages depended predominantly on books. Though literacy was of course far rarer then than now, reading was in one way a more important ingredient of the total culture.

In his play *The Life of Galileo* Bertolt Brecht has Galileo say (in David Edgar's translation):

The old is finished and the new is here. Where faith has always sat, there now sits doubt, and says: let's junk the textbooks, and see things for ourselves.

The motto of the Royal Society, formally constituted at Gresham College in 1660 and given a Royal Charter in 1662, is *nullius in verba* (take nobody’s word for it).

The rejection of testimony and tradition was not restricted to scientists. Alchemists also stressed the importance of experimentation. For example, the alchemist Petrus Severinus wrote in his book *Idea Medicinae Philosophicae* (Basle, 1571):

Sell your lands, your houses, your clothes and jewellery; burn up your books. Instead buy yourselves stout shoes, travel to the mountains; search the valleys, the deserts, the sea-shores and the deepest recesses of the earth; mark the distinctions between several kinds of animal, plant and mineral. Lastly, buy coal, build furnaces, work with fire. Thus and thus only will you attain to knowledge of things and their properties.

Empiricism became the dominant philosophy in Britain. The most influential British empiricists were John Locke (1632–1704), George Berkeley (1685–1753) and David Hume (1711–1776). Hume thought that knowledge only consisted of that which could be based on experience. He famously wrote in his *Enquiry Concerning the Human Understanding* (1748):

When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume; of divinity or school metaphysics, for instance; let us ask, *Does it contain any abstract reasoning concerning quantity or number?* No. *Does it contain any experimental reasoning concerning matter of fact and existence?* No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.

The marginalisation of testimony and tradition extends to the present day in science and it even distorts the research being done in AI and robotics. That empiricism is ripe in AI is illustrated by John Pollock who, in his book *Cognitive Carpentry*, p. 52, writes:

The starting point for belief formation is perception. Perception is a causal process that produces beliefs about an agent’s surroundings.

He is even more explicit in his article “Defeasible Reasoning” (1987), p. 486:

Perception represents the basic source of human knowledge. Nonintellectual mechanisms put us into various perceptual states, and being in those perceptual states constitutes a *prima facie* reason for conclusions about the world around us.

4 Empiricism

Bertrand Russell begins his paper “The Limits of Empiricism” (1936) with the definition of empiricism found in the *Encyclopaedia Britannica*, namely that empiricism is ‘the theory that all knowledge is derived from sense experience’. This definition is problematic for various reasons. Consider whose knowledge is being talked about. It

cannot be that all knowledge is derived from *my* sense experience. But to allow that all knowledge derives from *our* sense experience entails that empiricism self-destructs, because your observation statements are testimony to me and my observation statements are testimony to you. The following argument, for example, is not valid: ‘Albert says that he saw a blue car, therefore there was a blue car in Albert’s visual field’. Furthermore, people do reject other people’s testimony, even eyewitness testimony, sometimes correctly and sometimes not. If you go into any large bookshop, you will easily find many books on alien abduction. These contain eyewitness reports of people who have been abducted by aliens, yet most of us would consider these eyewitness reports to be false.

Another example of some people’s disinclination to believe eyewitness testimony—this time wrongly so—involves the pioneering deep-sea explorer William Beebe. In 1934 he made the deepest dive that had been made up to that time. His primitive bathysphere dived to a depth of half a mile. Beebe carefully described in his diary the strange creatures that he observed, but the life-forms that he wrote about were thought so outrageous by the scientific community that his observations were discounted. He gave many public lectures about what he had seen. Although many members of the general public were fascinated by his accounts of deep-sea creatures, the scientific community of his day was dismissive of his claims. Only in recent years, when more people have seen the same creatures that he saw, has his reputation been restored. (An account of Beebe’s dive can be found in his book *Half Mile Down* [2].)

5 Assertion Evaluation

5.1 Introduction

Hopefully, I have convinced you that you have acquired most of your knowledge by accepting other people’s assertions. Accepting the importance of testimony and tradition in acquiring information, however, creates various problems. Here is one such epistemological problem:

Every day you hear and read several hundred statements. How do you decide which of them to accept and which to reject?

The AI/Cognitive Science version of this problem is as follows:

An android would be in the same position. It would hear and read several hundred statements every day. How would you go about modelling the android’s ability to evaluate assertions.

In order to solve this problem we need to give some account of our assertion-evaluation mechanism (see Fig. 1). The assertions we encounter can be about anything. Here are some examples:

- The speed of light in a vacuum is 299,792,458 metres per second.
- General Sikorski’s death, on the 4th of July 1943, when his plane crashed into the sea off the coast of Gibraltar, was no accident.

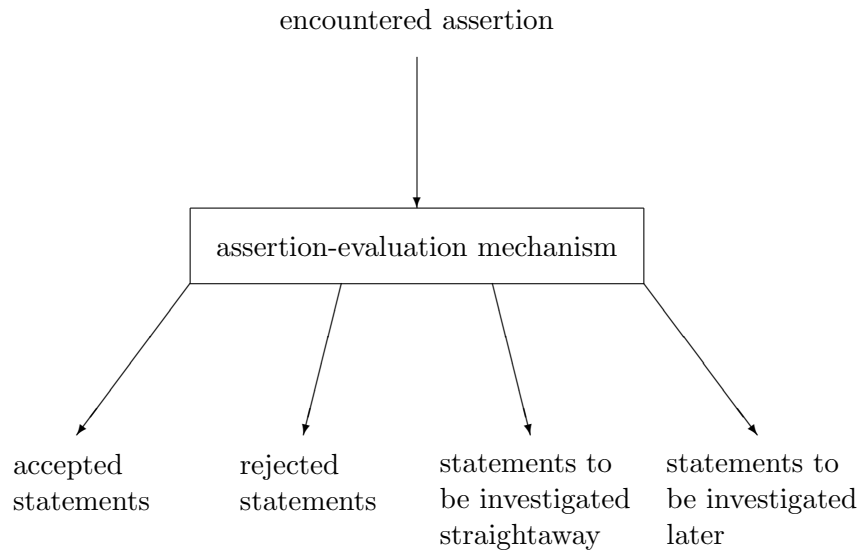


Figure 1: The assertion-evaluation mechanism.

- **Fermat’s Last Theorem:** There do not exist non-zero natural numbers x , y , z such that $x^n + y^n = z^n$, for $n > 2$.

Having read these statements do you now believe them? If so, why? If not, why not?

Before I continue, I just want to make clear that assertion evaluation is different from some of the things that you may already have been taught in AI. For example, assertion evaluation is different from each of the following:

- natural-language processing
- speech-act theory
- text summarisation
- belief-revision

Some work on assertion-evaluation was done in the early days of AI [1, 5], but this was not developed. Recently, the topic has been re-introduced into AI [7, 8, 9, 10, 21].

5.2 Some Related Non-AI Work

Some philosophers and historians have discussed testimony. I will briefly review this work here. Thomas Reid, for example, 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). This ‘is a disposition to confide in the veracity of others, and to believe what they tell us.’

Price wrote about testimony in lecture 5 of his book *Belief* [18]. Much of his discussion centres around the awkwardly phrased principle, ‘What there is said to be (or have been) there is (or was) more often than not’ [18, p. 116], but at one point he says [18, 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.’

More recently, Quine and Ullian [19, p. 14] have written about testimony and they think that our acceptance or rejection of testimony is governed by a number of meta-beliefs such as:

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

It should be noted, however, that people disagree about what constitutes a respected encyclopedia. I doubt whether Quine and Ullian would regard either *Harper’s Encyclopedia of Mystical and Paranormal Experience* or *The Encyclopaedia of Occult, Paranormal and Magick Practices* as being a respected encyclopedia! Quine and Ullian’s meta-beliefs relate to how we choose between two pieces of information that come from different sources, whereas Ried’s principle of credulity and Price’s principle of trust relate to a single piece of information. The historian Dibble [6] has formulated several meta-beliefs similar to Quine and Ullian’s that relate specifically to historical information. Here are some examples of these:

[Testimony] about specific details is likely to be more accurate than testimony about general conditions. [6, p. 204]

[Testimony] recorded shortly after an event took place is likely to be more accurate than testimony recorded long afterwards. [6, p. 204]

[Testimony] about ideologically relevant events which is addressed to people who share the witness’s beliefs and values is likely to be more accurate than testimony addressed to audiences which do not share the witness’s ideology. [6, pp. 204–205]

[D]isinterested testimony is likely to be more accurate than politically charged testimony [6, p. 205].

[T]estimony to words spoken in one’s native language is likely to be more accurate than testimony to words spoken in another language [6, p. 205].

[U]nprompted testimony is likely to be more accurate than prompted testimony [6, p. 205].

[T]estimony recorded before many versions of the event in question have been heard is likely to be more accurate than testimony recorded after the witness has heard a variety of versions. [6, p. 205]

Even considering a small number of people who have written about testimony we find a lot of disagreement about the way in which assertions are evaluated. The following are the things that could be used:

- principle (Reid)
- proposition (Price)
- maxim (Price)
- methodological rule (Price)
- meta-belief (Quine and Ullian, Dibble)

5.3 The Acquisition Rule

My suggestion is that, in evaluating testimony, we make use of the following defeasible rule, ‘Believe what you read and what you are told by others.’ I call this the *acquisition rule*. Defeasible rules are ones that hold unless overridden by some other principle. The idea of defeasibility arose in ethics and legal philosophy. An example of a defeasible rule in law is the principle not to kill another human being. This, for example, is overridden in a just war. Although it is not directly relevant, for information I list the main conditions that legal philosophers have thought a war must satisfy for it to be a just one. (These are taken from the article “War, just” in *The Oxford Companion to Philosophy*.)

- The war must be waged by a legitimate authority.
- The war can only be waged for a just cause.
- The war must be a last resort.
- There must be a formal declaration of war.
- There must be a reasonable hope of success.
- The means employed should be proportional to the end aimed at.
- The war must be a response to aggression.
- It is not permissible to kill non-combatants.

Another example of a defeasible rule is the principle not to harm another human being. This is, for example, overridden by a surgeon’s duty to save a patient’s life.

first	defeasible rule to believe others (perception, memory)
second	re-evaluations of beliefs and rejected assertions using various methodologies; critical assessment of new information using various methodologies

Table 1: The two-stage model.

5.4 The Two-stage Model

Our evaluation of information cannot be very sophisticated because of the quantity of information we are subjected to. Thus, the defeasible rule to believe others is not infallible. We reject true information and believe falsehoods. Therefore, the defeasible rule to believe others cannot be the whole story. I therefore have proposed a two-stage model of belief-acquisition (see Table. 1). Faced with an encountered assertion there are four things we can do with it:

- (i) We can decide to accept it straightaway and add it to our belief-system.
- (ii) We can choose to reject it straightaway, that is to say, not to believe it. It should be noted that this is not the same as accepting the negation of the assertion.
- (iii) We can decide to begin an investigation into the truth of the assertion straightaway. This happens when the content of the assertion is important to us, but we have some reservations about accepting or rejecting it straightaway.
- (iv) We decide to suspend judgement about the assertion until we have the time to investigate it fully. This happens when, as in (iii), the assertion's content is important to us, but we have some reservations about accepting or rejecting it straightaway and realise that such investigation may be time-consuming.

5.5 Overriding Factors

In this subsection I want to consider some of the factors that may cause us to override the defeasible rule to believe others. I begin by noting that there are many sources of information. The most important ones are the following:

- speaking
- books
- journal articles
- media
 - radio
 - newspapers
 - television

- Internet

Some overriding factors apply to all assertions equally, whereas others are specific to a particular source. Here are some examples for you to think about.

- Consider the statement, ‘David Hume lived for a while in Paris.’ This is from the article, “Hume, David”, by Professor Justin Broackes of Brown University in *The Oxford Companion to Philosophy* [13]. Knowing what you do, do you now believe this statement? If so, why? If not, why not?
- Now, consider this statement, ‘The Nasca lines on the plains of Peru are giant runways for space-craft.’ This is from Erich von Däniken’s book *Chariots of the Gods?* [22]. Knowing what you do, do you now believe this statement? If so, why? If not, why not?
- Do you believe the following story (from Holroyd’s book *Prelude to the Landing on Planet Earth*, p. 18)?

Late one night [in October 1975] I, my wife and a friend watched a television programme in the sitting-room of my home. The programme had a soporific effect on the other two, but I watched it through and at the end walked across the room and switched off the set. I then turned and spoke to the others and our friend woke up, and at that moment a potted plant on the mantelpiece behind me literally jumped out into the room and landed at my feet.

If you do believe it, think why? Similarly, if you do not believe it, think why you don’t. Perhaps you believe part of it, but not all of it.

- Do you believe the following story (from Andrija Puharich’s book *Uri* (1974), p. 193)?

I returned to New York on November 2, 1972. Uri [Geller] and Shipi stayed on in London and returned to Munich on November 4. As they were flying over the area of Schweinfurt in a Boeing 747, Uri saw that the Nikon F camera at his feet had levitated to his waist level . . .
 . Uri felt that this was a sign for him to photograph a UFO.

If you do believe it, think why? Similarly, if you do not believe it, think why you don’t. Perhaps you believe part of it, but not all of it.

- In his article “Defeasible Reasoning”, published in the prestigious journal *Cognitive Science*, vol. 11 (1987), pp. 481–518, John Pollock writes, ‘Perception and memory provide the starting point for reasoning.’ Knowing what you do, do you now believe this statement? If so, why? If not, why not?
- In his article “The Search for Meaning and its Biobehavioral Correlates”, published in the little-known journal *Ultimate Reality and Meaning*, vol. 7 (1984),

pp. 133–155, Frank Robert Vivelo writes, ‘We attempt to impose upon a seemingly chaotic universe an order we find satisfactory and, therefore, reassuring.’ Knowing what you do, do you now believe this statement? If so, why? If not, why not?

In their notorious paper “Peer-review Practices of Psychological Journals” [15], the psychologists Peters and Ceci resubmitted 12 articles to the journals that originally published them only altering the names of authors and institutions (for example, ‘Harvard’ became ‘Northern Plains Center for Human Potential’). The contents of the papers were not changed. One of these papers was accepted, 8 were rejected and 3 were detected as resubmissions. The commonest reason given for rejection was that the article in question contained ‘serious methodological flaws’. Many people were horrified by the experiment conducted by Peters and Ceci. Most academics like to think that articles are accepted or rejected because of their content and not because their authors belong to a prestigious university department. Most academics like to think that Lakatos got it right when he wrote [11, p. 427]:

[If] I want to judge the merits of Einstein’s 1905 paper, it does not matter at all that it was written by Einstein. It would have had exactly the same value if typed out by a monkey.

5.6 The Assessment Component

The way in which I unpack the acquisition rule is to represent it as an ordered set of rules all of which, except the last, are conditional ones. I call such a set of rules an *assessment component*. The first two rules are:

- (1) If I find assertion X intriguing or its content is important to me, but I do not feel that I can accept or reject it straightaway and the effort required to investigate it further is not great, then begin such an investigation straightaway.
- (2) If I find assertion X intriguing or its content is important to me, but I do not feel that I can accept or reject it straightaway and the effort required to investigate it further might be quite considerable, then postpone such an investigation until it is convenient to carry it out.

The last rule in the assessment component is the non-defeasible rule to believe the assertion in question. It should be noted that I am not assuming that an individual’s assessment component never changes. Here is an example of an assessment-component rule:

- (h) If assertion X is uttered by an actor during the performance of a stage play, then reject X .

5.7 Altering the Assessment Component

Consider the following assessment-component rule:

- (i) If assertion X is made by someone whose ideology is radically different from mine, then reject X .

If we had this rule, it would cause us to reject the following statements from the www.comunist-party.org.uk website belonging to the Communist Party:

The capitalist monopolies and their political representatives put profit before people and before the earth's environment. Capitalist exploitation and imperialism intensify inequalities of race and gender. The need for popular resistance and class struggle, for the working class to take state power in fact, is as great as ever.

However, it would also cause us to reject the information, also contained on the Communist Party website, that the Communist Party HQ is located at Ruskin House, 23 Coombe Road, Croydon, London, CR0 1BD, and that its phone number is: 020 8686 1659. We would be more sensible to change our rule to the following:

- (i') If assertion X is made by someone whose ideology is radically different from mine, then reject X , unless the content of X is unlikely to be contaminated by that different ideology.

The two main principles we use to alter our assessment component are these:

- (A) Minimise the number of future true assertions rejected.
- (B) Maximise the number of future false assertions rejected.

Using principle (A) we replace a rule, 'If $P(X)$, reject X ', with a rule, 'If $P(X)$ and $Q(X)$, reject X '. Using principle (B) we replace a rule, 'If $P(X)$, reject X ', with a rule, 'If $P(X)$ or $Q(X)$, reject X '.

An alternative approach to that presented above is to have the assessment component made up out of numerical rules like this, 'If $P(X)$, then the acceptability of X decreases by some number r (between 0 and 1)'. We begin by giving other people's assertions an initial acceptability value of 1 and if this falls below some number q (between 0 and 1), the assertion is rejected.

References

- [1] Robert P. Abelson and J. Douglas Carroll. Computer simulation of individual belief systems. *American Behavioral Scientist*, 8(9):24–30, May 1965.
- [2] William Beebe. *Half Mile Down*. Harcourt Brace, New York, 1934.
- [3] 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>.

- [4] 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.
- [5] Kenneth Mark Colby and David Canfield Smith. Dialogues between humans and an artificial belief system. In Donald E. Walker and Lewis M. Norton, editors, *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI69)*, pages 319–324, Boston, 1969. Mitre Corporation.
- [6] Vernon K. Dibble. Four types of inference from documents to events. *History and Theory*, 3:203–221, 1963. The year of this paper could be 1964 or 1963–1964.
- [7] Antoni Diller. Evaluating information found in journal articles. In Ángel Nepomuceno, José F. Quesada, and Francisco J. Salguero, editors, *Logic, Language and Information: Proceedings of the First Workshop on Logic and Language: Instituto de Lógica, Lenguaje e Información, Universidad de Sevilla, Sevilla, 29, 30 de noviembre y 1 de diciembre de 2000*, pages 71–78. Kronos, Sevilla, 2000.
- [8] Antoni Diller. Everyday belief-acquisition. In Gabriela P. Henning, editor, *Argentine Symposium on Artificial Intelligence (ASAI2000) Proceedings: Tandil, September 5–7, 2000*, pages 221–232. Sociedad Argentina de Informática e Investigación Operativa (SADIO), Buenos Aires, 2000. This paper is also available on the Internet at the following URL: <http://www.cs.bham.ac.uk/~ard/papers/asai2000.html>.
- [9] Antoni Diller. Acquiring information from books. In Max Bramer, Alun Preece, and Frans Coenen, editors, *Research and Development in Intelligent Systems XVII: Proceedings of ES2000, the Twentieth SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence, Cambridge, December 2000*, pages 337–348. Springer, London, 2001. This paper is also available on the Internet at the following URL: <http://www.cs.bham.ac.uk/~ard/papers/es2000.html>.
- [10] Antoni Diller. Assessing information heard on the radio. In Mieczysław A. Kłopotek, Sławomir T. Wierzchoń, and Krzysztof Trojanowski, editors, *Intelligent Information Processing and Web Mining: Proceedings of the International IIS:IIPWM'05 Conference held in Gdańsk, Poland, June 13–16, 2005*, Advances in Soft Computing, pages 426–430. Springer, Berlin, 2005.
- [11] Yehuda Elkana, editor. *The Interaction Between Science and Philosophy*. The Van Leer Jerusalem Foundation Series. Humanities Press, Atlantic Highlands (New Jersey), 1974.
- [12] Rosemary Ellen Guiley. *Harper's Encyclopedia of Mystical and Paranormal Experience*. HarperCollins, New York, 1991.
- [13] Ted Honderich, editor. *The Oxford Companion to Philosophy*. Oxford University Press, Oxford, 1995.

- [14] Brian Lane. *The Encyclopaedia of Occult, Paranormal and Magick Practices*. Warner Books, London, 1996.
- [15] Douglas P. Peters and Stephen J. Ceci. Peer-review practices of psychological journals: The fate of published articles, submitted again. *The Behavioral and Brain Sciences*, 5:187–195, 1982.
- [16] John L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.
- [17] John L. Pollock. *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT Press, Cambridge (MA) and London (England), 1995.
- [18] Henry H. Price. *Belief: The Gifford Lectures, 1960*. George Allen and Unwin, London, 1969.
- [19] W. V. Quine and J. S. Ullian. *The Web of Belief*. McGraw-Hill, New York, second edition, 1978.
- [20] Bertrand Russell. The limits of empiricism. *Proceedings of the Aristotelian Society*, XXXVI:131–150, 1936.
- [21] Paul Thagard. Testimony, credibility, and explanatory coherence. *Erkenntnis*, 63(3):295–316, December 2005.
- [22] Erich von Däniken. *Chariots of the Gods? Unsolved Mysteries of the Past*. Souvenir Press, London, 1969.