Modulating Video Credibility via Visualization of Quality Evaluations

Nicholas Diakopoulos
School of Communication and Information
Rutgers University
4 Huntington St., New Brunswick, NJ, USA
diakop@rutgers.edu

Irfan Essa
School of Interactive Computing
Georgia Institute of Technology
75 5th St. NW, Atlanta, GA, USA
irfan@cc.gatech.edu

ABSTRACT

In this work we develop and evaluate a method for the syndication and visualization of aggregate quality evaluations of informational video. We enable the sharing of knowledge between motivated media watchdogs and a wider population of casual users. We do this by developing simple visual cues which indicate aggregated activity levels and polarity of quality evaluations (i.e. positive / negative) which are presented in-line with videos as they play. In an experiment we show the potential of these visuals to engender constructive changes to the credibility of informational video under some circumstances. We discuss the limitations, and future work associated with this approach toward video credibility modulation.

Categories and Subject Descriptors

H.5 [**Information Interfaces and Presentation**]: user interfaces, multimedia information systems – *video, evaluation methodology*; H.4 [**Information Systems Applications**]: communications applications

General Terms

Design, Human Factors

Keywords

Video Annotation, Credibility, Visualization, Mechanical Turk

1. INTRODUCTION

Can you trust the information you get on a daily basis online? Where did it come from and who produced it? What biases of selection have contributed to that information? And what kinds of expertise did the person have who produced that information? The problem of information quality including aspects of credibility, validity, and accuracy is pervasive in contemporary media, especially as we begin considering user generated content, advertisers, and advocacy groups [17].

Oftentimes referred to as media *watchdogs*, web sites such as Politifact and FactCheck have evolved to address issues of

information quality by combing through the media and engaging in fact-checking and re-contextualization of news and other media reports. For high profile video events such as the State of the Union address given by the president of the U.S., there is a considerable demand for this type of watchdogging activity. For instance, recent coverage by news outlets like PBS included annotated transcripts and video snippets showing analysis from experts and journalists[1]. One of the major issues with such analytic presentations as are found on Politifact, Factcheck, and PBS is that, especially for video, the analysis is divorced from the video itself, making the multimedia context difficult to understand in relationship with the textual analysis.

While most methods of watchdogging are labor intensive, another method of coping with information quality encompasses harnessing social information processing systems [15] which seek to filter information and identify quality by aggregating the recommendations and ratings of many users through passive (e.g. through usage) or active (e.g. through voting or active rating) metrics of recommendation. Recent work on video annotation systems has combined the notion of watchdogging with social information procession and shown the benefit of collaborative evaluation of information quality with respect to enhanced understanding of context, comprehensiveness, and different perspectives by users [5]. But the effort associated with using such systems is still substantial and unwarranted for casual users.

In this work we develop and evaluate methods for the *in-context syndication* of video watchdog information to a less engaged class of users. Our goal is to enable sharing of the knowledge of interested watchdogs such as journalists with a wider population of users and in the process modulate perceptions of information quality. We do this by developing simple visuals that indicate aggregated activity levels and polarity of evaluations (i.e. positive / negative) shown in-line with videos as they play. More interested users can interact with and drill into the visualization for the details of the evaluations including tags, sources, and comments. In order to understand the influence of this visualization on casual video consumption we also evaluate its impact on the credibility of the information presented in the video as compared to a control presentation of the video.

2. RELATED WORK

Information quality, including such aspects as reliability, credibility, accuracy, and validity has been studied in a variety of contexts such as Wikis [30], social media [15], and traditional

---

[1] http://www.pbs.org/newshour/interactive/speeches/1/annotated-state-of-the-union/

news stories [14]. While some aspects of information quality are objectively verifiable (e.g. validity), others such as credibility (i.e. belief) are *perceived* and *subjective* notions of quality and as such can be modulated on an individual level [8]. Belief in particular can be thought of as a person's estimate of the *subjective probability* or certainty that a proposition is true [32]. The focus of our work here is the design and evaluation of visual cues which may engender constructive changes to perceptions of belief in informational video (e.g. by cueing people to poor quality information in video). An extensive review of the research and communication theories associated with attitude and belief change can be found in [19, 23, 32].

Recent work looking at Wikipedia has suggested that users' perceptions of trustworthiness and credibility of information can be impacted by detecting and then visualizing edit activity and reputation information using relatively simple visual dashboards [13, 24, 31]. Other work on Wikipedia has looked at visualizing the trustworthiness of segments of articles based on edit history metrics [1, 2]. Nakamura et al. postulate that credibility can be modulated using social annotation data showing the polarity of time-stamped textual responses to video information [21].

These approaches toward visualizing information quality often vary in the source of the annotations that they use. For instance, the data used by Nakamura as well as in other video response work by Ayman et al. [6, 26] utilizes short text messages that are associated to the video by the public as it is playing. Automatic text analysis (e.g sentiment analysis) is then used to determine the reaction of the message to the video content. Algorithms for automatically evaluating the information quality of content have also been employed by Adler [1, 2] as well as Murakami [20]. While there are certainly many benefits to employing automatic analysis, Nakamura's implementation also exposed several difficulties when dealing with unstructured video comments and sentiment detection including an inability to discern whether the sentiment of comments was in response to the original video or to other comments.

Some of these difficulties are avoided with more explicit video evaluation information such as that collected by the Videolyzer system [5], which includes hierarchically organized quality tags, sourcing, and free text comments. However the visual complexity of Videolyzer and its orientation toward motivated bloggers and journalists means that it is inappropriate for casual users to benefit from its rich annotation information. Here we consider a model where videos would be manually annotated using a structured tool. This would leverage existing journalistic practices by for instance FactCheck to add these annotations. But then these annotations would be syndicated to more casual users via simplified and aggregated representations of the annotations, so as to share the benefit of the manual annotation process with as wide an audience as possible. Our work is most similar to Nakamura's [21] with the addition of more interactive capabilities and layers of *structured* annotations (comments, tags, sources / evidence) in the system as well as an experimental evaluation of the effect of in-context visualizations on credibility.

3. VISUALIZATION DESIGN

In the development of our visualization we drew on work in dashboard design [7] and traditional broadcast graphics, which contextualize video information with maps, names, and titles, but for the most part do not provide any notion of social quality evaluation. Our design goal was to distill a detailed hierarchical and collaborative evaluation of quality into a set of simple cues which could be useful to viewers' assessment of a video's quality. Design decisions included both what data to visualize as well as whether that data should be immediately visible or only visible upon engagement and further interaction.

3.1 Visual Cue Selection

Prior work on discussion visualization suggests a range of quantifiable metrics for the characterization of the structure and content of online discussions such as *size (i.e. breadth, depth, number of messages and contributors), recency, activity level, anonymity, stability,* and *tone* [3, 13, 28]. The ability to detect these features automatically rests both on the degree of structure in the commenting system as well as the robustness of content analysis algorithms (e.g. sentiment or affect recognition).

In order to reduce consumption bandwidth as well as to maximize the potential for showing cues relevant to credibility we organized cues into three levels of successive detail. We chose to focus the initial visualization on aggregate measures: *activity level* and *annotation polarity,* with interactions revealing additional information such as *use of sources, number of contributors,* and ultimately individual *quality tags, comments,* and evidence *sources.*

Activity level, an honest signal of interest, indicates areas of the video that have generated more or less discussion and thus might be worth investigating [22]. Polarity shows whether people have evaluated a section of video as positive or negative. Evidence and sources were included because of their expected impact on credibility evaluations [10, 18]. Finally, the number of contributors was included in order to indicate if the activity or polarity of annotations in one area was the result of one person or a diversity of opinion. Our purpose in this paper is not to study the
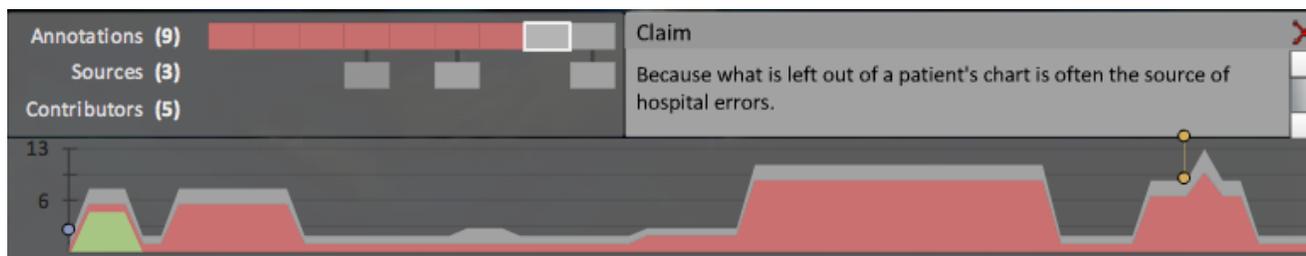

Figure 1. A close up of the graph and interactive elements from the experimental video player. It shows a stacked and colored graph of the annotations over the length of the video, which is time aligned to the navigation of the video. Hovering over the graph shows the panel at left and clicking the elements there expand in the panel at right.

individual effects of these cues but rather to understand if, taken together, they can impact the perceived credibility of the video.

3.2 Visual Design

Layered over the bottom of the video, the graphic (Figure 1) depicts the activity and polarity of annotations as a stacked line graph which is time-aligned to the timeline of the video. Negative annotations are red, positive are green, and neutral are gray. As the video plays, the timeline thumb advances and intersects the graph to show the relevant part of the graph.

Interaction with the graph reveals two additional layers of information, which are shown in panels that pop up. The first extra layer (Figure 1. top left) shows a row of annotations, with each annotation "chip" colored by its polarity (again red is negative, green positive, and gray neutral). An annotation that has an evidence source is visually connected to another chip in the "sources" row. The number of contributors to these annotations is also shown in text. These chips can be further drilled into and when selected open a secondary panel (Figure 1 top right) which shows the details of the annotation, whether that be the text of a comment, a tag, or a link to a supporting source. The user can scroll through and read the entire message there. All of these visuals roll-up from the bottom of the video and are designed to be tightly integrated with watching the video itself.

3.3 Interaction Design

The visuals provided in our experimental video player are designed to give a simple overview of the annotation activity of the video while also allowing for more in depth interaction ("details on demand") such as drilling into the two detail panels. The user sees the first detail panel by running their cursor (hovering) over the graph. The panel tracks the cursor and is overlayed above the graph (see Figure 2b.). If the use clicks the graph at this point it will pin the panel above, allowing the user to transition to interacting with the information in the first detail panel. Clicking the graph again will un-pin the panel and it will begin tracking the cursor again. Clicking an annotation in the detail panel then animates and expands the second panel to the right of the first one (see Figure 2c.). There is limited interaction with the second panel but the user can scroll through and read the message or close it. There are also standard video controls available for non-linear navigation (dragging the timeline thumb) as well as play and pause controls.

4. EXPERIMENT

We were interested in understanding to what extent the visualization we developed could modulate people's credibility evaluations of the video. We conducted an experiment comparing participants' belief ratings between the experimental video player and a control version, which did not have any additional graphics.

Each participant in the study completed a background questionnaire to collect data about political viewpoint, English fluency, and their interest levels in a variety of political issues. Then the participant was randomly assigned to one of the two conditions (between subjects experimental design). The participant was then exposed to three informational videos in succession. The ordering of the videos was counterbalanced to mitigate any potential ordering effects associated with seeing the experimental condition three times. A blue and yellow mapping was provided for color blind users. In the experimental condition,
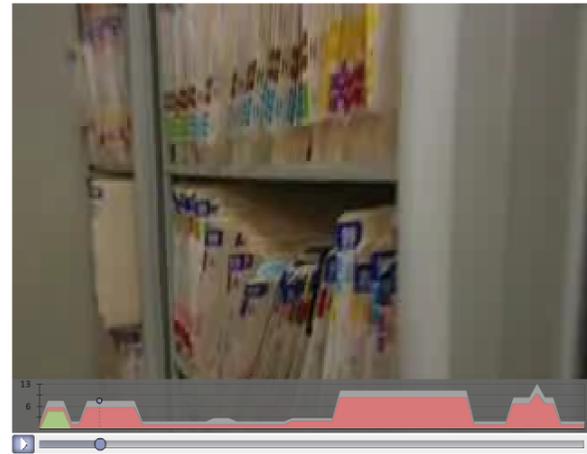
(a)

(b)

(c)

Figure 2. The experimental video player showing (a) the player with color stacked graph at the bottom, (b) the hover panel which shows when the user tracks their cursor over the graph, and (c) the individual annotations, tags, and sources when the user clicks into the information on the hover panel.

participants were told that the video had been evaluated by eight independent journalists and that these evaluations were accessible via the graphic.

Each video was followed by a brief questionnaire designed to assess reactance to the overall video and to its individual claims in terms of belief. After the last video, participants in the experimental condition completed a final questionnaire to elicit information about the user experience. The entire experiment took about 12 minutes on average to complete.

4.1 Participants and Data Filtering

Participants were recruited using Amazon Mechanical Turk (AMT), a marketplace where users complete micro tasks for small payments. One hundred and four people participated in the study for $0.25 each. As the amount of noise in user ratings can be substantial on AMT [6, 12, 13, 27, 29] results were filtered based on responses to three control questions. These control questions, specific to the content of each video rather than general as in [12], were easy to answer correctly if the video was watched. If any of these answers were incorrect, the questionnaire data from that participant was not included in our analysis.

Theories of attitude change dictate that reception and comprehension of information are essential [32]. We therefore excluded data according to users' self-reported English fluency (less than 5 on a 7 point scale was excluded). Also, if a participant indicated that they had seen a particular video before, their data was not considered in the analysis of that video since we wanted to focus on the effect of the visualization on people's first exposure to a video. Twenty-seven percent of responses were filtered out of the analysis using the above methods.

In addition to the verifiable content questions we also used interaction logs as a method for observing and filtering data based on participants' degree of engagement with the graphic in the experimental condition. This provided objective information about the level of attention that participants offered the video and graphic beyond that of answering the content questions correctly.

4.2 Content Selection and Preparation

We selected produced news packages because they generally tend to be more visually interesting and information dense than user generated content. The videos were chosen for a diversity of topics and include an ABC Medical Minute podcast about the use of electronic medical records, a NYT video report about traffic and congestion in New York City, and a CurrentTV documentary promo about the Mexican drug war. All videos were cropped and re-encoded to remove any trace of a corporate logo which might confound credibility assessments of the content. The content was annotated by an experimenter on a granular, statement-by-statement basis using Videolyzer [5] and the semi-structured information quality ontology available there [4].

The ABC and NYT videos were annotated mostly negatively with 86 percent of annotations negative whereas the CurrentTV video was annotated mostly positively with 90 percent of annotations positive. Annotations included marking claims in the videos, adding reactions such as agree, disagree, or hedge, adding rationale, tagging things like relevancy or validity on claims, and adding links to sources backing up rationale. Twenty-four of the 101 annotations had sources such as web pages and news stories supporting them. Examples of some of the annotations added to the ABC video on medical errors are shown in Table 1. This

1. Claim from video: "Because what is left out of a patient's chart is often the source of hospital errors"	
	Tag: Over emphasis
	Reason: "According to this Washington post article from 2008, the most common source of medical error was failure to recognize bed sores before they became a problem."
	Source: (link to article)
2. Claim from video: "Hospitals can reduce errors by going paperless, that is using electronic technology to better track patients' care."	
	Reaction: Hedge
	Reason: "A key point in the study described was that it's not just the technology that needs to be in place, but also that the doctors are trained to use it and use it consistently in their day to day practice.

Table 1. A sampling of the types of annotations that were added to the video including claims from the video, quality tags, and reactions. The polarity of the annotations is determined by the semantics and relationships among the tags.

method of annotating content is clearly arduous and in a real consumption context it would not work for live broadcasts, however as evidenced by recent video annotation productions in the mainstream media, such annotations could easily be obtained by the day after a large video event[2].

4.3 Results

4.3.1 User Interest

We first assessed how people responded to the graphs that were displayed in the experimental video player by examining the interaction logs. We logged user interface operations such as hovers and clicks on the graph and panels. Figure 3a shows a graph of the aggregate hover activity over the length of the NYT video for all users in the experimental condition. The figure also shows the average hover activity as a gray water mark. Figure 3b shows how the NYT video was annotated and seen by users in the study. As we can see by comparing Figure 3a with 3b, the areas that attracted the most hover activity were those with annotations visible in the graph. This indicates that these peaks attracted more attention by users than other areas of the video and that activity level is a good cue for drawing attention to salient sections of the video.

4.3.2 Video Credibility

For both of the negatively annotated videos, ABC and NYT, we found there to be an effect of the graphic on credibility ratings, with stronger effects observed for users who engaged the graphic more. To assess the degree of engagement of users we considered whether or not they had pinned the graphic in order to interact

with the detail pane (high engagement) or whether or not they had hovered their mouse over the graphic to get more information (medium engagement). If they did neither of these we considered them low engagement users. An average of 46% of users were classified as low engagement across the three different videos.

The ABC video showed a significant effect for the visualization on the overall credibility of the information in the video ($F(1, 49) = 9.43$, $p =.003$) for high engagement users. There were no effects for medium engagement or low engagement users. For the NYT video the effect was strongest for high engagement users ($F(1, 49) = 13.25$, $p = .001$), and weaker for medium engagement ($F(1,54)=5.01$, $p=.029$) or low engagement ($F(1, 75) = 4.60$, $p = .035$) users. In the case of the CurrentTV video, which was positively annotated, there were no significant effects observed. Means for the different videos and engagement levels are shown in Figure 4. Our analysis incorporates covariates collected from the background survey which would be expected to affect belief ratings such as the user's political viewpoint and topical knowledge and interest [32].

We also considered people's belief ratings on individual claims within the video. We found significant effects only for high engagement users and only on claims that had been negatively annotated. Claims that had mixed annotations (positive and negative) or that were positively annotated saw no effect on their belief ratings. For the ABC video, two of the three negatively annotated claims saw an effect for high engagement users ($F(1,49) = 4.65$, $p = .034$ and $F(1, 49) = 8.09$, $p = .006$). For the NYT video, one of the three negatively annotated claims saw a strong effect ($F(1, 49) = 9.56$, $p = .003$) and another claim saw a very weak effect ($F(1, 49) = 3.07$, $p = .084$) for high engagement users.

4.3.3 User Experience

Consistent with what others studying video graphics have found [11] there was a mixed response to having the additional interactive graphics at the bottom of the screen. Several people complained of there being too much information or of being distracted from the video by the graphics. The mean value for agreement with the statement, "I found the information graphics at the bottom of the video distracting" was 3.71 (1 = disagree, 7 = agree) indicating that most people were not largely distracted.

However, some people still had distraction problems, and as one person put it, "*It [the graphic] just distracted me and I actually barely looked at it because I was trying to watch the video.*"

Overall most people didn't appear to have too much difficulty with understanding the graphics. The mean for agreement with the statement: "I had difficulty making sense of the information graphics at the bottom of the screen" was 3.83 (1 = disagree, 7 = agree). The mean for agreement with the statement: "I found the information graphics at the bottom of the screen frustrating" was 3.08 (1 = disagree, 7 = agree). These descriptive statistics indicate that for the most part people did not have major problems with comprehending or using the graphics.

But while the graphics weren't overly distracting or incomprehensible, most users reported that they would still prefer not to have the additional graphics there. The mean agreement rating for the statement: "I liked having the interactive graphic at the bottom of the video." was 3.50 (1 = disagree, 7 = agree) and the mean agreement rating for the statement "I would prefer a video site that had a video player with information graphics similar to the one I saw here." was only 3.33 (1 = disagree, 7 = agree).

On the other hand qualitative feedback does reveal some of the more positive reactions to the graphics. We hand coded the qualitative responses and found that 50% of them were positive reactions. There were several comments that indicated that the graphics and the comments helped in understanding the video and were interesting to see while watching the video. Some thought it provided additional context and appreciated seeing the sources used to back up comments. For instance, one participant wrote, "*It gave the claims in the video more context, and allowed me to interpret what I was hearing better. Claims that I found hard to believe correlated with negative annotations.*" Another participant remarked, "*I liked the information it provided, especially the citing and links to sources.*" Several users found the granularity of the commenting to be novel. One user said,

> "*I like that you could see the key areas that comments were based on. That way when someone leaves a comment you can trace it to a specific part of the video and know exactly what they are referencing.*"
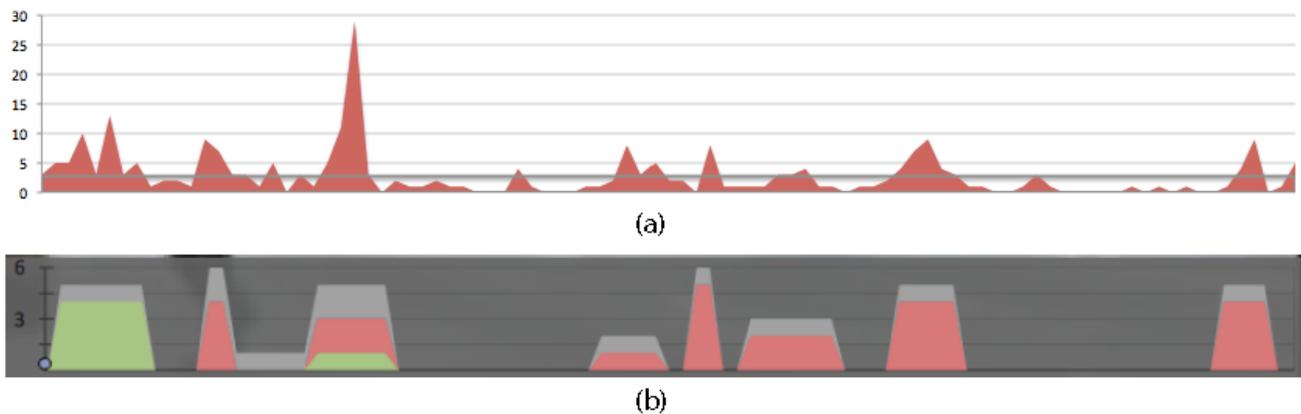
Figure 3. (a) Shows the total number of hover operations across the duration of the NYT video (red) as well as the mean number of hover operations as a gray watermark. (b) shows where the annotations were made on the video as they would have been seen by the user.
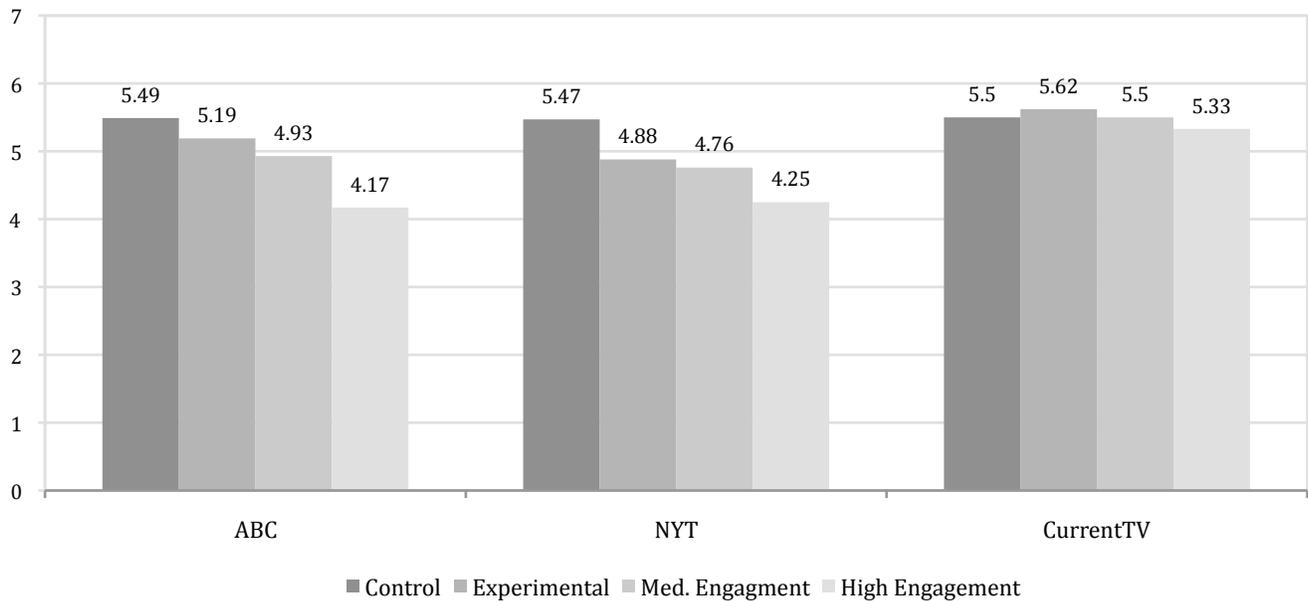
Figure 4. Mean agreement on a scale from 1 (disagree) to 7 (agree) with the statement: "Overall I consider the information presented in the video to be credible."

Taken together these results suggest that the graphics may have been distracting, intrusive, or not of sufficiently high value for some participants whereas for others the additional information and granular analysis was useful and welcome. Designers interested in integrating information visualization into inline video graphics should consider making such visuals optional so that users who do not meet the interest threshold can adapt the interface and have a more satisfactory experience.

5. DISCUSSION AND CONCLUSIONS

The strength of the effect of our visualization on credibility evaluations varied with the degree of engagement of users. Users who interacted more with the negative graphics reported more severe credibility ratings. These results are consistent with theories of attitude change and persuasion [9, 32] which predict the importance of information saliency and processing in facilitating attitude change. The more attention users gave to the graphic the more they assimilated what it was saying in terms of the quality of the information in the video.

For the CurrentTV video, which was annotated in a majority (90%) positive way there was no observed change in credibility evaluations. Theory predicts that people are more likely to scrutinize disconfirming evidence and to accept confirming evidence at face value [16]. Credibility ratings of the CurrentTV video were already high in the control condition (5.50) therefore annotations that bolstered the prevalent high belief were accepted and not scrutinized. In the ABC and NYT videos, belief ratings in the control condition were also high (5.49 and 5.47) but the negative annotations were scrutinized to a greater degree because they disagreed with prevalent beliefs.

The magnitude of the effect we observed would seem to indicate that for a majority of people such graphics have only a modest influence on their judgment of credibility. They do also attract attention to the most salient regions of the video. From the standpoint of journalism or any other truth-seeking discipline

however, any constructive change to beliefs that coincides with quality information is something to strive for.

Our results indicate that saliency and attention to the graphics we created are essential to their impact on users. Saliency could be ratcheted up by using more vivid colors or by incorporating animation to get even more attention. However, incorporating other methods to make the graphics salient could also serve to degrade the user experience by increasing distraction. At the same time, presenting annotations that are inconsistent with users' prior beliefs should naturally receive more attention. This suggests that annotations that contradict prior beliefs will have the most impact when presented in this way.

We operate under the assumption that the creators of the annotations are benign: professionals such as journalists whose ethical standards dictate that they act truthfully and honestly in their assessment of the video. However, there is always a danger in developing persuasive technology since it could be abused or misappropriated. For instance, having an open annotation system could engender gaming and bias of the annotations for personal or institutional gain. In syndication of annotations designers should carefully consider the motivation and self-interests of participants as well as decision factors like *when* to share annotations. If annotations were syndicated before they themselves could be evaluated by the community this could contribute to spreading false or misleading information.

The source of the annotations is something that warrants further attention and study. In our experiment, participants were told that the video annotations were made by journalists. This raises the issue of ecological validity since the annotations were in fact controlled by the experimenter. Outside of the laboratory such a video annotation system as Videolyzer could be open to both citizen journalists and professional journalists. One scenario would allow all annotations to be treated equally in rendering the visualization whereas another would privilege or limit annotation to a set of "trusted" sources such as professional journalists. In

practice users could even choose to filter the annotations along any facet of the source of the annotation data (e.g. political orientation, age, gender) in order to visualize how that slice of users responded. It's as yet unclear how such additional filtering capabilities would affect the credibility of the video. In general such real-world deployment variations are exciting avenues for future work.

The use of Mechanical Turk to recruit participants for the experiment does introduce additional confounds into the study. In particular, we know that the community of users on Mechanical Turk is international [25]. The videos chosen were North American in their content and focus. Would an international audience have the same response to the videos and visualizations as a domestic audience? Cultural differences in media understanding or in color coding could have impacted results and were not incorporated in our analysis.

In future work we would also like to assess the role of different types of video content. For instance, both Pirolli et al. [24] and Kittur et al. [13] found differences in credibility and trust levels when considering skeptically and controversially categorized content on Wikipedia. More work needs to be done to measure the potential for our method to impact credibility across a broader range of video content and in more ecologically valid video consumption scenarios.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Adler, B.T. and Alfaro, L.d., A Content-Driven Reputation System for the Wikipedia. in *Proceedings of the 16th International World Wide Web Conference (WWW)*, (2007).

[2] Adler, B.T., Chatterjee, K., Alfaro, L.d., Faella, M., Pye, I. and Raman, V., Assigning Trust to Wikipedia Content. in *In WikiSym 2008: International Symposium on Wikis*, (2008).

[3] boyd, d., Lee, H.-Y., Ramage, D. and Donath, J., Developing Legible Visualizations for Online Social Spaces. in *HICSS*, (2002).

[4] Diakopoulos, N. and Essa, I., An Annotation Model for Making Sense of Information Quality in Online Video. in *International Conference on the Pragmatic Web*, (Uppsala, Sweden, 2008).

[5] Diakopoulos, N., Goldenberg, S. and Essa, I., Videolyzer: Quality Analysis of Online Informational Video for Bloggers and Journalists. in *Conference on Human Factors in Computing Systems (CHI)*, (Boston, MA, 2009).

[6] Diakopoulos, N. and Shamma, A., Characterizing Debate Performance via Aggregated Twitter Sentiment. in *Conference on Human Factors in Computing Systems (CHI)*, (Atlanta, GA, 2010).

[7] Few, S. *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, 2006.

[8] Fogg, B.J. *Persuasive Technology: Using Computers to Change What We Think and Do*. Morgan, 2003.

[9] 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*, (2003), ACM, 722-723.

[10] Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P. and Treinen, M., What makes Web sites credible?: a report on a large quantitative study. in *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, (2001), ACM, 61-68.

[11] Josephson, S. and Holmes, M.E., Clutter of Content? How on-screen enhancements affect how TV viewers scan and what they learn. in *Proc. Symposium on Eye Tracking Research & Application*, (2006), 155-162.

[12] Kittur, A., Chi, E.H. and Suh, B., Crowdsourcing User Studies With Mechanical Turk in *Proceedings of CHI*, (2008), 453-456.

[13] Kittur, A., Suh, B. and Chi, E.H. Can You Ever Trust a Wiki? Impacting Perceived Trustworthiness in Wikipedia *CSCW*, San Diego, CA, 2008, 477-480.

[14] Lampe, C. and Garrett, R.K., It's all News to Me: The Effect of Insturments on Ratings Provision. in *Proceedings of HICSS*, (2007).

[15] Lerman, K. Social Information Processing in News Aggregation. *IEEE Internet Computing*. 16-28.

[16] Lord, C.G., Ross, L. and Lepper, M.R. Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, *37* (11). 2098-2109.

[17] Manjoo, F. *True Enough: Learning to Live in a Post-Fact Society*. John Wiley & Sons, 2008.

[18] McCroskey, J. The effects of evidence in persuasive communication. *The Journal of Western Speech Communications*, *31*. 189-199.

[19] Miller, K. *Communication Theories: Perspectives, Processes, and Contexts*. McGraw-Hill, 2004.

[20] Murakami, K., Nichols, E., Matsuyoshi, S., Sumida, A., Masuda, S., Unui, K. and Matsumoto, Y., Statement Map: Assisting Information Credibility Analysis by Visualizing Arguments. in *Workshop on Information Credibility on the Web*, (2009).

[21] Nakamura, S., Shimizu, M. and Tanaka, K., Can Social Annotation Support Users in Evaluating the Trustworthiness of Video Clips? in *Workshop on Information Credibility on the Web (WICOW)*, (2008), 59-62.

[22] Pentland, A.S. *Honest Signals: How They Shape our World*. MIT Press, 2008.

[23] Perry, D.K. *Theory and Reserach in Mass Communication: Contexts and Consequences (2nd Edition)*. Lawrence Erlbaum Associates, 2002.

[24] Pirolli, P., Wollny, E. and Suh, B., So You Know You're Getting the Best Possible Information: A Tool that Increases Wikipedia Credibility. in *Proc. CHI*, (2009), 1505-1508.

[25] Ross, J., Irani, L., Silberman, M.S., Zaldivar, A. and Tomlinson, B., Who are the Crowdworkers? Shifting

Demographics in Mechanical Turk. in *Conference on Human Factors in Computing Systems (CHI) alt.chi*, (2010).

[26] Shamma, A., Kennedy, L. and Churchill, E., Tweet the Debates. in *ACM Multimedia Workshop on Social Media (WSM)*, (2009).

[27] Sheng, V.S., Provost, F. and Ipeirotis, P.G., Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers. in *KDD*, (2008).

[28] Smith, M.A. and Fiore, A.T., Visualization Components for Persistent Conversations. in *Proc CHI*, (2001), 136-143.

[29] Snow, R., O'Connor, B., Jurafsky, D. and Ng, A.Y. Cheap and Fast - But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks *Conference on Empirical Methods in Natural Language Processing*, 2008.

[30] Stvilia, B., Twidale, M., Smith, L. and Gasser, L. Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, *59* (6). 983-1001.

[31] Suh, B., Chi, E.H., Kittur, A. and Pendleton, B.A., Lifting the Veil: Improving Accountability and Social Transparency in Wikipedia with WikiDashboard. in *Proc. CHI*, (2008), 1037-1040.

[32] Wyer, R.S., Jr., and Albarracín, D. Belief formation, organization, and change: Cognitive and motivational influences. in D. Albarracín, B. T. Johnson and Zanna, M.P. eds. *The handbook of attitudes*, Erlbaum, 2005.