Screenshot Journey Auditor: A Tool to Support Analysis of Smartphone Media Consumption Journey Using Screenshot Data

Chen-Chin Lin
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
spencer0929.cs04@nctu.edu.tw

Jian-Hua Jiang Chen
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
novel860329@gmail.com

Rebecca Ping Yu
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
rpyu@nycu.edu.tw

Wan-Yun Yu
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
mwyu@nycu.edu.tw

Yung-Ju Chang
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
armuro@nycu.edu.tw

ABSTRACT

Taking long series of screenshots for capturing and studying smartphone users' phone usage and media consumption has recently attracted research attention due to its advantage of capturing rich contextual information from users' phone use journeys. However, that approach creates a high volume of screenshots that take very considerable time and effort to inspect and annotate, especially when the granularity of analysis is low: such as when distinguishing among media-content units (e.g, single FB posts) and detecting events in them. We therefore developed *Screenshot Journey Auditor (SJA)*, a web application that identifies individual social media posts, and detects news items and other events of interest in them. It then visualizes users' journeys 'flow' among these media-content units. SJA also enables researchers/coders to collaboratively correct detections online. We evaluated SJA with five coders and received positive feedback on how the detections and visualizations made the analysis process more efficient and informative.

CCS CONCEPTS

• **Human-centered computing** → **Collaborative and social computing**.

KEYWORDS

Screenshot; Mobile phone; News consumption; Analysis tool; collaboration

ACM Reference Format:

Chen-Chin Lin, Jian-Hua Jiang Chen, Rebecca Ping Yu, Wan-Yun Yu, and Yung-Ju Chang. 2022. Screenshot Journey Auditor: A Tool to Support Analysis of Smartphone Media Consumption Journey Using Screenshot Data. In *Companion Computer Supported Cooperative Work and Social Computing (CSCW'22 Companion), November 8–22, 2022, Virtual Event, Taiwan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3500868.3559456

CSCW'22 Companion, November 8–22, 2022, Virtual Event, Taiwan

ACM ISBN 978-1-4503-9190-0/22/11.
https://doi.org/10.1145/3500868.3559456

1 INTRODUCTION

As mobile devices become more pervasive, media consumption has gradually shifted to them from desktops [14]. Using smartphones, people may access a huge variety of media content, prominently including news in various formats [4, 6, 7, 11]. How smartphone users discover and encounter news via various channels has attracted considerable interest [13] from researchers using a variety of methods, including surveys [23, 24], browser-history extraction [1, 9], app-log collection [5, 6], and interviews [2, 12]. However, these methods are either retrospective (i.e., surveys, interviews) and therefore subject to recall error and reconstruction bias [10, 22], or else strictly limited in terms of how much contextual information they can capture, making interpretation difficult [20]. In addition, people have an increasing variety of channels whereby they can access and consume news on their phones, one of which comprises social-media platforms [21]. However, neither browser histories nor app logs can capture news consumption across multiple apps. Therefore, neither approach allows researchers to gain a full picture of phone users' media consumption.

To address these limitations, many studies have taken device screenshots at regular intervals to capture in-situ and cross-app behaviors [18] that can then be reconstructed [19]. For example, Hu and Lee [8] developed a desktop application, ScreenTrack, that captures its users' software, document, and web-page use based on screenshots. Similarly, Brinberg et al. [3] collected screenshots every five seconds and used them to analyze temporal, textual, graphical, and topical features of what appears on people's screens, with the wider aim of describing the idiosyncratic nature of individuals' daily digital lives. However, a grand challenge to this method is the amount of time and effort it takes to inspect and code series of thousands of screenshots manually, especially when the research team is interested in capturing a number of different events. Moreover, when the granularity of these events is low (e.g., checking whether a user opens an external link in a post), inspecting nearly every screenshot is – though necessary – highly burdensome and tedious. Randomly sampling only a portion of collected screenshot data is a conceivable means of reducing the time and effort for data analysis, while avoiding potential systematic bias. However, adopting it would mean sacrificing data, and probably, failing to discern some patterns that would only be prominent if one inspected a series of similar screenshots rather than unique individual ones. Using machine-learning techniques to detect events of interest is a promising approach to reducing inspection effort. However, if a pre-trained model does not exist, researchers will have to build a training dataset themselves, and this again makes laborious data coding unavoidable. In addition, a human-in-the-loop mechanism and interface would be necessary to correct false detections before the model is mature.

In sum, analyzing screenshots to understand media consumption is appealing and may provide researchers with fresh insights that may not identified in prior research. However, the challenges and costs of analyzing such large datasets are likely to make them hesitant to adopt this approach.

To help address this problem, we conducted a case study of news consumption on Facebook, based on 648,726 screenshots from 88 different Facebook users' smartphones, and used its outcomes as the basis for our development of Screenshot Journey Auditor (SJA). SJA is a system that incorporates a *media-engagement recognizer (MER)* that uses machine learning to automatically separates media units (e.g., individual FB posts); determines the unit of the user's main focus (i.e., the post they are currently reading/watching); classifies that unit as a news article or not; and detects if specific events involving that unit (e.g., opening an external link, viewing comments) occur. SJA then incorporates a media journey visualizer that visualizes the outcome of the detection, i.e., displays the user's journey among media-content units, with the occurrence of the aforementioned events highlighted via a web interface. SJA also allows researchers and coders to collaboratively audit such user journeys, inspecting for and correcting false detections and feeding those findings back into the dataset.

For our case study, we engaged six coders who passed Institutional Review Board training and thus became official team members to regularly use SJA to audit our study participants' news-consumption journeys. We also recruited five graduate-student evaluators to use SJA to code the screenshots generated by the team members. The feedback from the five evaluators indicated that the system significantly reduced the labor that would otherwise have been required to inspect the screenshots and find/code the specified events. SJA also helped them establish a clear picture of the context of each news-consumption instance, and helped them to easily make sense of those instances. Below, we introduce the features of the system and its implementation.

2 SCREENSHOT JOURNEY AUDITOR

SJA's main role in screenshot analysis is to allow researchers to efficiently spot media-engagement instances and other events of interest, such as viewing comments or clicking an external link. To fulfill this role, it needs to identify such instances among screenshots. Next, it needs to present these instances to researchers in a form that allows them to correct the detection outcomes, such that the model behind each recognizer can then be improved iteratively. The following three subsections explain each of these three processes and their required components in detail.

2.1 Media-engagement Recognizer

The purpose of our MER is to identify instances of users engaging with specific media units by both consuming information and taking actions based on it. In our case study of news consumption on Facebook, MER needed to identify individual posts within news feeds, and any actions associated with each of them, based on screenshots.

For each screenshot, it first detects boundaries between posts, and then generates a list of individual posts in that screenshot. Next, it extracts the text of each individual post. Boundary and text information are both crucial to determining whether the user dwelled in the same post or moved to the next, since they reflect whether old text has disappeared and new text has appeared, and whether the user has been scrolling upward or downward. Boundary information is also used to determine which post a user was attending to at the time the screenshot was taken.

To detect boundaries, we build a model based on YOLO[1], a real-time object-detection system. To extract text, we used optical character recognition (OCR)[2] provided by Google Cloud Vision API[3]. Our model also allowed us to detect whether users opened a new link or were viewing comments. To determine whether a post was a news item, MER detected whether the text extracted from the post contained a source name that matched a news-sources whitelist, which we established based on news items that had appeared on Google News[4], Yahoo News[5] and OpView[6] that contained the majority of news outlets and sources in Taiwan.

2.2 Media-journey Visualizer

A second key component of SJA is our media-journey visualizer (MJV), which allows researchers and coders to view and audit participants' 'travel' among the media-content units captured in screenshots. Each such journey is represented as a chronological series of color bars, with earlier events on the left and later ones on the right, with each bar representing one screenshot (Fig. 1). The color of the bar changes when the attended media unit (e.g., a Facebook post), defined as the unit that is the main focus of the relevant screenshot, changes. A particular media unit is deemed to be the main focus of a screenshot if it occupies the largest proportion of an *attention area* in that screenshot. An attention area is defined as the span of effective vision on the phone screen, based on previous psychological research on human vision and attention using eye movement methodology [16, 17]. We elected to use attention area instead of the full screen for calculating the proportion because a mobile eye-tracking study has shown that people's attention on their phones is not on the full screen but rather within this area and on one unit at a time [15]. To determine the specific region of the attention area on smartphones, we used a visual-angle calculator

[1]https://pjreddie.com/darknet/yolo/
[2]https://cloud.google.com/vision/docs/ocr
[3]https://cloud.google.com/vision
[4]https://news.google.com.tw/
[5]https://tw.news.yahoo.com/
[6]OpView is an online data service that provides access to online news media, public Facebook, pages, YouTube, and other online media data (e.g., PTT—the largest bulletin board system in Taiwan, blogs, and forums). This is offered by eLand, a Taiwanese consumer insights company (https://www.opview.com.tw/product-insight).

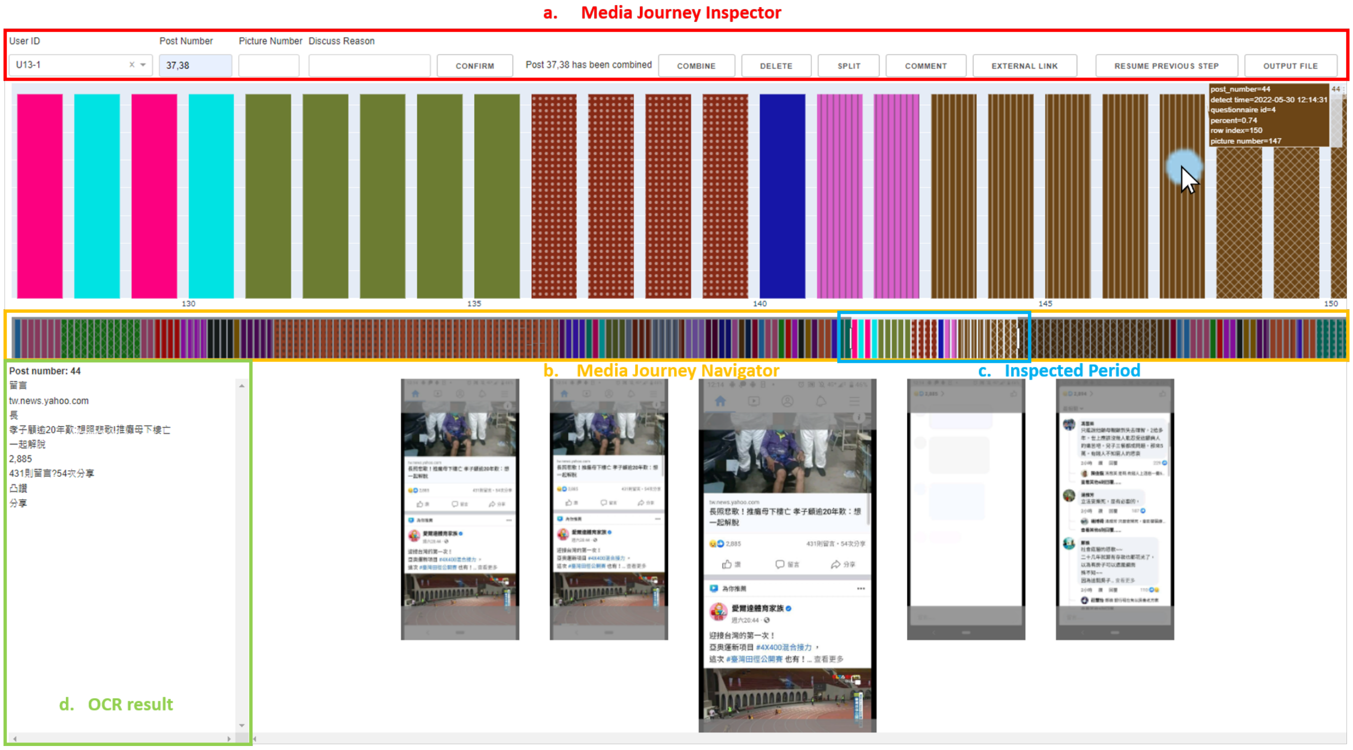

Figure 1: The Screenshot Journey Auditor (SJA) contains a media-engagement recognizer (MER) that detects users' engagement with news items, and a media-journey visualizer (MJV) that visualizes the detection outcomes. Different posts' bars are filled with various colors and patterns so that researchers can quickly see how the sampled user's main focus transitioned between posts. An entire journey is presented in the middle of the page, allowing the user to zoom in to a specific periond within the journey. A journey-auditing interface (JAI) is provided to enable coders and researchers to correct or adjust the detection outcomes. It should be noted that the screenshot in this figure is not from any real participant, to protect the participants' privacy.

developed for SR research, [7] which established that the attention area occupied the central 74% of the screen (measured vertically).

Given that each journey contains a large number of screenshots, MJV visualized events that were detected by MER to help the coders and researchers inspect events of interest. In our case study of news consumption on Facebook, we were especially interested in capturing three characteristics of events: 1) whether the post was a news item (filled with parallel lines in Fig. 1), 2) whether the user opened an external link (filled with dots), and 3) whether the user was viewing comments (filled with crosshatching). When a researcher needs to inspect a specific screenshot carefully, hovering the cursor on a bar causes its corresponding screenshot and those of its neighbors (two before and two after it, respectively) to appear below it, along with a message window showing that screenshot's metadata. Also, the areas above and below the attention area are grayed out, to render the attention area more prominent. The user's entire media journey is presented in the middle of the page, allowing the coders and researchers to navigate along the journey to find specific regions of interest they want to zoom in on for more careful inspection.

[7]https://www.sr-research.com/visual-angle-calculator/

2.3 Journey-auditing Interface

The third component of SJA is its journey-auditing interface (JAI), which enables researchers and coders to correct detection errors if they find any when inspecting screenshots. They can combine, delete, split, comment, and mark events of interest on the bar. Whenever researchers and coders make changes, the system immediately updates the database so that all the other users of the system can also see such changes in real time. In addition, SJA allows researchers and coders to improve its model's training dataset, as a means of enhancing MER's accuracy both when separating posts and when identifying events of interest. Finally, the researchers and coders can click JAI's *Output file* button to download a results file for follow-up analysis.

2.4 System Implementation

We developed SJA in Python[8] and used the library Dash[9], which can embed graph plotting that utilizes Plotly[10] and add HTML components, for building the visualization interface. To allow coders to

[8]https://www.python.org/.
[9]https://dash.plotly.com/.
[10]https://plotly.com/python/.

use SJA remotely, we build it with Flask[11], a micro web-application framework that can be connected to by multiple people.

3 USER EVALUATION

We conducted a user study with five graduate-student evaluators, all of whom had prior experience of conducting HCI research, to ascertain whether, how, and how much SJA helped them to inspect and code media journeys. We gave them task scenarios in which they were coders who needed to code occurrences of smartphone users' news consumption on Facebook, the actions such users performed in these instances, and the manual of SJA which determined the patterns of events. We did not tell evaluators what events and how many of them occurred among the screenshots.

The tasks involved a media journey represented by 233 screenshots collected from the researchers' phones that we imitate a normal user's news consumption, which contained 15 events: 11 of news consumption, two of comment viewing, and two of opening an external link. They first performed the tasks without SJA (merely inspecting screenshots), and then with SJA, using the same set of screenshots. The purpose of this was to let them experience the differences between these two conditions, so that they could reflect on them in a debriefing session. All five evaluators gave very positive feedback about the convenience of SJA, and deemed that the system substantially reduced their time and burden. As C2 put it: *"It's pretty nice that I can easily understand which posts the user stays longer versus scrolls over quickly"*. The system's event markers feature was also praised: with C1 saying, *"It is convenient to have the dot and grid on the bar. I can dive deeper to figure out what kind of post the user was more likely to click on, or what kind of post they were interested in."* Additionally, three participants mentioned that SJA provided them with a holistic view of the entire journey, as well as detailed behavior in each screenshot. As C1 noted: *"It's not only a forest I can see, but also each of the trees."* However, the evaluators also mentioned some SJA weaknesses. For example, some perceived that the latency of loading screenshots was longer than expected, due to the large volume of screenshots being loaded. They also expressed a hope that the recognizers could be made more accurate to reduce their effort of correcting detection errors.

4 FUTURE WORK AND CONCLUSION

We have presented Screenshot Journey Auditor and demonstrated how it can reduce the time and effort expended by researchers and coders when inspecting and analyzing smartphone users' news consumption on Facebook as captured by screenshots. Our future work will include developing media-unit detection capability for a wider range of platforms (e.g., Instagram and YouTube); increasing the range of detectable types of events; improving detection accuracy; and enhancing the system's visualization quality and other aspects of its usability.

ACKNOWLEDGMENTS

This research was supported in part by the National Science and Technology Council, R.O.C (110-2221-E-A49 -164 -). We thank the help from Tsai-Wen Tsai, Chuah Zheng Xuan, Tzu-Liang Wang, Bryan Wang, Chia-Chen Wu.

[11] https://flask.palletsprojects.com/en/2.1.x/.

REFERENCES

[1] Frank Bentley, Katie Quehl, Jordan Wirfs-Brock, and Melissa Bica. 2019. Understanding Online News Behaviors. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300820

[2] Pablo J Boczkowski, Eugenia Mitchelstein, and Mora Matassi. 2018. "News comes across when I'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New Media & Society* 20, 10 (2018), 3523–3539. https://doi.org/10.1177/1461444817750396 arXiv:https://doi.org/10.1177/1461444817750396

[3] Miriam Brinberg, Nilam Ram, Xiao Yang, Mu-Jung Cho, S. Shyam Sundar, Thomas N. Robinson, and Byron Reeves. 2021. The idiosyncrasies of everyday digital lives: Using the Human Screenome Project to study user behavior on smartphones. *Computers in Human Behavior* 114 (2021), 106570. https://doi.org/10.1016/j.chb.2020.106570

[4] Marios Constantinides, John Dowell, David Johnson, and Sylvain Malacria. 2015. Exploring Mobile News Reading Interactions for News App Personalisation. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 457–462. https://doi.org/10.1145/2785830.2785860

[5] Kristin Van Damme, Cédric Courtois, Karel Verbrugge, and Lieven De Marez. 2015. What's APPening to news? A mixed-method audience-centred study on mobile news consumption. *Mobile Media & Communication* 3, 2 (2015), 196–213. https://doi.org/10.1177/2050157914557691 arXiv:https://doi.org/10.1177/2050157914557691

[6] Kristin Van Damme, Marijn Martens, Sarah Van Leuven, Mariek Vanden Abeele, and Lieven De Marez. 2020. Mapping the Mobile DNA of News. Understanding Incidental and Serendipitous Mobile News Consumption. *Digital Journalism* 8, 1 (2020), 49–68. https://doi.org/10.1080/21670811.2019.1655461 arXiv:https://doi.org/10.1080/21670811.2019.1655461

[7] Nir Grinberg. 2018. Identifying Modes of User Engagement with Online News and Their Relationship to Information Gain in Text. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1745–1754. https://doi.org/10.1145/3178876.3186180

[8] Donghan Hu and Sang Won Lee. 2020. *ScreenTrack: Using a Visual History of a Computer Screen to Retrieve Documents and Web Pages*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376753

[9] Martijn Kleppe and Marco Otte. 2017. Analysing and understanding news consumption patterns by tracking online user behaviour with a multimodal research design. *Digital Scholarship in the Humanities* 32 (06 2017), ii158–ii170. https://doi.org/10.1093/llc/fqx030

[10] Tetsuro Kobayashi and Jeffrey Boase. 2012. No Such Effect? The Implications of Measurement Error in Self-Report Measures of Mobile Communication Use. *Communication Methods and Measures* 6, 2 (2012), 126–143. https://doi.org/10.1080/19312458.2012.679243 arXiv:https://doi.org/10.1080/19312458.2012.679243

[11] Dmitry Lagun and Mounia Lalmas. 2016. Understanding User Attention and Engagement in Online News Reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* (San Francisco, California, USA) *(WSDM '16)*. Association for Computing Machinery, New York, NY, USA, 113–122. https://doi.org/10.1145/2835776.2835833

[12] Danielle Lottridge, Katie Quehl, Frank Bentley, Max Silverman, Melissa Ong, Michael Dickard, Brooke White, and Ethan Plaut. 2022. Ubiquitous News Experienced Alone: Interviews with Americans and Their Devices. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 128 (apr 2022), 29 pages. https://doi.org/10.1145/3512975

[13] Judith Möller, Robbert Nicolai van de Velde, Lisa Merten, and Cornelius Puschmann. 2020. Explaining Online News Engagement Based on Browsing Behavior: Creatures of Habit? *Social Science Computer Review* 38, 5 (2020), 616–632. https://doi.org/10.1177/0894439319828012 arXiv:https://doi.org/10.1177/0894439319828012

[14] Nic Newman, Richard Fletcher, Anne Schulz, Simge Andi, Craig T Robertson, and Rasmus Kleis Nielsen. 2021. Reuters Institute digital news report 2021. *Reuters Institute for the Study of Journalism* (2021).

[15] Jakob Ohme, Ewa Maslowska, and Cornelia Mothes. 2021. Mobile News Learning — Investigating Political Knowledge Gains in a Social Media Newsfeed with Mobile Eye Tracking. *Political Communication* 0, 0 (2021), 1–19. https://doi.org/10.1080/10584609.2021.2000082 arXiv:https://doi.org/10.1080/10584609.2021.2000082

[16] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372.

[17] Keith Rayner. 2009. Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62, 8 (2009), 1457–1506. https://doi.org/10.1080/17470210902816461 arXiv:https://doi.org/10.1080/17470210902816461 PMID: 19449261.

[18] Byron Reeves, Nilam Ram, Thomas N. Robinson, James J. Cummings, C. Lee Giles, Jennifer Pan, Agnese Chiatti, Mj Cho, Katie Roehrick, Xiao Yang, Anupriya Gagneja, Miriam Brinberg, Daniel Muise, Yingdan Lu, Mufan Luo, Andrew Fitzgerald, and Leo Yeykelis. 2021. Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. *Human–Computer Interaction* 36, 2 (2021), 150–201. https://doi.org/10.1080/07370024.2019.1578652 arXiv:https://doi.org/10.1080/07370024.2019.1578652 PMID: 33867652.

[19] Adam Rule, Aurélien Tabard, and Jim Hollan. 2017. Using Visual Histories to Reconstruct the Mental Context of Suspended Activities. *Human–Computer Interaction* 32, 5-6 (2017), 511–558. https://doi.org/10.1080/07370024.2017.1300063 arXiv:https://doi.org/10.1080/07370024.2017.1300063

[20] Yoonmo Sang, Jee Young Lee, Sora Park, Caroline Fisher, and Glen Fuller. 2020. Signalling and Expressive Interaction: Online News Users' Different Modes of Interaction on Digital Platforms. *Digital Journalism* 8, 4 (2020), 467–485. https://doi.org/10.1080/21670811.2020.1743194 arXiv:https://doi.org/10.1080/21670811.2020.1743194

[21] Elisa Shearer and Amy Mitchell. 2021. News Use Across Social Media Platforms in 2020. *Pew Research Available at: https://www.pewresearch.org/journalism/2021/01/12/news-use-across-social-media-platforms-in-2020/* (2021).

[22] Arthur A Stone, Christine A Bachrach, Jared B Jobe, Howard S Kurtzman, and Virginia S Cain. 1999. *The science of self-report: Implications for research and practice*. Psychology Press.

[23] Harsh Taneja, James G. Webster, Edward C. Malthouse, and Thomas B. Ksiazek. 2012. Media consumption across platforms: Identifying user-defined repertoires. *New Media & Society* 14, 6 (2012), 951–968. https://doi.org/10.1177/1461444811436146 arXiv:https://doi.org/10.1177/1461444811436146

[24] David Tewksbury. 2006. What Do Americans Really Want to Know? Tracking the Behavior of News Readers on the Internet. *Journal of Communication* 53, 4 (01 2006), 694–710. https://doi.org/10.1111/j.1460-2466.2003.tb02918.x arXiv:https://academic.oup.com/joc/article-pdf/53/4/694/22330549/jjnlcom0694.pdf