Can You Ever Trust a Wiki? Impacting Perceived Trustworthiness in Wikipedia

Aniket Kittur[+], Bongwon Suh[*], Ed H. Chi[*]

[+] Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15232
nkittur@cs.cmu.edu

[*] Palo Alto Research Center Inc.
3333 Coyote Hill Road
Palo Alto, CA 94304 USA
{suh, echi}@parc.com

ABSTRACT

Wikipedia has become one of the most important information resources on the Web by promoting peer collaboration and enabling virtually anyone to edit anything. However, this mutability also leads many to distrust it as a reliable source of information. Although there have been many attempts at developing metrics to help users judge the trustworthiness of content, it is unknown how much impact such measures can have on a system that is perceived as inherently unstable. Here we examine whether a visualization that exposes hidden article information can impact readers' perceptions of trustworthiness in a wiki environment. Our results suggest that surfacing information relevant to the stability of the article and the patterns of editor behavior can have a significant impact on users' trust across a variety of page types.

Author Keywords

Wikipedia, wiki, collaboration, stability, trust, visualization, social computing.

ACM Classification Keywords

H.5.3 [Information Interfaces]: Group and Organization Interfaces – Collaborative computing, Computer-supported cooperative work, Web-based interaction, H.3.5 [Information Storage and Retrieval]: Online Information Systems, K.4.3 [Computers and Society]: Organizational Impacts – Computer-supported collaborative work

INTRODUCTION

Large-scale collaborative co-creation has become a highly successful paradigm for creating, finding, and consolidating content online. One of the largest and most popular examples is Wikipedia, an online encyclopedia in which any reader can also contribute, with their changes immediately visible to subsequent visitors. Wikipedia has enjoyed tremendous success and popularity; it is ranked in the top 10

most used websites by Alexa.com, and includes over 2 million articles in the English Wikipedia alone.

However, Wikipedia is sometimes viewed with skepticism by readers and contributors alike due to its mutable nature and user-generated content. One Wikipedian we surveyed succinctly stated the view that has been echoed by many other users, readers, and news sources thus: "*Wikipedia, just by its nature, is impossible to trust completely. I don't think this can necessarily be changed.*"

More concretely, Denning et al. [2] enumerate a number of risks associated with the usage of Wikipedia, which are applicable to many other collaborative systems with user-generated content. These include:

1. Accuracy: Not knowing which content is accurate; often exacerbated by lack of references.
2. Motives: Not knowing the motives of editors, who may be biased for various reasons.
3. Expertise: Not knowing the expertise of editors.
4. Stability: Not knowing the stability of an article and how much it has changed since the last viewing.
5. Coverage: Spotty coverage of topics.
6. Sources: Cited information may come from hidden or non-independent source

One possibility is that distrust of wiki content is not due to the inherently mutable nature of the system but instead to the lack of available information for judging trustworthiness. Note that five of the six risks described above (all except for coverage) are not about the content itself but instead about the reader not having sufficient information to evaluate the trustworthiness of the content. Disclosing patterns of past performance and providing rich feedback about users and content are best practices for increasing trust online [12], suggesting that Wikipedia and other systems with rich transaction histories may be especially fruitful targets for increasing trust through aggregating relevant information, since every edit is recorded. However, although a number of researchers have mined Wikipedia's editing history to develop trust-relevant metrics, there is little extant evidence that such metrics can impact users' trust.

In this paper we explore the question of how users' perceptions of trustworthiness for a mutable wiki system can be changed through surfacing page-related information.

Both researchers and system designers can benefit from understanding the degree to which trust can be affected, the directions in which it can be altered, and the factors that are effective in changing it.

RELATED WORK

Trust is an important issue that has become increasingly studied in online environments. Fogg and Tseng [6] define trust as: "*a positive belief about the perceived reliability of, dependability of, and confidence in a person, object, or process.*" It has been the subject of considerable research due to its importance in virtually all online interactions, ranging from e-commerce transactions to online discussion groups [11][12].

Although many distrust Wikipedia, researchers have found its content to be of surprisingly high quality. A 2005 Nature study [7] found that Wikipedia had only slightly more errors than the Encyclopedia Britannica (approximately 4 to every 3) for a sample of science-related articles. Researchers have also found that vandalism is typically reverted very quickly, often on the order of minutes [9][14]. Many articles go through one or more formal vetting and review processes [15], with the resulting feedback incorporated into improving the content. Furthermore, the history of every edit made to every article is available to anyone; thus there is at least the potential for high social transparency. Visualizations of article history have been developed to help improve this transparency [13][14].

There have also been a number of studies developing explicit trust metrics in Wikipedia. Adler and de Alfaro [1] derive author reputation from the survival of an author's edits over time. Zeng et al. [16] use dynamic Bayesian networks to calculate the evolution of trust in an article using as input editing status of authors and inserted and deleted text. Dondio et al. [3] use a combination of factors to produce an overall article trustworthiness score.

However, these studies fail to address the question of whether and to what degree surfaced information can actually affect users' perceptions in a mutable environment such as a wiki. To begin to address this gap we quantify the effect that surfaced information has on users' perceptions of the trustworthiness of content.

SURFACING TRUST

We designed a visualization of the history information of Wikipedia articles that aggregates a number of trust-relevant metrics. Our purpose in this study was not to disentangle the effects of each individual metric, but instead to determine whether and to what degree trust can be impacted. Metrics were chosen that were likely to affect trust based on past research and best practices [3][5][12], which suggested two key principles compatible with Wikipedia's history information: disclosing the past history of the *content* and of the *author*.

We developed high-trust and low-trust versions of the visualization (see Figure 1) by manipulating the following metrics:

- Percentage of words contributed by anonymous users. Anonymous users with low edit-counts often spam and commit vandalism [1].
- Whether the last edit was made by an anonymous user or by an established user with a large number of prior edits.
- Stability of the content (measured by changed words) in the last day, month, and year.
- Past editing activity. Displayed in graphical form were the number of article edits (blue), number of edits made to the discussion page of the article (yellow), and the number of reverts made to either page (red). Each graph was a mirror image of the other, and showed either early high stability with more recent low stability, or vice versa.

EXPERIMENT 1A

We collected user ratings of the trustworthiness of content that included either a high-trust or low-trust visualization. The content of the page in both conditions itself was identical; the only difference was which version of the visualization was included. To ensure that the visualization was salient all participants answered questions requiring them to attend to the information represented, including:

- In what time period was this article the least stable?
- How stable has this article been for the last month?
- Who was the last editor?
- How trustworthy do you consider the above editor?

Following the above questions, users were told that the article was being considered for inclusion in a collection of trustworthy articles, and to rate on a 7-point scale how strongly they felt that the article should be included in the collection. All hyperlinks in the article were disabled (including those to the edit history) to reduce the likelihood that a user would mine this information from the article edit history.

Design

We used a 2x2x2 design, selecting articles that were either high or low quality, highly controversial or uncontroversial, and included a high-trust or low-trust visualization (see Table 1). Every participant saw only one version of each page. Article content was held constant for all conditions, ensuring that any effects would be due to the visualization.

High quality articles were randomly selected from the pool of Wikipedia "GA-class" articles, which have been peer-reviewed to meet a high level of quality[1]. Low quality articles were selected from the pool of "B-class" articles. Articles of high quality were approximately matched in length, as were articles of low quality.

---

[1] GA-class articles are the highest quality class for which exist both high- and low-controversial articles.

	High quality	Low quality
Controversial	George W. Bush Abortion	Scientology and celebrities Pro-life feminism
Uncontroversial	Volcano Shark	Disk defragmenter Beeswax

Table 1. Articles tested in each condition. For each article there was a high- and low-trust version of the visualization.

The degree of controversy of an article was measured using the technique described in [9], based on the number of revisions to an article labeled by users as controversial. We selected the two GA-class articles (high quality) with the highest measured controversy, and two B-class articles (low quality) in similar topics (famous people and abortion).
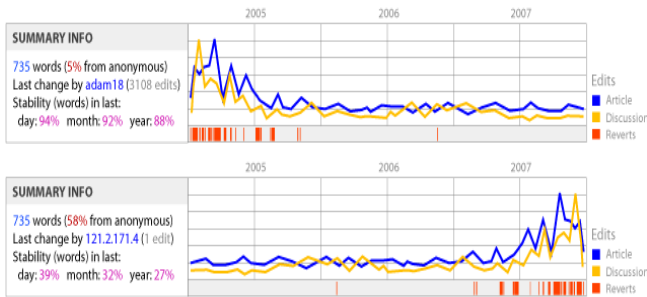

Figure 1. Examples of trust visualizations including both content-relevant and author-relevant information.

Participants and Procedure
Users were recruited via Amazon's Mechanical Turk web service using procedures similar to [10]. The Mechanical Turk system presented users with rating tasks in a random order; the nature of the system as a market meant that users could choose the number of tasks to complete. A total of 127 participants completed 433 ratings for 7 cents per rating (an average of 3.4 ratings per participant and 13.5 ratings per article-condition).

Results
There was a significant effect of the visualization such that articles with high-trust visualizations were rated as more trustworthy than articles with low-trust visualizations ($F(1, 425) = 84.69$, $p < .001$). There were also main effects of quality and controversy, with high-quality articles rated as more trustworthy than low-quality articles ($F(1, 425) = 25.37$, $p < .001$), and uncontroversial articles rated more trustworthy than controversial articles ($F(1, 425) = 4.69$, $p = .031$). There was also an interaction between controversy and quality such that high quality articles were rated equally trustworthy whether controversial or not, while low quality articles were rated lower when they were controversial than when they were uncontroversial (see Figure 2).
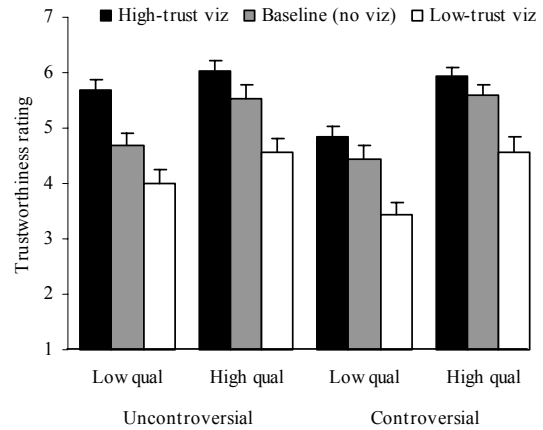

Figure 2. Ratings of page trustworthiness broken out by visualization condition, controversiality, and quality.

Importantly, there were no significant interactions between the trust visualization condition and either quality or controversy. As shown in Figure 2, the drop in perceived trustworthiness between the high-trust and low-trust conditions was not affected either by the quality of the article or by the degree of controversy. These results provide encouraging evidence of the robust effect of surfacing trust-relevant information across a variety of page types.

EXPERIMENT 1B: Baseline condition
One interpretation of the previous results is that surfacing trust information increases trust in the high-trust condition and decreases trust in the low-trust condition. However, the data are also consistent with alternative views. For example, if users have already discounted the article in its base condition as being untrustworthy, a low-trust visualization may not affect ratings (since it is already considered untrustworthy) but a high-trust visualization could have a large impact (since it contradicts prior expectations).

To determine the directions in which trust can be influenced we ran a baseline, no-visualization condition. We replaced the four visualization-specific questions with verifiable questions relating to the content of the article (e.g., how many sections / images / references it had) to match as closely as possible the level of processing across conditions and to provide verifiable input as to whether users had evaluated and processed the article.

Participants
Users were recruited via Amazon's Mechanical Turk web service. A total of 126 participants completed 240 ratings for 7 cents per rating (an average of 1.9 ratings per participant and 15.8 ratings per article-condition).

Baseline results
The results are shown in Figure 2 between the high- and low-trust ratings. Analysis revealed trustworthiness ratings in the no-visualization "baseline" condition to be significantly higher than the low-trust condition ($t(391) =$

5.43, $p < .001$), and significantly lower than the high-trust condition ($t(390) = 3.41$, $p < .01$). These results suggest that surfacing trust-relevant information can have an impact in both the positive and negative directions.

CONCLUSION AND FUTURE WORK

Surfacing trust-relevant information had a dramatic impact on users' perceived trustworthiness, holding constant the content itself. The effect was robust and unaffected by the quality and degree of controversy of the page. Trust could be impacted both positively and negatively. High-trust information increased trustworthiness above baseline and low-trust information decreased it below baseline.

These results suggest that the widespread distrust of wikis and other mutable social collaborative systems may be reduced by providing users with transparency into the stability of content and the history of contributors [13]. Given the right information, readers can make more informed judgments of the trustworthiness of content, which may increase overall trust in the system.

Future work is needed to untangle which factors are most important and how they can be best represented, since each surfaced element competes for user attention and takes time to process. There may be other negative effects of surfacing information as well: Erickson & Kellogg highlight the importance of some information remaining private, or at least not highly salient [4]. Serious thought and care is needed in aggregating and surfacing information that may lead to manipulation of the system for personal gain.

It will also be important to explore other factors that may affect user trust. One such important factor is *external credibility* [6]. Some wiki-based encyclopedia systems, such as Scholarpedia.org and Citizendium.org, are already trying to achieve greater external credibility through the involvement of experts, though it is too early to determine the success of such approaches.

Future work also includes application to systems making use of the findings in the current study. Using visualizations to promote the awareness of rules and increase accountability could benefit not only end-users but also improve the effectiveness of active participants [4]. The work described here provides both empirical data and theoretical evidence that can improve systems surfacing content and user information in Wikipedia, such as our early efforts described in [13].

Finally, more work is needed to understand how well our findings generalize to other collaborative systems with user-generated content. These results could have useful applications in influencing trust for other systems with rich transaction history information, such as discussion groups, social networks (e.g., Facebook), and user-rated content sites (e.g., Digg).

REFERENCES

1. Adler, B. T., and de Alfaro, L. A content-driven reputation system for the Wikipedia. In *Proc. WWW 2007*, ACM Press (2007), 261-270.

2. Denning, P., Horning, J., Parnas, D., and Weinstein, L. Wikipedia risks. *CACM 48* (2005), 152.

3. Dondio, P., Barrett, S., Weber, S., and Seigneur, J. Extracting Trust from Domain Analysis: a Study Case on the Wikipedia Project. In *Proc. IEEE ATC 06*, IEEE Press (2006), 362-373.

4. Erickson, T., and Kellogg, W. Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM TOCHI 7* (2000), 59-83.

5. Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., and Tauber, E. R. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. In *Proc. 2003 Conf. on Designing for User Experiences,* ACM Press (2003), 1-15.

6. Fogg, B.J. and Tseng, H. The Elements of Computer Credibility. In *Proc. CHI 1999*, ACM Press (1999), 80-87.

7. Giles, G. Internet encyclopaedias go head to head. *Nature*, *438*, 7070 (2005), 900-901.

8. Kantowitz, B.H., Hankowski, R.J., and Kantowitz, S.C. Driver acceptance of unreliable traffic information in familiar and unfamiliar settings. *Human Factors 39* (1997) 164–176.

9. Kittur, A., Suh, B., Chi, E., and Pendleton, B. A. He says, she says: Conflict and coordination in Wikipedia. In *Proc. CHI 2007,* ACM Press (2007), 453-462.

10. Kittur, A., Chi, E., Suh, B. Crowdsourcing User Studies With Mechanical Turk. In *Proc. CHI 2008,* ACM Press (2008).

11. Resnick, P., Kuwabara, K., Zeckhauser, R., and Friedman, E. Reputation systems. *CACM 43* (2000), 45-48.

12. Shneiderman, B. Designing trust into online experiences. *CACM 43* (2000), 57-59.

13. Suh, B., Chi, E., Kittur, A., Pendleton, B. Lifting the Veil: Improving Accountability and Social Transparency in Wikipedia with WikiDashboard. In *Proc. CHI 2008,* ACM Press (2008).

14. Viégas, F. B., Wattenberg M., and Dave K. Studying cooperation and conflict between authors with history flow visualizations. In *Proc. CHI 2004*, ACM Press (2004), 575-582.

15. Viégas, F. B., Wattenberg M., and McKeon, M. M. The hidden order of Wikipedia. In *Proc. HCII* (2007).

16. Zeng, H., Alhossaini, M., Ding, L., Fikes, R., and McGuinness, D. L. Computing trust from revision history. In *Proc. Intl. Conf. on Privacy, Security and Trust* (2006).