

How Not to Be Fooled by Cardboard Rabbits

Heuristics for Epistemic Reliability Judgments in the World Wide Web

Dario Taraborelli

Department of Psychology
University College London
Gower Street
London WC1E 6BT
United Kingdom
d.taraborelli@ucl.ac.uk

Abstract. In the past decade a large number of studies have addressed the issue of *what makes information on the Web credible*. The problem of understanding how we select reliable sources of information from the *World Wide Web* and how we estimate their credibility is not only of interest to marketers, information architects, policy makers and computer scientists: it is a key issue for scientists interested in studying human cognition and decision-making strategies in particular.

In this paper I will argue that the study of information search skills can provide an extraordinary opportunity for cognitive scientists to investigate processes mediating *knowledge acquisition by epistemic deference*. I will shortly review some of the major methodological proposals that have been developed so far study how users judge source reliability in the *World Wide Web* and I will propose an alternative approach inspired by the idea that—as cognitively evolved organisms—human beings are designed to select decision strategies that are as effortless as possible. Given the impact of *Web* technologies on our cognitive skills and on our perception of the *Web* as a medium, I will argue that users are likely to develop simple heuristics to address in a computationally tractable way the problem of how to select reliable sources of information.

I will discuss a number of general arguments supporting the hypothesis that epistemic reliability judgments in the *World Wide Web* are grounded in heuristics and propose a methodological framework and research directions to empirically test this hypothesis.

Keywords. information foraging; predictive judgments; evaluative judgments; heuristics; epistemic deference; information scent; information foraging; ecological rationality; modularity;

1 Judgments of Epistemic Reliability

Possessing reliable knowledge and being able to identify reliable sources of information are skills essential not only for our survival but also for our ability to

deal efficiently with the problems raised by our interactions with the physical and social environments. As evolved cognitive organisms, we negotiate demanding cognitive problems by selecting parsimonious strategies that provide us with sufficiently accurate solutions. *Epistemic deference* is one of such strategies.

1.1 Epistemic Reliability and Deference

Social epistemology has introduced the concept of “epistemic deference” to refer to those processes of knowledge acquisition in which a subject (the *deferrer*) relies on an external source (the *deferee*) in order to extend her knowledge to facts with which she has no direct acquaintance. Relying on experts in order to make a decision is a typical example of process in virtue of which we adopt an epistemic deference stance towards an other individual to extend our individual knowledge beyond its natural boundaries. Epistemic deference is a constitutive trait of language competence, as the capacity by which we can “entertain thoughts through the language that would not otherwise be accessible to us” [1] as when – for example – we use the term “arthritis” in a conversation without exactly knowing the precise reference of this term. Deference to an external source of knowledge, insofar as we trust that source for epistemic matters, allows us to extend our beliefs to facts which we do not thoroughly understand. As such, deference is a principle found everywhere in human cognition and one of the basic strategies used to bootstrap knowledge and language acquisition in young children.

As information consumers, we *defer* to external sources of knowledge in situations in which either (i) we lack reliable knowledge on a given subject matter that is required to make a decision (in which case deference is a *necessary* condition) or (ii) in cases in which deference provides a convenient, *sufficient* solution to meet the requirements of a given decision problem. The massive availability of information in the *World Wide Web* is making deferential practices of both kinds a constitutive part of our belief-formation and decision-making strategies.

The crucial question is then how to identify *trustworthy deferees* whenever we engage in information search in the *Web* in order to acquire knowledge. It is often the case that we rely on background knowledge and previous experience in order to decide whether a source is trustworthy or not: *experienced credibility* [2] is one of the most common grounds of deferential behavior. In the general case, though, we have no prior information on the reliability of external sources and we need to estimate it.

1.2 Evaluative Judgments of Epistemic Reliability

Studies of credibility and perception of epistemic reliability in the *Web* have mostly focused on *evaluative judgments*, i.e. judgments subjects make to express the trustworthiness of a source on the basis of extensive inspection of the content provided by the source, and not much on *predictive judgments* or judgments about the expected reliability of a source prior to its extensive inspection [see 3, 4].

Arguably, the main reason why the study of evaluative judgments has been privileged in the literature is that *in ideal conditions*, whenever users are required to estimate the credibility of a source, they are not subject to particular constraints of time or cognitive effort to produce this judgment. I will come back later on the issue of cognitive constraints on reliability judgments and concentrate in this section on evaluative judgments. The study of evaluative judgments of source credibility has proved central to address two issues:

1. how easily information acquired by deferring to external sources can be integrated into one's system of beliefs;
2. how prone such information is to revision.

Mainstream theories of persuasion [5, 6, 7] suggest that among the factors affecting the likelihood of subsequent amendment or revision of an evaluative judgment, the *amount* of processed information and the *degree of involvement* play a crucial role. Judgments based on small amount of information or in conditions of low involvement are more likely to be subsequently revised [8, 9]. It might seem natural to assume, then, that in the case of deference-mediated knowledge acquisition in the *Web*, humans tend to invest a large amount of information processing effort in identifying trustable sources of information.

Web Credibility studies [2, 10] have collected large datasets on evaluative judgments of source reliability by asking users to verbally report on the quality of visited websites. The results [11] have allowed to identify factors that affect, in turn;

1. the likelihood of specific source features to be noticed by the user;
2. the attribution of positive or negative values to those features that are noticed.

Verbal reports and *questionnaire*-based methodologies have been the most popular approach to studying judgments of credibility and source reliability at least since two decades [12] and are still the preferential approach adopted to investigate credibility in the *World Wide Web*. It should be noted that this approach is not limited to judgments of *Web* source credibility decontextualized from the specific task in which they are usually produced: even when researchers have looked at judgments of epistemic reliability in real human-computer interaction tasks, they have tended to privilege verbal reports (or “think-aloud” protocols) as the main source of empirical evidence over other possible kinds of data [4].

1.3 Beyond Verbal Reports to Study Epistemic Reliability

The use of verbal reports to understand *Web* credibility relies on the assumption that introspection is the best way to determine factors affecting judgments and decisions on the epistemic reliability of a source. As, for example, [4] observes,

[t]he method used in this study is premised on the assumption that the users can identify and discuss the characteristics and features of information objects that influence their judgments of information quality and cognitive authority. (p.150)

However useful verbal reports may prove to study evaluative judgments in de-contextualized conditions, they face a number of major limitations:

- verbal reports assume that subjects are *conscious* of the factors affecting the selection of a specific source as credible or epistemically reliable;
- reports relying on extensive inspection of a *Web* source can hardly account for the kind of processes in which users engage when they are involved in real information search tasks, which are usually constrained by available time, by the specific goal of the task and by further conditions on the cognitive effort a user is willing to invest in the task;
- qualitative studies based on verbal reports may take for granted that *a posteriori* evaluative judgments are immune to preference reversal that might occur when the evaluation is embedded in a specific task (which biases a user’s perceived utility of the different items);
- building a taxonomy of factors affecting credibility judgments on the basis of qualitative verbal reports can bias the results in favor of an arbitrary class of variables chosen by the experimenter.

These limitations intrinsic to verbal report studies make it urgent to develop a complementary methodology to investigate user information search behavior in ecologically valid conditions and under the typical constraints of real information acquisition tasks.

1.4 Predictive Judgments of Reliability and Information Scent

Information foraging studies [13, 14, 15] have started drawing the attention of the research community on factors that affect *predictive judgments* of the profitability of a source of information as opposed to *a posteriori* evaluations.

Why are predictive judgments underrepresented in current research on *Web* credibility? One of the possible reasons is that – since attitudes formed by less cognitively demanding means (those processes that tend to be classified as “peripheral routes to persuasion” [6]) are usually taken to be less predictive of behavior – researchers may have assumed that these are less representative of processes underlying deferential practices than attitudes based on careful, more cognitively demanding evaluative judgments.

The question boils down to understanding what is the average level of involvement typical of information search behavior on the web. As Fogg observes:

Web users typically spend small amounts of time at any given site or individual page, and are thus likely to develop strategies for assessing credibility quickly. One could argue that people typically process Web information in superficial ways, that using peripheral cues is the rule of Web use, not the exception. From a user perspective, there are too many competitors on the Web for deep credibility evaluation. Even the words people use to describe Web use – “visiting sites” and “surfing the Web” – suggest lightweight engagement, not deep content processing. Research has yet to examine the relationship between engagement level and credibility assessments online- [10, p.15]

The question, we argue, is not only a matter of user engagement, but also a matter of constraints that apply to epistemic reliability judgments in the context of information acquisition tasks. Fogg’s quotation evokes a number of such constraints.

First of all, *time constraints*. In a world in which online content is becoming massively and constantly available, user interactions with the *Web* naturally tend to become shorter, more frequent and increasingly mediated by search engines. *Information snacking*[16] can be seen as the application to the *Web* of a known principle of situated cognition that states that organisms tend to externalize the solution of demanding cognitive problems to the environment and use the environment as an external cognitive device to reduce cognitive workload.[17]

The second major class of constraints comes from *epistemic pollution* [18]. The larger the volume of available relevant information, the more urgent is the need of effective source selection skills. In conditions in which the number of potential competitors is overwhelming, *a posteriori* evaluative judgments simply become untractable and accurate predictive judgments are the only viable solution.

The sum of time constraints and task constraints as those raised by epistemic pollution is the main rationale in favor of the hypothesis according to which the problem of producing judgments of source reliability in the *World Wide Web* is likely to reduce to the problem of selecting appropriate heuristics, i.e. sufficiently reliable predictive strategies based on surface evaluation rather than time consuming and cognitively demanding *a posteriori* evaluation processes.

It should be noted that the problem of epistemic pollution is not solved by postulating that reliability judgments are taken in charge by heuristic processes. Fighting epistemic pollution is mainly a matter of detecting cheaters (e.g. sources of unreliable information) on the basis of surface cues, i.e. developing strategies to identify diagnostic cues for unreliable sources. As Nielsen observes, “in information foraging terms, information pollution is like packing the forest with cardboard rabbits” [19]: in terms of heuristics, the question becomes then how to select good strategies to predict a real rabbit from a fake.

2 Heuristics for Epistemic Reliability

I have introduced in the previous section the main rationale for assuming that credibility judgments in the context of real information search tasks in the *World Wide Web* are likely to be delegated to heuristics. In this section I will focus on some broader theoretical implications of this hypothesis.

2.1 Proximity

One of the main limitations of verbal-report studies is the fact that they have neglected the role of predictive judgments of reliability based on proximal information representing distal sources. The *World Wide Web* is rich of proximal cues that refer (in a more or less reliable way) to distal sources. These cues have been

referred to in the information foraging literature as constituting *information scent*, defined as a proximal indicator of the profitability of a distal source [14]. The hypothesis suggested by information foraging studies is that information seekers base the choice of optimal navigation patterns on the perceived strength of *information scent*.

Being able to use proximal information to reliably infer properties of distal sources is one of the most important goals of evolved perceptual systems. In sufficiently constrained environments, multiple classes of proximal cues can be taken as reliable predictors of distal properties. Consider for example the amount of proximal information about *spatial depth* that is available to an active observer [20]. Much as proximal information can be relied upon to accurately estimate depth in our physical environment, so it is plausible to assume that the problem facing *Web* users seeking reliable information is a matter of understanding whether proximal cues are sufficiently diagnostic of epistemically reliable distal sources given the ecology of the *Web*. This, I assume, is possible only under the condition that this ecology is stable and sufficiently constrained.

2.2 Stable Environments

The *ecology of the Web* has been the object of extensive studies in the information science literature [21, 22, 23]. The existence of strong ecological regularities constrains the way in which users learn the structure of the *Web* and determines to a large extent their preferential strategies in information search. It is plausible to assume that information seekers are *situated* in this environment and rely on ecological regularities they have learnt in order to select effective solutions for negotiating source selection problems.

Web technologies (such as *search engines*) aim at improving our information retrieval skills by reducing the cognitive effort required to solve particularly demanding tasks. In this sense, they tend to favor the selection of simple, effortless and automatic strategies over more costly processes. By enriching the user ecology with highly informative cues, technology aims at reducing information processing requirements on the user.

As suggested in the Ecological Rationality literature [24, 25], stable environments offer conditions that favor the selection of shallow, effortless and relatively rigid computational strategies. These are typical features of *modular solutions* to the problem of negotiating cognitively demanding problems.

2.3 Modularity

Defendants of the modularity hypothesis argue that modularity arises as a good solution in stable environments whenever a tractability problem is presented to an organism [26, 27, 28]. As Carruthers observes,

[t]he mind is realized in processes which are computational, operating by means of algorithms defined over sentences or sentence-like structures. But computational processes need to be *local*—in the sense of having a

restricted access to background knowledge in executing their algorithms— if they are to be tractable, avoiding a “computational explosion”. And the only known way of realizing this, is to make such processes modular in nature—[27].

If epistemic deference has to be cognitively advantageous, solutions to the problem of estimating a source’s reliability must be computationally tractable (i.e. result in costs that are still profitable with respect to the expected benefits). I will call this a *cognitive affordability constraint* on deferential strategies. The selection of deferees is a paradigmatic case of problem that has to be solved in a computationally tractable way by setting limits to background knowledge, in order to avoid computational explosion. In the case of reliability judgments, this means finding sufficiently local criteria for estimating the reliability of a source, that do not draw in turn on further reliability judgments and so on.

If local inferential strategies can be identified that reliably yield the expected results, then the basic conditions are present for the selection of a modular solution to the problem.

3 Empirical Research Directions

I have reviewed some of the broad implications of the hypothesis according to which reliability judgments in the *Web* can in principles be underpinned by highly specialized heuristics exploiting proximal cues. The question that needs to be answered on empirical grounds is then whether – given the ecology of the *World Wide Web* – there actually are heuristics based on information scent that users can adopt to predict the epistemic reliability of a source. In this section, I will sketch a proposal that we are currently implementing to turn this research programme in directions for empirical research.

3.1 Information Scent for Reliability in Impoverished Displays

The first study we are currently implementing aims to investigate correlations in judgments of epistemic reliability based on information scent on the one hand, with judgments based on inspection of the target sources on the other hand. The experiments focus on subjects’ performance in judging the reliability of sources referred to by search engine results. Subjects are divided in a “proximal” and a “distal” group. Each group of subjects is required to provide answers to unfamiliar questions (e.g. *What’s the name of the smallest dinosaur?*, *Who invented the Java programming language?*, *How many electrons has the iron’s atom?*, *When was Charles Darwin Born*) on the basis of information available in *Web* sources listed in search engine results. Subjects are instructed to provide answers by picking up what they judge to be the most reliable source of information. They are explicitly told that the results have been pre-selected so that the target sources only differ in the *reliability* and *accuracy* of the information they provide (but not in their *relevance* to the task assignment). The “distal” group

is then required to rank the reliability of the sources on the basis of a full inspection of their content. The “proximal” group, on the contrary, is required to predict and rate the reliability of the same sources on the basis of proximal information displayed in the search engine result page. Proximal information is then systematically manipulated in order to mask specific cues (i.e. titles, URLs, snippets), thus producing *impoverished proximal displays*. As soon as sufficient data are collected, regression analysis will be used to determine whether judgments based on proximal cues can be used to predict the reliability ranking of the actual sources. If subjects can predict the reliability of a source from the proximal information available in the search engine result page, this suggests the effectiveness of heuristics strategies.

3.2 Information Scent with Explicit Reputational Cues

A second set of experiments aims to study a complementary issue, namely the impact of *explicit reputational cues* on judgments of epistemic reliability. The *Web* offers a plethora of *visual reputational cues* that mark items as “popular” (or rank some items as “more popular” than others). In the general case, subjects have no reasons to trust the validity of such reputational cues other than trusting the provider of these cues. A second research direction is then to study to which extent judgments of epistemic reliability are affected by displaying reputational cues in the proximal stimuli and to what extent the overall trust of the subject in the system providing these explicit representations of popularity modulates their judgments. Different test conditions are contrasted in which subjects are instructed that the displayed reputational cues result from (1) algorithmically generated ratings, (2) ratings provided by single reviewers or (3) collaborative ratings provided by a user community. The design of the experiment is identical to the one adopted in the first study, the only difference is that, instead of masking specific types of proximal cues and presenting impoverished search engine displays, search engine results will be enriched with explicit reputational cues in order to estimate potential biases.

4 Conclusions

This paper has fleshed out the main rationale, some theoretical implications and potential research directions for studying processes underlying epistemic reliability judgment in the *World Wide Web* as a class of capabilities depending on the selection of highly specialized heuristics. I have suggested the conditions under which such heuristics are likely to emerge and the advantage they offer to the *Web* user insofar as they allow predictive judgments of reliability in situations that rule out extensive, cognitively demanding inspections of the distal sources. As suggested in the rationale for this research programme, a study of heuristic processes is complementary to the study of *a posteriori* evaluative judgments. It is not likely to account for *any* kind of epistemic reliability judgment in which users engage when searching for information in the *Web*, but only for those cases in which:

- the engagement of the user need not be too strong (i.e. can tolerate revisions and amendment of erroneous inferences based on deceiving predictions)
- effective heuristics can be built on the basis of the constraints afforded by the *Web* ecology.

I have argued that such cases could be more common and crucial than the current literature on *Web* credibility has realized, as they may provide a more plausible account of reliability judgments in real-world conditions.

Bibliography

- [1] Recanatani, F.: *Oratio Obliqua, Oratio Recta*. MIT Press, Cambridge, MA (2000)
- [2] Fogg, B.J., Tseng, H.: The elements of computer credibility. In: CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM Press (1999) 80–87
- [3] Hogarth, R.: *Judgment and choice. The psychology of decision*. John Wiley & Sons, Inc., New York (1987)
- [4] Rieh, S.Y.: Judgment of information quality and cognitive authority in the web. *Journal of the American Society for Information Science and Technology* **53**(2) (2002) 145–161
- [5] Eagly, A.H., Chaiken, S.: *The psychology of attitudes*. Harcourt Brace Jovanovich., New York (1993)
- [6] Petty, R.E., Cacioppo, J.: The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology* **19** (1986) 123–205
- [7] Petty, R.E., Wegener, D.T.: The elaboration likelihood model: Current status and controversies. In Chaiken, S., Trope, Y., eds.: *Dual-Process Theories in Social Psychology*. Guilford Press, New York (1999)
- [8] Eagly, A.H., Chaiken, S.: Attitude strength, attitude structure, and resistance to change. [29]
- [9] Petty, R.E., Haugtvedt, C., Smith, S.M.: Elaboration as a determinant of attitude strength: Creating attitudes that are persistent, resistant, and predictive of behavior. [29]
- [10] Fogg, B.J., Soohoo, C., Danielson, D.R., Marable, L., Stanford, J., Tauber, E.R.: How do users evaluate the credibility of Web sites? A study with over 2,500 participants. In: DUX '03: Proceedings of the 2003 conference on Designing for user experiences, New York, NY, USA, ACM Press (2003) 1–15
- [11] Fogg, B.J.: Prominence-interpretation theory: explaining how people assess credibility online. In: CHI '03: CHI '03 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM Press (2003) 722–723
- [12] Wathen, N.C., Burkell, J.: Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology* **53**(2) (2002) 134–144
- [13] Pirolli, P., Card, S.: The evolutionary ecology of information foraging (1997)
- [14] Pirolli, P., Card, S.: Information foraging. *Psychological Review* **106**(4) (1999) 643–675
- [15] Pirolli, P.: Rational analyses of information foraging on the web. *Cognitive Science* **29** (2005) 343–373
- [16] Nielsen, J.: Alertbox: When search engines become answer engines. Website (August 16 2004)
- [17] Clark, A.: *Natural born cyborgs: Minds, technologies and the future of human intelligence*. Oxford: Oxford University Press (2003)

- [18] Sterelny, K.: Cognitive load and human decision, or, three ways of rolling the rock up hill. In Carruthers, P., Laurence, S., Stich, S., eds.: *The Innate Mind: Culture and Cognition*. Cambridge University Press, Cambridge (2006)
- [19] Nielsen, J.: Alertbox: Information pollution. Website (August 11 2003)
- [20] Wexler, M., van Boxtel, J.J.: Depth perception by the active observer. *Trends in Cognitive Science* **9**(9) (September 2005) 431–438
- [21] Pitkow, J.E., Recker, M.: Results from the First World-Wide Web User Survey. *Computer Networks and ISDN Systems* **27**(2) (1994) 243–254
- [22] Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN Systems* **27**(6) (1995) 1065–1073
- [23] Huberman, B.A.: *The Laws of the Web : Patterns in the Ecology of Information*. The MIT Press (April 2003)
- [24] Todd, P.M.: Fast and frugal heuristics for environmentally bounded minds. In Gigerenzer, G., Selten, R., eds.: *Bounded Rationality*. MIT Press, Cambridge, MA (1999)
- [25] Bullock, S., Todd, P.M.: Made to measure: Ecological rationality in structured environments. *Mind and Machines* (1999)
- [26] Sperber, D.: In defense of massive modularity. In Dupoux, E., ed.: *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*. MIT Press, Cambridge, MA (2002)
- [27] Carruthers, P.: Moderately massive modularity. In O’Hear, A., ed.: *Mind and Persons*. Cambridge University Press, Cambridge (2003)
- [28] Carruthers, P.: Simple heuristics meet massive modularity. In P. Carruthers, S. Laurence, S.S., ed.: *The Innate Mind*. Oxford University Press, Oxford (2006)
- [29] Petty, R.E., Krosnick, J.A., eds.: *Attitude strength: Antecedents and consequences*. Erlbaum, Mahwah, NJ (1995)