

Tilburg University

Reliability of information on the Internet

Vedder, A.H.; Wachbroit, R.S.

Published in:
Ethics and Information Technology

Publication date:
2004

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Vedder, A. H., & Wachbroit, R. S. (2004). Reliability of information on the Internet: Some distinctions. *Ethics and Information Technology*, 5, 211-215.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Reliability of Information on the Internet: Some Distinctions*

Anton Vedder^α and Robert Wachbroit^β

In: *Ethics and Information Technology* 2003; 5: p. 211-215 [Please note that this article was actually published in 2004, with (c) 2004.]

Abstract

In this contribution, we identify and clarify some distinctions we believe are useful in establishing the reliability of information on the Internet. We begin by examining some of the salient features of information that go into the determination of reliability. In so doing, we argue that we need to distinguish content and pedigree criteria of reliability and that we need to separate issues of reliability of information from the issues of the accessibility and the usability of information. We then turn to an analysis of some common failures to recognize reliability or unreliability.

Epistemology, ethics, information, Internet, reliability, truth

1. INTRODUCTION

The problem of determining the reliability of information on the Internet has become increasingly important as more and more people get their information from the Internet. The general issue of the meaning and determination of reliability has of

* This article is an elaboration of a paper presented by the first author under the same title on the Computer Ethics: Philosophical Enquiry Conference on Computer Ethics in the Post-September 11 World, held July, 25th-27th in Boston, and published in: Grodzinsky, Frances S., Richard A. Spinello, Herman T. Tavani (eds.) *Computer Ethics in the Post-September 11 World: Proceedings for Computer Ethics Philosophical Enquiry*. Boston: Boston College, 2003, 144-157. The authors wish to express their gratitude to the organizers and audience of this conference for their helpful comments and suggestions, especially to James Moor. They are equally grateful to those who discussed previous versions of this paper on the occasion of workshops organized at Tilburg University and the University of Twente (Netherlands). The paper is one of the results of a collaborative research project on ethical and legal aspects of the quality of medical information on the Internet of Tilburg University, Netherlands, and the University of Maryland, USA, funded by the Netherlands Organization for Scientific Research.

^α Anton Vedder is an Associate Professor of Ethics and Law at the Faculty of Law of Tilburg University, the Netherlands.

^β Robert Wachbroit is a Research Scholar at the Institute for Philosophy and Public Policy at the University of Maryland and an Adjunct Associate Professor of OB/GYN in the University of Maryland School of Medicine.

course a long and ancient history. Not only is there an extensive philosophical literature analyzing the meaning of “reliability,” but also the history of scientific method is a history of efforts to codify rules for determining reliability. Our aims regarding information on the Internet, however, are much more modest. We would like to suggest some distinctions and a general framework which, we believe, may help identify how the Internet as a source of information can be improved. Whether this framework can or should be applied to situations other than information on the Internet is an issue we do not address here.

To begin with, we should be clear that by the phrase “reliable information” we mean justified information, information that we would be justified in believing, information that we can trust. It does not necessarily mean, at least directly, information that is true. For example, people in the mid-Nineteenth Century were justified in believing the principles of Newtonian mechanics because the best available evidence and the best available methods at the time largely supported these principles. Even though we now know that these principles are false (and so we are not justified in believing them), that does not change what people 150 years ago were justified in believing. Indeed, the history of science can be seen as a history of efforts to show that claims – though responsibly deemed trustworthy and reliable at the time – were later shown to be false. This should not be surprising since, as our knowledge grows, our evidence becomes more extensive and our investigations become more refined. Consequently, discovering that a belief is false does not necessarily mean that, at an earlier time, people were not justified in believing it or that it was wrong to trust it. What is reliable, trustworthy, justified is a matter of what we already know.

The point of emphasizing that reliability is about justification rather than about truth is that it focuses our problem. Determining whether some information is reliable is a matter of determining whether it is properly justified. It is also a matter of characterizing epistemic responsibility – e.g., when is it irresponsible to rely upon some information obtained from the Internet

In what follows we will try to identify and clarify some distinctions we believe are useful in identifying or establishing the reliability of information on the Internet.

We begin by examining some of the salient features of information that go into the determination of reliability. In so doing, we argue that we need to separate issues of reliability of information from the issues of the accessibility or the usability of information. We then turn to an analysis of some common failures to recognize reliability or unreliability, particularly when search engines are used. This analysis will suggest some directions for improving the current situation and so directions for making the Internet more trustworthy as a source of information.

2. CONTENT AND PEDIGREE

Determining the reliability of some information can be a complicated matter, and it can sometimes involve energetic investigation, sophisticated instrumentation, and extensive background knowledge. Nevertheless, in these cases as well as in the more mundane determinations of reliability, the various considerations that are called upon can be grouped into two broad categories – content criteria and pedigree criteria.

By “content criteria” we mean the conditions or criteria of reliability that are a function of the content of the information itself. Perhaps the most significant criteria of this group are the criteria of evidence: whether some information is reliable will depend upon the available evidence and how that evidence supports the information. The specification of appropriate evidence criteria is the work of several disciplines, including research methodology, probability and statistics, experimental design, etc. Evidence criteria thus range from common-sense judgments of what counts as evidence to sophisticated techniques of confirmation. But despite their obvious importance to the determination of reliability, typically only people familiar with the subject matter of the information can make use of evidence criteria, since often only they will know what evidence is available or be in a position to judge how good that evidence is.

Other examples of content criteria are logical criteria and, arguably, subject-matter criteria. The specification of logical criteria is straightforward: if some information is inconsistent or entails an absurdity – e.g., that the world began ten

minutes ago – then the information cannot or should not be believed. The specification of subject-matter criteria is more controversial but nevertheless also plays a role. For example, information about the bathing habits of Martians would be dismissed out of hand by many as simply unreliable because of the subject-matter. No one could plausibly claim now to know such things.

In general, most people cannot base their assessments of reliability on content criteria. The logical and subject-matter criteria are weak, serving more to rule that some information is unreliable rather than establish reliability. And the evidence criteria, as we noted, might not be available to many insofar as the subject matter of the information is specialized and unfamiliar. (For some epistemologists, particularly those taken with the force of philosophical skepticism, content criteria are the only criteria for reliability. Plainly, we do not accept this conclusion since philosophical skepticism is not the problem here.)

Many of us determine reliability by “pedigree criteria” – the conditions or criteria of reliability that are a function of the source of the information. If the information is provided by an authoritative source or a source that has been reliable in the past, then the information itself is considered reliable. For example, we usually regard medical information provided by a physician as reliable precisely because of the physician’s expertise – i.e., his or her acknowledged authority in providing such information.

It is important to note that the use of pedigree criteria is pervasive, even among people who are familiar with the subject matter of the information and so are capable of using content criteria such as evidence criteria. Scientists, scholars, and professionals typically get information about their specialty from peer-reviewed journals; the information is deemed reliable because it appears in an authoritative journal. Even researchers employing evidence criteria will often have to use pedigree criteria as well. For example, in performing an experiment, the scientist may have to rely upon the information provided by an instrument whose reliability is based on the authority of the manufacturer of the instrument. The scientist might rely upon the

temperature readings provided by his instrument because he or she knows that the manufacturer of that instrument provides reliable, laboratory-grade instruments.

For the most part, pedigree criteria are established by credibility-conferring institutions. Perhaps foremost among these are the academic institutions. Universities, medical schools, law schools, etc., not only teach but also certify that people who successfully emerged from their programs can speak with authority on specific subject matters. They are identified as experts and are provided with credentials to that fact. Peer-reviewed scholarly journals present another example. Their authority is grounded in an institution – typically an academic institution or a professional society – that can be counted upon to implement correctly peer review prior to publication. To a much lesser extent, the newspaper institutions are yet another example, but only insofar as their reputation is based on balanced reporting. We can expect that such newspapers will make appropriate efforts to ensure the reliability of their information on pain of losing their reputation and so their readership. People rely on the information provided and read these newspapers accordingly. But this is of course not true of all newspapers. Many are known for their ideological biases, are deemed less reliable, and so are read accordingly.

Nearly all credibility-conferring institutions have an online presence in the form of at least a web site, indicating that pedigree criteria may also be found on the Internet. This in turn might suggest that the problem of determining reliability is no greater online than offline. Indeed, one might argue that simply because some information is printed on newsprint does not mean, nor is taken to mean by most people, that the information provided must be as reliable as information from *The New York Times*. Similarly, simply because some information is located on the Internet does not mean that we cannot distinguish its reliability from information provided on *The New York Times* web site. As we shall see later, some of the problems in recognizing pedigree criteria on the Internet are distinctive and so deserve their own analysis. But before turning to that, we should briefly examine and separate certain other features of information – i.e., accessibility and usability – which are sometimes linked with reliability.

3. ACCESSIBILITY AND USABILITY

As we noted earlier, reliability is a property or characteristic of information. On the other hand, the accessibility or the usability of information is a relational property that depends upon the person who is trying to access or use the information: some information might be accessible or usable to one person but inaccessible to another. Logically, this entails that reliability is distinct from these other features since the reliability of information is not relational in this way: if some information is reliable for one person at a particular time, it is reliable for any person at that time.

Nevertheless, it is worth noting the impact of these features on the issue of reliability. The Internet has had an enormous impact on the accessibility of information. Many people are able to disperse information on the Internet and even many more are able to find information on the Internet. People need not go to libraries anymore; they need not order books and journals and try to get them home in a cumbersome way. Complete libraries, books and journals are often available by clicking a button. This increase in accessibility in part motivates concerns about reliability. The relatively easy access to so much information appears to challenge or overwhelm our pre-Internet habits of determining reliability. These habits will be examined in the next section.

The usability of information – whether a particular person can use the information – is clearly a matter of that person's background and resources. An individual cannot make much use of medical information for diagnosing some symptoms if he cannot understand all of the terms words in the information or if he lacks some needed diagnostic instrument. To some extent, accessibility and usability can shade into each other. Plainly one cannot use information one does not have access to; and one might question whether one has access to some information if that information is not comprehensible.

Increasingly, however, accessibility and usability are being distinguished, at least as far as the Internet is concerned. Accessibility has become understood more in

terms of connectivity: information is accessible insofar as one can electronically connect, e.g., to a web site containing that information. Usability has become understood more in terms of web design. How should a web site be constructed to make it more “user-friendly”? This can involve such considerations as the placement of information, the design of menus, and the use of graphics. Like accessibility, usability is important because there is little point to determining the reliability of unusable information.

4. RECOGNIZING RELIABILITY

As we noted earlier, determining the reliability of information consists to a large extent in recognizing that the appropriate pedigree criteria are satisfied. Failures to recognize the reliability of reliable information can take many forms: (a) The appropriate pedigree criteria might be missing or inadequate. For example, the information might be presented without documentation or complete references. (b) Even though adequate pedigree criteria are present, the underlying credibility-conferring institution might not be recognized as such by a particular user. For example, someone might not know that *The Lancet* is one of the top medical journals in England, and so not recognize the pedigree that information from that source would have. (c) Pedigree criteria can sometimes appear to conflict with each other or with content criteria. Perhaps the most common example of this is when we come across information from a credible source that is shocking, surprising, or conflicts with other beliefs.

Although this list is not intended to be a complete list of the types of failures to recognize reliability, it does indicate that the fault can sometime lie with the provider of information and sometimes with the receiver or user of information. Responding to concerns about the information provider might seem straightforward. We can set standards or guidelines – e.g., citation formats and source referencing that adequately display the appropriate pedigrees. Responding to concerns about the receiver of information is a different matter. The ability to recognize reliable information depends

upon the individual's background knowledge and habits of rationality regarding the assessment of criteria of reliability. Information providers have to meet certain standards if they want their information to be recognized as reliable, but this obligation can only go so far. Users of information must also exercise an epistemic responsibility, which doesn't cease when they go online.. The Internet is not fool-proof.

Theoretically, the problems regarding determining the reliability of information seem no greater online than offline. The problems of inadequate pedigree criteria, of failing to recognize legitimate credibility-conferring institutions, of conflicting criteria exist whether or not we are dealing with the Internet as an information source. Nevertheless, as a practical matter, recognizing reliable information on the Internet might be more challenging. As one of us has pointed out extensively elsewhere,¹ the very possibility of adequately recognizing pedigree criteria is often lacking where the Internet is concerned. Often, a content provider is anonymous or merely has a virtual identity. Generally, the influence of individuals in providing information on the Internet is diminishing, whereas the influence of intelligent systems is increasing. Also, the lack of traditional intermediaries (such as libraries, librarians, specialized publishers) may have a negative influence on the capabilities of information seekers to assess the reliability of information. These kinds of factors, i.e. the lack of information about content providers, the diminishing human influence in the provision of information, and the lack of traditional intermediaries, are responsible for the fact that an information seeker often lacks clues or any indication whatsoever about the character, background, and institutional setting of the content provider. Further complicating the problem is the phenomenon of globalization, which is inherent to the Internet. Even when the recipient has some information about the content provider, the individual might be unable to estimate the credibility of that provider, simply because the individual will often not be acquainted with the relevant backgrounds and institutional settings from completely different cultures. The

¹ Vedder (2001 and 2002). In these and other publications, Vedder used to refer to content criteria as "primary" criteria while he referred to pedigree criteria as "secondary" criteria. Because of the – unintended – suggestion of an hierarchical order associated

recognition procedures and traditions that make up the institutional basis of the application of pedigree criteria may be different in different cultures. A recipient from one culture may not recognize the procedures and traditions of the culture's provider.

One might respond that many of these problems are merely transitional – as the Internet becomes more established, they will become less of a concern – or they are no worse than the problems one confronts offline. As we noted earlier, more and more credibility-conferring institutions – universities, scholarly journals, professional associations, etc. – are establishing an online presence. Insofar as information providers are anonymous or have virtual identities, they cannot expect their information to be recognized as any more reliable than hearsay, rumor, or gossip. Furthermore, the phenomenon of globalization, while increasing the variety of easily accessible information sources, arguably does not significantly alter the user's epistemic responsibility. Coming across an unfamiliar credibility-conferring institution is much like coming across an unfamiliar newspaper or a book from an unknown author and publisher. It is commonsense to be careful and cautious with regard to information from unfamiliar sources.

Before leaving this section, a few comments should be made regarding search engines because most people who get information from the Internet do so via search engines. People typically get information from the Internet by seeking it out: a word, phrase, or question is put to a search engine which will then normally identify relevant web sites. The question we need to address here is whether search engines confer credibility. Is being a result of a search engine procedure an appropriate pedigree insofar as reliability is concerned?

It should be clear that search engines do not by themselves confer credibility. Being ranked highly by a search engine on a particular topic may indicate relevance; it does not necessarily indicate reliability. It is difficult to argue this point in detail since the ranking algorithms of most search engines are a closely held secret and it is often unclear how relevance is exactly defined. Nevertheless, the aim of these algorithms is to capture relevance by measures such as how often the phrase occurs on the web site

with it, the latter formulation was discarded.

or how popular is the web site for that topic (how many other sites have links to that web site). Relevance is no more a mark of reliability than being a frequently borrowed book in a library is a mark of its reliability.

One could argue that in some cases, relevance is tied to reliability insofar as other users of the Internet have been epistemically responsible. Suppose most people get their information regarding a particular topic from a web site whose reliability is recognized from standard content and pedigree criteria. Depending upon the search engine, this web site might well rank highly on searches on that topic precisely because of its popularity. We should therefore not be surprised if reliable web sites – or more often, official web sites – are frequently ranked high in relevance. Nevertheless, this occasional association between relevance and reliability is contingent. Reliability should not be based on search engines as such.

We should note that search engines can be useful in assessing credibility. For example, a search that is restricted to servers of credibility-conferring institutions will presumably retrieve only reliable information. Nevertheless, it is the restriction to these institutions that apparently ensures credibility. The search engine itself does not add to the credibility assessment.

5. CONCLUDING REMARKS

How can we improve (the recognition of) the reliability of information on the Internet? The short answer is to foster the online presence of established credibility-conferring institutions. Gagliardi and Jadad (2002) and Kunst (2002) have shown that many efforts to ensure reliability exclusively through a system of labels or certifications or through mentions of sources, last update, contact address of the content provider, etcetera, tend to get bogged down. In any case, such systems of certifications are either unnecessary or inadequate. If enough of the normal content or pedigree criteria are present, then the further “pedigree” of a “certificate of reliability” adds nothing. If not enough content or pedigree criteria are present, no mere certificate of reliability will

fill the gap. In other words, determining the reliability of information on the Internet is not fundamentally different from determining the reliability of information offline. We would therefore do well to bring as many of the established offline resources to bear as is appropriate.

In a similar vein, the standards that people use to judge the reliability of information should be the same regardless of whether the information provider is online or offline. This means that people should exercise the same care and caution on the Internet as they would for any information if it came from an unknown source. Being read on the Internet is by itself no more an indication of reliability than, e.g., being heard on the telephone is by itself an indication of reliability.

Nevertheless, we should also acknowledge that people might be more likely to misjudge reliability when the information is on the Internet. The evident sophistication of computers, network connections, etc., can lead some people to believe that not just any nonsense can appear on the Internet. Presumably, as people become more familiar with the Internet – and so less in awe of it – this sort of credulity will disappear. Credulity regarding search engines is a different matter. Search engines are certainly marvelous research tools, whose use will likely increase as people become more familiar and comfortable with the Internet. But, as we noted earlier, people should not draw any conclusions about reliability from generated relevance rankings, unless they can assume that the range of servers searched by the search engine has been restricted to the servers of trusted institutions.

References

Gagliardi, A., A. Jadad (2002) Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *British Medical Journal* 324, 569-573

Kunst, H. (2002) Accuracy of information on apparently credible websites: survey of five common health topics". *British Medical Journal* 2002; 324 581-582.

Vedder, A. (2001) Misinformation through the Internet: Epistemology and Ethics. In: *Ethics and the Internet* (ed. Anton Vedder), Intersentia , Antwerpen, Groningen, Oxford, 2001, p. 125-132.

Vedder, A. (2002) What people think about the reliability of medical information on the Internet, in: *The transformation of organisations in the information age: Social and ethical implications* (eds. I. Alvarez, T. Ward Bynum, J. Alvaro de Assis Lopes, S. Rogerson), Universidade Lusitana, Lisbon, 281-292.