Media and Communication
2024 • Volume 12 • Article 8577
https://doi.org/10.17645/mac.8577

ARTICLE

Open Access Journal

Quantifying Data-Driven Campaigning Across Sponsors and Platforms

Michael M. Franz [1], **Meiqing Zhang** [2], **Travis N. Ridout** [3], **Pavel Oleinikov** [2,4], **Jielu Yao** [5], **Furkan Cakmak** [6], **and Erika Franklin Fowler** [7]

[1] Department of Government and Legal Studies, Bowdoin College, USA
[2] Wesleyan Media Project, Wesleyan University, USA
[3] School of Politics, Philosophy, and Public Affairs, Washington State University, USA
[4] Quantitative Analysis Center, Wesleyan University, USA
[5] East Asian Institute, National University of Singapore, Singapore
[6] Wesleyan Media Project, Wesleyan University, USA
[7] Department of Government, Wesleyan University, USA

Correspondence: Travis N. Ridout (tnridout@wsu.edu)

Submitted: 30 April 2024 **Accepted:** 3 September 2024 **Published:** 31 October 2024

Issue: This article is part of the issue ”Data-Driven Campaigning in a Comparative Context: Toward a 4th Era of Political Communication?” edited by Katharine Dommett (University of Sheffield), Rachel Gibson (University of Manchester), Sanne Kruikemeier (Wageningen University & Research), Sophie Lecheler (University of Vienna), Esmeralda Bon (University of Manchester), and Stephanie Luke (University of Sheffield), fully open access at https://doi.org/10.17645/mac.i457

Abstract

Although modern data-driven campaigning (DDC) is not entirely new, scholars have typically relied on reports and interviews of practitioners to understand its use. However, the advent of public ad libraries from Meta and Google provides an opportunity to measure the scope and variation in DDC practice in advertising across different types of sponsors and within sponsors across platforms. Using textual and audiovisual processing, we create a database of ads from the 2022 US elections. These data allow us to create an index that quantifies the extent of DDC at the level of the sponsor and platform. This index takes into account both the number of unique creatives placed and the similarity across those creatives. In addition, we explore the impact of sponsor resources, the office being sought, and the competitiveness of the race on the measure of DDC sophistication. Ultimately, our research establishes a measurement strategy for DDC that can be applied across ad sponsors, campaigns, parties, and even countries. Understanding the extent of DDC is vital for policy discussions surrounding the regulation of microtargeting and data privacy.

Keywords

data-driven campaigning; digital campaigning; election campaigns; political advertising

© 2024 by the author(s), licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

For decades, campaigns have used data to make decisions about effective campaign practices, though the sophistication in data use has ramped up considerably in recent years. So while data-driven campaigning (DDC) is not new, academic studies of the phenomenon have typically relied on reports and interviews with practitioners in order to understand how it is practiced. Examples of this include Anstead's (2017) study of the 2015 general election campaign in the UK, Kreiss's (2016) work on the use of technology in campaigns in the US, and Kefford's (2021) book on such practices in Australia. As a result, scholars are starting to get a picture of how DDC is practiced, but the picture is an incomplete one. Most reports on DDC practices center on the most prominent elections in each country, ignoring regional and local campaigns, and most studies focus on a single country (the five-country study by Dommett, Kefford, and Kruschinski [2024] is one exception). Clearly, more research is needed to consider variation in DDC practice across countries, parties, candidates, and offices, but given that most research on the practice of DDC relies on in-depth interviews or participant observation, it may take a long time for such work to emerge.

What we propose attempts to rapidly widen the scope of contexts in which DDC is studied, relying on the Meta and Google ad libraries to create quantitative measures of DDC. Specifically, we focus on two measures—the number of unique creatives that were placed in each race and the similarity of the ad content—to create an overall indicator of the extent to which campaigns engage in DDC. We take advantage of such data from the 2022 US elections given the heavy spending on advertising in the US, the high professionalization of campaigns in the US (Farrell et al., 2001; Maarek, 2011), and the large number of campaigns occurring at the same time. That said, our approach could be used for any number of races in the large number of countries for which Meta and Google data are available.

We first establish wide variation in our sophistication measure across campaigns, and scores that make sense given what we know about those campaigns. We also show that the DDC sophistication measure is higher in Senate races than in House races and higher in more competitive races. Moreover, Democratic Party affiliation predicts greater DDC sophistication for Meta ads, and higher campaign resources are associated with greater DDC in Google data.

2. Defining DDC

Definitions of DDC and related concepts are not always spelled out directly in research in the area, but they tend to take two approaches. The first focuses on specific practices that are thought to be indicative of DDC, such as targeting and testing. Baldwin-Philippi (2017, p. 628) writes:

> Data-driven campaigning involves two main features: targeting, or deciding which messages go to what potential voters at what time during the campaign, and testing, or empirically measuring how well messages perform against one another and using that information to drive content production and further targeting.

Dommett, Barclay, and Gibson (2024), by contrast, offer a definition that focuses on whether "data is used to inform decision-making" regarding "voter communication, resource generation and/or internal organization" (p. 2). They make the reasonable argument that focusing on specific practices does not guarantee that data were used in campaign decision-making. One could target messages, for instance, based on "rule of thumb" or based on "what we've always done" as opposed to utilizing data in the decision of how to target.

In spite of this drawback noted by Dommett, Barclay, and Gibson (2024), we take the Baldwin-Philippi (2017) approach, focusing on one particular instantiation of DDC: the tailoring of ad messages. Admittedly, while our approach gives up something in terms of understanding exactly how the data are used, it has the advantage of allowing an examination of DDC in multiple campaign contexts.

3. Studying DDC

There are multiple ways to study the practice of DDC, with each having advantages and disadvantages. The most common approach is interviews with campaign strategists and practitioners. Kreiss (2016), for example, conducted interviews with more than 50 staffers who had worked on digital technology in Democratic and Republican campaigns in the US. Kefford (2021) was even more ambitious in his study of how parties in Australia use campaign technologies, conducting over 150 interviews with both campaign managers and volunteers that stretched over several years. Anstead's (2017) research on DDC in the 2015 general election in the UK involved 31 interviews with political practitioners. Dobber et al. (2017) explored "political behavioral targeting" in the Netherlands, relying on eleven semi-structured interviews with Dutch campaign leaders involved in the 2017 national elections. A couple of studies of data campaigning in Canada also depended on interviews. For example, Montigny et al. (2019) looked at discourse surrounding the regulation of data-driven practices in Quebec using 45 semi-directed interviews. Munroe and Munroe (2018) took an in-depth look at how data are used and perceived in one constituency campaign in British Columbia through a dozen elite interviews with campaign managers, staff, and volunteers. Most ambitiously, Dommett, Kefford, and Kruschinski (2024) drew on 329 interviews with campaign and party operatives, pollsters, and data brokers for their five-country study of DDC.

Sometimes, these interviews are supplemented by participant observation. Kefford (2021) worked for both Labor and Green campaigns in Australia, and Munroe and Munroe (2018) relied on participant observation in addition to their interviews. Observational and archival data were also relied on in the Dommett, Kefford, and Kruschinski (2024) study, and another study used country experts to make comparisons across six countries (Kefford et al., 2023).

The advantage of such approaches is the ability to get an in-depth understanding of how campaigns use data to inform decision-making. The drawback is that interviews and participant observation require a lot of time to accomplish, making it more difficult to compare across multiple races, parties, types of office, localities, or countries.

A couple of studies have leveraged more quantitative measures. For instance, Vliegenthart et al. (2024) conducted a 25-country survey of people's perceptions of DDC, finding that people assign a range of acceptability to various practices, but this study does not focus on how DDC is practiced. In addition, Beraldo et al. (2021) traced Facebook political ad targeting patterns in the Dutch 2021 general elections by enrolling participants in a user audit program and collecting the political posts they were exposed to through a browser extension. This methodology promisingly combined data from surveyed political attitudes and the same individuals' online feeds, but the user audit deployment is costly and the insights that could be drawn are contingent on the selected sample of users.

Perhaps the best example of the use of quantitative measures to examine the practice of DDC is a study by Kruschinski and Bene (2022), which examined parties' online practices in European Union elections across 28 countries. They used several measures of Facebook activity by country, including the mean number of "organic" posts by party, the mean number of unique sponsored posts by party, the mean number of unique ads by party, and the number of duplicate versions of ads. They suggested that using many different versions of paid ads "is an indication for the use of more sophisticated digital marketing strategies, like using different targeting options on the same message or testing the messages with different audience segments" (Kruschinski & Bene, p. 61). For instance, the average number of creatives is over 6,800 in Germany and over 1,000 in the UK compared to 19 in Ireland and just 15 in Latvia. The inference, then, is that DDC was much more advanced in Germany and the UK than in most other European countries.

Our approach is similar to that of Kruschinski and Bene (2022) in that we utilize data on political advertising obtained from the Meta and Google ad libraries to create a quantitative indicator of DDC: the tailoring of political ads. But our research differs as well. First, we focus not just on the number of ad creatives produced but the similarity across those creatives as well. Second, our focus goes beyond the country level and the party level to consider the use of ad tailoring by specific campaigns.

4. Variations in DDC

Dommett, Kefford, and Kruschinski (2024) proposed a framework for studying DDC that recognizes variation across several levels, including the system level, the regulatory level, and the party level. System-level factors include the electoral system, the system of government (such as unitary vs. federal), the party system (how many parties), and the media system. Regulatory-level factors include regulations on parties, regulations on campaigns, data and privacy regulation, and regulations on media. Party-level variables include the financial resources of the party, the organizational structure of the party, the party's ideology, and elite and grass-roots attitudes toward campaigning. What is lacking in this model is variation across campaigns, something that is particularly pertinent to the study of DDC in the US and other countries that have candidate-centered, rather than party-centered, campaigns.

In this research, we ask two questions. First, how can we quantify the level of sophistication in DDC through online advertising data? Second, what factors explain variation in the sophistication of DDC?

To the first point, we need to be clear. Our measure of DDC sophistication is really a measure of the extent to which campaigns tailor ads to cater to specific audiences, which we believe is an important indicator of whether a campaign is utilizing practices at the most sophisticated end of DDC, such as targeting ads at specific groups of voters or ad testing. Indeed, tailoring ads by creating many different versions of a message seems pointless if they are not being targeted.

While we hope to make a methodological contribution in advancing a measure of DDC sophistication that can easily be applied across multiple campaigns, we also hope to make a more substantive contribution by explaining variation across campaigns in this measure. This investigation will also help establish the validity of our measure if those factors that we theorize predict DDC sophistication, in fact, do predict them.

Thus, in trying to explain variation across campaigns, we consider several factors. The first is the office (whether House or Senate). Because Senate constituencies (entire states) are typically larger than House districts, they require more resources to reach out to a larger number of voters. And tailoring ads to specific parts of the electorate becomes more efficient when larger numbers of people fit that demographic. As such, we would expect to see more tailoring of ads in larger Senate constituencies:

> H1: Senate races should exhibit higher DDC sophistication than House races.

The second factor we consider is the candidate's party, whether Democratic or Republican. As Kreiss (2016) has documented, the parties do not invest equally in the use of digital technologies. While this technological advantage can vary over time, most accounts suggest that Democrats are more advanced in the use of digital campaign technologies (Baldwin-Philippi, 2015; Kreiss & Jasinski, 2016):

> H2: Democratic sponsors should exhibit higher DDC sophistication than Republican sponsors.

Third, we consider the competitiveness of particular races, as more competitive races may raise the stakes and drive campaigns to engage in more data-intensive practices:

> H3: More competitive races should exhibit higher DDC sophistication.

Finally, we consider the resources available to a campaign. Greater financial resources, as Dommett, Kefford, and Kruschinski (2024) point out, can provide access to more knowledgeable staff and can help purchase data and technology—both key for effective DDC:

> H4: More well-financed races should exhibit higher DDC sophistication.

We also consider the type of sponsor: a candidate, party, or interest group. We have no strong expectation that sponsorship will affect DDC sophistication, but given that groups and parties are generally active in multiple campaigns—and candidates are active in a single one—we believe it wise to control for sponsorship. Finally, we also control for the ad's mode of communication, whether image, text, or video since the modality can affect the number of features that could be tailored.

5. Data

We collected political advertising data in text, image, and video formats from the Meta Ad library and the Google Transparency Report. Our analysis focuses on election ads in the US during the general election phase of the 2022 election cycle, but our approach to quantifying DDC is applicable to any country for which Meta and Google publish digital advertising data.

The ad libraries of Google and Meta make available the metadata of each ad placed on Google Search, Google Display Network, YouTube, Gmail, Facebook, and Instagram. For text ads, these include the spending range, dates being run, advertiser or page names, impressions, demographic information of audiences reached, title, and text. However, obtaining the content of audiovisual ads required more processing. We downloaded political ads in raw images and videos and extracted creative content from the original media files. Specifically, we obtained transcripts of the audio from video advertising through the Google Cloud Speech-to-Text API (https://cloud.google.com/speech-to-text) and detected text in images and videos through the Amazon Rekognition service (https://aws.amazon.com/rekognition).

Covering the 2022 US election, our dataset captures election ads that ran during the general election period. Within the universe of political ads placed on Google and Meta's platforms during this period, we focus on federal election activity. An ad is identified as a federal election ad if (a) it mentioned or featured at least one federal candidate or sitting senator or (b) it was sponsored by a federal candidate or national party.

6. Method

We consider two indicators of ad tailoring, a practice often employed in DDC. The first is the number of unique creatives produced by a campaign, and the second is the level of content similarity across a sponsor's unique creatives.

6.1. Unique Creatives

A greater number of unique creatives implies that the campaign was more likely to be (a) tailoring advertisements to specific audiences for purposes of targeting and/or (b) testing variations of ads to assess their effectiveness. Thus, we expect the number of unique creatives to be positively associated with a more sophisticated use of DDC.

We define a unique creative as a unique combination of textual and audiovisual elements. A unique creative is different from other unique creatives in at least one of the available creative elements we extracted from ads, including creative body, title, link caption, creative description, audio transcription from videos, overlaid text to audiovisual media, and other non-textual audiovisual elements/imagery. All forms of textual information were obtained from the public ad libraries or extracted using audiovisual recognition tools, as was previously described. We represented the unique audiovisual information in image and video ads using the SHA-256 checksum value of the media files. Checksum is a cryptographic hash code computed from the bytes of a file that serves as the "fingerprint" of the file. Minimal variations in media content, be they textual or audiovisual, would generate different checksum values. Therefore, if two creatives share the same creative body, overlaid text, video transcription, and other textual elements but differ in checksum values, that implies they vary in audiovisual elements.

We quantify unique creatives by concatenating variables representing the textual and audiovisual elements listed above and dropping the exact duplicates. This results in 153,952 unique creatives out of 377,721 ads on Meta's platforms and 28,683 unique creatives out of 84,225 ads on Google's platforms. We then count the number of unique creatives of each federal race and sponsor.

6.2. Text Similarity

The second indicator of DDC that we consider is content similarity across a sponsor's unique creatives. We expect that a sophisticated data-driven campaign would vary its messages depending on the audience and therefore present greater creative variations. To capture this variation, we computed pair-wise similarity scores between all pairs of unique creatives at the sponsor level using a state-of-the-art semantic textual similarity measure (Reimers & Gurevych, 2019). We concatenated the identifying fields (creative body, title, caption, video transcription, optical character recognition text, checksum, etc.) and encoded them using a pre-trained sentence transformer model (Reimers & Gurevych, 2019). The pre-trained model used is "all-MiniLM-L6-v2" (https://www.sbert.net/index.html). The resulting corpus embeddings represent the creative content in each ad in the vector spaces and allow us to compute the textual similarity, represented by cosine similarity scores, between any given pairs of creatives. Unique creatives are represented by the embeddings of all the textual fields and checksum values described earlier. Each sponsor should have *N choose* 2 pairs of unique creatives. The similarity score is the cosine similarity between the vectors representing a pair of unique creatives and has a range of [0, 1]. A score of 0 indicates the least similar, and 1 the most similar. A pair of exact duplicates should have a similarity score of 1.

Because ads in different media formats (text, image, or video) can vary significantly in messaging and visual effects, we aggregated unique creatives by both sponsor and media type and calculated average pairwise textual similarity scores at the sponsor and media level.

We expect that a lower average textual similarity score for a sponsor implies higher creative variation and thus more tailoring of messages. That is, the campaign may be creating an ad to address the concerns of parents of young children, another ad directed at voters who care about the environment, and a third ad that seeks the votes of senior citizens on Medicare. By contrast, higher similarity scores suggest that a campaign, perhaps from lacking staff resources, is reusing content in multiple ads. The caveat here, of course, is that a campaign that is doing a lot of ad testing—which one might associate with DDC—might produce more ads that are high in similarity. Ultimately, we rely on some real-world examples of campaigns known for their use of DDC to validate the interpretation of similarity scores, which we show in the results section.

6.3. Measure of DDC Sophistication

Taking into account the information given by both unique creatives and content similarity, we formulate an index to measure the sophistication of a campaign's tailoring of messages. A campaign sponsor's sophistication index SI is given by:

$$SI = x' \times \left(1 - \overline{y}_{i,j}\right)$$

Here, x' represents the sponsor's normalized number of unique creatives, rescaled to a value between 0 and 1, and $\overline{y}_{i,j}$ represents the average pairwise similarity between these unique creatives. The number of unique creatives x are normalized using min-max normalization, which can be expressed as:

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)}$$

Here, x is the original value of a sponsor's number of unique creatives and $\min(X)$ and $\max(X)$ represent the minimum and maximum of the entire sample. The normalized value of x, that is x', has a range of [0, 1]. Sponsors with the lowest number of unique creatives have a value of 0, and sponsors with the largest number of unique creatives have a value of 1.

Similarly, $y_{i,j}$ is the pairwise similarity between a sponsor's unique creatives i, j for all possible pairs of unique creatives, and $\bar{y}_{i,j}$ is the mean of $y_{i,j}$. Because we expect lower average similarity to indicate more sophistication, the second factor of the index subtracts $\bar{y}_{i,j}$ from 1. $(1 - \bar{y}_{i,j})$ also has a range of [0, 1].

Therefore, the sophistication index SI also has a range of [0, 1]. The higher the value, the more sophisticated the campaign's tailoring of messages—again, an indicator of DDC. We calculated the sophistication index for each race, sponsor, and media format.

6.4. Office Being Sought and Race Competitiveness

After identifying the candidate sponsors for digital ads, we created indicators of whether a federal candidate campaigned for a House or Senate seat in the 2022 general election based on Federal Election Commission (FEC) data.

Furthermore, we measured race competitiveness using data from the final Cook Political Report predictions (Wasserman, 2022). Races were placed on a four-point scale from the least to the most competitive (1 = *safe Democratic/Republican*, 2 = *Democratic/Republican favored*, 3 = *lean Democratic/Republican*, 4 = *too close to call*).

6.5. Type of Campaign Sponsor

Ad sponsors included candidates' campaigns, political parties (e.g., National Republican Congressional Committee, Iowa Democratic Party), and other political and business groups (e.g., super political action committees [PACs], nonprofit organizations). Using data from OpenSecrets (a non-profit that tracks spending in American elections) and the FEC, we attempted to match advertising sponsors to known federal spenders, creating indicators for whether the spender was a campaign, party, or group. The latter category includes groups we identify with the help of OpenSecrets and other spenders who are not candidates and not parties but for whom we have no further information, either in the FEC data or from OpenSecrets.

6.6. Party Affiliation

Our indicators of whether an ad sponsor favored a Democratic or Republican candidate were optimized using a combination of FEC and OpenSecrets sourced data, manual coding, and sponsor-level party classifiers. For candidate sponsors, we obtained their party affiliation from the FEC data. For party and group sponsors, we manually coded their party affiliation based on the ad sponsor names (e.g., New Jersey Democrats for Universal Healthcare) and disclaimers, cross-checked with OpenSecrets data. We used the FEC-provided and manually coded party affiliation for federal spenders known to the FEC and OpenSecrets. For digital ads sponsored by organizations for which we do not have manually coded data, which amounts to 23% of the entities on Meta and 22% on Google, we relied on inferences from our sponsor-level party classifiers. The classifiers took vectorized textual fields of ads (including sponsor and page names) as features and trained with a random forest model. Validated by our manually compiled and coded party affiliation, the party classifiers achieved an overall accuracy of 85.7% for sponsors in the Meta dataset and 82.1% for those in the Google dataset.

6.7. Campaign Resources

We used sponsors' political ad spending on television, obtained from the Wesleyan Media Project, as a measure of campaign resources. The non-partisan Wesleyan Media Project provides information on political ad spending in the US based on data from an ad-tracking firm. Expenditures on traditional media, as opposed to digital ad spending, were used to avoid endogeneity since FEC reporting information would include resources used on digital advertising, and campaign finance reports are inconsistent in how they describe and report spending on advertising in general. For this reason, we draw on measures of spending on television as a proxy for the resources at any given advertiser's disposal by utilizing Wesleyan Media Project data from Kantar's Campaign Media Analysis Group. The Campaign Media Analysis Group provides estimates at the ad-level of the cost of each television ad airing based on market, station, daypart, and program information. We then merged information on total TV ad spending with our digital data. Because not all sponsors who spent on digital platforms spent on television, we limited our analysis to the 333 sponsors on Meta platforms and 368 advertisers on Google platforms who also spent on TV advertising in the model that incorporated this variable.

7. Results

7.1. Descriptive Results

We calculated the number of unique creatives for each sponsor using the method described in the previous section. The distribution of the number of unique creatives per sponsor is right-skewed due to a long tail of outside group sponsors who placed a very small number of unique creatives (see Figure 1a and 1c). The average number of creatives per sponsor was 13 on Meta and 13 on Google. However, several sponsors had several thousand unique creatives. Leading sponsors by the number of unique creatives on Meta were super PACs and Senate candidates. Sponsors with the most unique creatives on Meta include Americans for Prosperity Action (759 creatives), Colorado Senator Michael Bennet (758), the Congressional Leadership Fund (455), Georgia Senator Raphael Warnock (446), Ohio Senate candidate Tim Ryan (412), and House Majority PAC (388). Google sponsors with the most unique creatives were the Congressional Leadership Fund (720), the Democratic Congressional Campaign Committee (205), House Majority PAC (204), US House candidate Mary Peltola (190), and Colorado Senator Michael Bennet (168).

We calculated the average pairwise similarity at the sponsor-media level. The average pairwise similarity score centers around 0.80 for Meta and 0.65 for Google. Scores for Meta are left-skewed, while scores for Google are approximately normally distributed across much of the similarity scale except for a spike towards the most similar end (see Figure 1b and 1d). This bucket mostly consists of sponsors who placed a small number of highly similar unique creatives.

We examined some real-world examples of campaigns known for their use of DDC to evaluate the meaning of the average text similarity measure. For example, Georgia Senator Raphael Warnock is widely acknowledged to have run a sophisticated re-election campaign in a focal race in 2022, managed by a rising political strategist in the Democratic Party, Quentin Fulks. His campaign raised over $150 million (Giorno, 2022) and won a tight re-election race. The average pairwise similarity score of unique creatives in his Meta campaigns is lower (at 0.70 for video ads, 0.71 for image ads) than the sample median (0.80 for video ads, 0.80 for image ads).

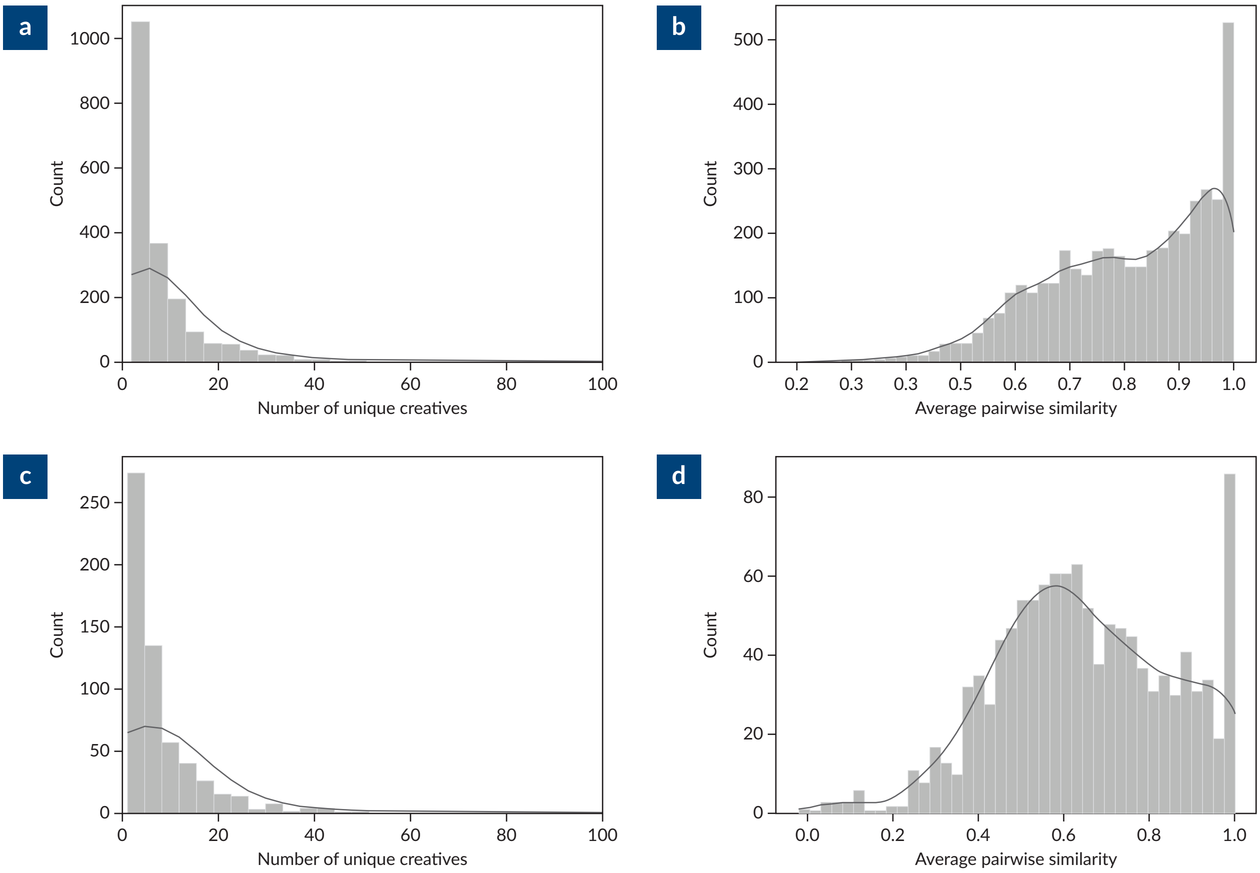

Figure 1. Distribution of unique creatives and ad similarity on Meta and Google: The number of unique creatives at the sponsor level for Meta (a) and Google (c); and distribution of average pairwise similarity at the race–sponsor–media level for Meta (b) and Google (d). Note: To illustrate the distribution of unique creatives on Meta (a) and Google (c), we limited the range of the x-axis to 100 due to the right-skewness of the distributions.

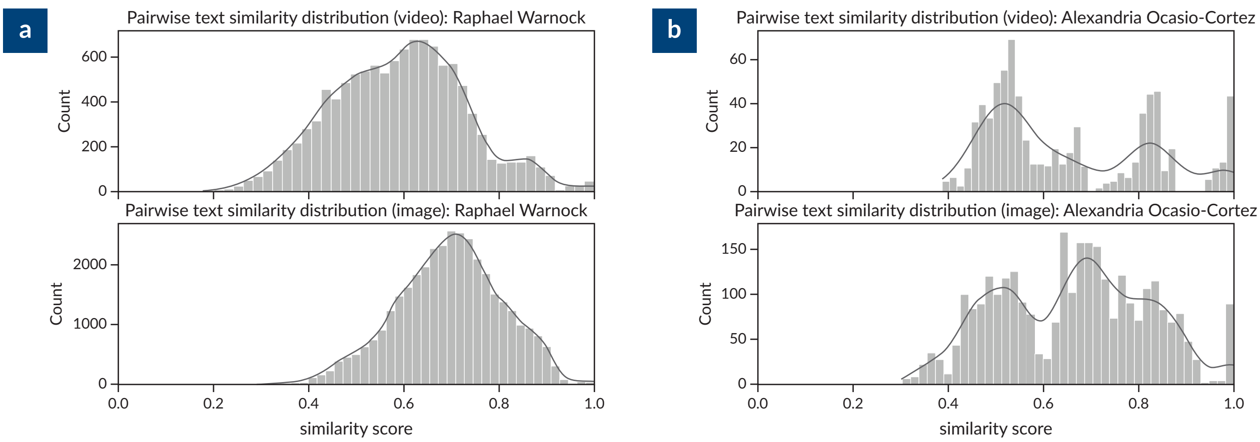

Figure 2. Distribution of pairwise content similarity of two federal candidates' digital campaigns on Meta where (a) refers to Senator Raphael Warnock and (b) to Representative Alexandria Ocasio-Cortez.

Pairwise similarity of Warnock ads has a unimodal distribution that spreads widely across the similarity scale, indicating abundant variation in those creatives (Figure 2a).

Another politician known as a sophisticated digital campaigner is New York's 14th District Representative Alexandria Ocasio-Cortez (Freedlander, 2021). Her campaign on Meta displays a multimodal distribution of pairwise similarity (see Figure 2b). Although the average similarity is lower than the sample median, there exist clusters of highly similar creatives (near-duplicates), representing minor variations around the same messages.

Therefore, average text similarity adds information but is not a sufficient indicator in itself of the use of tailoring ads. Our sophistication index takes into consideration both the number of unique creatives and average text similarity, and the top campaigns in this index align with well-funded, sophisticated, or creative campaigns known in the real world (see Tables 1 and 2). These include candidates running in highly competitive US Senate races, such as Mark Kelly in Arizona, Tim Ryan in Ohio, Mandela Barnes in Wisconsin, and John Fetterman in Pennsylvania. All of these candidates were able to raise multiple millions of dollars as well. While Senate candidates dominate the lists, many House candidates on the list are ones who ran in competitive races, such as Jeff Jackson in North Carolina and Jay Chen and Mike Garcia, both in California.

Table 1. DDC sophistication by campaign on Meta (top campaigns).

Sponsor (race)	Sophistication index	Media type	Unique creatives	Average text similarity
Raphael Warnock (Senate–GA)	0.305	Image	273	0.695
Michael Bennet (Senate–CO)	0.237	Image	239	0.729
Raphael Warnock (Senate–GA)	0.226	Video	152	0.591
Jeff Jackson (NC–14th District)	0.195	Image	166	0.679
Mark Kelly (Senate–AZ)	0.169	Image	123	0.622
Adam Schiff (CA–30th District)	0.166	Image	148	0.691
Tim Ryan (Senate–OH)	0.165	Video	112	0.592
Jay Chen (CA–45th District)	0.143	Image	94	0.580
Marco Rubio (Senate–FL)	0.135	Image	118	0.686
Mark Kelly (Senate–AZ)	0.133	Video	85	0.567
Jake LaTurner (KS–2nd District)	0.131	Image	68	0.462
Michael Bennet (Senate–CO)	0.124	Video	113	0.697
Mandela Barnes (Senate–WI)	0.123	Image	118	0.713
Elise Stefanik (NY–21st District)	0.120	Image	106	0.687
Mike Garcia (CA–27th District)	0.115	Image	117	0.729
John Fetterman (Senate–PA)	0.112	Video	74	0.579
Katie Porter (CA–47th District)	0.112	Image	107	0.711
Kim Schrier (WA–8th District)	0.108	Image	156	0.809
Catherine Cortez-Masto (Senate–NV)	0.106	Image	84	0.650
Christian Castelli (NC–6th District)	0.104	Image	84	0.657

Table 2. DDC sophistication by campaign on Google (top campaigns).

Sponsor (race)	Sophistication index	Media type	Unique creatives	Average text similarity
Raphael Warnock (Senate–GA)	0.420	Video	98	0.397
Mary Peltola (AK–1st District)	0.387	Image	140	0.613
Mark Kelly (Senate–AZ)	0.359	Video	94	0.464
Kevin Porter (FL–11th District)	0.294	Video	65	0.362
Maggie Hasan (Senate–NH)	0.209	Video	57	0.481
Michael Bennet (Senate–CO)	0.208	Video	55	0.463
Catherine Cortez Masto (Senate–NV)	0.201	Video	54	0.472
Marco Rubio (Senate–FL)	0.170	Text	41	0.409
Mary Peltola (AK–1st District)	0.158	Video	50	0.550
Mark Kelly (Senate–AZ)	0.155	Text	35	0.364
Catherine Cortez Masto (Senate–NV)	0.153	Text	32	0.315
Elise Stefanik (NY–21st District)	0.146	Image	44	0.528
Mike Garcia (CA–27th District)	0.142	Text	30	0.318
Herschel Walker (Senate–GA)	0.138	Text	49	0.594
Mark Robertson (NV–1st District)	0.135	Image	62	0.693
Josh Harder (CA–9th District)	0.134	Video	39	0.511
Katie Porter (CA–47th District)	0.134	Video	36	0.469
John Fetterman (Senate–PA)	0.132	Video	39	0.517
Danny O'Connor (OH–12th District)	0.125	Video	26	0.303

7.2. Regression Results

We estimated a model predicting the DDC sophistication index, which combines both the number of unique creatives and the similarity of creatives. We first report in Figure 3 analyses that examine only candidate sponsors on Meta platforms. The first model considers all candidates, while the second model, because it includes a measure of campaign resources that is based on television ad spending, is restricted to only those candidates who aired television ads.

We found first that Senate candidates are more sophisticated in their use of data campaigning than House candidates, consistent with H1. This supports our expectation given that Senate campaigns are typically better financed, which gives them the resources needed to tailor ads, and have larger constituencies than House candidates, which means producing unique creatives for segments of the electorate is more worthwhile.

Model estimates also show that Democratic campaigns engage in more sophisticated data campaigning than Republican campaigns, a finding that supports H2. This is consistent with Kreiss' (2016) observation that Democrats, at times, have invested more in digital technologies than have Republicans. Democrats also tend to have more support from tech elites (Broockman & Malhotra, 2017; Miles, 2002), including technology support for political campaigning (see for example Nix, 2024). Third, we found that more competitive races are associated with greater tailoring, consistent with H3. This makes sense given that the stakes are higher in more competitive races. Investment in DDC practices is not necessary if one is sure to win or sure to lose.

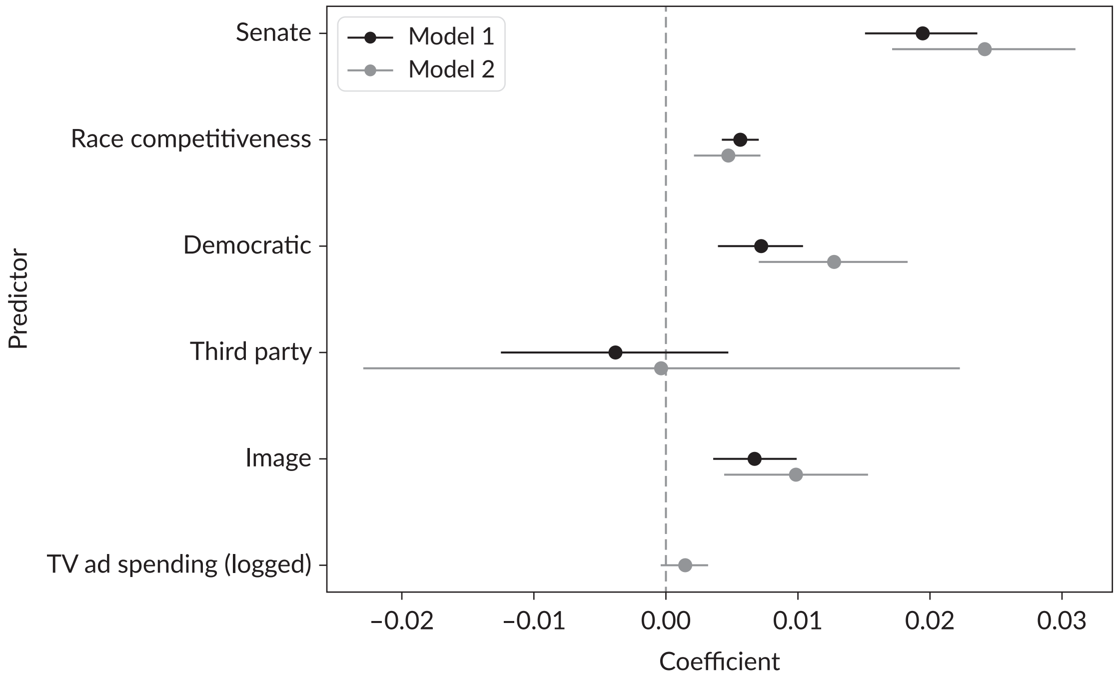

Figure 3. OLS regression estimates predicting campaign sophistication on Meta platforms (federal candidates). Notes: Error bars represent 95% confidence intervals; model 2 only analyzed the subset of candidates who invested in TV ads; full regression table can be found in Table 1 of the Supplementary File.

Our second model in Figure 3 tests the influence of campaign resources, finding that greater resources (as measured by TV ad spending) are positively associated with the use of ad tailoring. This finding is consistent with H4.

Next, we expand the analysis beyond candidates to consider groups, parties, and other sponsors on Meta as well. The unit of analysis here is the sponsor–race. Our findings, reported in Figure 4, are consistent with the analysis that included only candidates: Senate races see higher sophistication in the use of campaign data for tailoring, as do more competitive races, and Democratic sponsorship also predicts higher sophistication. Greater resources are associated with greater sophistication ($p < 0.10$), as model 2 reveals. This prompts the question of whether Meta provides a more equalizing campaigning platform and moderates the resource imbalances between well and poorly-funded candidates, as prior research has suggested (Fowler et al., 2021). Interestingly, we also find that parties and groups are less likely than candidates to engage in sophisticated ad tailoring, which makes sense given that groups and parties often must pay attention to multiple races while candidates are focused on a single one.

We next turn to the analysis of the Google data, starting with candidate-sponsored ads. Models 1 and 2 estimates (Figure 5) show that Senate races are associated with greater DDC sophistication than House races, consistent with H1. The first model also lends support to H2, as Democratic sponsors do more tailoring than Republican sponsors, though this relationship is only statistically significant at the 0.1 level in model 2, which has a smaller sample size due to the elimination of sponsors who did not advertise on television. Third, we find that greater competitiveness is associated with greater DDC sophistication, consistent with H3. Moreover, we find support for H4 in model 2, which shows that greater campaign resources positively predict greater ad tailoring on Google's platforms.

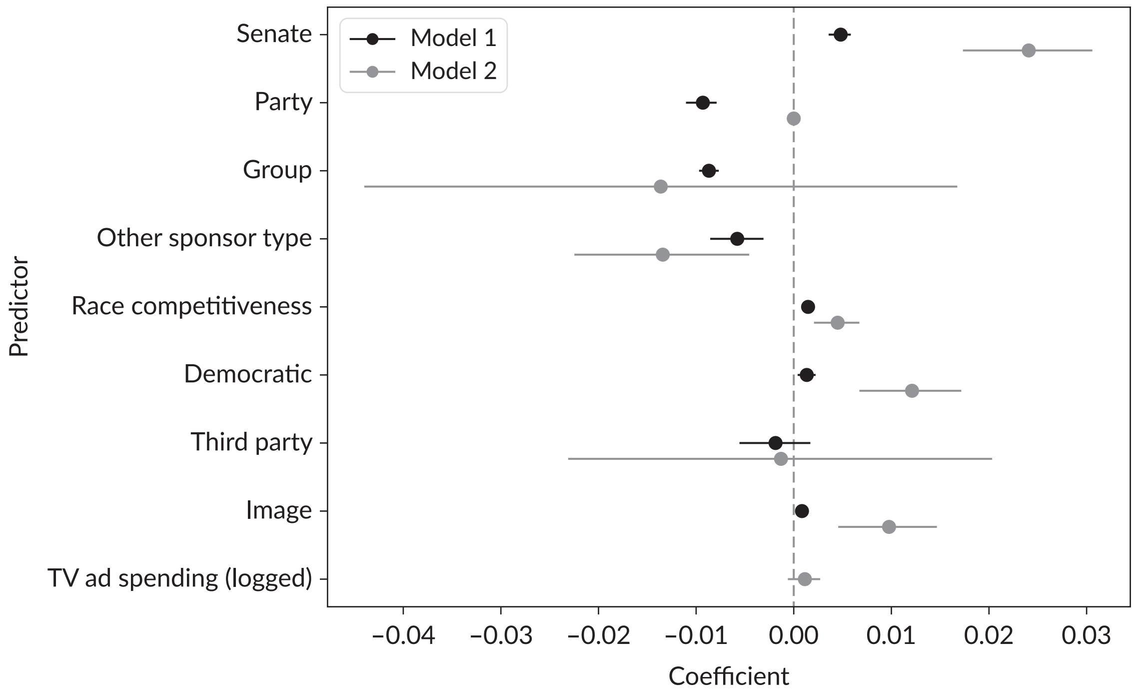

Figure 4. OLS estimates predicting campaign sophistication on Meta platforms (all House and Senate race sponsors). Notes: Error bars are 95% confidence intervals; model 2 only analyzed the subset of sponsors who invested in TV ads; ads placed by non-candidate sponsors such as national parties and groups were included separately for each race where they sponsored ads; "Other sponsor type" includes government offices and coordinated efforts between a group and campaign or party; full regression table can be found in Table 2 of the Supplementary File.

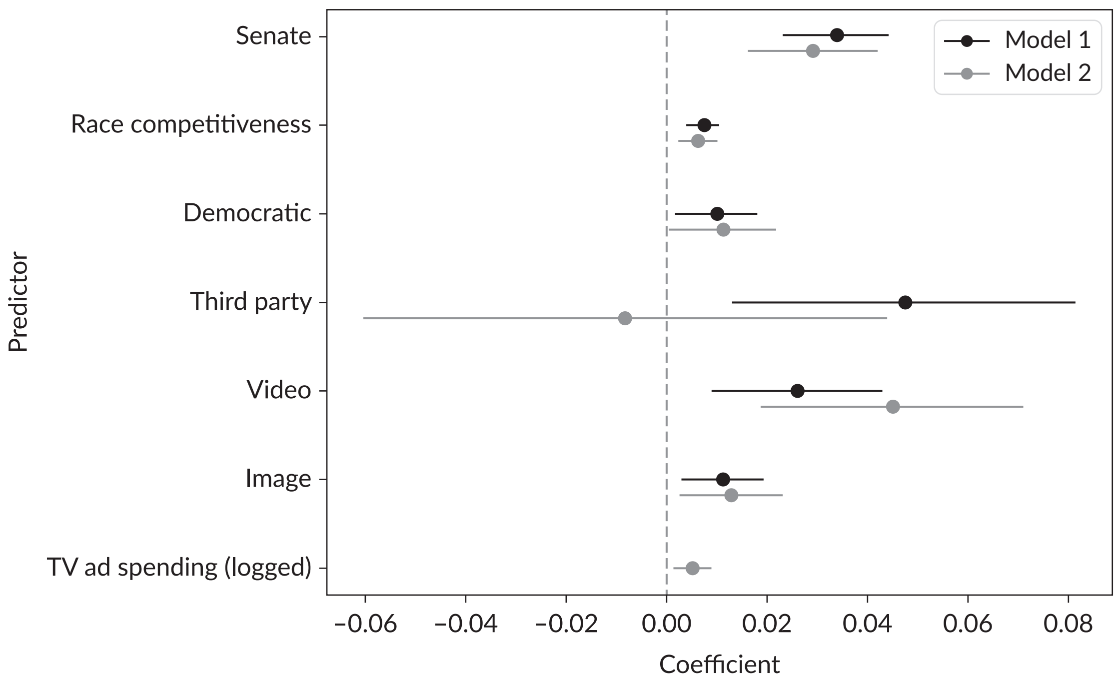

Figure 5. OLS estimates predicting campaign sophistication on Google platforms (candidates). Notes: Error bars represent 95% confidence intervals; model 2 only analyzed the subset of candidates who invested in TV ads; full regression table can be found in Table 3 of the Supplementary File.

Finally, the estimates in Figure 6 come from a model that examines all sponsors who placed ads on Google in US House and Senate races. Our story remains fairly consistent. Senate races feature more ad tailoring, and greater competitiveness is associated with more tailoring as well. Greater resources are associated with greater tailoring (model 2), and Democratic sponsors are more likely to tailor ads than Republican sponsors.

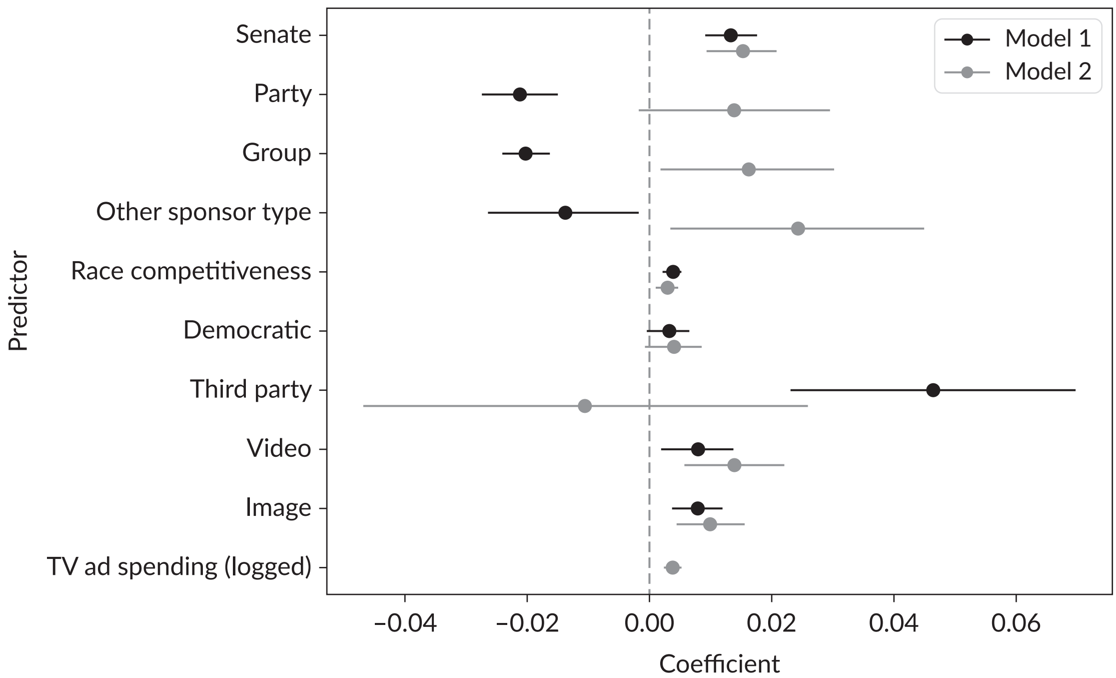

Figure 6. OLS regression estimates predicting campaign sophistication on Google platforms (all sponsors of Senate and House races). Notes: Error bars represent 95% confidence intervals; model 2 only analyzed the subset of candidates who invested in TV ads; ads placed by non-campaign sponsors such as national parties and groups were included separately for each race where they sponsored ads; "Other sponsor type" includes government offices and coordinated efforts between a group and campaign or party; full regression table can be found in Table 4 of the Supplementary File.

8. Conclusion

DDC is often studied qualitatively. Thus, studies using interviews or participant observation provide a deep understanding of how campaigns use data to make decisions, but such studies also necessarily tend to focus on a small number of cases. Our approach, which relies on publicly available data from platforms' digital ad libraries, offers a relatively simple way to study variation in ad tailoring, a key practice associated with DDC, across ad sponsors, races, parties, and even across countries. Our measure of DDC relies on the number of unique ad creatives employed by each sponsor, following the lead of Kruschinski and Bene (2022), in addition to the degree of similarity across those creatives.

Importantly, we find that tailoring of ads is more likely in Senate than House races, reflecting, we believe, both (a) the greater need to use a more data-driven approach in a larger, more heterogenous electoral district and (b) the greater practicality of making content for particular segments of the electorate when those segments are larger. We also find, in most cases, greater tailoring among Democrats and those backing Democrats, and we find greater tailoring when the race is more competitive. This suggests that when the stakes are higher—even when controlling for campaign resources—tailoring of ads is more likely. In addition, we found that having greater campaign resources predicts greater tailoring of ads. This is consistent with the conclusions of others who find an important role for campaign resources in using DDC (e.g., Dommett, Kefford, & Kruschinski, 2024; Ridout, 2024).

Obviously, these findings apply to a singular context—US Senate and House races—but one key contribution here is that our measure of ad tailoring, which we use as an indicator of DDC, could easily be generated using ad data from the dozens of countries found in the Meta and Google ad libraries, facilitating cross-campaign, cross-party, and cross-national research. This work can also complement the rich qualitative approach that is often the focus of DDC research.

The development of our sophistication index is an initial effort to quantify digital ad tailoring across larger samples, but it is not perfect. It only takes into account the degree of within-campaign variation in creative content. It is limited in characterizing campaign sophistication with respect to message quality, especially evaluated in relation to its target audience. The creative content we extracted from image, video, and text ads serves great potential for developing measures of campaign sophistication in relation to message effectiveness. For example, one could evaluate the connections between issue framing in the ads and the geographic locations and demographic groups being targeted. Investigation of the strategic use of visual communication in audiovisual ads will also enrich the evaluations of campaign sophistication. Social media platforms arguably engage in more personalized and de-professionalized political campaigning (Enli, 2017; Enli & Skogerbø, 2013), and audiovisual media is a productive vessel for personalized messages. Our regression results suggest greater ad tailoring for image and video media compared to text ads. This motivates future content analysis that looks at the stylistic choices, emotional cues, and patterns of candidate and opponent appearances in image and video ads placed on social media platforms.

It is important to understand how commonly campaigns use a more sophisticated variant of DDC and the contexts in which they do so. An understanding of this reality should inform any proposal to regulate the use of data in campaigns, such as limits on microtargeting and rules for data privacy. Knowing the prevalence of ad tailoring might compel regulators to consider additional rules for digital ad platforms, such as transparency requirements on ad targeting or within-ad disclaimer rules for ads targeted to certain constituencies. At the moment in the US, regulations on digital ads are very limited. Knowing more about how campaigns employ DDC for digital ad platforms might facilitate a conversation among policy-makers about needed rules to give citizens a clearer understanding of how election advertisers are reaching different types of voters.

Acknowledgments

We gratefully acknowledge the work of Breeze Floyd, Laura Baum, and numerous Wesleyan Media Project undergraduate research assistants for their efforts in coding advertising. We also thank Saray Shai, Sebastian Zimmeck, and Manolis Kaparakis for all of their support. Wesleyan Media Project work in 2022 was supported by the National Science Foundation under grants 2235006, 2235007, and 2235008. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Funding

National Science Foundation grants 2235006, 2235007, and 2235008.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability

Data that support the findings of this study are stored on figshare and publicly available upon user registration (https://www.creativewmp.com/data-access). Code for data analysis associated with the current submission is available at https://github.com/Wesleyan-Media-Project/quantifying-ddc-paper/tree/main

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Anstead, N. (2017). Data-driven campaigning in the 2015 United Kingdom general election. *The International Journal of Press/Politics*, *22*(3), 294–313.

Baldwin-Philippi, J. (2015). *Using technology, building democracy: Digital campaigning and the construction of citizenship*. Oxford University Press.

Baldwin-Philippi, J. (2017). The myths of data-driven campaigning. *Political Communication*, *34*(4), 627–633.

Beraldo, D., Milan, S., de Vos, J., Agosti, C., Sotic, B. N., Vliegenthart, R., Kruikemeier, S., Otto, L., Vermeer, S. A. M., Chu, X., & Votta, F. (2021). Political advertising exposed: Tracking Facebook ads in the 2021 Dutch elections. *Internet Policy Review*, *11*.

Broockman, D., & Malhotra, N. (2017, September 22). America's tech entrepreneurs are Democrats not libertarians, but they may be changing the Democratic Party. *The Washington Post*. https://www.washingtonpost.com/news/monkey-cage/wp/2017/09/22/americas-tech-entrepreneurs-are-democrats-not-libertarians-but-they-may-be-changing-the-democratic-party

Dobber, T., Trilling, D., Helberger, N., & De Vreese, C. H. (2017). Two crates of beer and 40 pizzas: the adoption of innovative political behavioural targeting techniques. *Internet Policy Review*, *6*(4), 1–25.

Dommett, K., Barclay, A., & Gibson, R. (2024). Just what is data-driven campaigning? A systematic review. *Information, Communication & Society*, *27*(1), 1–22. https://doi.org/10.1080/1369118X.2023.2166794

Dommett, K., Kefford, G., & Kruschinski, S. (2024). *Data-driven campaigning and political parties: Five advanced democracies compared*. Oxford University Press.

Enli, G. (2017). Twitter as arena for the authentic outsider: Exploring the social media campaigns of Trump and Clinton in the 2016 US presidential election. *European Journal of Communication*, *32*(1), 50–61.

Enli, G. S., & Skogerbø, E. (2013). Personalized campaigns in party-centred politics: Twitter and Facebook as arenas for political communication. *Information, Communication & Society*, *16*(5), 757–774.

Farrell, D. M., Kolodny, R., & Medvic, S. (2001). Parties and campaign professionals in a digital age: Political consultants in the United States and their counterparts overseas. *Harvard International Journal of Press/Politics*, *6*(4), 11–30.

Fowler, E. F., Franz, M. M., Martin, G. J., Peskowitz, Z., & Ridout, T. N. (2021). Political advertising online and offline. *American Political Science Review*, *115*(1), 130–149.

Freedlander, D. (2021). *The AOC generation: How millennials are seizing power and rewriting the rules of American politics*. Beacon Press.

Giorno, T. (2022). *Walker-Warnock U.S. Senate race in Georgia most expensive in 2022 cycle as runoff intensifies*. OpenSecrets. https://www.opensecrets.org/news/2022/11/walker-warnock-u-s-senate-race-in-georgia-most-expensive-in-2022-cycle-as-runoff-intensifies

Kefford, G. (2021). *Political parties and campaigning in Australia*. Springer.

Kefford, G., Dommett, K., Baldwin-Philippi, J., Bannerman, S., Dobber, T., Kruschinski, S., Kruikemeier, S., & Rzepecki, E. (2023). Data-driven campaigning and democratic disruption: Evidence from six advanced democracies. *Party Politics*, *29*(3), 448–462.

Kreiss, D. (2016). *Prototype politics: Technology-intensive campaigning and the data of democracy*. Oxford University Press.

Kreiss, D., & Jasinski, C. (2016). The tech industry meets presidential politics: Explaining the democratic party's technological advantage in electoral campaigning, 2004–2012. *Political Communication*, *33*(4), 544–562.

Kruschinski, S., & Bene, M. (2022). In varietate concordia?! Political parties' digital political marketing in the 2019 European Parliament election campaign. *European Union Politics*, *23*(1), 43–65.

Maarek, P. J. (2011). *Campaign communication and political marketing*. Wiley.

Miles, S. (2002). *How to hack a party line: The Democrats and Silicon Valley*. University of California Press.

Montigny, E., Dubois, P. R., & Giasson, T. (2019). On the edge of glory (...or catastrophe): Regulation, transparency and party democracy in data-driven campaigning in Québec. Internet Policy Review, *8*(4), 1–19. https://doi.org/10.14763/2019.4.1441

Munroe, K. B., & Munroe, H. D. (2018). Constituency campaigning in the age of data. *Canadian Journal of Political Science/Revue Canadienne de Science Politique*, *51*(1), 135–154.

Nix, N. (2024, June 10). Tech workers use corporate advertising tricks to turn out Democratic voters. *The Washington Post*. https://www.washingtonpost.com/technology/2024/06/10/start-up-advertising-voters-democrats-election-2024

Reimers, N., & Gurevych, I. (2019). *Sentence-bert: Sentence embeddings using siamese bert-networks*. arXiv. https://doi.org/10.48550/arXiv.1908.10084

Ridout, T. N. (2024). Political advertising and data-driven campaigning in Australia. *Media and Communication*, *12*, Article 8462. https://doi.org/10.17645/mac.8462

Vliegenthart, R., Vrielink, J., Dommett, K., Gibson, R., Bon, E., Chu, X., de Vresse, C., Lecheler, S., Matthes, J., Minihold, S., Otto, L., Stubenvoll, M., & Kruikemeier, S. (2024). Citizens' acceptance of data-driven political campaigning: A 25-country cross-national vignette study. *Social Science Computer Review*, *42*(5), 1101–1119. https://doi.org/10.1177/08944393241249708

Wasserman, D. (2022). *Final House ratings: A clear GOP edge, and a wide range of possible gains*. The Cook Political Report. https://www.cookpolitical.com/analysis/house/house-overview/final-house-ratings-clear-gop-edge-and-wide-range-possible-gains

About the Authors

Michael M. Franz is professor of Government and Legal Studies at Bowdoin. He studies elections, campaign finance, and interest groups. He is author or co-author of five books, including *Political Advertising in the United States* (2021, Routledge). He is also the co-director of the Wesleyan Media Project, a collaboration of researchers that tracks spending on political advertising.

Meiqing Zhang is a postdoctoral researcher at Wesleyan Media Project and Quantitative Analysis Center, Wesleyan University. She studies political communication, political culture, and computational social science.

Travis N. Ridout is Thomas S. Foley distinguished professor of Government and Public Policy in the School of Politics, Philosophy, and Public Affairs at Washington State University. He also serves as co-director of the Wesleyan Media Project. His research focuses on political advertising, elections, and campaigns.

Pavel Oleinikov is an associate professor of the practice at the Hazel Quantitative Analysis Center at Wesleyan University in Middletown, Connecticut, and a member of the Wesleyan Media Project. His areas of work include computational social science, data engineering, and remote sensing.

Jielu Yao is a research fellow at the East Asian Institute of the National University of Singapore and was a postdoctoral researcher at the Wesleyan Media Project. Her research interests include political methodology, computational political science, and political communication.

Furkan Cakmak is an independent scholar. Previously, he was a postdoctoral researcher in quantitative analysis at Wesleyan University where he worked with the Wesleyan Media Project, which tracks political advertising, and collaborated with the Quantitative Analysis Center.

Erika Franklin Fowler is professor of Government at Wesleyan University and co-director of the Wesleyan Media Project. She specializes in large-scale analyses of political and health-related communication, and her work on the content and effect of messaging has been published in political science, communication, law/policy, and medical journals.