Political Communication

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/upcp20

What Did We Learn About Political Communication from the Meta2020 Partnership?

Kevin Munger

To cite this article: Kevin Munger (2025) What Did We Learn About Political Communication from the Meta2020 Partnership?, Political Communication, 42:1, 201-207, DOI: 10.1080/10584609.2024.2446351

To link to this article: https://doi.org/10.1080/10584609.2024.2446351

POLITICAL COMMUNICATION
2025, VOL. 42, NO. 1, 201–207
https://doi.org/10.1080/10584609.2024.2446351

Routledge
Taylor & Francis Group

What Did We Learn About Political Communication from the Meta2020 Partnership?

Kevin Munger

Department of Social and Political Sciences, European University Institute, Florence, Italy

ABSTRACT

What did we learn from the Meta2020 research studies? Despite the ambition of this project, tensions emerge between producing precise statistical estimates and generating broadly applicable knowledge about social media's societal impacts. There are crucial challenges in translating theoretical concepts into measurable interventions on social media platforms exacerbated by the difficulties in conducting big, consensus-based science. While the Meta2020 studies represent the pinnacle of social science research on social media, they also expose the limitations of empirical social science in understanding and regulating rapidly evolving digital platforms. I argue that proactive regulation of social media is necessary, given the demonstrated lack of understanding of these platforms' effects, even by the platforms themselves. This analysis contributes to ongoing discussions about the role of academic-industry partnerships in studying and governing social media platforms.

KEYWORDS

Facebook; Instagram; experiments; Meta2020; US presidential election

The 2020 Election research partnership between Meta and a team of 17 independent academics is the most ambitious social science of social media to date. There have been five papers published so far (Allcott et al., 2024; Guess et al., 2023a, 2023b; Nyhan et al., 2023; González-Bailón et al., 2023), involving large-scale field experiments manipulating key elements of the Facebook and Instagram platforms or analyzing the network structure of the platforms in unprecedented detail. However, the fact that each of these is simply "a paper," from the perspective of a literature review or the intuitive sense in which academics think about evidence, does them a disservice. Social media platforms have their own "physics," defined by their infrastructural affordances. To an extent that is impossible in the physical world, these studies were able to experimentally manipulate something like gravity.

My aesthetic appreciation for these experiments cannot be overstated. They are simply beautiful. Thinking as I sometimes do of social science experiments as performance art, the craft and the vision on display is deeply satisfying. The most important dimension in evaluating the papers produced by this collaboration, however, is: "What did we learn?"

In the narrow sense, what we learned was numbers. A gigantic research apparatus was spun up, millions of dollars spent, thousands of hours by some of the best researchers in the field – and what was produced were numbers (or, really, ranges of numbers). For example, "The point estimate for the effect of Facebook deactivation

CONTACT Kevin Munger kevinmunger@gmail.com Department of Social and Political Sciences, European University Institute, Via della Badia dei Roccettini, 9, Fiesole, Florence, FI 50014, Italy

on Trump vote is a reduction of 0.026 units ($p = 0{:}015$, $Q = 0{:}076$, 95% CI bounds = $-0{:}046$, $-0{:}005$)" (Allcott et al., 2024). But in fact, according to the language in the papers, even these numbers are less important than the binary fact of whether the range of numbers intersects 0, once doing a statistical adjustment: "This effect falls just short of our preregistered significance threshold of $Q < 0{:}05$" (Allcott et al., 2024).

We learned a lot of those numbers and the relevant comparisons with zero. Including all of the analyses of heterogeneity in all of the appendices, there are hundreds of such comparisons.

The meta-scientific question that I take up in this article is: so what?

In Munger (2019, p. 1) I argued that "I no longer believe that one-off field experiments are as valuable a tool for studying online behavior as I once did. The problem is that the knowledge they generate decays too rapidly because the object of study changes too rapidly."

How do we generalize our findings from one place to another, from the past to the future? How do we ensure that this generalization is done at the same level of rigor of the original study? If we cannot, the rigor of the original study is simply performative – beautiful, I maintain, but not scientifically useful.

The traditional approach has been replication – we know we can transport knowledge if other people can replicate the study and produce the same result. As the scope of rigorous social science has expanded, this has become obviously impossible: we cannot "replicate" the Meta 2020 research collaboration. Facebook and Instagram are already dramatically different than they were in 2020.

Notably, the experiment in which some subjects had their algorithmic recommendation feeds replaced by old-school chronological feeds was conducted before the dramatic shift to a heavily algorithmic Instagram feed in the summer of 2021 in response to pressure from TikTok. Concerns about Instagram's feed have only really arisen in recent years; this study, conducted in 2020 but published in 2023, is already a historical case study.

So if we cannot replicate these studies, how do we rigorously apply their results? External validity/generalizability is the most important question in social science methodology – especially when studying something that changes rapidly. These papers do not address this problem except to specifically punt on it. These hedges are the least beautiful part of these papers. They're so boilerplate that they seem pro forma. And they all boil down to saying, "This study doesn't tell you what you actually want to know."

Consider Guess et al. (2023b p. 403): "our design cannot speak to 'general equilibrium' effects, because doing so would imply making inferences about societal impact."

To which my question is, again: what does it speak to?

Or Nyhan et al. (2023, p. 143): "replications in other countries with different political systems and information environments will be essential to determine how these results generalize."

Given that these proposed replications are *impossible*, this is an admission that we cannot know how these results generalize.

Allcott et al. (2024) best illustrates my point: "readers should be cautious about generalizing beyond our specific sample and time period . . . our results can inform readers' priors about the potential effects of social media in the final weeks of high-profile national elections."

There is a massive, even majestic amount of care and rigor put into the production of these papers and the statistical infrastructure to evaluate them – and essentially zero attention paid to the question of how humans synthesize the information contained in those papers. In order to begin rigorously answering the question at issue (what did we learn?) we have to know the answer to another question: What did we know?

What Did We Know, Theoretically?

These papers have thus far been published in *Science*, *Nature*, and *PNAS*. Publications in these venues face a tight space constraint, and the theoretical discussion in each case is circumscribed. Each paper has an appendix which is over 200 pages long – but as far as I could tell, none of this essentially unlimited space is given over to theoretical discussion. Instead, we find dozens of additional tables and details about implementation.

So – given the ability to manipulate platform physics, how did the research team select which fundamental constants to manipulate?

One of the experiments decreased subjects' exposure to "like-minded sources" (Nyhan et al., 2023): "To evaluate a potential response to concerns about the effects of echo chambers, we . . . reduced exposure to content from like-minded sources during the 2020 US presidential election by about one-third." (p. 137)

This discussion of social media "echo chambers" reveals a lack of rigor applied to the step of synthesizing existing theoretical knowledge and developing a shared theoretical ontology. We need to know what an echo chamber is before we can know what echo chambers do, or if social media is an echo chamber. This step, as I argue in Munger (2024), should be thought of as *poetic validity*: the extent to which the natural-language phrase we use to represent a concept "rings true."

I specifically examine the poetic validity of the idea that "social media is an echo chamber." So – how do the Meta 2020 papers use the phrase "echo chamber"? Nyhan et al. (2023) equates them with "exposure to content from like-minded sources." In contrast, González-Bailón et al. (2023) equates them with "social curation processes": "social curation processes (echo chambers) increase the probability that people will surround themselves with ideologically compatible information."

Taken together, the theoretical question in these two papers seems to be whether social curation processes cause biased media diets–or rather, *whether echo chambers cause echo chambers.*

There is not even internal poetic validity within this project and coauthor team! It makes it difficult to know what we knew before the studies were conducted – and thus, what we learned.

What Did We Know, Ontologically?

The treatment in Allcott et al. (2024)'s deactivation experiment is ontologically clear: you stop using Facebook or Instagram entirely. The other three experiments which have been published so far involve a more complicated modification to the architecture of the recommendation algorithm. In each case, the control condition was the standard newsfeed, and the respective treatment conditions were:

- A chronological feed, arranging content in the order it was posted

- A feed which removed all content which had been reshared
- A feed which reduced the amount of content from "like-minded" sources by one third

Each of these interventions was motivated by theoretical concerns about the effects of social media. But what exactly are these treatments? In order to operationalize the human-language theories (like "content from like-minded sources causes XYZ"), the academics and especially the Facebook research scientists needed to do a massive amount of work to understand what was going on at the level of the technical infrastructure. Nobody had ever *measured* these concepts in real-time and at scale before – let alone intervened on them.

A common refrain from the Meta2020 team is that they had to "build the plane while flying it" – that they were scrambling to make everything work, that they did not have the luxury of planning things in advance. It is an "agile" business cliche, but we would benefit by taking it seriously: *how did they build the plane?* This is the knowledge I really wish the team had prioritized as an output.

Because a detailed understanding of how the mechanics of the newsfeed algorithm works is the other key input toward understanding what we learned. There is a certain alienness to the machine-language categories behind these Meta2020 algorithmic interventions. Social scientists simply did not understand what the newsfeed is.

We wouldn't have had very informed priors about what these treatments would cause if we did not have the capacity to understand what these treatments are. We can definitely update our priors about the effects of these specific treatments – but it is less clear how we should update our priors about the human-language beliefs about the effects of "online echo chambers on social media."

Much of this discussion has been about the experiments resulting from the collaboration. In part, this is because there have been four of them published so far, compared to just one of the descriptive papers. Wagner (2023, p. 391) makes the observation that "it was faster to analyze the experiments . . . [than] the nonexperimental papers."

But the entire purpose of the descriptive papers is to translate the massive machine lifeworlds powering Facebook and Instagram into our language. This still had to be blunt; we define the categories that we care about, and the data have to be compressed into this one specific dimension (in the case of the only paper published so far, the dimension is "ideology"). However, this still requires more attention paid to the act of understanding what the platforms actually are – there is no flashy intervention to distract from this fundamental task.

The Challenges of Big Science

Sociologically, "building the plane" required figuring out how to get these 32 researchers to work together. The scope of the research was simply incommensurate with what a lone academic or small team of collaborators can accomplish. However, there are challenges and downsides to this "industrial organization" of science.

The most obvious is time. There is no denying that this knowledge would have been more relevant had the papers been published sooner. Indeed, as of October 2024, the majority of papers still have not been published. Consensus-based science simply takes longer.

Publishing in venues like *Science* and *Nature* caused an even greater delay – by involving high standards for peer review, but mostly by preventing the circulation of pre-prints or working papers. Perhaps, this was necessary to overcome skepticism based on this academic-industry partnership. Even still, if 4 years is the minimum time to publication, this represents a significant argument against academics pursuing this kind of research.

There are also inevitably disagreements over scientific practice. The standard scientific approach to this kind of problem is cost–benefit analysis. But this approach seems somehow profane in the hallowed context of Science. Scientists have strong commitments to methodological standards. How can we use methods that have been identified as producing bias, or inflating false-positive rates?

These latent tensions in social scientific practice are clarified in a large, consensus-based project. My reading of the evidence suggests that these tensions were not fully resolved, producing sentences like the following:

> [the effect on] self-reported vote choice is large enough to be meaningful in a close election, although it is not significant at our preregistered threshold. (Allcott et al., 2024)

The most ambitious social science of social media to date was *underpowered to detect changing the results of a close US Presidential election*.

Now, the experiment was only underpowered because the authors elected to

(1) follow strict rules about pre-registration and statistical adjustments for multiple comparisons ; while also
(2) conduct a large number of tests.

Each of these steps makes sense. (1) is a coherent response to the false positives behind the "Replication Crisis" coming out of social psychology: we were simply running too many tests, with too many researcher degrees of freedom. (2) follows from the practical insight that it is far cheaper at the margin to add more outcome variables when conducting a massive, expensive RCT. The ratio of the cost of measuring additional variables to the cost of conducting the field experiment is very low; so why not just measure as many outcome variables as you can?

Well, the reason not to do (2) is that there is a "penalty" in terms of power due to (1).

Neither side wanted to budge – so we get an estimated effect of Facebook deactivation which was "large enough to be meaningful in a close election" but which the study is unable to distinguish from zero.

The methodological tension is not unique to the Meta2020 collaboration, but is clarified and intensified by the industrial organization of the enterprise. My meta-scientific impulse, elaborated in Munger (2023), is to put a fine point on this contradiction. (1) is an example of science reform conducted in a culturally positivist way, bending toward public and verifiable displays of rigor. (2), in contrast, is the reality of conducting expensive experiments. My conclusion is that we cannot do science reform by commandment. We need an actual cost–benefit analysis – we need metascience.

In this case, my sympathies lean toward prioritizing (2). As discussed above, no one has ever done anything at this scale and depth before. Strong theoretical expectations about what would happen strike me as unjustified.

Given a flat organizational structure and a team of 32, it is easy to see how the number of outcome variables would flourish. As, in my opinion, it should – if you're running a multi-million dollar experiment, you should try to learn as much as you can! But strict adherence to (1) means that it is unclear that we learned anything at all.

This methodological debate is not merely academic. It has direct ethical implications. The impetus for the specific interventions deployed is widespread public concern that social media is "causing" various ills, from negative mental health to political polarization. It is thus *clearly in Meta's interest to find that social media causes nothing*–which they did, more or less, because of (1).

Ethics

The larger "ethical question" is whether these experiments should have been conducted at all. This is an obvious question, and has thus been discussed at length; the consensus, among scholars who were not invited to participate and therefore will not accumulate thousands of citations to their Google Scholar account, seems to be "no," for one reason or another.

From my perspective, the relevant question to me is not whether this work was ethically Bad vs. ethically Acceptable (and thus whether it should be banned or allowed) – it is at the other margin. That is, is this work Ethically Acceptable or Ethically Obligatory? Should it be legally mandated? Should the US FTC or the EU, under the Digital Services Act, force Meta to replicate these experiments moving forward?

This leads to my final point about the Meta2020 collaboration. This is such a monumental undertaking that we see the stark limitations of studying this or that aspect of Facebook without direct access to Facebook's architecture. If this is social science, what have we been doing?

But also – every one of the papers published as part of this collaboration is evidence of something that no one knew before. Not even at Meta. You can add the phrase "Meta Didn't Know" to the beginning of every paper title in this series and it becomes a news headline.

No one is keeping track of any of this. Any conspiratorial account of Meta doing this or that political thing is mistaken; it is far too expensive, so they cannot afford to know. The political/regulatory question is *whether they should be required to know*.

The Meta2020 collaboration required an unprecedented financial commitment by Meta to pull off. Wagner (2023, p. 391) says that Meta spent "more than $20 million and redirect[ed] the time of dozens of employees to serve the project." Let us call it $25 million. This is compared against $134 billion in revenue for 2023 (Meta, 2024). That is .019% of their annual revenue devoted to this 6-year project.

Is that the right amount? How does this compare to drug companies, or car companies, or investment banks? Does Meta know enough about what their product is, or what it does? I think that the Meta2020 collaboration demonstrates that the answer is no.

I have made some critiques of the theoretical and methodological decisions made as part of the project. There is room for improvement. But I recognize that according to prevailing standards of empirical social science, the studies conducted as part of the Meta2020 are exemplary of the best social science of social media we can plausibly expect to conduct.

And my conclusion is that we need to recognize the limitations of social science. The only way to effectively regulate something like social media is proactively—and empirical,

quantitative social science can only ever be reactive. The Meta2020 collaboration has demonstrated that this kind of proactive regulation is possible, though expensive. So the most important question is now a political one.

Can we afford to be governed by a newsfeed that no one understands?

Disclosure Statement

No potential conflict of interest was reported by the author(s).

References

Allcott, H., Gentzkow, M., Mason, W., Wilkins, A., Barberá, P., Brown, T. O., Cisneros, J. C., Crespo-Tenorio, A., Dimmery, D., Freelon, D., González-Bailón, S., Guess, A. M., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., Nair-Desai, S., Nait El Barj, H. . . . Stroud, N. J. (2024). The effects of Facebook and Instagram on the 2020 election: A deactivation experiment. *Proceedings of the National Academy of Sciences*, *121*(21), 12121e2321584121. https://doi.org/10.1073/pnas.2321584121

González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T. O., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., Iyengar, S., Kim, Y. M., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C. V., Settle, J., Thorson, E. . . . Stroud, N. J. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, *381*(6656), 392–398. https://doi.org/10.1126/science.ade7138

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R. . . . Stroud, N. J. (2023a). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, *381*(6656), 398–404.

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T. O., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J. . . . Stroud, N. J. (2023b). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, *381*(6656), 404–408. https://doi.org/10.1126/science.add8424

Meta. (2024). Meta reports Fourth Quarter and full year 2023 Results; initiates quarterly dividend. *Meta Investor Relations* https://investor.fb.com/investor-news/press-release-details/2024/Meta-Reports-Fourth-Quarter-and-Full-Year-2023-Results-Initiates-Quarterly-Dividend/default.aspx

Munger, K. (2019). The limited value of non-replicable field experiments in contexts with low temporal validity *social Media+ society*. *Social Media + Society*, *5*(3), 532056305119859294. https://doi.org/10.1177/2056305119859294

Munger, K. (2023). Temporal validity as meta-science. *Research & Politics*, *10*(3). https://doi.org/10.1177/20531680231187271

Munger, K. (2024). *The YouTube apparatus*. Cambridge University Press.

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler Tucker, J. A (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137—144.

Wagner, M. W. (2023). Independence by permission. *Science*, *381*(6656) , 388–391.