Gobo: A System for Exploring User Control of Invisible Algorithms in Social Media

Rahul Bhargava**
MIT Media Lab
rahulb@media.mit.edu

Anna Chung*
MIT Media Lab
 awchung@mit.edu

Snehalkumar `Neil' S. Gaikwad**
MIT Media Lab
gaikwad@mit.edu

Alexis Hope*
MIT Media Lab
ahope@mit.edu

Dennis Jen*
MIT Media Lab
dsjen@mit.edu

Jasmin Rubinovitz*
MIT Media Lab
jasrub@gmail.com

Belén Saldías-Fuentes*
MIT Media Lab
belen@mit.edu

Ethan Zuckerman**
MIT Media Lab
ethanz@mit.edu

*All authors contributed equally to this research. Authors are listed in alphabetical order.
** Correspondence authors.

ABSTRACT

In recent years, there has been an unprecedented growth in content that is shared and presented on social media platforms. Along with this growth, however, there is an increasing concern over the lack of control social media users have on the content they are shown by invisible algorithms. In this paper, we introduce Gobo, an open-source social media browser system that enables users to manage and filter content from multiple platforms on their own. Gobo aims to help users control what's hidden from their feeds, add perspectives from outside their network to help them break filter bubbles, and explore why they see certain content on their feed. Through an iterative design process, we've built and deployed Gobo in the wild and conducted a pilot study in the form of a survey to understand how the users respond to the shift of control from invisible algorithms to themselves. Our initial findings suggest that Gobo has potential to provide an alternate design space to enhance control, transparency, and explainability in social media.

INTRODUCTION

Invisible filtering algorithms control what hundreds of millions of users see or don't see on social media platforms [2, 3, 12]. The algorithms power content organization mechanisms, such as news feeds, and play an important role in how users perceive and share information about politics, culture, sports, and other topics. Despite their computational efficiency, there have been several concerns over the black-box nature of invisible algorithms.

Percentage of users who addressed control, transparency, or explainability

—— % Biggest complaints of social media   —— % What drove users to try Gobo   —— % How Gobo helped its users
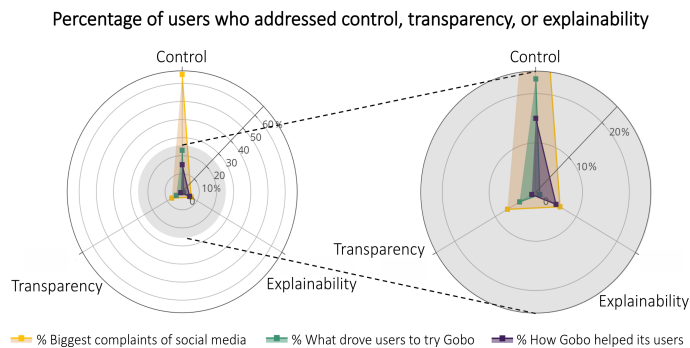
Figure 1: Users' feedback regarding Gobo's design dimensions. The left-side figure shows that about 65% of users' biggest complaints about social media feeds are related to a lack of control. The right-side figure is a zoomed-in view of the left and shows that users are potentially interested in trying tools that provide more control over their feed.

What users' think about social media platforms
(sample responses from the pilot study)

- "**Lack of control** over what I see; content showing up in my timeline hours or days after they were initially posted; not knowing when content was posted and where it originated."

- "**Lack of democratic participation** in order to define rules, decide about new design and assigning roles within the platform/community. poor privacy, data lock-in effects."

- "I **hate how they censor** things and am dying to find a better platform that does not support censorship. I quit Pinterest because of this. [...]"

- "It's an **echo chamber**. You have to work to see opposing views."

- "Resignation that **the control** always seems to be too much or too little."

Firstly, the algorithms heavily depend on digital footprints of users' activities to filter relevant news or posts from their social networks. This often leads to filter bubbles where the users don't see the filtered content that might provide them with an alternate perspective on a topic [1, 8, 12]. In fact, many social media users aren't even aware that the algorithms shape their use of the service [9]. Secondly, the algorithms have exhibited biases and have led to discriminatory outcomes against marginalized voices. For example, an algorithm behind YouTube's "restricted mode" had filtered out non-harmful LGBTQ content [5, 15]. Thirdly, the algorithms fail to explain how content is filtered and presented to the users. The lack of explanations lead to a breakdown in trust in social media platforms as shared by a participant from a prior study [9]--- *"After the study, I stopped using Facebook because I felt the way the Feed items were curated had, in some ways, broken the expectations between myself and Facebook [...] By neither showing me everything nor making their actions explicit, I felt like I was being lied to."*

To date, there has been little agreement on how one should design for increasing transparency and explainability of invisible filtering algorithms and creating a shared mental model between the system and social media users. Although extensive research has been carried out on misinformation and filter bubbles [7, 12, 16, 17], no study or open-source system exists which systematically evaluates how control should be shared between algorithms and social media users. In this paper, we introduce Gobo, an open-source social media browser system that enables users to manage and filter content from multiple social media platforms on their own. We have been building and deploying Gobo in the wild for two years, using iterative design and a needfinding process [13]. Gobo aggregates distributed feeds from users' multiple social media accounts and lets the users decide which ones to see. Using Gobo's interface, the users can set filters or "rules"– seriousness, rudeness, virality, gender and brands– to hide posts from their feed. Gobo also introduces "filtering-in" design features along with "filtering-out" to provide a wide array of perspectives from media outlets that users don't follow (Fig. 2). Gobo presents an exciting opportunity to advance our understanding of how social media users respond to the shift of control from invisible algorithms to themselves.

In our pilot survey study, the responses of the users (n=137) indicate that Gobo's design interventions can potentially help address their needs regarding control, transparency, and explainability. Our initial findings suggest that the main reason users engaged with Gobo was to regain control. Upon using it, they also encountered more explainability about the posts than they expected.

We envision Gobo will be instrumental in informing and studying alternate design interventions for increasing transparency and accountability of invisible algorithms. Towards this vision, we have made Gobo open-source and accessible globally. To date, the platform has over 7,000 registered users.
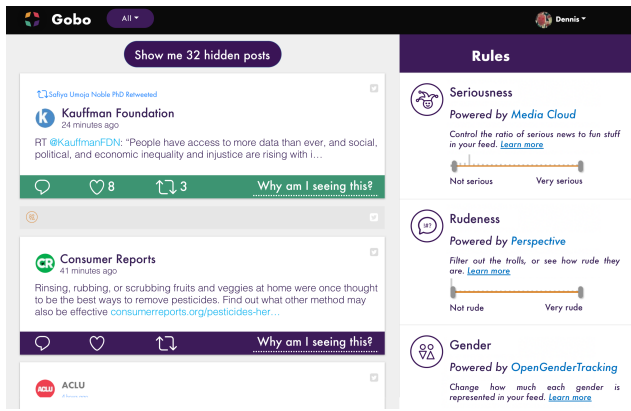
Figure 2: Gobo's feed interface indicates a user's posts from multiple social media accounts. The posts with a green bar indicate content that is added into a user's feed. Collapsed posts marked with a rule icon indicate content that is hidden from a user's feed. The side bar highlights different Rules users can set for "filtering-in" or "filtering-out".
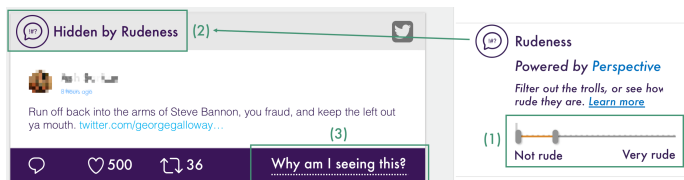

Figure 3: Gobo's Rule's interface enables users to adjust filters. For example, a user can control the Rudeness slider (1) to hide rude or not rude posts. The interface labels each hidden post with the rule it was hidden by (2). The interface explains to the user why a certain post is hidden or shown through the "Why am I seeing this?" option (3).

[1]https://www.perspectiveapi.com

GOBO: SYSTEM DESIGN

To accomplish the design goal of control, transparency, and explainability, we designed interface mechanisms for creating the shared mental model between Gobo and its users.

Control: Design for Increasing Human Autonomy

"Feeling of control" is one of the most important aspects in human-agent interaction [11]. As noted by one of the users: *"I want to be in charge of my social networks, and try to use algorithms in my favor, not against me."* Gobo provides users with meta and micro level controls. At the meta level, Gobo presents a centralized view of distributed information by allowing users to browse all of their social media feeds in one place. Existing social media platforms lack interoperability; users cannot move their feeds easily between multiple services. This distributed information often results in high cognitive overload on the users. To address this, Gobo regularly retrieves posts through platform-supplied APIs and processes each post through a set of semantic analysis tools and algorithms. The current version of the system is compatible with popular social media platforms, such as Twitter and Facebook. However, Facebook's API is particularly limited, as we can only access public page content.

At the micro level, Gobo combines human and machine insights to produce the feed. In contrast to existing social media platforms, it lets users control their feeds with a set of "rules." These rules can either help the users in "filtering-out" i.e. hiding content from the feed and "filtering-in" i.e. adding content from accounts they do not already follow. To enhance the usability, we have rendered the rules as toggles or sliders which a user can adjust to hide or add posts from their timeline (Fig 3).

The "filtering-in" design influences the user's feed by adding content. For example, the "politics" rule allows users to add content from social media accounts from the US political "left," "right," or "center." We curated accounts in each category based on Media Cloud's 2016 election report [4]. On the other hand, the "filtering-out" design influences the user's feed by hiding content. For example, the "rudeness" rule is powered by Perspective[1], which scores the toxicity of posts. The user can then adjust a slider to see "more rude" or "less rude" posts on their feed, and based on how Perspective's algorithm scores the posts, certain posts will be hidden. To understand how the users engage with rules and the effects of rules on the users' feed, Gobo's current rule toolbox provides an initial set of filters for Seriousness, Rudeness, Gender, Politics, Tech Perspectives, Obscurity, Brands, and 2020 US Election News. In future deployments, we plan to let users create their own custom rules.

Transparency and Explainability: Design for Creating Shared Mental Model to Create Trust

Transparency and explainability play an important role in establishing trust between a system and its users [14]. To increase transparency, Gobo's interface communicates which posts are added and which are hidden.
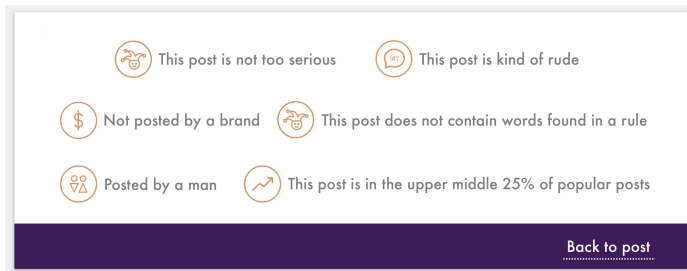
Figure 4: Gobo provides explanation when users click the "Why am I seeing this?" option. The interface shows how each rule has analyzed the post. In future deployments we plan to understand how users interpret different explanations.

GOBO, what's in the name?

Gobos are filters cut from sheets of metal and placed in front of a light to shine a particular pattern on a curtain or other surface. They are often used on a stage production to create patterns of light coming through a window or the leaves of a forest. In the social media context, we have created Gobo to let the users experiment with the effects of filters on their feeds. The system is available at: http://gobo.social

We have made Gobo **open-source and accessible** to researchers, practitioners, and communities across the world. Source code: https://github.com/mitmedialab/gobo

The interface visually differentiates added posts by a green bar (Fig. 2). Gobo labels them with the rule icons that caused the post to be hidden. Gobo presents two modes for viewing hidden posts. First, the interface collapses hidden posts and labels them with the rule icons that caused the post to be hidden. Second, Gobo provides an alternate view for users to see hidden posts. In this view, each hidden post is shown and labeled with the rule it was hidden by. While the users of popular platforms don't know what posts are hidden from them [10], Gobo gives users the option to explore all hidden posts to move towards a shared understanding.

Most users don't know why they see certain posts on their feeds [7]. To address this, Gobo presents a flippable card with an explanation at the back, describing how each rule interpreted the post and reasons why it was included in the feed. In addition to explaining how a post was interpreted by a rule, Gobo's interface describes the mechanism and tools used to power each rule (Fig. 4).

PRELIMINARY EVALUATION AND DISCUSSION

To understand how Gobo users respond to the shift of control from invisible algorithms to themselves, we conducted a pilot study in the form of a survey (n=137 with over 80% using social media for 6+ years) and coded the responses to identify themes. The survey was sent out to all registered Gobo users and included questions about experiences with social media and Gobo. Two researchers independently coded the responses with a broad range of categories and then grouped them in control, transparency, and explainability design dimensions. They coded issues with less prominence as others, which included being aware of one's own social media behavior. The strength of an inter-annotator agreement between the independent coders is good (Fleiss' κ=0.60) [6]. We resolved the disagreements through deliberation. To account for multiple labels for a comment, we report average percentages between both coders. Fig. 1 illustrates users' feedback regarding Gobo's design dimension. Our initial findings suggest that upon using Gobo users gained more explainability than they expected. One user reported that "*I do want more control and it seems possible to achieve this way […]*", whereas another user said "*It's made me more reflective about what is showing up and where it's coming from*". The thematic findings support the idea of defining the broader design space of transparent and explainable social media feeds where humans will have more control. Note that the scope of this study was limited in terms of population in the survey. More research is needed to better understand the effect of Gobo's design mechanisms and their implications.

In the upcoming deployments, we are experimenting with different interface techniques and measuring their effectiveness in increasing transparency and explainability on a large population. We envisage this work will help inform the future iterations of Gobo and designs for transparent and explainable social media feeds. We invite researchers and communities across the world to join us in this endeavor.

REFERENCES

[1] Engin Bozdag. Bias in algorithmic filtering and personalization. Ethics and information technology 15.3, 2013: 209-227.

[2] Nicholas Diakopoulos. Algorithmic accountability: Journalistic investigation of computational power structures. Digital journalism 3.3, 2015: 398-415.

[3] Nicholas Diakopoulos. Algorithmic accountability reporting: On the investigation of black boxes. (2014).

[4] Robert Faris Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election.", 2017.

[5] Megan Farokhmanesh, YouTube is still restricting and demonetizing LGBT videos — and adding anti-LGBT ads to some, The Verge 2018 https://www.theverge.com/2018/6/4/17424472/youtube-lgbt-demonetization-ads-algorithm URL Accessed 06/26/2019

[6] Joseph L. Fleiss and Jacob Cohen. "The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability." Educational and psychological measurement 33, no. 3 (1973): 613-619.

[7] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, Jonathan L. Zittrain, The science of fake news. Science 359, no. 6380 (2018): 1094-1096.

[8] Solomon Messing, and Sean J. Westwood. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. Communication research 41.8 (2014): 1042-1063.

[9] Eslami Motahhare, Aimee Rickman, Kristen Vaccaro, Amirhossein Aleyasen, Andy Vuong, Karrie Karahalios, Kevin Hamilton, and Christian Sandvig. I always assumed that I wasn't really that close to [her]: Reasoning about Invisible Algorithms in News Feeds. In Proceedings of the 33rd annual ACM conference on human factors in computing systems, pp. 153-162. ACM, 2015.

[10] Eslami Motahhare, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, and Alex Kirlik. First i like it, then i hide it: Folk theories of social feeds. In Proceedings of the 2016 cHI conference on human factors in computing systems, pp. 2371-2382. ACM, 2016.

[11] Donald A. Norman. How might people interact with agents. Communications of the ACM 37, no. 7 (1994): 68-71.

[12] Eli Pariser. The filter bubble: What the Internet is hiding from you. Penguin UK, 2011.

[13] Dev Patnaik and Robert Becker. Needfinding: the why and how of uncovering people's needs. Design Management Journal (Former Series) 10, no. 2 (1999): 37-43.

[14] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, 2016.

[15] Chase Ross Tweet. Do not let YouTube Get Away With This https://twitter.com/ChaseRoss/status/1001922360600137728/photo/1 URL Accessed 06/26/2019

[16] Nicholas Valentino, Antoine J. Banks, Vincent L. Hutchings, and Anne K. Davis. Selective exposure in the Internet age: The interaction between anxiety and information utility. Political Psychology 30, no. 4 (2009): 591-613.

[17] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. Science 359, no. 6380 (2018): 1146-1151.