

# Visual Representations of Credibility on Wikipedia: Less is More

*Jessica Richman*  
Stanford University  
jessica.richman@gmail.com

*Leslie Wu*  
Stanford University  
lwu2@cs.stanford.edu

## ABSTRACT

We present the results of a between-subjects ( $n=93$ ) experiment that investigates the effect of different visual design representations on user perceptions of credibility of Wikipedia articles. Users were recruited through Amazon.com's Mechanical Turk system. Participants were randomly assigned to a different condition, corresponding to the visual representation, and were asked a fact-based question about a semi-fictitious Wikipedia article. Participants furthermore reported their perceptions of credibility. An analysis of completion times and responses suggests that simpler visual representations have greater impact on users' perceptions of credibility than more complex visual representations.

**ACM Classification:** H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

**General terms:** Documentation Experimentation Human Factors Measurement

**Keywords:** Wikipedia credibility trust representations

## INTRODUCTION

According to a recent Pew Internet Project survey, more than a third (36%) of the online American adult population consults Wikipedia [5]. In spite of possible concerns regarding the accuracy of articles on Wikipedia, users still cite Wikipedia's convenience as an important reason for using Wikipedia as a data source [5].

Researchers have created tools which provide better social transparency on Wikipedia, whether on a per-article or editorial basis [7], or to the public as a whole [<http://wikiscanner.virgil.gr>]. Other researchers have attempted to derive metrics of trust that may correlate to credible or trustworthy articles on Wikipedia [3, 4] as well as to display these metrics visually [1].

While some efforts have been made to unpack the notion of credibility in computer environments [6], little work has been done to investigate the efficacy of information displays or tools that attempt to change how users assess credibility online. In what ways does the design of these displays effect users' behavior or perceptions of credibility on Wikipedia? In this paper, we present the results of a web-based experiment, administered via Mechanical Turk, that investigates these effects.

## EXPERIMENTAL DESIGN

We first ran a survey of Wikipedia users ( $n=41$ ) to better understand how users currently assess the credibility of

Wikipedia, and then used the results of this survey to drive the design of our second experiment.

We designed an experiment where users were shown a Wikipedia article and then asked to answer a fact-based question). Users were told that we were "designing a new keyword search system for Wikipedia" and then shown a series of three "related search results". They were asked to rate the credibility of the article, and had the option of clicking on links to find additional information outside of Wikipedia.

After accepting the task presented, users were randomly assigned (through a partitioned MD5 hash of their worker IDs) to one of fifteen conditions (or the null condition), corresponding to a high or low rating for trustworthiness. The article was based on an existing Wikipedia entry, modified to create the semi-fictitious persona of "Sigmund Morris", a Canadian politician. Above the text of the article, all users were shown the same three related (also semi-fictitious) "search result" links, and above these links, users were shown one of fifteen visual representations (or no visual aid at all).

We asked users to determine what year "Sigmund" was born, and to assess and then report the credibility of the article using a slider which ranged from "Not Credible" to "Credible". We used two different dates of birth for "Sigmund" to differentiate between those who entered the date from the Wikipedia page ("1919") versus those who used the additional links ("1924"). Users were promised a bonus for the "right" answer. We tracked completion times, numeric answers, and measures of credibility.

## RESULTS

### Complex vs. Simple

Our most significant result was that choice of representation had a significant effect on the perception of credibility of Wikipedia articles. Simpler visual representations had greater impact. Of the sixteen configurations we tested, the ones with the greatest difference in perceived credibility between high and low trust measures were the numbers (90 and 10) and the smileys (smile and frown) (see Figure 1). This was contrary to our expectation that providing more information would be helpful to users in deciding whether or not a page was credible.

### Completion Times

When presented with a Wikipedia page, users tended to determine if it was credible very quickly. Of those who an-

swered “1919” for the birth year, 75% answered in less than three minutes. Of those who answered “1924” for the birth year, 100% answered that that it was less than 20% credible within two minutes. Moreover, completion times appeared a power law distribution, with 10% taking more than six minutes; of the other 90%, the mean response time was 97 seconds. We take this to indicate that credibility assessments occur very rapidly; this may be why simpler measures had greater impact.

### SURVEY

On a 7-point Likert scale (1 = most trustworthy, 7 = least trustworthy), participants were asked to rate the overall credibility of Wikipedia. The most common responses were 3 and 5 (23 of 41 responses). The mean answer was 3.8. This may indicate that while users do not trust the site as a whole, they do trust particular articles as they use them. As one respondent noted, “I have never doubted the credibility of a particular page, but I am somewhat dubious about completely trusting the site as a whole.” This may point to a need for future work on the differences between site and page credibility.

### RELATED WORK

Rieh and Danielson present a comprehensive, multidisciplinary framework for understanding credibility, noting that many researchers decompose credibility into the key dimensions of trustworthiness and expertise [6]. Credibility has been studied in the context of the web, and as users increasingly depend on information sources such as Wikipedia, researchers have begun to call into question the credibility of these online sources [2].

Recently, researchers have attempted to construct systems which automatically analyze, and subsequently display, metrics of trust [1]—the dimension of trustworthiness—and which aggregate and visualize the social life of Wikipedia data, making somewhat visible the dimension of expertise [7]. Other researchers have suggested that critical readings of Wikipedia should consider not only content quality, but “collaborative working” dynamics [3] such as dissonance amongst authors.

### FUTURE WORK

This work could be expanded in several ways. In our user survey, we identified two major types of tasks performed by users on Wikipedia: specific fact inquiries (finding a date, place, definition, etc.) and general inquiries (learning general information about a topic). Future studies could examine the effects of credibility metrics for general inquiries (perhaps by using a multiple choice questionnaire format to verify comprehension). Another possibility would be to include different types of pages (e.g., controversial, disputed). Also important may be the effects of learning on visual representations of credibility. By viewing a series of web pages, users may learn the system and be able to better appreciate more complex tools. Further, additional studies could be performed with a much larger number of participants to more robustly demonstrate these effects.

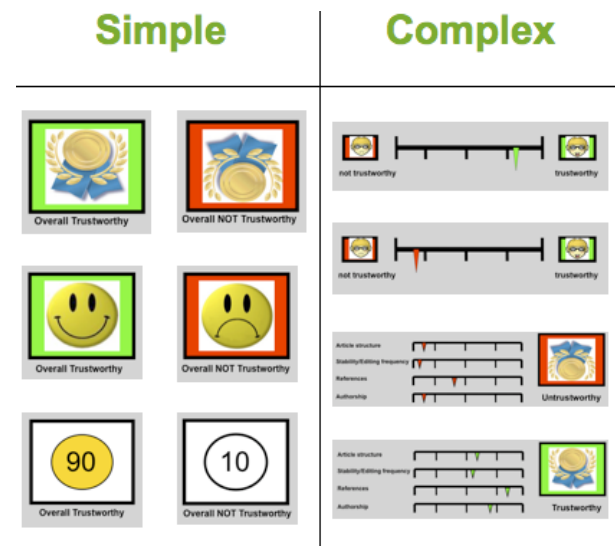


Figure 1: A sample of simple vs. complex representations of credibility.

### ACKNOWLEDGMENTS

The authors would like to acknowledge Scott Klemmer and Joel Brandt for their superb guidance and Ed Chi for sharing his experiences with WikiDashboard.

### REFERENCES

1. Adler, B.T., J. Benterou, K. Chatterjee, L. de Alfaro, I. Pye, V. Raman. Assigning Trust To Wikipedia Content. *Technical Report UCSC-CRL-07-09, School of Engineering, University of California, Santa Cruz, CA, USA*, November 2007.
2. Denning, P., Horning, J., Parnas, D., and Weinstein, L. 2005. Wikipedia risks. *Commun. ACM* 48, 12 (Dec. 2005), 152-152
3. Dondio, P., Barrett, S., Weber S., Seigneur, J. et al. Extracting trust from domain analysis: a study on Wikipedia, *IEEE ATC*, Wuhan, China, 2006.
4. Dondio, P., Barrett, S. Calculating the Trustworthiness of Wikipedia Articles using DANTE Methodology. *LADIS eSociety conference*, Dublin, Ireland, 2006.
5. [http://www.pewinternet.org/pdfs/PIP\\_Wikipedia07.pdf](http://www.pewinternet.org/pdfs/PIP_Wikipedia07.pdf)
6. Rieh, S.Y. & Danielson, D.R. Credibility: A multidisciplinary framework. B. Cronin (Ed.), *Annual Review of Information Science and Technology*, Vol. 41. Medford, NJ: Information Today, 2007, pp. 307-364.
7. Suh, B.; Chi, E. H.; Kittur, A.; Pendleton, B. A. Lifting the veil: improving accountability and social transparency in Wikipedia with WikiDashboard. *Proceedings of the 26th Annual ACM Conference on Human Factors in Computing Systems; 2008* (April 5-10; Florence, Italy). NY: ACM; 2008; pp. 1037-1040.