Comprehension Quizzes Countering Clickbait

Headlines can almost never tell you everything you need to know from an article, whether through bad-faith intentions like clickbait generating more clicks or the simple fact that a couple of lines can't possibly capture a complex issue. Despite this, many users consume articles through headlines alone, leaving themselves vulnerable to internalizing mis- or disinformation without even realizing it. In designing a solution for this problem space, I consulted relevant literature and projects in order to shape the form it would come in. I found a large body of work dedicated to this subject, from broad views of social media's effects on science communication to a Google Chrome extension specifically tailored to Reddit browsing habits, aptly named RTFA. The resulting design policy is a platform-implemented comprehension quiz, required to be completed by users before sharing or commenting on an article. The quiz will be an automated form of crowdsourcing, where users generate questions which other users must answer before comment or sharing access for the article is granted. Through sitting down in qualitative testing with users, I have determined that this is a viable framework and refined it according to user feedback.

The Problem

From my time on Reddit, I noticed that users would often make statements based purely on the headline, when unsupported or even disproved in the article. A common response to this on Reddit is to RTFA: an exasperated way of saying "Read The Article."

At best, this behavior muddies discussion by creating a need for corrections and rebuttals to ensure that everyone is on the same page. At worst, it spreads misinformation and reinforces misinformed views.

Just a week ago, I came across the perfect example of the dangers of this habit, shown in figure 1. The title and the comment together give the impression that Honeywell engaged in subterfuge to undermine the US, and more or less got away with it. Users who read this comment without exploring further into the thread amend their knowledge of the company to include this misinformation, and can even amplify it by sharing it with others. The potential effects of this are subtle, yet dangerous; a user could internalize such misinformation into not only their perception of Honeywell, but of their ideas towards China and Sino-American relations as well. When the subject is something more important, like controversial legislature or political news, this can easily have disastrous effects on one's capability to maintain an informed worldview over time.

Figure 1: RTFA

The Literature

In reviewing relevant literature in order to properly form my solution to address this problem of misinformation through headline surfing, three papers were especially influential to my design process. In "A little bit of knowledge: Facebook's News Feed and self-perceptions of knowledge," Anspach uses the news snippets given by Facebook's News Feed to demonstrate that users who read only a headline, or even a short summary, gain as much knowledge as someone who read nothing at all, but feel as if they know as much as someone who read the entire thing (2019). Aside from anecdotal evidence from my Reddit sessions, I wanted to see if there was empirical evidence that this was a widespread occurrence. "Consumers and Curators: Browsing and Voting Patterns on Reddit" (Jhaver et. al 2019) shows that of 41,000 analyzed

posts, 73% were voted on without before the user viewed its content, and one in three users almost never open the link before voting on it. Finally, "Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator" (Jhaver et. al 2019) highlighted the overwhelming demand on moderation teams, and the demand for automated tools. This showed me that any solution for a problem like this had to be implemented with minimal additional overhead for the mod team.

Solution Development

In addition to the three papers listed above, my research brought me to two solutions for aspects of this problem space. The Center for Media Engagement found that quizzes following news articles serve three purposes: "They're enjoyable, they help people to learn, and they encourage people to spend more time on a webpage" (Stroud 2013). A Google Chrome extension called RTFA prevents Reddit users from viewing a posts' comments until they've actually clicked on the link, designed with intentions along the same lines as my own. With these two solutions and three papers in mind, I landed at a mandatory quiz users must take to ensure they've actually read the article for my solution.

My initial design was highly oriented around the goal of automation. It consisted of a bot utilizing NLP to analyze and automatically generate a comprehension quiz based on any given article, which users would then have to respond to before interacting with the post. This direction eliminates the problems associated with collecting user-generated questions, it poses technical problems outside of the scope of this research. Additionally, it lacks the critical reflection aspect of users having to look at a text and decide what's important for other users to come away with.

The final version of the quiz is a crowdsourced platform-implemented policy. Once an article is posted, the first X users ("generators") to interact with the post must supply a question and its answer before continuing, shown in figure 2. After X responses have been collected, all subsequent users ("responders") are shown a quiz consisting of a randomly selected question and a multiple choice response with the correct answer alongside three other user generated, randomly selected answers, shown in figure 3.
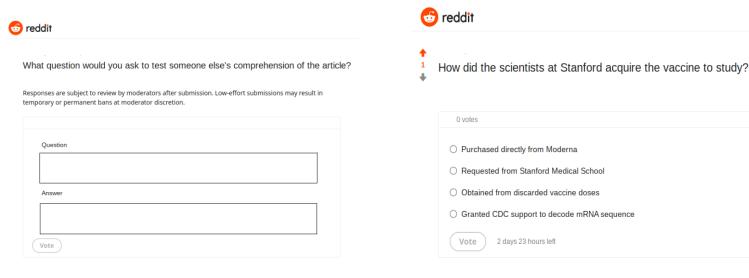
Figure 2: Generator Quiz

Figure 3: Responder Quiz

This design incorporates three design principles: First, the design should be effective in forcing users to actually read the article. The platform-wide implementation ensures users are on the same--or at least adjacent--page, while maintaining a consistent user experience across the site. Second, the quiz should not dramatically raise the barrier of participation for users. To this end, the quiz appears in a very simple format for both generators and responders, containing only the necessary elements for the task. For responders, multiple choice answers are provided to bypass the frustrations of fill-in-the-blank responses, which would be difficult with a couple words and almost impossible for longer sentences. Third, the implementation of the quiz on the platform should require minimal moderator oversight. The defining feature of this policy, its crowdsourced nature, is directly intended to address this concern by ensuring moderators don't have to manually generate questions for each article submitted to their community. Similarly, randomly selecting answers shown to responders takes the burden of assembling the quiz off of the moderators' shoulders.

Testing

My testing process was highly iterative and revolved around collecting qualitative data from users. As a placeholder for a working implementation of the quiz, I created a Google Forms sheet ("survey") for users to interact with. I individually sat down with a total of six users evenly divided across three iterations. All testers were of similar age, around 21; there were four males and 2 females; and all had varying, but at least some, experience with Reddit. To understand the user interaction process better, I created a list of questions ("questionnaire") to assess the participants' knowledge of the subject, their impression of the survey, and most importantly the mental process they went through in responding to it. This first questionnaire was divided into pre- and post-survey sections, with users reading the survey and giving their impressions in the pre-survey and then completing and recounting their experience in the post-survey. The survey

itself consisted of the question generation prompt and several auxiliary questions regarding the user's assessment of the article and its headline. This iteration highlighted the importance of a minimally intrusive quiz, as both participants indicated that they would require some form of incentive to feel compelled to take the quiz. Participants also brought up interesting points on the form generated questions should take, ie open or closed-ended.

Figure 4 (left): Survey V1
Figure 5 (above): Survey V3

In the second iteration, I significantly cut down the length of both the survey and the questionnaire. First round users indicated that looking over the survey beforehand caused them to read the article differently, so I removed the pre-survey portion of the questionnaire to more closely emulate the experience a user in the wild would encounter. In the survey, all auxiliary questions were removed so that it would mirror the quiz generators would take. Following user feedback from this iteration, I made adjustments to the specific wording of the test, and included a description of the quiz's goal to help the user better frame their questions.

For the third and final iteration, I removed the questionnaire and added a section for users to include an answer with their question. This is the version which most closely resembles the final quiz. Users interacting with this version reported an understanding of the desired outcomes as well as a comfortableness with generating a question and answer set. At this point, the overall structure of the quiz proved defensible and is ready for small-scale implementation or further explored in other variations.

Critical Reflection

In addition to the previously mentioned takeaways from each iteration of testing, working on this design taught me a lot about the process of testing in general. As a computer science student, all of my experience in testing has been highly centered around collecting quantitative data. Even in HCI areas like designing UX, my testing has revolved around numbers like ratings or time on task. Working on this project has been my first experience in conducting qualitative interviews where I talk with a user and figure out how they feel and think while interacting with my design. This proved much more useful than a quantitative approach, as I was able to adjust the design to the specific responses I received across iterations. I was able to not only observe the participants entire interaction, but engage in back-and-forth questioning and follow up on points as well. From the first iteration, users responded well to the idea of the design; however the length of the first iteration was negative. This was addressed in the next versions by first cutting down the survey and questionnaire and then removing the latter completely. Overall I learned a great deal in regards to what testing in this space looks like, how to conduct an interview, and how to interpret and utilize user feedback.

The intricacy of this problem space is underlined by the fact that even after a semester of research, there remains a complex and intriguing task with many avenues left to be explored. The current question and answer format, where generators submit a question and its answer and responders are shown a multiple choice question, requires further testing to see if this is the optimal design . Randomly selected wrong answer choices might be nonsensical or identical, yet manually ensuring reasonable answers either raises the barrier of participation for the user or increases strain on the moderation team. As minimizing moderator burden is a core foundation of this design, any changes made in opposition to that goal should be thoroughly analyzed. Aside from the Q&A format, there are several ways in which this design could turn into a moderator headache. it is yet to be seen how the policy would fare against trolls; if the generated questions themselves require vetting, that could be another job of the moderators' plate. A possible solution to this could again lie in crowdsourcing, with a second round of users ("checkers") pruning troll or low-effort responses. Perhaps the most difficult question left to consider is how to incentivize such a policy. A reward too mundane could lead to users feeling unmotivated to spend the effort in completing the process and thus foregoing it all together, negatively affecting site participation rates; a reward too appealing could lead to users abusing the system to obtain as many rewards as possible. Further research could even show that the platform-wide mandatory policy isn't ideal, and that the design is better suited to an optional implementation by subreddits or even on a post-by-post basis.

Conclusion

While this design is by no means exhaustive, it is a much-needed framework for countering misinformation through headline browsing. Solutions in this space have the potential to recreate the landscape of social article consumption, taking us another step closer to informed discussion.

The value of this design lies in its implications for three distinct groups of information agents. On an individual level, it solidifies a user's knowledge by forcing them to recall information and think critically on the article. For the moderation team, the semi-automated nature of the design--and the fact that the manual half falls on users, not mods--ensures that this solution doesn't become a "solution" which only serves to further inundate already flooded mod queues. Finally, it strengthens the subreddit community by enforcing informed discussion throughout posts, the benefits of which can clearly be seen in strictly-moderated yet beloved subreddits like /r/AskHistorians.

Works Referenced

1. Anspach, N. M., Jennings, J. T., & Arceneaux, K. (2019). A little bit of knowledge: Facebook's News Feed and self-perceptions of knowledge. Research & Politics. https://doi.org/10.1177/2053168018816189
1. Glenski, Maria, et al. (2017). Consumers and Curators: Browsing and Voting Patterns on Reddit. IEEE Transactions on Computational Social Systems, vol. 4, no. 4, 2017, pp. 196–206., doi:10.1109/tcss.2017.2742242.
2. Imartel (2019). RTFA. Github. https://github.com/lmartel/RTFA
3. Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. (2019). Human-Machine Collaboration for Content Regulation: The Case of Reddit Automoderator. ACM Trans. Comput.-Hum. Interact. 26, 5, Article 31 (July 2019), 35 pages. https://doi.org/10.1145/3338243
4. Stroud, Natalie Jomini, Scacco, Joshua, and Muddiman, Ashley. (2013). How online quizzes can benefit newsroom websites. Center for Media Engagement. https://mediaengagement.org/research/online-polls-and-quizzes