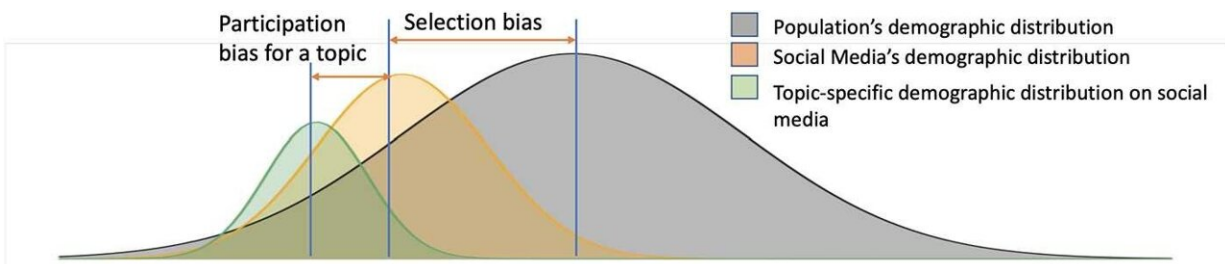


Study cautions that opinions drawn from social media are skewed by vocal minorities

September 18 2023, by Harini Barath



A pictorial illustration of the participation bias, selection bias for a social media platform, and the underlying population distribution. Credit: *EPJ Data Science* (2023). DOI: 10.1140/epjds/s13688-023-00405-6

Surveys have been a time-tested mechanism that allows policymakers to gauge the pulse of public opinion on a wide range of issues.

In recent years, [social media users](#) have publicly weighed in on many issues pertaining to policy, and this wealth of data is being considered by policymakers as a viable alternative to traditional surveys that are both time-consuming and expensive to collate.

However, it's important to be aware that data from social media is riddled with biases, according to Neeti Pokhriyal, an American Association for the Advancement of Science (AAAS) Science and Technology Policy Fellow at the National Science Foundation who was a

computer science postdoc and visiting scholar at Dartmouth, and Soroush Vosoughi, assistant professor of computer science.

Unlike surveys, which are designed to collate opinions from diverse groups that closely reflect the country's demographics, it is well established that the demographics on [social media platforms](#) are not truly representative of the larger population.

For example, more [young people](#) use social media than seniors, who jumped on the bandwagon later. Fewer than half of those 65 and older use [social media sites](#), while more than 80% of those under the age of 50 are regular users, according to 2021 data from the Pew Research Center.

What's more, this varies across platforms. Snapchat and Instagram largely attract young users, while Facebook has the highest share of older users. These are some well-known factors that make data from these sources biased.

There's another, lesser-known bias—[quantified for the first time in a recent paper](#) co-authored by Pokhriyal, Vosoughi, and Professor of Government Benjamin Valentino—known as participation bias. The paper is published in *EPJ Data Science*.

This bias arises not from who is on a platform, but from who among them are active, vocal participants on that platform, says Pokhriyal. And this varies based on the topics being discussed.

"Even if you have everyone on Twitter, they may only participate in certain topics—ones that they find interesting or maybe feel comfortable talking about in public," says Vosoughi. So, he says, when a small group is very vocal about a particular issue, their opinions get over-represented in the data.

While participation bias has been studied in survey science, it has not been analyzed in the digital context. To put their finger on how much participation bias there is, the researchers built a [computational model](#).

Their model looks at social media data and based on data collected from existing representative surveys on the same topic, it estimates the demographics of the population that could have participated in the discussion on social media. The difference between the model's estimate and the platform's actual demographics reveals the participation [bias](#) for that topic.

In their paper, they perform a [case study](#) on the topic of gun control in the U.S., comparing data from X—known as Twitter at the time of their analysis—with [survey data](#) from several polling institutions such as NPR, Newshour, and Marist.

Demographic data from Pew shows that men and women are equally represented on Twitter and its users lean Democratic. In discussions about gun control, however, the model estimates that Republicans and men are weighing in more heavily.

"We're hoping that this kind of research can help put what we see on social media into context, and also to make it easier to track changes in public opinion without the need to run repeated, expensive surveys," says Valentino.

The model is also designed to account for noise in social media data such as posts generated by bots, says Pokhriyal, while also acknowledging that their model only works when there is survey data available.

Conducting surveys is resource-intensive process, and in recent years, has seen a dip in the numbers of willing participants. Digital data can help policymakers supplement survey findings, says Pokhriyal, but only

if the existing biases can be adequately accounted for.

More information: Neeti Pokhriyal et al, Quantifying participation biases on social media, *EPJ Data Science* (2023). [DOI: 10.1140/epjds/s13688-023-00405-6](https://doi.org/10.1140/epjds/s13688-023-00405-6)

Provided by Dartmouth College

Citation: Study cautions that opinions drawn from social media are skewed by vocal minorities (2023, September 18) retrieved 28 April 2024 from <https://phys.org/news/2023-09-cautions-opinions-drawn-social-media.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.