

The Upworthy Research Archive: A Time Series of 32,488 Experiments in U.S. Advocacy

J. Nathan Matias* and Kevin Munger†

September 9, 2019

Firms routinely conduct thousands of behavioral experiments per year in media [2], politics [6, 8], activism [10], and the technology industry [5, 12]. Except for a small number of researchers with ties to experimenting firms, scientists have been largely unable to test causal hypotheses at a similar scale, advance methods for high volume experimentation and analysis, or train researchers for that kind of work [14]. To broaden scientific research and education, we are preparing a dataset of 32,488 A/B tests conducted by an influential advocacy organization over two years, from 2013 to 2015. In this abstract, we present the context of the experiments, describe the dataset we are planning to make available, and summarize several possible approaches to using this data to advance education and scientific discovery.

A/B Testing at Upworthy, 2013-2015

Upworthy.com is a company started in 2012 by *The Filter Bubble* author Eli Pariser [19] and cofounder Peter Koechley to reach large audiences with “stuff that matters.” By the end of 2013, Upworthy was being called the fastest-growing media company in the world [9, 10]. A crucial component of their growth strategy was to develop of enticing “packages” of a headline, a subheading, and an image—similar to what viewers were shown when the article was shared on social media. To determine which package would be most influential, Upworthy adopted A/B testing practices already common in the technology industry and political campaigns. Their content management infrastructure was optimized to deploy packages, measure responses, and compare the probability of a viewer clicking on different potential packages for the same story.

Upworthy was a central actor in the history of U.S. media from 2013-2015. As many publishers adopted headlines similar to some of Upworthy’s most widely-spread stories, people attributed Upworthy’s success to the idea of ‘clickbait.’ In March of 2015, Koechley apologized and pledged to change, saying ‘sorry we kind of broke the internet last year’ [18]. As Upworthy

* Assistant Professor, Cornell University Department of Communication. Email: nathan.matias@cornell.edu

† Assistant Professor, Penn State University Department of Political Science. Email: kmm7999@psu.edu

changed its editorial practices and business model, social media platforms also changed their algorithms. In 2016, the social network Facebook adjusted their recommender algorithms to reduce the visibility of clickbait, a change that Upworthy praised [20]. In 2017, Upworthy merged with Good Worldwide [11].

Upworthy conducted its experiments on the homepage and article pages of Upworthy.com, randomly assigning different readers to see different packages for the same story in recommendations to readers. Site engineers and editors reported in interviews that they allowed only one experiment per page to limit interference between experiments. The content management system recorded the number of participants that were shown a given package (impressions) and the number that clicked on the package (clicks). After a period of time, an editor would review a dashboard that reported the results and either conduct an additional experiment (potentially with new packages) or choose which package to finalize for a given article. Editors sometimes finalized a package other than the best performing one. From that decision point, Upworthy would only display the final chosen package.

Archive of Upworthy A/B Tests

The current dataset consists of 32,488 tests conducted from January 24, 2013 through April 14, 2015, just after the company announced editorial shifts. For each test, the dataset includes viewer responses to each package in an experiment. The dataset includes over 150 thousand packages with a median of 4 packages per test. Together, these packages received over 538 million impressions and over 8 million clicks. Each test included a median of 14,342 impressions and a median of 201 clicks per test.¹

Each package includes the following: the experiment ID, the time the package was created; the headline, subhead, and social media preview text; the preview image; and the number of impressions and clicks received by the package during the test.

Privacy and Research Ethics

As field experiments have become common in business, democracy, and the social sciences, experimenters have faced legitimate concerns about data privacy and research ethics [3].

Participant privacy is a major risk in any scientific dataset involving people. Attempts to anonymize information can often be circumvented [17]. While sophisticated techniques may protect privacy while enabling scientific study [4], it is much safer to avoid storing or sharing any individual-level data at all. In this project, we prioritize participant privacy over all else. We only include aggregate results from each experiment; our dataset includes no personal information of any kind.

¹We are still validating this dataset, so the final counts may be different from the one reported here.

Our privacy decisions shape our approach to research ethics. While academic researchers in the U.S. are required to follow Common Rule requirements on research ethics, non-academic experimenters such as headline writers are not required to obtain committee review for each individual news headline or for the design of their headline tests [7]. We have been informed by the Cornell University IRB that this project is not governed by U.S. research ethics regulations, since (a) the dataset does not include any information about individual people, and (b) the studies were conducted several years ago by Upworthy alone, with no involvement from us or our institutions.

By working with aggregate data, we also prevent ourselves or anyone else from contacting the people who participated in these studies.

Advancing Science with the Upworthy Research Archive

As a uniquely large dataset of causal studies, we hope The Upworthy Research Archive could be used to advance fields including communication, political science, psychology, media studies, computer science, marketing, and business.

For example, the pace of change of online behavior poses novel challenges for social scientists. Although researchers are familiar with the challenges of generalizing research across culture and geography, scientists are still reckoning with the challenge of generalizing research results over time [16]. This dataset could enable a rigorous demonstration of the challenges posed by knowledge decay in the quickly-changing online world.

We also hope that this data can contribute to advances in quantitative methods as used by scientists and practitioners. Surveying the state of applied econometrics, Athey reports that “the use of multiple experiments... provides fertile ground for future work” [1]. So far, access to large corporate archives of experiments has been limited to a small number of scientists. These constraints have also limited educators in our teaching of large-scale experimentation [14]. By publishing The Upworthy Research Archive, we aim to help level the playing field of research and teaching in this crucial area.

Releasing this dataset openly all at once would be scientifically irresponsible. In the rush for discoveries, research findings would likely be overfit to this specific case and would lead people astray. Rather than publishing all of the data at once and inviting a scramble for over-fit, over-confident results, we plan to develop a process for pre-registration and cross-validation of hypotheses associated with this dataset. We are drawing lessons and inspiration from the Fragile Families Challenge [13] and Psychological Science Accelerator [15].

We are excited to announce this new resource and welcome suggestions for managing its usefulness for science, education, and the common good.

References

- [1] ATHEY, S., AND IMBENS, G. W. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives* 31, 2 (2017), 3–32.
- [2] BULIK, M. Which Headlines Attract Most Readers? *The New York Times* (June 2016).
- [3] DESPOSATO, S. *Ethics and experiments: problems and solutions for social scientists and policy professionals*. Routledge, 2015.
- [4] DWORK, C., AND ROTH, A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science* 9, 34 (2014), 211–407.
- [5] FABIJAN, A., DMITRIEV, P., OLSSON, H. H., AND BOSCH, J. The Evolution of Continuous Experimentation in Software Product Development. In *International Conference on Software Engineering (ICSE)* (2017).
- [6] GREEN, D. P., AND GERBER, A. S. *Get out the vote: How to increase voter turnout*. Brookings Institution Press, 2019.
- [7] GRIMMELMANN, J. The Law and Ethics of Social Media Experiments.
- [8] ISSENBERG, S. *The victory lab: The secret science of winning campaigns*. Broadway Books, 2013.
- [9] KAMENETZ, A., AND KAMENETZ, A. How Upworthy Used Emotional Data To Become The Fastest Growing Media Site of All Time, June 2013.
- [10] KARPF, D. *Analytic activism: Digital listening and the new political strategy*. Oxford University Press, 2016.
- [11] KELSEY, S. Upworthy to merge with Good Worldwide, newsrooms to consolidate-POLITICO Media. *Politico* (Jan. 2017).
- [12] KOHAVI, R., DENG, A., FRASCA, B., WALKER, T., XU, Y., AND POHLMANN, N. Online Controlled Experiments at Large Scale. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2013), KDD ’13, ACM, pp. 1168–1176.
- [13] LIU, D., AND SALGANIK, M. Successes and struggles with computational reproducibility: Lessons from the fragile families challenge.
- [14] MATIAS, J. N. Teaching Large-Scale Digital Experimentation to Undergraduates and Graduate Students.

- [15] MOSHONTZ, H., CAMPBELL, L., EBERSOLE, C. R., IJZERMAN, H., URRY, H. L., FORSCHER, P. S., GRAHE, J. E., MCCARTHY, R. J., MUSSER, E. D., AND ANTFOLEK, J. The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science* 1, 4 (2018), 501–515.
- [16] MUNGER, K. Temporal validity in online social science.
- [17] NARAYANAN, A., AND FELTEN, E. W. No silver bullet: De-identification still doesn't work. *White Paper* (2014), 1–8.
- [18] O'REILLY, L. Upworthy cofounder Peter Koechley at Guardian Changing Media Summit - Business Insider. *Business Insider* (Mar. 2015).
- [19] PARISER, E. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.
- [20] UPWORTHY. Upworthy is totally on board with Facebook's algorithm change. For reals., Oct. 2016.