Quantifying attention via dwell time and engagement in a social media browsing environment

Hause Lin*
Levene Graduate School of Business
University of Regina
Regina, CA S4S 0A2
hle952@uregina.ca

Ziv Epstein*
The Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02143
zive@mit.edu

Gordon Pennycook
Levene Graduate School of Business
University of Regina
Regina, CA S4S 0A2
gordon.pennycook@uregina.ca

David Rand
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02143
drand@mit.edu

Abstract

Modern computational systems have an unprecedented ability to detect, leverage and influence human attention. Prior work identified user engagement and dwell time as two key metrics of attention in digital environments, but these metrics have yet to be integrated into a unified model that can advance the theory and practice of digital attention. We draw on work from cognitive science, digital advertising, and AI to propose a two-stage model of attention for social media environments that disentangles engagement and dwell. In an online experiment, we show that attention operates differently in these two stages and find clear evidence of dissociation: when dwelling on posts (Stage 1), users attend more to sensational than credible content, but when deciding whether to engage with content (Stage 2), users attend more to credible than sensational content. These findings have implications for the design and development of computational systems that measure and model human attention, such as newsfeed algorithms on social media.

1 Introduction

In our current attention economy [Wu, 2017], digital ecosystems and social media environments are designed to grab and vie for users' attention. Social media platforms often leverage vulnerabilities in human psychology to distract users and exploit their attention [Lorenz-Spreen et al., 2020], leading many to argue that the ways digital platforms quantify and extract value from users' attention have led to a crisis in attention [Hwang, 2020, Wu, 2017].

However, it remains unclear how attention actually operates in digital ecosystems. Indeed, understanding how attention can be detected, modeled and influenced by computational systems is crucial for promoting better digital ecosystems. In this paper, we draw on research and methods from cognitive science, psychology, AI, and human-computer interaction to propose and validate a model for attention for digital ecosystems.

*These authors contributed equally.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

1.1 Measuring and quantifying attention online

Decades of attention research has led to many insights into how attention operates in many contexts [Buschman and Kastner, 2015, Simon, 1971]. Digital environments such as social media offer yet another context that requires new ways of measuring, quantifying, and understanding attention (see Lorenz-Spreen et al. [2020]). The digital advertising industry is among the first to systematically measure and quantify attention online [Hwang, 2020]. By standardizing user engagement and attention with metrics like the number of clicks and dwell time, it turned advertising into an online marketplace where user attention is commodified and traded via real-time bidding [Hwang, 2020, Wang et al., 2016], and these attention metrics have also been used to predict purchase intentions and behaviors of (even anonymous) website visitors [Mokryn et al., 2019].

Although researchers in different fields such as collaborative filtering and information retrieval have long recognized the value of quantifying attention and engagement via dwell time [Resnick et al., 1994], it is only in the last decade where work has been done to measure and model dwell time for different types of digital content [Lamba and Shah, 2019], and use it determine whether digital content is useful for and relevant to individual users Liu et al. [2011], Yi et al. [2014]. Surprisingly, little work has has been done to integrate dwell time and engagement data, which provide different yet complementary measures of attention. As far as we are aware, the only work is by Lagun and Lalmas [2016], who proposed a four-level taxonomy (bounce [<10s dwell time], shallow engagement, deep engagement, complete engagement [dwell and interact]). However, because this taxonomy does not offer details into whether and how attention operates differently across levels, it does not provide insights into how to optimize different types of attention.

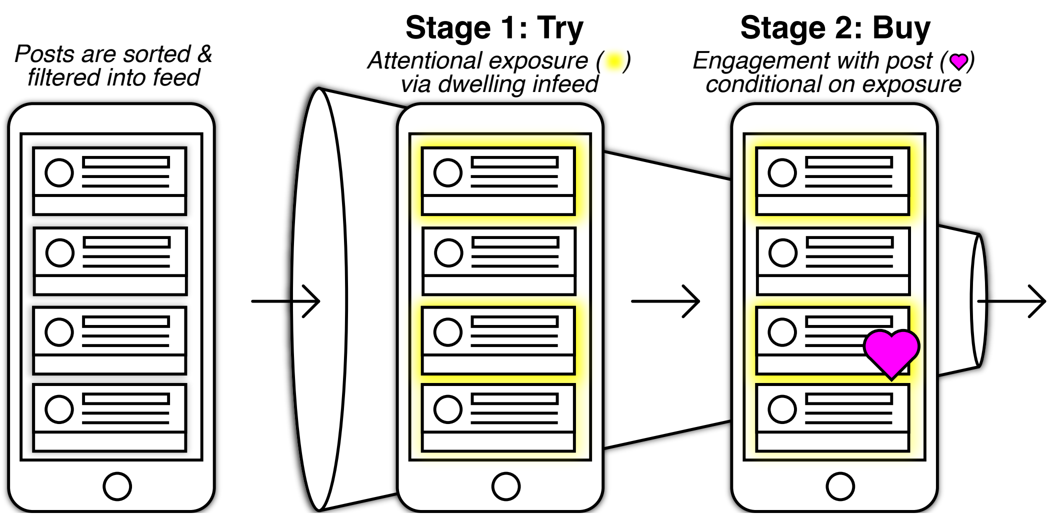

Figure 1: Two-stage model of attention in social media environments (Try + Buy).

1.2 Understanding and dissociating attention with a two-stage try-buy model

Here, we propose a two-stage model (Try + Buy) that integrates different ways of conceptualizing attention in social media environments (Figure 1). Users are initially exposed to content in an algorithmically-generated newsfeed (Stage 1), and then engage with content conditional on having been exposed to it (Stage 2). Crucially, our model jointly considers distinct attention dynamics at two different stages. First, the extent to which users attend to a piece of content reflects the amount of "trying" (Stage 1), which can be quantified continuously via dwell time. Second, engagement behavior such as sharing or liking content reflects "buying" (Stage 2).

Previous work focused largely on Stage 2—what causes people to engage with or "buy" content in digital ecosystems and what are the consequences [Chen et al., 2021, Salganik et al., 2006]. However, some work highlights how people have to first sample or "try" content before they decide whether to "buy" it [Krumme et al., 2012, Van Hentenryck et al., 2016], and crucially, different processes like social influence may operate differently in the "try" versus "buy" stages (see also Epstein et al. [2021] for a review). Crucially, unlike previous work that used two-stage models to primarily define and measure content quality [Wu et al., 2018, Abeliuk et al., 2017] and jointly predict "trying" and "buying" [Zhou et al., 2018], this paper focuses on dissociating the "try" and "buy" stages and examining how attention operates differently in these stages afforded by systems like social media environments.

An important implication of our "try-buy" model is that systems that optimize for dwell time (e.g., TikTok [Smith, 2021, Team, 2021]) versus engagement (e.g., Facebook [Oremus et al., 2021]) focus on different attention dynamics by design, which in turn may lead to different information environments. Thus, in this paper, we use dwell time and engagement data to provide insights into how attention operates in the "try" and "buy" stages, and reflect on how algorithm designers may optimize for these signals responsibly.

2 Methods

2.1 Dwell time for posts on social media feed

We recruited a convenience sample of Americans (N=644), of which 628 completed the survey on desktop computers (n=483) or mobile devices (n=145), using the recruitment platform Prolific. (compensation: $9 USD/hour; total amount: approx. $1700). Our participants had mean age of 35.7 (46.5% female, 66% white). At the start of the survey, participants provided informed consent and were routed to Yourfeed, a website we designed that displays content in a scrolling feed layout [Epstein and Lin, 2022]. The user interface mirrors the appearance of commonly used social media sites, such as Facebook or Twitter. Participants saw a modal that said "Thank you for participating! Next, you will see a social media newsfeed, configured just for you. Please browse this newsfeed like you usually would for social media. For each post, indicate whether you would consider sharing it with your network."

Crucially, this platform measured dwell time by considering how much time a participant spent on each post, which was determined based on how much each post was in the visible area of the browser window (also known as viewport time; [Lagun and Lalmas, 2016]). Occasionally, two posts could be fully visible in the browser window—in such cases, we assumed participants were viewing both because it was impossible to determine exactly which post they were looking at. This design detail reflects a trade-off between internal and external validity [Lin et al., 2021], and our platform was designed specifically to mimic the user interfaces of existing social media platforms like Twitter and Facebook, which often have two or more posts fully in view.

The website displayed 120 actual and recent social media posts to each participant in a scrollable feed, and participants could click to share or/and like any post (these behaviors were hypothetical engagement decisions, and did not affect what other participants saw). Of the 120 posts shown, 90 were randomly sampled from a set of 200 political and non-political news items [Epstein et al., 2022, Pennycook et al., 2021], half of which are true and half false. The other 30 were randomly sampled from a set of 76 opinion and mundane news items. The mundane posts were sourced from tabloid sites (e.g. The Sun, Daily Mail) and opinion posts are opinion pieces from reputable sources (e.g., New York Times Economist). All posts contained both an image and text (see Appendix A for example posts).

2.2 Feature ratings for each post

In addition to the task described above, we also conducted a separate rating survey to obtain out-of-sample post-level features for each of the 276 posts used. We recruited participants (N=1248) from the recruitment site Lucid to rate these posts (compensation: $9 USD/hour; total amount: approx. $1800), and included in our analyses only participants (N=872) who passed two attention checks. After providing informed consent, participants rated 40 randomly selected posts (of 276) on one of eight dimensions: 1) If you were to see the above article on social media, how likely would you be to share it?, 2) Are you familiar with the above headline (have you seen or heard about it before)?, 3) What is the likelihood that the above headline is true?, 4) Assuming the above headline is entirely accurate, how favorable would it be to Democrats versus Republicans?, 5) How provocative/sensational is this headline?, 6) How informative is this headline?, 7) How surprising is this headline?, and 8) How impactful is this headline?

We computed the mean across participants to compute a single estimate for each post feature (an average of 15.06 ratings per post per feature). We then use these ratings below to examine how they might be associated with the two stages of the try-buy model. Note that true posts were rated as significantly more true than false headlines ($b = 1.11$, $p < 0.001$) and were more likely to be shared by participants [Epstein et al., 2022]. Moreover, posts participants indicated they were more likely to share in this rating survey were also posts that were shared more frequently by participants in the actual experiment described above ($r = 0.26$, $p < 0.001$).

2.3 Dwell time preprocessing

Following Lin et al. [2022], we excluded posts whereby dwell times were longer than 30s. We then excluded the first three and last three posts because dwell times could not be determined precisely when participants were reading the instructions at the start or deciding whether to submit and proceed to the next phase of the study at the end of feed.

Because scrollable social media feeds introduce dependencies between dwell time and engagement (i.e., to engage with a post, people have to slow down and click the share/like buttons; but when they do not want to engage with a post, they do not have to slow down or click any button), we dissociated the motor and attentional components that contribute to dwell time. To do so, we fitted a Bayesian hierarchical mixed-effects model to predict dwell time as a function of the number of times participants engaged with any given post. The participant-specific coefficients provided an estimate of the time it takes for each participant to engage once with a post ("movement time"), and we adjusted dwell times by subtracting "movement time" to eliminate the motor component of dwell time, which should leave us with primarily the attentional component. After which, we excluded posts with dwell times shorter than 0.15s because it is unlikely that participants could attend to and evaluate post features so rapidly [Lin et al., 2022]. Data and code to reproduce the experimental results can be found here.

3 Results

3.1 Post features and dwell correlations

To investigate whether dwell times correlated with the eight post features, we computed the mean dwell times (across participants) for each post and correlated them with post features. As shown in Figure 3, the features correlated with each other, and three correlated significantly with dwell (e.g., surprising and true posts had longer and shorter dwells, respectively). Crucially, these correlations suggest dwell captures attention exposure and dynamics and the amount of "trying."

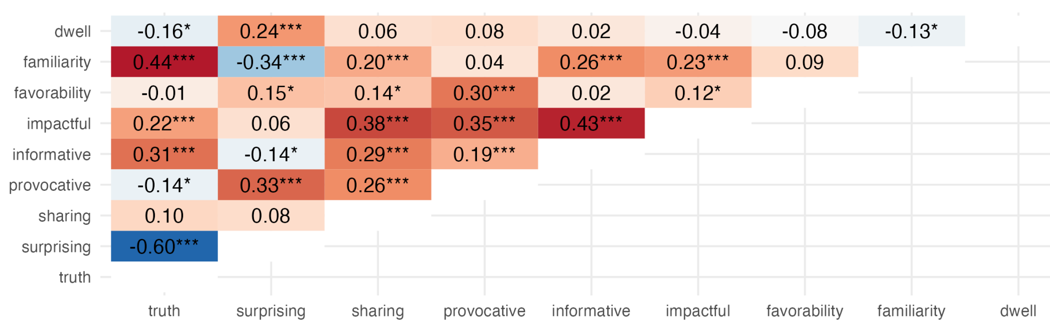

Figure 2: Correlations between post features and dwell.

We then performed principal component analysis (PCA), which revealed that the first two components, together, explained more than half the variance in the data (PC1: 29%; PC2: 25%). As shown in Table 1, relative to PC1 which has large positive weights for the truth, informative, and familiarity features, PC2 has a negative weight for the truth feature and large positive weights for the provocative and surprising features. Thus, PC1 seems to capture variance related to "credibility," whereas PC1 captures variance related to "sensationalism " (see Appendix A for the top posts for each component).

Table 1: PCA component weights and variance explained

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
familiarity	0.43	-0.17	0.28	0.26	0.68	-0.27	0.32	-0.02
favorability	0.12	0.31	0.81	-0.06	-0.42	-0.04	0.20	-0.09
impactful	0.43	0.28	-0.24	-0.20	0.05	0.66	0.36	-0.27
informative	0.45	0.07	-0.35	-0.40	-0.29	-0.63	0.15	-0.00
provocative	0.17	0.52	0.12	-0.29	0.38	-0.03	-0.67	0.01
sharing	0.35	0.28	-0.19	0.79	-0.28	-0.04	-0.23	-0.00
surprising	-0.26	0.54	-0.13	0.05	0.14	-0.06	0.41	0.66
truth	0.44	-0.37	0.13	-0.12	-0.16	0.30	-0.19	0.70
variance	0.29	0.25	0.12	0.09	0.08	0.07	0.06	0.04
cumulative variance	0.29	0.54	0.66	0.75	0.83	0.90	0.96	1.00

As with the dwell-feature correlations (Figure 3), PC1 (the "credibility" component) correlates negatively (marginally significant) with dwell (r = -0.11, p = 0.063), but PC2 (the "sensationalism" component) correlates positively with dwell (r = 0.17 p = 0.005). Together, these correlations suggest that more sensational posts were associated with *more* "trying," but more credible posts were associated with *less* "trying." For a breakdown of the relationships between the component scores and dwell times separately posts that had been engaged with or not, see Figure 3. Given these results, we focus on the two PCA components (instead of the 8 features) in the analyses that follow.

Figure 3: Relationships between principal component scores and dwell for posts that had been engaged with or not. Each dot is one post.

3.2 Evaluating the two-stage model with dwell and engagement analyses

Next, we examined what influenced the extent to which participants "tried" each post by fitting a fixed-effects linear regression to predict dwell time (Table 2). For posts participants had engaged with (i.e., shared and/or liked), dwell time was longer (b = 0.31, p < .001). Participants also dwelled longer on posts with higher PC1 ("sensationalism" component) scores (b = 0.04, p < .001), but less on posts with higher PC2 ("credibility" component) scores (b = -0.02, p = 0.017). There was also an interaction effect, such participants dwelled even longer on sensational posts they engaged with (b = 0.05, p < .001). Thus, consistent with the results in the previous section, we find that whether participants had engaged with a post and the post's credibility and sensationalism influenced how much participants "tried" each post.

Table 2: Fixed-effect regression predicting log(dwell) as a function of engagement (no: -0.5, yes: 0.5), credibility (PC1 z-scored), and sensationalism (PC2 z-scored)

	Estimate	SE	t value	Pr(>|t|)
engage	0.311	0.025	12.348	0.000
credibility	-0.017	0.007	-2.405	0.017
sensationalism	0.038	0.008	4.729	0.000
engage:credibility	0.010	0.011	0.901	0.368
engage:sensationalism	0.048	0.013	3.694	0.000

Having shown that dwell time captures attentional exposure and dynamics and that it serves as a measure of "trying," we turn to the question of what features are associated with decisions to "buy" by fitting a fixed-effect logistic regression to predict engagement (i.e., yes: 1, no: 0). As shown in Table 3, longer dwell times were associated with an increased probability of engagement ($b = 0.36$, $p < .001$). That is, the more participants "tried," the more likely they were to "buy."

However, in contrast to the model predicting dwell times (i.e., extent of "trying") above (Table 2), participants were more likely to engage with credible posts ($b = 0.21$, $p < .001$), but less likely to engage with sensational posts ($b = -0.22$, $p < .001$). Moreover, dwell interacted with sensationalism, such that sensational posts with longer dwell times were more likely to be engaged with ($b = 0.06$, $p = 0.003$). In other words, participants were more and less likely to "buy" credible and sensational posts, respectively. But when they dwelled longer on sensational posts, they were also more likely to then engage with these posts.

Table 3: Fixed-effect logistic regression predicting engagement as a function of log(dwell) (z-scored), credibility (PC1 z-scored), and sensationalism (PC2 z-scored)

	Estimate	SE	t value	Pr(>|t|)
dwell	0.355	0.029	12.355	0.000
credibility	0.212	0.049	4.361	0.000
sensationalism	-0.221	0.047	-4.711	0.000
dwell:credibility	0.011	0.020	0.538	0.590
dwell:sensationalism	0.062	0.021	2.921	0.003

4 Discussion

In this paper, we introduce a two-stage model of attention to conceptualize and understand how attention operates in social media environments. Using an analytic approach informed by our try-buy model, we find dissociations between the "try" and "buy" stages: in the "try" stage, attention (dwell) was focused on sensational posts and not credible posts. Conversely, in the "buy" stage, attention (engagement) was focused on credible posts and not sensational posts. However, our experiments used data about hypothetical engagement. While past work has shown that self-reported news sharing in surveys correlates with actual sharing on Twitter [Mosleh et al., 2020], future work should replicate these findings with actual engagement data from social media.

Nevertheless, our model and results have important implications for how attention is modeled and leveraged by AI systems in human-computer interactions: For one, algorithmic systems that explicitly optimize for dwell time may prioritize sensational content over credible content and therefore inadvertently proliferate misinformation. Conversely, while optimizing for engagement may indeed surface credible content, we found that people were more likely to engage with sensational content after dwelling more on them, which could create a positive feedback loops that drives the spread of misinformation [Hao, 2021]. Future work is needed to apply our findings to the adaptive dynamics of optimized newsfeed algorithms, and how to align such algorithms with human values by more rigorously evaluating optimization metrics [Dmitriev and Wu, 2016], learning complex multi-variate objectives from stakeholders [Stray et al., 2021] and directly giving users control of the algorithms instead of trying to infer their desires [Ekstrand and Willemsen, 2016, Bhargava et al., 2019].

References

Tim Wu. *The attention merchants: The epic scramble to get inside our heads*. Vintage, 2017.

P Lorenz-Spreen, S Lewandowsky, CR Sunstein, and R Hertwig. How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature Human Behaviour*, 4(11): 1102–1109, 2020.

Tim Hwang. *Subprime attention crisis: Advertising and the time bomb at the heart of the Internet*. FSG originals, 2020.

Timothy J Buschman and Sabine Kastner. From behavior to neural dynamics: An integrated theory of attention. *Neuron*, 88(1):127–144, 2015.

Herbert A Simon. Designing organizations for an information-rich world. In *Computers, Communications, and the Public Interest*, pages 37–52. Johns Hopkins University Press, Baltimore, MD, 1971.

Jun Wang, Weinan Zhang, and Shuai Yuan. Display advertising with real-time bidding (rtb) and behavioural targeting. *arXiv*, page 1610.03013v2, 2016.

Osnat Mokryn, Veronika Bogina, and Tsvi Kuflik. Will this session end with a purchase? inferring current purchase intent of anonymous visitors. *Electronic Commerce Research and Applications*, 34:100836, 2019.

Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186, 1994.

Hemank Lamba and Neil Shah. Modeling dwell time engagement on visual multimedia. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1104–1113, 2019.

Chang Liu, Jingjing Liu, Nicholas Belkin, Michael Cole, and Jacek Gwizdka. Using dwell time as an implicit measure of usefulness in different task types. *Proceedings of the American Society for Information Science and Technology*, 48(1):1–4, 2011.

Xing Yi, Liangjie Hong, Erheng Zhong, Nanthan Nan Liu, and Suju Rajan. Beyond clicks: dwell time for personalization. In *Proceedings of the 8th ACM Conference on Recommender systems*, pages 113–120, 2014.

Dmitry Lagun and Mounia Lalmas. Understanding user attention and engagement in online news reading. In *Proceedings of the ninth ACM international conference on web search and data mining*, pages 113–122, 2016.

Cathy Xi Chen, Gordon Pennycook, and David Rand. What makes news sharable on social media. *PsyArXiv*, 2021.

Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856, 2006.

Coco Krumme, Manuel Cebrian, Galen Pickard, and Sandy Pentland. Quantifying social influence in an online cultural market. *PloS one*, 7(5):e33785, 2012.

Pascal Van Hentenryck, Andrés Abeliuk, Franco Berbeglia, Felipe Maldonado, and Gerardo Berbeglia. Aligning popularity and quality in online cultural markets. *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)*, 110(1):398–407, 2016.

Ziv Epstein, Matthew Groh, Abhimanyu Dubey, and Alex Pentland. Social influence leads to the formation of diverse local trends. *Proceedings of the ACM on Human-Computer Interaction*, 5 (CSCW2):1–18, 2021.

Siqi Wu, Marian-Andrei Rizoiu, and Lexing Xie. Beyond views: Measuring and predicting engagement in online videos. In *Twelfth international AAAI conference on web and social media*, 2018.

Andrés Abeliuk, Gerardo Berbeglia, Pascal Van Hentenryck, Tad Hogg, and Kristina Lerman. Taming the unpredictability of cultural markets with social influence. In *Proceedings of the 26th international conference on world wide web*, pages 745–754, 2017.

Tengfei Zhou, Hui Qian, Zebang Shen, Chao Zhang, Chengwei Wang, Shichen Liu, and Wenwu Ou. Jump: A joint predictor for user click and dwell time. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence. AAAI Press*, pages 3704–3710, 2018.

Ben Smith. How tiktok reads your mind. Web, December 2021. URL `https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html`.

WSJ Team. Inside tiktok's algorithm: A wsj video investigation. Web, July 2021. URL `https://www.wsj.com/articles/tiktok-algorithm-video-investigation-11626877477`.

Will Oremus, Chris Alcantara, Jeremy Merrill, and Artur Galocha. How facebook shapes your feed: The evolution of what posts get top billing on users' news feeds, and what gets obscured. Web, October 2021. URL `https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works/`.

Ziv Epstein and Hause Lin. Yourfeed: Towards open science and interoperable systems for social media. *arXiv preprint arXiv:2207.07478*, 2022.

H Lin, KM Werner, and M Inzlicht. Promises and perils of experimentation: The mutual-internal-validity problem. *Perspectives on Psychological Science*, 16(4):854–863, 2021.

Ziv Epstein, Hause Lin, Gordon Pennycook, and David Rand. How many others have shared this? experimentally investigating the effects of social cues on engagement, misinformation, and unpredictability on social media. *arXiv preprint arXiv:2207.07562*, 2022.

Gordon Pennycook, Jabin Binnendyk, Christie Newton, and David G Rand. A practical guide to doing behavioral research on fake news and misinformation. *Collabra: Psychology*, 7(1):1–13, 2021.

Hause Lin, Gordon Pennycook, and David Rand. Thinking more or thinking differently? using drift-diffusion modeling to illuminate why accuracy prompts decrease misinformation sharing. *PsyArXiv*, 2022.

Mohsen Mosleh, Gordon Pennycook, and David G Rand. Self-reported willingness to share political news articles in online surveys correlates with actual sharing on twitter. *Plos one*, 15(2):e0228882, 2020.

Karen Hao. How facebook got addicted to spreading misinformation. *MIT Technology Review*, 2021.

Pavel Dmitriev and Xian Wu. Measuring metrics. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 429–437, 2016.

Jonathan Stray, Ivan Vendrov, Jeremy Nixon, Steven Adler, and Dylan Hadfield-Menell. What are you optimizing for? aligning recommender systems with human values. *arXiv preprint arXiv:2107.10939*, 2021.

Michael D Ekstrand and Martijn C Willemsen. Behaviorism is not enough: better recommendations through listening to users. In *Proceedings of the 10th ACM conference on recommender systems*, pages 221–224, 2016.

Rahul Bhargava, Anna Chung, Neil S Gaikwad, Alexis Hope, Dennis Jen, Jasmin Rubinovitz, Belén Saldías-Fuentes, and Ethan Zuckerman. Gobo: A system for exploring user control of invisible algorithms in social media. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, pages 151–155, 2019.

A Appendix

Interpreting the PCA components

We checked the posts qualitatively (Figure 4), which corroborated our interpretations of PC1 and PC2. For example, the top two PC1 posts were from reputable mainstream new sources: "New York City Mandates Vaccines for Its Workers to ‘End the COVID Era’" (New York Times) and "President Biden’s oil price two-step won’t lower your gas prices" (Washington Post). However, the top two PC2 posts were from unreliable or fake news sources: "Democrats Introduce Bill To ‘Euthanize Seniors’ To Save Social Security" (Daily World Update) and "920 Women Lose Their Unborn Babies After Getting Vaccinated" (The True Defender).

Figure 4: Top headlines for PC1 (left) and PC2 (right).