



Is first- or third-party audience data more effective for reaching the ‘right’ customers? The case of IT decision-makers

Nico Neumann¹ · Catherine E. Tucker² · Kumar Subramanyam³ · John Marshall³

Received: 15 July 2022 / Accepted: 27 July 2023 / Published online: 24 October 2023
© The Author(s) 2023

Abstract

Often marketers face the challenge of how to communicate best with the customers who have the right responsibilities, influence or purchasing power, especially in business-to-business (B2B) settings. For example, B2B marketers selling software and IT need to identify IT decision-makers (ITDMs) within organizations. The modern digital environment in theory allows marketers to target individuals in organizations through specifically designed third-party audience segments based on deterministic prospect lists or probabilistic inference. However, in this paper we show that in our context, such ‘off-the-shelf’ segments perform no better at reaching the right person than random prospecting. We present evidence that even deterministic attribute information is flawed for ITDM identification, and that the poor campaign results can be partly linked to incorrect assignment of established prospect profiles to online identifiers. We then use access to our publisher network data to investigate what would happen if the advertiser had used first-party data that are less susceptible to the identified issues. We demonstrate that first-party demographics or contextual information allows advertisers and publishers to outperform both third-party ITDM audience segments and random prospecting. Our findings have implications for understanding the shift in digital advertising away from third-party cookie tracking, and how to execute digital marketing in the context of broad privacy regulation.

Keywords Digital targeting · User profiling · Data brokers · Third-party data

JEL classification M3 · O3 · L8

✉ Nico Neumann
n.neumann@mbs.edu

Extended author information available on the last page of the article

1 Introduction

Deploying targeted communication to reach very specific prospects and customers has always been one of the main challenges of marketing. This is particularly true for a business-to-business (B2B) context, where marketers must identify individuals who actually have decision-making authority in an organization with potentially thousands of employees. To execute their targeted branding and advertising campaigns, B2B companies have therefore progressively used digital channels: An eMarketer report (Ryan, 2020) estimates that B2B marketers spent \$8.14 billion on digital ads in 2020, up 22.6% from the previous year.¹

The increase in digital ad spending both in B2B and business-to-consumer [B2C] settings has been fuelled by the emergence of ‘off-the-shelf’ segments that marketers can now buy from information aggregators, publishers and digital platforms to target a variety of customer audiences. For example, B2C marketers tend to purchase demographic, behavioral or intent data for distinct categories (e.g. ‘in market for cars’). In a similar way, B2B marketers can buy specifically-tailored segments (e.g. ‘health-care professionals’ or ‘IT decision-makers’) derived from firmographics or synthesized name lists based on data partnerships. Collectively, this type of audience segments, created from multiple sources and provided by external parties, is referred to as ‘third-party data.’

While third-party data has created multiple new targeting opportunities for marketers and a flourishing market for data brokers, recent studies demonstrate that there is high performance heterogeneity across leading audience data vendors and networks, and that the average profile accuracy is often disappointing (Trusov et al., 2016; Neumann et al., 2019). Neumann et al. (2019) further recommend considering the odds of naturally finding the target audience in the population (without targeting), and show that the extra costs of buying third-party data are often not justified for the most common B2C segments.

However, despite an emerging literature stream on digital profiling and targeting, several critical questions remain unanswered. First, Neumann et al. (2019) focus on the most popular B2C segments (e.g. gender or ‘sports interest’) where the odds of finding the right individual by chance are still quite high (i.e. 50% for gender or 67% for sports interest, page 923). The cost-benefit trade-off for targeting in such scenarios may be unfavourable for third-party data usage, but most marketers will wonder: How does third-party data fare (economically) for more specific audiences that are rarely found in the population and harder to define?

Second, the existing profiling and data-targeting studies do not distinguish between different types of third-party data. Some vendors offer deterministic profile data, while others use probabilistic inference to create or enrich profile attributes (De Bruyn & Otter, 2022).

Third, while prior studies demonstrate considerable variance in data-targeting performance, little is known about why. What factors influence successful identification of individuals of interest beyond algorithmic inference issues? In particular, marketers

¹ According to the eMarketer report, the trend to rely more on B2B targeted ads began pre-pandemic, but increased during the global COVID-19 outbreak as more commerce moved online and in-person meetings became infeasible.

need to know to what degree targeting outcomes are influenced by inaccurate [inference of] profiles versus correct profiles that were assigned to the wrong individual in the media buying process (De Bruyn & Otter, 2022).

Finally, all marketers – B2B and B2C – face the challenge of how best to communicate with the desired audiences when third-party targeting becomes less feasible due the phasing out of third-party web cookies. This raises the question of what alternative targeting methods may be the most promising alternatives in a privacy-centred world.

Our research addresses these four research gaps. We first use a large-scale field study in the United States to analyze the accurate reach of purely deterministic data segments versus (black-box) probabilistic data segments for a B2B context. Specifically, we focus on the tech sector — the largest and fastest growing B2B category in 2020 (Lebow, 2021) — and investigate the proportion of IT decision-makers (“ITDMs”) that can be reached when relying on third-party data targeting on a digital publisher network. To establish the efficacy of third-party targeting for context, we benchmark the performance results of two deterministic and three probabilistic data segments from leading data brokers with random prospecting on the same digital publisher network (our baseline without targeting).

Second, while we find that deterministic data segments perform marginally better than probabilistic data segments, our results show that both types of these third-party audiences do not lead to better outcomes than random prospecting for our context. Since deterministic data is not subject to problematic probabilistic inference as shown in prior studies (Trusov et al., 2016; De Bruyn & Otter, 2022), we also analyze our deterministic data in more detail. We find that essential information pieces of prospects, such as the job function or location, are not correct for the majority of people.

Third, we investigate what factors influence the correct digital identification of ITDMs and individuals in general. We present suggestive evidence that poor third-party data targeting can be linked to wrong cookie-identifier matching, which is a fundamental step when using deterministic data segments. We further show that the likelihood of identifying an ITDM depends on people’s background, firm size, and digital footprint, such as availability of online signals and use of devices.

Fourth, we use the results of our findings on factors affecting identification of ITDMs to examine the performance of potential alternative targeting strategies that would rely on first-party publisher network data. Such a strategy has the advantage of not relying on third-party cookies and is less prone to the identified issues around possibly obsolete or flawed deterministic data and cookie-matching. In particular, we show that targeting based on first-party demographic data and content-interest (contextual) information that publishers could gather themselves from consumers (with consent) outperforms specifically-designed third-party segments at finding ITDMs (both in terms of efficacy and efficiency).

Our work offers three main contributions. First, to our knowledge, this is one of the first papers to study digital advertising for targeting of B2B audiences, using one of the most popular B2B categories (information technology). In such situations, where the odds of reaching the right individuals preclude the use of mass marketing, managers are highly dependent on successful targeting and require guidance on the most promising digital strategies.

Second, our research contributes to the academic literature on data quality and success drivers of digital profiling and targeting (Murthi and Sarkar, 2003; Trusov et al., 2016; Lin & Misra, 2022). We present evidence for digital targeting issues beyond probabilistic inference and show that fragmentation bias does not only affect marketing measurement and attribution, but also branding and advertising campaigns (Lin & Misra, 2022). Our investigation into probabilistic and deterministic data segments and what people- and process-specific factors generally affect targeting campaigns, offers practical guidelines for marketers and media planners.

Third, our research presents useful insights into a world of ad targeting without third-party cookies. Marketers and regulators require knowledge of effective alternatives to current targeting practices. The driving force is the current privacy initiatives from leading technology companies, such as Apple Intelligent Tracking Prevention or the phasing out of third-party cookies by many internet browsers, that have and continue to create an environment where the currently dominant targeting tactics based on third-party data become less and less feasible.

Our paper is structured as follows. We first provide a literature review about customer targeting, while outlining our research method for investigating ITDMs in the subsequent section. We then summarize the descriptives for our sample as well as the reach-accuracy analyses for our initial targeting field tests (Study 1). Next, we present an investigation of deterministic data accuracy using two additional studies. Study 2 validates the deterministic segment information using other available online sources, while Study 3 examines the profile-cookie matching process on campaign outcomes in another field test with profile onboarding services. In Study 4, we investigate which people and process factors help describe being an ITDM using correlational analysis. This is followed by our final test (Study 5) of alternative demographic and content-related targeting tactics based on publisher first-party data. We conclude by presenting a brief benefit-cost analysis and a discussion of why our results matter to marketing managers and policymakers, including possible limitations and future research directions.

2 Contribution to the existing literature

Targeting diverse customer groups with differential promotional and advertising elements has played a key role in tactical marketing plans for a long time. Therefore, it is no surprise that there is a multitude of studies documenting the benefits of targeting for marketing communications (Chen et al., 2001; Narayanan & Manchanda, 2006; Goldfarb, 2014). Following the early work on why marketers may want to use targeting for communication efforts, a literature stream on the efficiency of customer targeting and customized content related to digital marketing emerged (Bucklin & Sismeiro, 2003; Ansari & Mela, 2003; Park & Fader, 2004; Manchanda et al., 2006; Hauser et al., 2009). This body of work expanded subsequently into all aspects of technology-driven targeting, such as contextual targeting (Goldfarb & Tucker, 2011), behavioral targeting (Summers et al., 2016), keyword targeting (Li et al., 2016), mobile targeting (Luo et al., 2014; Andrews et al., 2016; Chenet et al., 2017) and re-targeting (Lambrecht & Tucker, 2013; Johnson et al., 2017).

During the last decade, another wave of papers emerged, examining the specific pitfalls and opportunities of collecting and selling data based on consumers' browsing behavior. Bergemann and Bonatti (2015) develop an analytical data pricing model for selling the information encoded in third-party web cookies. Trusov et al. (2016) demonstrate that consumer profiling can be biased because many ad networks and data aggregators only obtain a partial view of customers' behavior across websites, leading to incorrect inferences about people's true website visiting behavior. The authors develop a probabilistic method of using cookie data to improve user profiling via a bias correction, allowing online businesses to make profile predictions when limited information is available. In a similar vein, De Bruyn and Otter (2022) present a Bayesian profiling approach to augmenting customer segments based on aggregate data. Coey and Bailey (2016) and Lin and Misra (2022) focus on a different aspect of online data issues and shed light on the consequences of identity fragmentation for successful customer identification and the resulting bias. Miller and Skiera (2017) and Johnson et al. (2020) investigate the value of cookies for publishers. Neumann et al. (2019) examine the accuracy of consumer profile data and document a strong heterogeneity in data accuracy, even among leading data providers for consumer ad targeting.

The slowly growing literature on digital segments and consumer profiling has focused on the consumer-related segments and broad business-to-consumer (B2C) contexts. However, often marketers wish to use data targeting to reach very specific audiences to find customers which can influence or make purchase decisions. This is particularly true for the case of business-to-business (B2B) settings, as the odds of finding the right person (e.g. key business decision makers in organizations) naturally in the population tend to be much lower than the odds of finding people with general consumer attributes (e.g. 'sports interest').

The seminal studies of Trusov et al. (2016) and De Bruyn and Otter (2022) revolve around probabilistic inferences of profile attributes, but less is known about the performance of deterministic data segments as well as other potential error sources of profiling and targeting. Most importantly, to leverage any established profile attribute of a person for media campaigns, the attribute information needs to be matched to the right individual using some identifier (e.g. web cookies). Prior research has demonstrated the far-reaching consequences of consumer identity fragmentation for marketing attribution and incrementality analyses, a problem that has been termed 'fragmentation bias' (Lin & Misra, 2022).

Finally, all marketers need to be aware that the trend towards privacy protection has led to the phasing out of third-party cookies by many key internet browsers, such as Mozilla Firefox, Apple Safari and Brave. Google, whose browser 'Chrome' has a market share of more than 65% (Nagpal, 2021), will also cease supporting third-party tracking (Google, 2021b, a). As a result, many of the third-party data targeting tactics that marketers use today will not be feasible anymore.

In sum, essential questions about efficient digital advertising tactics still remain unanswered:

1. How effective is third-party data targeting for audiences that are less common to find in the population?

2. How effective is deterministic versus probabilistic data?
3. What factors influence data targeting performance beyond probabilistic inference abilities?
4. What are the most promising targeting tactics for a future without third-party cookies and limited customer tracking?

Our study addresses these four questions in five studies that build on each other. Specifically, our work focuses on IT decision-makers (ITDMs), a widely used audience for which many data vendors have created specific ‘off-the-shelf’ third-party segments that marketers can buy.² The popularity and availability of ITDM segments allows us to investigate comparable audience data from multiple vendors and for deterministic and probabilistic data segments. In the next section, we outline our empirical focus and methodological details.

3 Study 1: examining third-party audiences for ITDM

The research objective of the first study study is to shed light on the efficacy of ITDM third-party, ‘off-the-shelf’ audience segments for a typical online advertising campaign. To be able to conduct a field test in which we can examine ITDM audiences in a realistic setting, we collaborated with a globally leading IT-product and -service brand in the United States from September 2019 until April 2020. The campaign goal for our research context was to reach the appropriate B2B audience (i.e. ITDM) in the domestic market using digital media for branding and prospecting purposes.

3.1 Audiences data types: deterministic versus probabilistic data

We selected five of 10 of the most reputable third-party B2B data brokers which offer specific ITDM audience data for the U.S. While we cannot reveal the identities of the data vendors – we will only refer to vendors A, B, C, D, E – it should be noted that the five selected firms are globally leading providers of B2B audience data, such as Oracle, Bombora, V12 Data, LiveRamp, DemandScience or 180bytwo.³

Two of the selected data brokers in our study, vendors A and B, provided so-called deterministic data for ITDMs. Deterministic data clearly identifies a person, for example through names or email addresses, and is typically directly provided by consumers. In our context, we obtain detailed ITDM lists from the two brokers, including name, location, role, e-mail and/or cellphone information for about 24 million people (e.g. Joe Average, Software Procurement Specialist at Delta, 5 East Ave, NY 10014, J.Average@delta.com).

Because deterministic data refers to a person’s offline profile and includes some personal identifier (mobile number, email or name with postal addresses) but the ecosystem of media-buying relies on online identifiers (e.g. web cookies), one

² The tech sector is the largest B2B advertising vertical, with an over 30% share of US B2B digital ad spend and one that has shown 50.3% annual growth in 2020 (Lebow, 2021).

³ We list these vendors as examples of leading data providers. There is no suggestion that all of these vendors were part of our study.

needs to link these two types of identifiers. In other words, media campaigns using deterministic-data segments also require an ‘onboarding’ process (also called ‘activation’) that matches the respective offline profile to web cookies or device identifiers that can be used to recognize the people of interest online. In our case, we want to be able to show ads to the users deemed as ITDM prospects once these enter a publisher website that is part of our campaign. Onboarding services are offered by companies with a large online reach and rich databases that allow linking different identifiers in scale. Often this onboarding is done via customer data or data management platforms (e.g. ZeoTap or Lotame), which have been designed to manage all marketing data for online campaigns and are integrated with demand side platforms (i.e., software to buy ads online, such as the Trade Desk). Many demand side and ad buying platforms also offer onboarding offline data without the need to first use a specialised platform just for data ingestion purposes. For example, Google and Facebook have their own login data as well as a sufficient market reach that allow providing onboarding services for deterministic data in order to buy ads from their inventory using non-cookie identifiers. In addition, there are third-party onboarding solutions that are available across several platforms and just offer identity matching as a standalone service (e.g. Signal or LiveRamp). We will return to the specific influence of the onboarding process on accurate targeting in Study 3, but for now merely wish to mention this essential process step for online targeting based on deterministic data.

The other three data vendors (C, D, E) in our field test provided probabilistic ITDM audiences. In their cases, we did not obtain any personal information, but only the respective cookies that the three vendors mark as belonging to the ITDM segments. Probabilistic data is based on different signals (e.g. IP location, operating system or page views) that are synthesized from various online sources and analyzed using proprietary black-box methods. Cookies are then grouped into segments based on their likelihood that a user may possess a certain attribute. The exact classification method is considered a trade secret and normally not made public. We highlight that probabilistic segments could in theory have some deterministic data elements that were used to infer the desired attribute that is offered to marketers. However, the key characteristic of probabilistic data segments is that the attribute of interest to marketers is still statistically derived (e.g. I obtain different information pieces and infer using machine learning or heuristics that a user is male), while deterministic data segments do not require this attribute-creation step (e.g. someone indicated in a database to be male and a broker obtained access to this database).

3.2 Validation method and instrument design

Because of the campaign goal of branding among likely prospects, the key metric for our test campaign is ‘reach,’ that is, whether the target audience of ITDMs, and not someone else, is exposed to display ads on a specific website that is part of a media campaign. To quantify reach of the right customers, we measure the proportion of actual ITDMs among the sample of targeted web cookies for different targeting tactics we pursue in our digital campaigns.

To identify the proportion of ITDMs for each of the five purchased ITDM segments and a baseline, we use a survey design and user-reported data to find out about people's ITDM responsibilities. This process is in line with previous work investigating the behavior and types of online users who were reached or are reachable in online campaigns (Trusov et al., 2016; Neumann et al., 2019). Self-reported data can be subject to errors, but using alternative digital metrics like click-through rates would not tell us whether the right individuals with IT responsibilities were reached. Any consumer may click on an ad, but that consumer might not have any influence or purchasing power in an organization, making that B2B ad unsuccessful.

The nature of a B2B setting, where multiple stakeholders tend to be involved, also requires a flexible definition of ITDMs, which is a vague label summarizing several business functions, often spread across different stakeholders. Therefore, we cannot employ a single binary measure that simply asks whether someone is an 'IT decision-maker.' Instead, we ask people three questions about possible IT responsibilities in their company. The three questions, developed in collaboration with IT consultants and operations specialists, cover the most essential functions of IT procurement:

1. IT needs identification: "I identify needs for new IT products in my company/department."
2. IT vendor selection: "I select/shortlist vendors for IT purchases in my company."
3. IT contract responsibility: "I sign contracts/make financial decisions for IT purchases in my company."

Since employees' involvement in IT purchases is rarely a single job and often involves team work, we ask respondents to rate how much each of the three tasks is their responsibility, using a five-point scale ("This is my main responsibility/significant portion of my role/half of my role/minor portion of my role/not part of my role.")

We also aim to examine factors that can influence the data quality of digital segments and describe our outcome of interest, which is why we collect data on several covariates. These cover basic demographic information (e.g., age tier, gender); work-related characteristics and firmographics (e.g., firm size, job role); and variables related to digital footprint (e.g., having a work laptop, online behavior, the browser linked to the cookie/used to complete the survey)⁴ and content interest (which online topics someone reads). More information on the questionnaire can be found in Appendix A.

3.3 Field test methodology

In the case of niche or highly specific audiences (such as being an ITDM), it is not always possible to obtain the natural distribution of the audience characteristic in the population to establish a baseline comparison for reaching the desired people, as Neumann et al. (2019) did to assess the targeting effectiveness of data providers. Moreover, marketers can rarely reach the entire population of a country leveraging online media. Marketers rather select for an online campaign a group of publishers and apps where they wish to show their ads. Thus, another plausible baseline for comparing the efficacy of any targeting tactic is simply the outcome that can be achieved randomly (without

⁴ The browser can be tracked online without the need to use a survey question.

targeting) for a given set of publishers where one wishes to buy media placements. We can then compare the proportion of reached target users (ITDMs) of different audience data vendors for the same publisher group in comparison to the benchmark of using no targeting (i.e., not buying third-party audience data in addition to the media placements). This approach allows controlling for the unique online population of the selected publishers of the ad campaign, which is often likely to be different from the U.S. census data.

The chosen baseline for our field test campaign is a selection of websites and apps in a digital publisher network. The publisher partners include a diverse group of content providers of all sizes, such as Time Magazine, Design Home, basketball-GM.com, JoyBits and Glu Mobile. We highlight that such a true field test of audience targeting always allows only relative comparisons of efficacy in relation to the baseline (the natural reach that can be achieved by the chosen publishers for the media campaign). In other words, we benchmark our five ITDM targeting tactics by establishing the outcomes above and beyond what the network of publishers would achieve without buying additional ITDM data segments (which would also incur additional data costs).

To be able to validate the reach of the baseline (random prospecting) and the reach of ITDM audiences (vendors A to E) on the digital publisher network, we need an advanced technological setup that only works if two conditions are met. First, the publisher network needs to be able to recognize which cookie was tagged as an ITDM by any of the five vendors. To achieve this, the publisher network needs to be able to sync their own cookies with as many cookies of the ITDM data vendors as possible (the cookies serve as identifiers and must be matched). But since our digital test publisher network does not have any direct integration with the respective data brokers, the advertiser (our IT-product and service brand) needs to step in as an intermediary and add a tracking pixel of the digital publisher network to their campaign tracking code. This step allows the digital publisher network to sync their cookies with any cookies the advertiser can access during their media campaign. Next the advertiser buys audience data from each ITDM vendor for as many websites as possible in an online advertising campaign. Thus the advertiser helps our test publisher network identify the cookies that belong to the total pool of cookies or users that each of the five vendors can access and deemed to be an ITDM.⁵ We highlight that this initial cookie-tagging campaign is independent of the ultimate test campaign (i.e. targeting people on our selected digital publisher network) that we try to use for our reach validation. The latter is a sample of each data broker's cookie pool across all their publishing partners across the U.S.. That is, only those vendor-tagged users who first were reached in the initial advertising campaign and then visit the content providers of our chosen network will be eligible for our survey. This step ensures that we can establish the reach of each vendor only for our chosen publishers where we wish to run (and test) the ultimate ad campaign.

Second, the respective publishers of our validation campaign need to be able to show surveys to the users who visit the app or website. To be able to serve our validation survey to the tagged or random users of the digital publisher network, we rely on

⁵ This can be an expensive requirement and its success depends on many unknown factors, such as the match rate of the population of cookies from one vendor and the desired number of cookies one wishes to identify on the chosen publisher network. To provide a representative large cookie pool sample for ITDM in the US, the advertiser bought over 102.6 million ad impressions for our study.

an online measurement company that is integrated with the backend of the network and uses ‘surveywall technology’ to collect user information. Surveywall technology enables finding the same user via web cookies across integrated publishers and works as follows.

The partners “lock” content which is only made accessible to visitors who complete a survey. Popular content locking options include ebooks, audio or article content, free Wifi-access, or an “ad-free” experience on the website. Users who were previously shown ads and hence were marked as ITDMs by data vendors were then shown our ITDM survey in the surveywall once these users visited a partner of the network.⁶

In contrast to the five purchased ITDM segments, for our baseline group, we served our survey randomly to online users of the network partners. It should be noted that the integration of the measurement company with publishers means we can replicate how ad serving process would look like for a real publisher network without actually serving an ad. This replicates some of the motivation behind the Ghost Ads methodology (Johnson et al., 2017), which similarly was focused on saving media costs. All survey answers were collected anonymously.

3.4 Descriptive statistics of field test sample

We were able to collect 1,249 responses across our five ITDM segments and 600 responses for our baseline group via the publisher network. Table 1 summarizes the most important descriptive statistics for our sample, cross-tabulated for the six groups. First, we find that between 3.7% and 8% of the respondents are senior executives, between 9.5% and 16.2% are managers, between 7.1% and 11.9% are associates/analysts, between 2.1% and 7.1% have an entry-level position, between 1.2% and 7.1% are interns/casuals/part-time workers, between 1.8% and 11.0% are students and between 48.3% and 67.3% indicated that they were not employed at the time of the survey.

Overall, the sample distribution of job roles appears reasonable, but the high proportion of ‘currently not employed’ (including our baseline) seems to stand out. While we did not specifically track ‘retirees,’ examining age tiers for this group suggests that many respondents may be retirees. This seems to be at least the case for the five purchased B2B segments, where a large percentage is 65 or older.

In addition, it is likely that several respondents may not want to reveal information about their job function or seniority and select ‘currently not employed’ to protect their privacy in this regard.⁷ We acknowledge that there is a potential survey response bias that is unavoidable for this type of research and campaign goal.

Since our main interest lies in benchmarking the relative reach of various digital ad targeting tactics (in comparison to the baseline), our conclusions about which

⁶ If users reload the page or do not complete the survey several times, then they are shown alternative options to access the content.

⁷ Because our data collection period overlapped with the beginning of the COVID19 pandemic, we also checked whether this could explain this pattern. However, comparing the March and April 2020 data with the earlier months showed no significant differences.

Table 1 Descriptive statistics of sampled online users and their digital behavior

| Audience | Sample (n) | Senior Executive (%) | Manager (%) | Analyst/Associate (%) | Entry level (%) |
|----------|------------------------------|----------------------|------------------------------|-----------------------|-------------------------|
| A | 565 | 8.0 | 13.5 | 11.9 | 6.0 |
| B | 81 | 3.7 | 13.6 | 7.4 | 3.7 |
| C | 42 | 0.0 | 9.5 | 7.1 | 7.1 |
| D | 170 | 5.3 | 11.2 | 7.7 | 5.9 |
| E | 391 | 5.4 | 10.2 | 9.2 | 2.1 |
| Baseline | 600 | 3.8 | 16.2 | 10.5 | 5.3 |
| Audience | Intern/Casual/ Part-time (%) | Student (%) | Not currently employed * (%) | Works in IT (%) | Works in operations (%) |
| A | 5.13 | 5.49 | 50.1 | 9.9 | 17.5 |
| B | 1.23 | 4.94 | 65.4 | 8.3 | 25.0 |
| C | 7.14 | 4.76 | 64.3 | 0.0 | 15.4 |
| D | 3.53 | 3.53 | 62.9 | 5.2 | 21.1 |
| E | 4.09 | 1.79 | 67.3 | 6.6 | 20.7 |
| Baseline | 4.83 | 11 | 48.3 | 6.1 | 25.0 |
| Audience | Small firm (%) | Mid-sized firm (%) | Large firm (%) | Male (%) | Female (%) |
| A | 37.1 | 27.1 | 35.8 | 41.8 | 58.2 |
| B | 45.9 | 12.5 | 41.7 | 30.9 | 69.1 |
| C | 46.2 | 7.7 | 46.1 | 42.9 | 57.1 |
| D | 38.5 | 29.9 | 31.6 | 37.6 | 62.4 |
| E | 45.5 | 25.6 | 28.9 | 44.5 | 55.5 |
| Baseline | 33.7 | 32.4 | 34.0 | 48.7 | 51.3 |
| Audience | Age 18-24 (%) | Age 25-34 (%) | Age 35-44 (%) | Age 45-54 (%) | Age 55-64 (%) |
| A | 4.8 | 9.2 | 10.3 | 14.9 | 25.5 |
| B | 7.4 | 6.2 | 2.5 | 11.1 | 29.6 |
| C | 7.1 | 4.8 | 2.4 | 14.3 | 28.6 |

Table 1 continued

| | | | | | |
|----------|---------------------------|-------------------------|------------------------------|-----------------------------|-----------------------------|
| D | 2.4 | 7.6 | 5.9 | 12.9 | 30.6 |
| E | 1.8 | 4.6 | 6.1 | 11.5 | 28.6 |
| Baseline | 10.8 | 19.2 | 18.2 | 15.7 | 16.5 |
| Audience | Age 65+ (%) | Daily internet use (%) | Daily social media use (%) | Shop online daily (%) | Play mobile games daily (%) |
| A | 35.4 | 62.7 | 60.9 | 25.5 | 44.2 |
| B | 43.2 | 63 | 60.5 | 27.2 | 48.1 |
| C | 42.9 | 76.2 | 69 | 28.6 | 40.5 |
| D | 40.6 | 60.6 | 59.4 | 25.9 | 42.4 |
| E | 47.3 | 59.3 | 58.3 | 24.8 | 42.2 |
| Baseline | 19.7 | 57.2 | 54.5 | 26.2 | 39.2 |
| Audience | Anti-tracking browser (%) | Work laptop (%) | Private laptop for work (%) | Work phone (%) | Private phone for work (%) |
| A | 13.6 | 38.3 | 19.1 | 13.1 | 23.4 |
| B | 28.4 | 25.0 | 25 | 20.9 | 12.5 |
| C | 7.1 | 15.5 | 0 | 0 | 30.7 |
| D | 22.9 | 42.1 | 12.2 | 8.6 | 23.0 |
| E | 15.1 | 30.7 | 12.3 | 14.2 | 16.5 |
| Baseline | 49.8 | 36.9 | 18.4 | 14.3 | 20.2 |
| Audience | Two laptops for work (%) | Two phones for work (%) | Not much online for work (%) | Read technology content (%) | Read business content (%) |
| A | 4.8 | 2.8 | 12.7 | 22.3 | 16.3 |
| B | 4.2 | 0 | 16.7 | 21.0 | 14.8 |
| C | 0 | 0 | 30.8 | 21.4 | 14.3 |
| D | 7.0 | 1.8 | 17.5 | 12.9 | 10.0 |
| E | 0.8 | 2.5 | 15.7 | 16.9 | 13.0 |
| Baseline | 7.4 | 3.3 | 31.2 | 15.0 | 12.8 |

Notes: Small firm = 1-99 employees; Medium firm = 100-999 employees; Large firm > 1,000 employees

* Category 'Not currently employed' includes retirees. Senior executives include C-suite/ Directors/ VPs
Additional information on some descriptives can be found in Appendix B

targeting method performs best are unlikely to be completely explained by differences in possible survey response biases.⁸

In terms of area of work, up to 9.9% of respondents indicated that they work in IT and between 15.4% and 25% in operations (for information on the other areas see Appendix B). For company sizes, we observe a relatively even distribution between small (1-99 employees), mid-sized (100-999 employees) and large companies (>999 employees). Likewise, our sample has a fair gender balance, with men representing 30.9%-48.7% across all six groups. This small skew toward women in our sample could be because women are typically more often online (Lambrecht & Tucker, 2019). As mentioned earlier, in terms of age, the five purchased ITDM segments (A-E) skew older, with more than half of the sample being 55 or older (vs. 36.2% for the baseline). We speculate that this age group pattern is an artifact of the objective to find our desired audience of business decision-makers, as commercial responsibility normally increases with seniority.

Regarding their digital footprint and online behavior, we find most respondents (54.5%-76.2%) use the internet or social media daily, around 24.8%-28.6% shop online daily and around 39.2%-48.1% play mobile games daily.⁹ Our browser statistics, which are tracked and not elicited, suggest that between 7.1% and 28.4% appear to use anti-tracking browsers for the ITDM segments versus 49.8% for the baseline group.¹⁰ Anti-tracking browsers would be Mozilla, Safari or other browsers which do not allow third-party cookies.¹¹ The observed browser differences therefore provide evidence that the baseline group consists of random online users who can be reached through the first-party cookies of the digital publishers. We further find that 15.5%-42.1% have a laptop provided for work, up to 25.0% use their private laptop for work, up to 20.9% have a mobile phone provided for work and 12.5%-30.7% use their private mobile phone for work. This results in up to 7.4% of people having two laptops and 3.3% having two mobiles they use for work. Between 12.7% and 31.2% also indicate that they do not spend much time online for work. Finally, 15%-22.3% read business content every week and 10%-16.3% read business content (for information on other content interests, see Appendix B).

3.5 Results: reach analysis of ITDM segments and random prospecting

Next we analyze how our five segments and the baseline differ in their reach metrics for ITDMs. Table 2 summarizes the descriptive statistics for our three outcome metrics: IT needs identification, vendor selection, and contract signing. We can see that across all five examined segments, over 83.5% of respondents have no IT product or purchase responsibilities as indicated by our three measured questions. All of these users were

⁸ We also present several robustness tests and will investigate the deterministic data with a second research method that is independent of survey responses.

⁹ About 49% of Americans play online games. See <https://www.globenewswire.com/news-release/2020/09/10/2091972/0/en/New-AdColony-and-DISQO-Study-Reveals-Half-of-US-Adults-Play-Mobile-Games-Daily.html>

¹⁰ In fact, Safari browsers were only found in the random baseline group.

¹¹ Mozilla released a new version that blocked 3rd party cookies by default in June 2019.

Table 2 Summary of outcome measures for IT decision-making responsibilities

| | Audience segment (%) | | | | | Baseline |
|--|----------------------|------|---------------|------|------|----------|
| | Deterministic | | Probabilistic | | | |
| | A | B | C | D | E | |
| I identify needs for new IT products in my company/ department | | | | | | |
| This is my main responsibility | 2.7 | 6.2 | 2.4 | 1.2 | 4.1 | 1.8 |
| This is a significant portion of my role | 2.5 | 3.7 | 0 | 4.1 | 1 | 2.8 |
| This is half of my role | 3.2 | 1.2 | 0 | 0 | 1 | 2.7 |
| This is a minor portion of my role | 8.1 | 4.9 | 7.1 | 4.1 | 3.6 | 6.2 |
| This is not part of my role | 83.5 | 84 | 90.5 | 90.6 | 90.3 | 86.5 |
| I select/ shortlist vendors for IT purchases in my company | | | | | | |
| This is my main responsibility | 2 | 3.7 | 2.4 | 0.6 | 3.1 | 2 |
| This is a significant portion of my role | 2.5 | 2.5 | 0 | 2.9 | 1.3 | 2 |
| This is half of my role | 1.8 | 1.2 | 0 | 1.2 | 1 | 2.7 |
| This is a minor portion of my role | 5.3 | 6.2 | 2.4 | 1.8 | 2.8 | 2.5 |
| This is not part of my role | 88.5 | 86.4 | 95.2 | 93.5 | 91.8 | 90.8 |
| I sign contracts/ make financial decisions for IT purchases in my company | | | | | | |
| This is my main responsibility | 2.1 | 3.7 | 2.4 | 0.6 | 3.6 | 2.2 |
| This is a significant portion of my role | 3.9 | 1.2 | 0 | 2.4 | 1.5 | 2.3 |
| This is half of my role | 1.4 | 2.5 | 0 | 2.4 | 1 | 1.8 |
| This is a minor portion of my role | 3.5 | 4.9 | 2.4 | 0 | 1.8 | 4.2 |
| This is not part of my role | 89 | 87.7 | 95.2 | 94.7 | 92.1 | 89.5 |
| ITDM = IT decision-maker (at least one responsibility) | 18.2 | 16.1 | 9.5 | 9.4 | 11 | 16 |
| MITDM = Major IT decision-maker (responsible for all 3 functions) | 8.7 | 11.1 | 4.8 | 5.3 | 6.4 | 6.8 |

associated with cookies that were marked and sold as ITDM to advertisers. The two deterministic data segments based on prospect lists (A,B) perform slightly better than the other three probabilistic data segments (C, D, E), but still barely better than the baseline alternative.

To shed more light on our specific challenge to find ITDMs and not people with related job functions, we also generated two ITDM classifications for each user based on consumers' stated IT responsibilities: 1) We formally define an ITDM as a person that has at least some responsibility for any of the three key IT operations functions we measured (even if only a minor portion of the job role). 2) In addition, we distinguish an ITDM from a 'major ITDM' (MITDM). An MITDM is a person that has at least some responsibility for all three key functions (even if only a minor portion of the job role).

Table 2 summarizes the results of our field studies for these two classifications. We find that the three probabilistic data segments have only 9.4%-11.0% ITDMs and 4.8%-6.4% MITDMs in their samples, which is lower than the proportions in

our baseline group (16% and 6.8% for ITDMs and MITDMs, respectively). The two deterministic data segments perform better, with ITDM proportions of 18.2% and 16.1% and MITDM proportions of 8.7% and 11.1% (vendor A and B, respectively).

To examine whether the observed differences in sample proportions for ITDMs and MITDMs are statistically significant, we conduct two-tailed Fisher's exact tests (FET). We find statistically significant differences for ITDM proportions across our five purchased B2B segments ($p=0.005$, FET), but not for MITDMs ($p=0.315$, FET). Furthermore, we do not find significant differences in ITDM or MITDM proportions among the three probabilistic-data vendors ($p=0.884$ and $p=0.920$, respectively, FET) and the deterministic-data vendors ($p=0.757$ and $p=0.531$, respectively, FET). We therefore pool our reach numbers across vendors and segment types into "Baseline", "Probabilistic data segment" (vendors C,D,E), "Deterministic data segments" (vendors A,B) and "Segment average without baseline" (vendors A,B,C,D,E). Figure 1 summarizes the average reach results for the four segment types, including confidence intervals. Comparing all ITDM segments (five segments pooled) with the baseline does not result in any statistically significant differences for our two ITDM classifications (ITDM and MITDM). When comparing the deterministic with the probabilistic data segments, we find that deterministic outperforms probabilistic for both ITDMs and MITDMs ($p<0.001$ and $p=0.053$, respectively, FET). However, the proportion comparison between deterministic data segments and our publisher-network baseline again yields no significant differences ($p=0.366$ for ITDM and $p=0.1738$ for MITDM, FET). Likewise, we find no differences between probabilistic data segments and the baseline for MITDMs ($p=0.558$, FET), while the former even result in lower ITDM proportions than the baseline ($p=0.005$, FET). In Appendix C, we show that these results also hold for the original three outcome variables and are independent from our chosen ITDM/ MITDM definition.

Overall, our findings reveal that relying on deterministic data segments is more effective than using probabilistic data segments for identifying ITDMs via digital ads. Nevertheless, even the deterministic data segments are not statistically better at reaching our target market than our baseline segment based on random serving. Is this finding unique to our chosen publisher network, which could have a higher share of business readers than other publishers or the U.S. population?

We have not found an official statistic for the number of ITDMs in the US population. However, we can easily obtain a plausible estimate for existing ITDMs to understand how different our publisher network partner may be to the adult population of our study context. According to the latest Census data, there are about 32.6 million businesses in the U.S.,¹² each of which must have at least one person responsible for any IT purchases. Given the adult population is about 258.3 million, this would correspond to a lower boundary of 12.6%. This is well in the range of our two ITDM definitions, being 6.8% and 16.0%.

While our type of validation field test only allows interpreting relative results as it depends on the publishers where a campaign is assessed, we can see that our results appear to be in line with the reach we would expect for random user selection.

¹² <https://www.census.gov/programs-surveys/susb.html>

4 Examining error sources of deterministic data segments

The finding that the average probabilistic (black-box) ITDM segment yields unsatisfactory performance expands the findings of Neumann et al. (2019) from B2C to a B2B (and similar) context where it is very challenging to find the right customer by chance. However, while results are slightly better when using deterministic data segments, we find a disappointingly low proportion of actual ITDMs in our field tests for this type of data too.

Two critical factors mainly influence the success of finding the right customers for deterministic data: 1) the accuracy of the actual profile information of the synthesized lists and 2) the accurate matching of the purchased profiles to web cookies against which media can be bought online. Without access to the back end of the data-broker process, it is not possible to examine exactly how much each factor contributes to a specific campaign. However, we subsequently present two further empirical tests to investigate whether each of the two factors is problematic.

4.1 Study 2: validating deterministic profile information

For the first test, we validate the profile information that is the foundation of the deterministic data segments. This test can also be seen as a robustness check for the survey-based reach results in Study 1 for the two deterministic data segments. In our case, we have the detailed profile information (i.e. names, email/cellphone number, location and job role) for each person from two different data vendors. To examine the accuracy of each piece of information, we took a random sample of 884 people

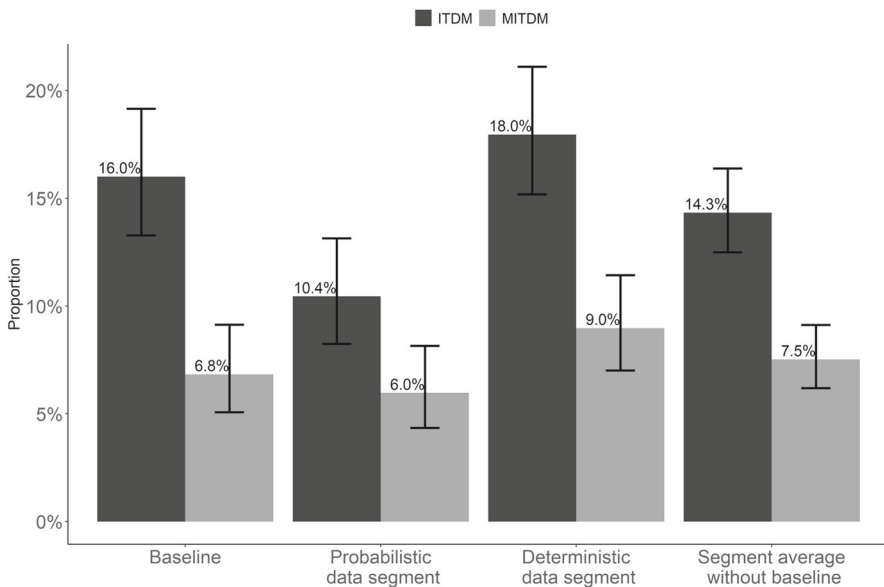


Fig. 1 Average reach results and confidence intervals for different B2B targeting

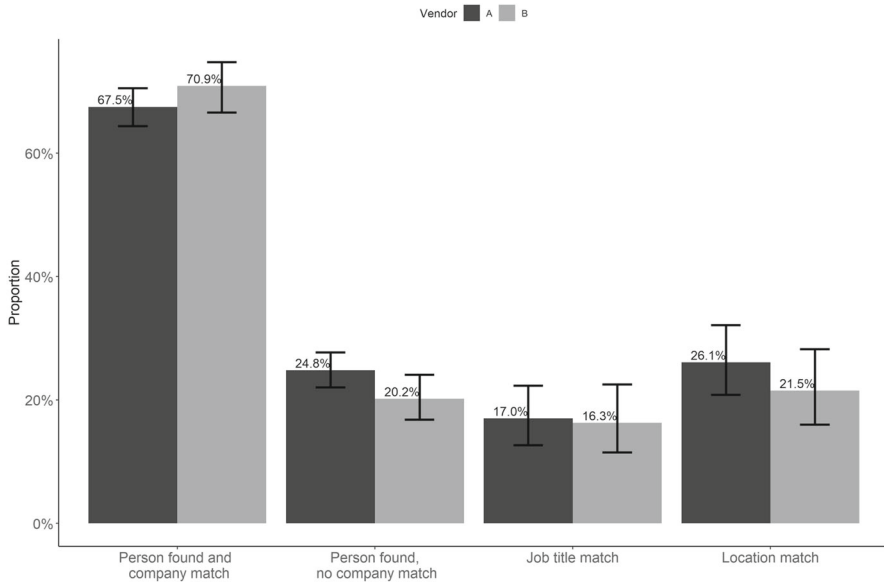


Fig. 2 Deterministic profile information validation (mean and confidence intervals)

from vendor A and 470 people from vendor B from the same media campaign lists that were used for the digital campaigns.

To identify whether a person could be found online, we first used a Google Search. This was followed by searches on LinkedIn, ZoomInfo, Rocket Reach and Signal Hire if we could not find the name of a listed person in Google. We then used LinkedIn to investigate the location and job function of matched people.¹³ We find name and company matches of 67.5%–70.9%, while 20.2%–24.8% of the search queries result in a person match with a different company entry (Fig. 2). Next we checked the job function/provided title as well as the location of the profiled individuals of the deterministic data lists. We find that only 16.3%–17.0% of the job titles and 21.5%–26.1% of the location information resulted in a match. We can only speculate about the underlying reason, one of which could be simply outdated data. Some errors could also stem from the use of IP addresses for tracking; these may be incorrect because companies use firewalls or because some private information may be confounded.

Independent of the underlying reason, our deterministic profile information validation suggests a significant level of inaccuracy, which seems pretty constant across the two vendors. Of particular concern to our goal of reaching ITDMs are the likely incorrect job functions: The average match rate of 16.3%–17% is close to the 16.1%–18.2% ITDM proportion uncovered by our survey results for deterministic data segments (Table 2).

¹³ We first used ClearBit to get a clean name and be able to account for various abbreviations of a company and primary location. The matching process was done via canonicalizations of search strings and various rules using Java custom code. Details are available on request.

4.2 Study 3: profile-cookie matching tests

The profile information of deterministic data is only one possible source of error for digital campaigns. As described earlier, any deterministic profile also needs to be matched to other identifiers of publishers or ad networks to be able to buy media against it. This matching process, which is carried out by some onboarding service provider (see Section 3.3), is another potential source of errors for online targeting using deterministic data segments. Specifically, identity fragmentation bias (Lin & Misra, 2022) — the fact that a single user has multiple devices and may use different browsers — could create challenges in matching the right person to identifiers. Put differently, it's obvious that the quality of matching different identifiers relies entirely on avoiding linking the wrong pieces together. Otherwise the wrong people are reached even though the original attribute was correct.

We next carry out a field test to explore whether the onboarding process itself could have contributed to the poor reach results of our deterministic data segments in our ITDM branding campaign. For privacy reasons, we cannot contact a web cookie through our publisher surveywall and validate the identities of the targeted cookies by asking: "Are you Joe Average from Chicago?" However, most platforms providing onboarding services for media targeting allow uploading user lists (in hashed format) and return the match rates from their database. We carry out two different match tests using these features and our ITDM prospect lists from the US.

For the first test, we upload 100,000 profiles (Name, Location, e-mail and, if available, cellphone number) to four selected platforms: Google, Facebook and two leading onboarding services that are typically used for third-party data targeting.¹⁴ We then examine in a second test what happens when we upload and intend to match the same 100,000 profiles but this time the uploaded lists lack a clear identifier and leave open the option of multiple matches to real people. To achieve this, we remove the address and mobile number, while we add five numbers to the username of the email addresses (e.g. Jane.Doe@work.com to Jane.Doe99987@work.com). We remind the reader that the email address will be hashed such that the onboarding platform will not be able to recognize any patterns in our modified emails and should be unlikely to find any matches in their database using emails as key identifier (unless someone actually uses the email 'Jane.Doe99987@work.com'). If there is no e-mail match between the uploaded list and the onboarder's database, then the only way to provide some match would be just using the name (which should be ambiguous). The results of this test using modified lists are summarized in the top rows of Table 3. We find that Google matches 12.3%, Facebook 20.0%, 'onboarder 1' 75.2% and 'onboarder 2' 55.3% of the original profiles. For the modified profile data, Google, Facebook and 'onboarder 2' report less than 1% matches, while 'onboarder 1' find 74.0% matches.

Next we repeat the same two-step procedure using 100,000 profiles based on the company's global customer relationship management (CRM) system. We additionally remove the last name for the modified data to further increase match difficulty and lower the odds in finding a single correct match. The results are summarized in the

¹⁴ We cannot reveal the identities of the two onboarders or which of the four tested services in this study was employed for the media campaigns in our previous field study.

Table 3 Summary of profile matching tests

| | Google | Facebook | Onboarder1 | Onboarder2 |
|----------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| US ITDM prospect-list data | | | | |
| Original | 12.3% [12.1%, 12.5%] | 20.0% [19.7%, 20.2%] | 75.2% [74.9%, 75.5%] | 55.3% [55.0%, 55.6%] |
| Modified | <1% - | <1% - | 74.0% [73.7%, 74.3%] | <1% - |
| Global CRM data | | | | |
| Original | 44.0% [43.7%, 44.3%] | 62.0% [61.7%, 62.3%] | 79.2% [78.9%, 79.5%] | 76.5% [76.3%, 76.7%] |
| Modified | <1% - | <1% - | 22.5% [22.2%, 22.8%] | 3.8% [3.7%, 3.9%] |

Notes: Proportion means and 95% exact confidence intervals (Clopper & Pearson, 1934). We are restricted by the number format that has been reported by the media partners. Therefore, no confidence intervals are reported for the values with '<1'

bottom rows in Table 3. We find that Google now reports 44.0% matches, Facebook 62.0%, 'onboarder 1' 79.2% and 'onboarder 2' 76.5% of the actual profiles. For the modified masked profiles (where we created wrong/ fake emails), Google and Facebook report less than 1% matches again, while 'onboarder 1 and 2' still find 22.5% and 3.8% matches, respectively. Thus, the CRM global test seem to be in agreement and confirm a similar pattern as our test on US ITDM prospect lists. While Google and Facebook show the results we would expect — that our modified data results in virtually no matches — the two onboarders partly still report some matches. Between the two onboarders we benchmarked, we also find strong differences in the match rates for the modified data. It is concerning that 'onboarder 1' still reports a large percentage of matches for profiles that should not be linkable to one single person. While this is only suggestive evidence that the user profiles are likely to be often mismatched in a campaign relying on some onboarding services, we see that the matchmaking process is subject to errors and, depending on the chosen provider, will contribute to poor digital targeting results.

5 Study 4: what features help finding IT decision-makers?

Study 1 illustrates that the custom ITDM segments are no more helpful for finding the desired audience than random prospecting from a digital publisher network. This result raises the question of how to best reach ITDMs via digital channels. Some covariates we measured in our survey can be used as robustness checks for our ITDM outcome metric (e.g. being a senior executive), while others are typical customer characteristics, some of which companies could access or buy for targeting purposes too (e.g. demographics). We next explore which features help finding online users who are ITDM. For this analysis, we first focus on demographic and firmographic features and then on digital behavior covariates.

5.1 Demographic and firmographic features

We investigate the associations of classic demographic and firmographic features on the likelihood of an online user being an (M)ITDM using the entire sample of 1,849 consumers for these analyses. Since we have individual data on each person i and whether someone is an (M)ITDM, we perform logistic regressions that determine the probability of being an (M)ITDM for our measured features:

$$Probability([M]ITDM_i | x_i) = \frac{\exp^{x_i'\beta}}{1 + \exp^{x_i'\beta}}, \quad (1)$$

$$x_i'\beta = \beta_0 + Demographics_i\beta_A + Firmographics_i\beta_B, \quad (2)$$

where β_0 is the model's constant and $Demographics_i$ is a row vector of fixed effects for different demographic consumer attributes, such as 'age 18-25' or 'gender,' while the row vector $Firmographics_i$ captures fixed effects about the person's background with respect to the job and company, such as firm size, job role and industry. We then estimate five different models with different groups of variables, four reduced model specifications and one full model (with all covariates included), using 'ITDM' and 'MITDM' as the dependent variable, respectively.

Table 4 summarizes the results of our demographic and firmographic features for our ten estimated models. First, we find that working in IT/operations (columns 1 and 6) and being an senior executive or manager (columns 2 and 7) significantly increases the likelihood of being either an ITDM or MITDM. We interpret these results, which theoretically show a positive association, as a robustness and data validity check.

We next focus on age and gender. We find that being older or a man is associated with a greater likelihood of being an ITDM or MITDM (columns 3, 8). In particular, being 35-64 (55-64) years old significantly increases the odds of being an ITDM or MITDM. This finding appears logical, as corporate responsibilities tend to grow with seniority. This effect disappears for people over 64, probably because of retirement. Further evidence is available in Appendix E. While not shown here, logit models suggest no significant association of gender and being a senior executive, but a significant association of being male and a manager for our data. Hence, the gender effect may reflect the well-documented disparity across gender in IT (Lambrecht & Tucker, 2019) as well as some potential gender bias in being promoted to manager roles. Finally, we find that people working for small and mid-sized companies are more likely to be an (M)ITDM than for large firms (columns 4 and 9). This finding appears plausible, given the odds are simply smaller at larger firms of being a person with ITDM responsibilities. If we include all the features in a model (columns 5 and 10), gender, firm size, work function (IT/ operations) and seniority (manager/executive) are the covariates with the most precise associations with being an (M)ITDM.

Table 4 Firmographic and demographic features of IT decision-makers

| Dependent variable: | Likelihood of online user being an ITDM | | | | | Likelihood of online user being an MITDM | | | | |
|---------------------|---|----------------------|----------------------|----------------------|----------------------|--|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Intercept | -2.079*** (0.078) | -2.706*** (0.107) | -2.310*** (0.308) | -2.652*** (0.108) | -3.041*** (0.364) | -2.804*** (0.106) | -3.651*** (0.164) | -3.279*** (0.439) | -3.710*** (0.176) | -4.130*** (0.499) |
| Works in IT | 2.500*** (0.292) | | | | 2.396*** (0.339) | 2.054*** (0.313) | | | | 1.709*** (0.367) |
| Works in operations | 1.673*** (0.184) | | | | 0.727*** (0.231) | 1.146*** (0.247) | | | | -0.041 (0.297) |
| Senior Executive | | 3.567*** (0.242) | | | 2.851*** (0.276) | | 3.552*** (0.258) | | | 2.657*** (0.308) |
| Manager | | 2.486*** (0.167) | | | 1.902*** (0.197) | | 2.254*** (0.229) | | | 1.631*** (0.270) |
| Age 25 to 34 | | | 0.394 (0.354) | | 0.034 (0.415) | | | 0.247 (0.510) | | -0.210 (0.565) |
| Age 35 to 44 | | | 0.635* (0.346) | | -0.083 (0.409) | | | 0.244 (0.510) | | -0.644 (0.570) |
| Age 45 to 54 | | | 0.710** | | -0.290 | | | 0.562 | | -0.631 |

Table 4 continued

| Dependent variable: | Likelihood of online user being an ITDM | | | | | Likelihood of online user being an MITDM | | | | |
|---------------------|---|-----------|-----------|-----------|-----------|--|----------|----------|----------|----------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Age 55 to 64 | | | (0.333) | | (0.398) | | | (0.474) | | (0.537) |
| | | | 0.562* | | -0.123 | | | 0.844* | | 0.073 |
| | | | (0.322) | | (0.381) | | | (0.449) | | (0.504) |
| Age 65 and over | | | -0.158 | | -0.290 | | | 0.087 | | -0.264 |
| | | | (0.327) | | (0.386) | | | (0.458) | | (0.515) |
| Male | | | 0.469*** | | 0.112 | | | 0.654*** | | 0.460** |
| | | | (0.133) | | (0.167) | | | (0.183) | | (0.216) |
| Mid-sized firm | | | | 1.763*** | 0.767*** | | | | 1.723*** | 0.919*** |
| | | | | (0.190) | (0.231) | | | | (0.280) | (0.323) |
| Small firm | | | | 2.525*** | 1.284*** | | | | 2.814*** | 1.667*** |
| | | | | (0.163) | (0.204) | | | | (0.222) | (0.268) |
| Observations | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 |
| Log Likelihood | -710.947 | -582.599 | -756.682 | -641.433 | -526.068 | -459.572 | -370.141 | -470.470 | -391.053 | -333.644 |
| AIC | 1,427.895 | 1,171.197 | 1,527.364 | 1,288.865 | 1,078.137 | 925.144 | 746.282 | 954.939 | 788.106 | 693.289 |

Note: *p < 0.1; **p < 0.05; ***p < 0.01

5.2 Digital behavior characteristics

Next we investigate the associations of behavioral characteristics regarding web-surfing and technology preferences and the likelihood of an online user being an (M)ITDM. We again perform logistic regressions (in line with Eq. 1), but this time we regress the following features on the the probability of being an (M)ITDM:

$$x_i'\beta = \beta_{00} + DeviceCharacteristics_i\beta_C + OnlineBehavior_i\beta_D, \quad (3)$$

where β_{00} is the model's constant and $DeviceCharacteristics_i$ is a row vector of fixed effects for different device characteristics, such as the types of PCs, browsers and mobiles used by an individual, whereas the row vector $OnlineBehavior_i$ pertains to a user's online behavior, such as the types of content that someone reads, or their indicated frequency of key activities (e.g. mobile gaming or social media usage). We again estimate five different models with different groups of variables, four reduced model specifications and one full model (with all covariates included) for ITDM and MITDM as dependent variables, respectively.

Table 5 summarizes our digital behavior characteristics for our ten models. We find first that using an anti-tracking browser has a significant and negative correlation with the likelihood of an online user being an MITDM (column 6), which does not hold when all covariates enter the model (column 10). We further test an interaction effect of using an anti-tracking browser and an indicator for probabilistic data segments. Probabilistic data segments have a main (correlational) effect with a negative sign, although this effect only reaches statistical significance for one ITDM analysis (column 1). This result suggests that probabilistic data segments are not always worse than random prospecting or deterministic data segments but that the classification likelihood strongly depends on the characteristics of the online user.

Interestingly, we find a negative sign for the interaction effect of probabilistic data segments and anti-tracking browser, which is negative but reaches statistical significance for ITDMs only (see columns 1 and 5).¹⁵ This finding suggests that tracking challenges affect probabilistic data segments more strongly regarding the likelihood of identifying an ITDM. This is expected, given that the creation of probabilistic data segments heavily relies on available online signals from third-party cookies that must be enabled in web browsers.

In addition, we find that having a personal laptop or phone that is used for work, or a specifically provided work phone or laptop, increases the likelihood of online users being an ITDM or MITDM (columns 2, 7). If people have two devices (phones or laptops), then we find a negative association, which is in line with the concept of identity fragmentation bias (Lin & Misra, 2022). That is, online users who use multiple devices are more likely to be classified incorrectly. This effect only reaches statistical significance for the ITDM analyses (columns 1 and 5). The finding that many effects have the expected sign but do not reach a high precision in the estimates may be linked to the reduced power given by the limited number of respondents who

¹⁵ We need to keep in mind that people behind the data-broker audiences do not know that their cookies were marked as ITDM (and are therefore targeted).

Table 5 Digital behavior characteristics of IT decision-makers

| Dependent variable: | Likelihood of online user being an ITDM | | | | Likelihood of online user being an MITDM | | | | | |
|---|---|----------------------|----------------------|----------------------|--|----------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Intercept | -1.544*** (0.090) | -2.663*** (0.103) | -2.130*** (0.128) | -2.190*** (0.098) | -2.744*** (0.181) | -2.331*** (0.121) | -3.294*** (0.136) | -2.678*** (0.164) | -2.950*** (0.135) | -3.098*** (0.222) |
| Anti-tracking browser | -0.130 (0.164) | | | | 0.061 (0.199) | -0.417* (0.243) | | | | -0.390 (0.274) |
| Probabilistic data segments | -0.415** (0.163) | | | | -0.127 (0.190) | -0.259 (0.213) | | | | 0.050 (0.236) |
| Anti-tracking browser X probabilistic data segments | -2.516** (1.027) | | | | -2.753** (1.073) | -1.597 (1.049) | | | | -1.584 (1.084) |
| Work laptop | | 1.912*** (0.180) | | | 1.848*** (0.191) | | 1.061*** (0.251) | | | 0.914*** (0.268) |
| Private laptop | | 2.616*** (0.265) | | | 2.506*** (0.282) | | 1.855*** (0.312) | | | 1.679*** (0.340) |
| Work phone | | 1.907*** (0.283) | | | 1.878*** (0.292) | | 1.577*** (0.310) | | | 1.637*** (0.326) |
| Private phone | | 0.704*** (0.240) | | | 0.630** (0.251) | | 0.907*** (0.286) | | | 0.801*** (0.306) |
| Two phones | | -1.423** (0.664) | | | -1.192* (0.696) | | -1.197* (0.654) | | | -1.011 (0.702) |
| Two laptops | | -2.080*** (0.504) | | | -2.256*** (0.515) | | -0.855 (0.527) | | | -0.760 (0.563) |
| Daily social-media user | | | -0.210 (0.182) | | -0.398* (0.216) | | | -0.642** (0.253) | | -0.649** (0.281) |
| Daily online browsing | | | 0.810*** (0.183) | | 0.129 (0.223) | | | 0.693*** (0.246) | | -0.086 (0.286) |

Table 5 continued

| Dependent variable: | Likelihood of online user being an ITDM | | | | Likelihood of online user being an MITDM | | | | | |
|-----------------------|---|-----------|--------------------|---------------------|--|----------|----------|----------------------|---------------------|---------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Daily online shopping | | | 0.454** (0.184) | | 0.557** (0.223) | | | 1.104*** (0.267) | | 1.214*** (0.302) |
| Daily mobile gaming | | | -0.356* (0.182) | | -0.126 (0.217) | | | -0.710*** (0.268) | | -0.479 (0.299) |
| Travel content | | | | 0.382** (0.176) | 0.182 (0.201) | | | | 0.312 (0.237) | 0.124 (0.258) |
| Home & Garden | | | | 0.163 (0.177) | 0.360* (0.202) | | | | -0.076 (0.247) | -0.005 (0.272) |
| Health & Fitness | | | | -0.283 (0.180) | -0.383* (0.206) | | | | -0.212 (0.243) | -0.260 (0.263) |
| Computers & Games | | | | -0.134 (0.184) | -0.162 (0.217) | | | | -0.371 (0.256) | -0.492* (0.292) |
| Food & Entertainment | | | | 0.006 (0.166) | -0.017 (0.188) | | | | -0.107 (0.229) | 0.017 (0.249) |
| Business | | | | 0.954*** (0.177) | 0.212 (0.212) | | | | 1.325*** (0.226) | 0.766*** (0.252) |
| Technology | | | | 0.744*** (0.183) | 0.596*** (0.211) | | | | 0.577** (0.245) | 0.329 (0.266) |
| Sports | | | | 0.256 (0.163) | 0.338* (0.189) | | | | 0.279 (0.215) | 0.344 (0.237) |
| Observations | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 |
| Log Likelihood | -761.289 | -595.977 | -760.894 | -727.578 | -568.669 | -476.692 | -408.922 | -468.071 | -449.755 | -382.600 |
| AIC | 1,530.577 | 1,205.955 | 1,531.788 | 1,473.155 | 1,181.339 | 961.385 | 831.844 | 946.143 | 917.509 | 809.201 |

Note: *p<0.1; **p<0.05; ***p<0.01

classify as MITDM and are online users exposed to probabilistic targeted ads within our publisher network.

When looking at general online behavior (columns 3, 8), daily online shopping/daily internet use is positively correlated with being an (M)ITDM. This finding could be a behavioral characteristic or be linked to the available digital footprint needed to build user profiles. Being an (M)ITDM is negatively correlated with daily social media use or mobile gaming, although these effects only reach significance for the models just looking at behavioral variables for daily mobile gaming (columns 3 and 8) and three of four models for daily social media usage (columns 5, 8 and 10). Speculatively, key decision-makers plausibly lack the time to engage daily in mobile gaming or social media. Reading business content weekly is positively correlated with being an (M)ITDM (columns 4, 9); reading travel or technology content is positively correlated with being an ITDM. When including all variables, having or using a laptop/phone for work, daily online shopping and reading business content appear to be the most precise behavioral descriptive covariates.¹⁶

6 Study 5: alternative targeting tactics

Our analysis of customer features describing the association with ITDM likelihood provides some possible alternative characteristics that could be used for digital targeting. Of course, not every firmographic proxy or customer covariate for our target audience can be bought in scale from data vendors or easily be collected. However, our analyses suggests two types of other targeting criteria that publishers can use to build their own segments and then offer them to potential advertisers: (1) Demographics (such as age and gender) and (2) an interest in content such as technology or business. Similar to our random cookie selection for our baseline benchmark, we can use the digital publisher network integration of our measurement company to examine how using the two criteria (demographics and content-interest) would fare as targeting mechanism for an ITDM campaign.

Moreover, we use a first-party data targeting approach for our next field test. We target only cookies whose self-reported data meet our alternative targeting criteria.¹⁷ Because our measurement company is fully integrated with the websites and apps of the digital publisher network, it can collect data on visitors and can later recognize and select/ target users with certain criteria (provided they do not delete their cookies). In other words, for our study we can again mimic the targeting process in the publisher network without actually serving an ad to the user behind the cookie that meets our criterion (but show our validation survey instead of an ad). Our approach, collecting information directly from its own users and then building segments based on this information, can be carried out by any publisher. Such a first-party data approach

¹⁶ We cannot include the additional covariate 'Not much online for work' in this analysis as it only applies to the sub-sample of employed users. In Appendix E, we present the same analysis just for employed respondents and include the respective covariate, which has a negative and significant effect on ITDM.

¹⁷ These users did not participate in our survey previously, but provided their demographic information in other surveywall surveys of the same digital publisher network.

reduces risks of poor data quality found in third-party demographic or interest-based data that were bought from data aggregators (Neumann et al., 2019).

6.1 First-party demographic and content-interest targeting for ITDMs

To investigate the efficacy of the two identified targeting alternatives in reaching ITDMs, we conduct a new field test. We target only cookies with an age tier of 45-64 years of age and who are male¹⁸ in one test and those who indicate they regularly read business or technology content in another test.

Our additional field test resulted in survey responses from a sample of 333 online users for demographic targeting and 103 for content interest. We again compare the proportions of (M)ITDMs with the results from our previous tests (see Fig. 3). We find that targeting by age/gender results in an average of 23.4%/16.2% ITDMs/MITDMs, whereas selecting users by technology- or business-content interest leads to 28.2%/14.6% and 41.8%/22.8% ITDMs/ MITDMs, respectively. Chi-squared tests suggest that these proportion differences in frequencies of observed (M)ITDMs are all statistically significant (see Table 6).

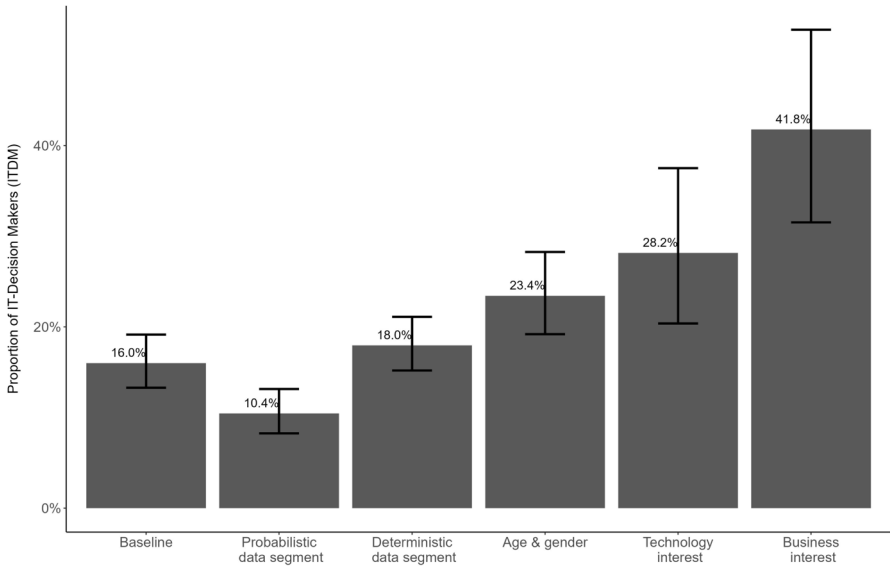
The field test suggests that these two alternative targeting tactics to reach ITDMs outperform the probabilistic data segments, random prospecting and the deterministic data segments. We carry out two robustness checks to validate the findings using different samples.

6.2 Robustness checks for alternative targeting results

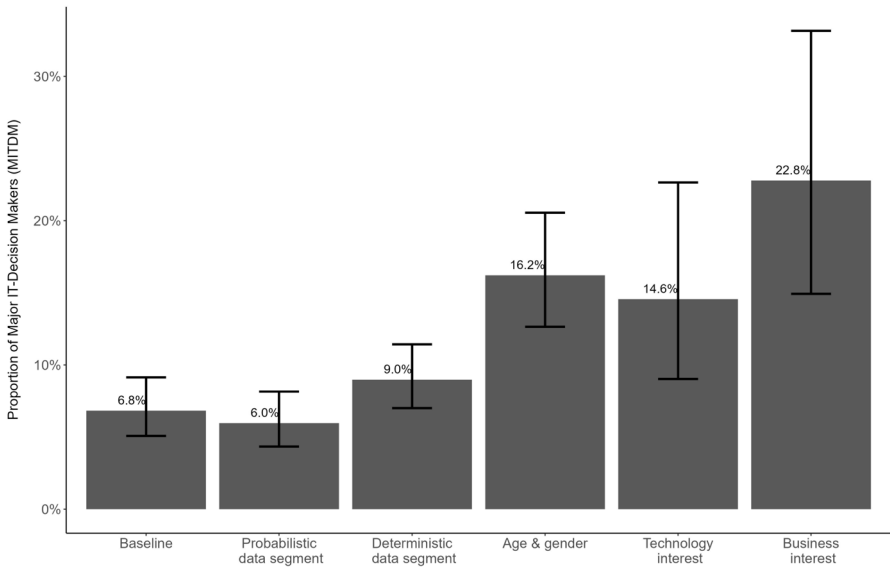
As a first robustness check, we repeat our analysis of how targeting alternatives fair using the random sample of 600 people that made up the baseline group. We can mimic our targeting tactics around age/gender and content interest by filtering only for those users who would meet the criteria of being male and 45-64 years of age (or alternatively reading business/technology content). The results of this alternative and second, independent sample are summarized in Fig. 4 and illustrate again that demographic and content-interest targeting based on the publisher's own data lead to greater ITDM and MITDM reach.

For our second robustness check, we replicate our analysis of how targeting alternatives fare in comparison to probabilistic and deterministic data segments as well as the baseline using only the sample proportion of respondents who was employed ($n=710$). The results are summarized in Fig. 5 and confirm the previous findings about the superior performance of demographic and content-interest targeting based on publisher first-party data.

¹⁸ This paper deliberately does not discuss the ethical merits of gender- and age-based targeting relative to alternatives. The authors acknowledge that this is often problematic as a practice. Some papers which discuss these are Datta et al. (2015) and Lambrecht and Tucker (2019).



(a) Results for ITDM



(b) Results for MITDM

Fig. 3 Reach comparison for different targeting methods (means and confidence intervals)

Table 6 ITDM reach for demographic and content-interest targeting

| Responsibility | Comparison | Proportion 1 (%) | Proportion 2 (%) | Chi-square statistic | Chi-square p-value | Odds Ratio | Fisher's exact p-value |
|-------------------------|---|------------------|------------------|----------------------|--------------------|------------|------------------------|
| IT decision-maker | Age & gender targeting vs. baseline | 23.4 | 16.0 | 7.78 | 0.005 | 1.60 | 0.006 |
| IT decision-maker | Age & gender targeting vs. off-the-shelf audiences | 23.4 | 14.3 | 15.97 | <0.001 | 1.83 | <0.001 |
| Major IT decision-maker | Age & gender targeting vs. baseline | 16.2 | 6.8 | 20.62 | <0.001 | 2.64 | <0.001 |
| Major IT decision-maker | Age & gender targeting vs. off-the-shelf audiences | 16.2 | 7.5 | 23.41 | <0.001 | 2.38 | <0.001 |
| IT decision-maker | Technology content interest vs. baseline | 28.2 | 16.0 | 8.88 | 0.003 | 2.05 | 0.005 |
| IT decision-maker | Technology content interest vs. off-the-shelf audiences | 28.2 | 14.3 | 13.97 | 0.001 | 2.34 | 0.005 |
| Major IT decision-maker | Technology content interest vs. baseline | 14.6 | 6.8 | 7.16 | 0.007 | 2.32 | 0.016 |
| Major IT decision-maker | Technology content interest vs. off-the-shelf audiences | 14.6 | 7.5 | 6.35 | 0.012 | 2.09 | 0.021 |
| IT decision-maker | Business content interest vs. baseline | 41.8 | 16 | 30.13 | <0.001 | 3.76 | <0.001 |
| IT decision-maker | Business content interest vs. off-the-shelf audiences | 41.8 | 14.3 | 41.7 | <0.001 | 4.28 | <0.001 |
| Major IT decision-maker | Business content interest vs. baseline | 22.8 | 6.8 | 22.39 | <0.001 | 4.01 | <0.001 |
| Major IT decision-maker | Business content interest vs. off-the-shelf audiences | 22.8 | 7.5 | 22.40 | <0.001 | 3.62 | <0.001 |

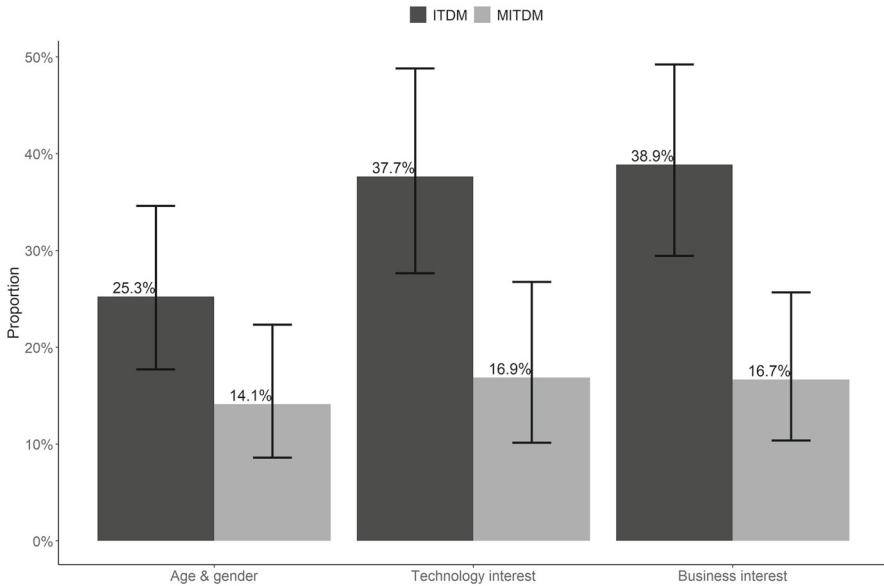
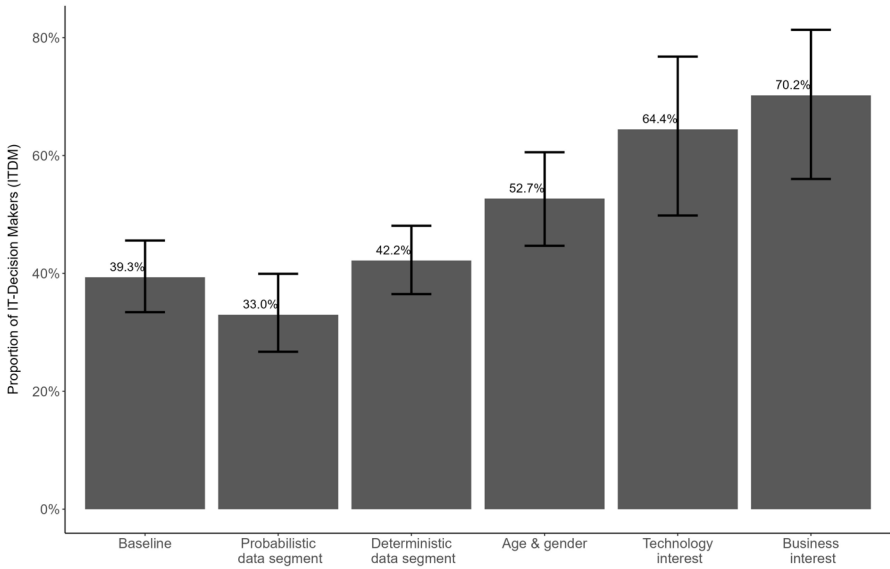


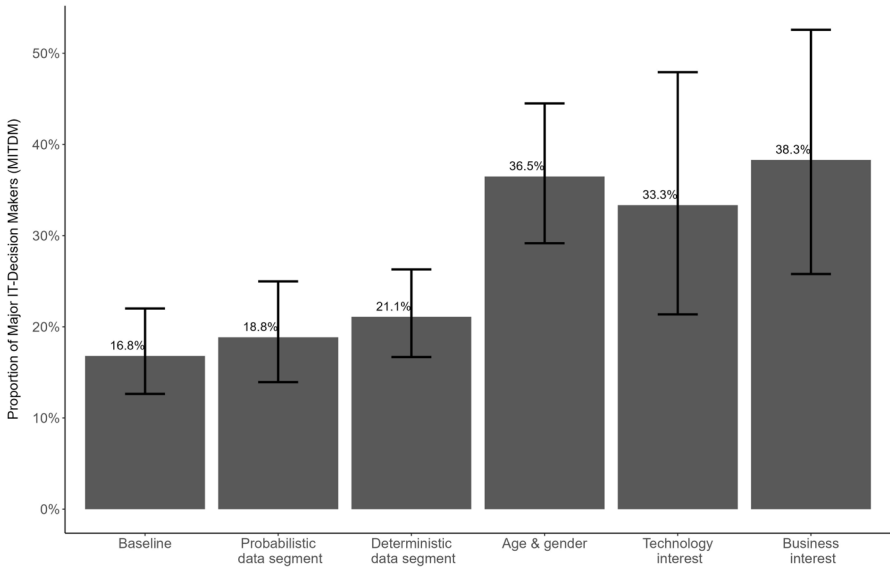
Fig. 4 Targeting methods replication for random baseline group (means and confidence intervals)

7 Cost-benefit analysis

In Study 5, we have shown that alternative targeting approaches outperform both the publisher network baseline and purchasable ‘off-the-shelf’ ITDM segments, including deterministic data segments, in reaching ITDM. A remaining question is whether the revealed performance differences are captured by the market prices for audiences and which segment provides the best benefit-cost ratio. We therefore compare the data costs that are directly associated with buying that segment relative to their benefits (Table 7). These costs are typically added as cost per mille (CPM) to a media campaign in addition to media and technology costs. Because our baseline in all our tests represents the propensity to find ITDM across our publisher network and not a purchasable data segment, we cannot use this as our benchmark for the benefit-cost analysis. Instead, we use the probabilistic ITDM segment, which is also the most common type of ITDM audiences bought by advertisers, as our anchor for comparing benefit-cost ratios of our targeting types. We synthesized typical CPM ranges for our four types of targeting based on external sources (Nylen, 2018), a leading data management platform and the input from the media team from the IT service provider. Probabilistic ITDM data costs range from US\$1.50-3.00, deterministic data from \$2.00-3.00, content-interest/contextual data from \$0.29-1.33 and demographic first-party data from \$1.00-1.50. Using the middle values of the range for our back-of-the-envelope estimation results in a CPM of \$2.25 for the probabilistic data segments, a \$2.75 CPM for deterministic data segments, 81 cents for contextual/content-interest targeting and \$2.50 for demographic first-party data (for two attributes each with a middle value of \$1.25). We then estimate different multipliers for CPM costs and (M)ITDM using the results from



(a) Results for ITDM



(b) Results for MITDM

Fig. 5 Reach comparison for different targeting methods and the sample of ‘employed’ (means and confidence intervals)

Table 7 Benefit-cost analysis for ITDM and MITDM targeting options

| Targeting type | CPM (\$) market range | CPM (\$) estimate | CPM multiplier | ITDM reach | ITDM multiplier | ITDM benefit-cost ratio | MITDM reach | MITDM multiplier | MITDM benefit-cost ratio |
|-----------------------------|-----------------------|-------------------|----------------|------------|-----------------|-------------------------|-------------|------------------|--------------------------|
| Probabilistic data | 1.50-3.00 | 2.25 | 1 | 0.104 | 1 | 1 | 0.06 | 1 | 1 |
| Deterministic data | 2.00-3.50 | 2.75 | 1.22 | 0.18 | 1.73 | 1.42 | 0.09 | 1.50 | 1.23 |
| Contextual/content interest | 0.29-1.33 | 0.81 | 0.36 | 0.35* | 3.37 | 9.35 | 0.187* | 3.12 | 8.66 |
| Demographics (1st party) | 1.00-1.50 | 2.5 | 1.11 | 0.234 | 2.34 | 2.11 | 0.162 | 2.7 | 2.43 |

* Mean of the technology and business content interest segments

Table 3. This procedure leads to a benefit-cost ratio (with third-party probabilistic as anchor = 1) of 9.35 (8.66) for contextual/ content-interest targeting, and 1.42 (1.23) for deterministic data targeting to reach ITDM (MITDM). Hence, considering both market costs and reach from our study, we find that contextual/content-interest targeting is nearly 9-10 times more (cost) efficient than using the specific ITDM segments (both probabilistic and deterministic). Demographic first-party data targeting is also about twice as (cost) efficient as the ITDM segments. While we chose probabilistic data segments as a basis for our analysis, we highlight that, given the cost differences, the same conclusion can be drawn when using a deterministic data segment as an anchor for the calculations.

8 Discussion and conclusion

In this paper, we investigate in five studies how best to reach IT decision-makers (ITDMs) for digital advertising campaigns when the campaign goal is prospecting or branding for a very specific audience. We benchmark several targeting tactics that marketers can use for this objective of reaching the ‘right’ customers.

For our context, we find that deterministic data segments outperform probabilistic data segments, but are not significantly better than the random prospecting baseline from our digital publisher network. Moreover, all our probabilistic data segments performed worse than the baseline comparison from the publisher network in our tests. Thus our findings expand the work of Neumann et al. (2019) on third-party (probabilistic black-box) targeting from common B2C segments to one of the most important B2B segments, where the odds of naturally finding the right customer in the population tends to be low. Furthermore, to the best of our knowledge, our study presents first evidence that even deterministic data segments are problematic and may not perform better than using no targeting (random prospecting).

After investigating what factors help find an ITDM online, we conduct further benchmarking tests to see how alternative targeting tactics based on first-party publisher data would fare in comparison to the third-party ‘off-the-shelf’ ITDM segments. We demonstrate that such publisher-data targeting based on age tiers and gender or based on technology-/business-content interest outperforms the random baseline and the custom, specifically-created segments (both deterministic and probabilistic) at reaching ITDMs. Our findings appear robust across all (parametric and nonparametric) analyses we carried out. When further considering data costs, we find that content-interest - which can be regarded as a form of contextual targeting - appears to be not only the most effective but also the most (cost-)efficient tactic to reach ITDM via digital advertising.

What explains the poor results for the ITDM segments in our tests? In total, we document three possible mechanisms that account for our results. First, we find evidence of identity fragmentation bias (Lin & Misra, 2022), which can affect profiling in two ways. On the one hand, original user profiles may be matched to the wrong individual

in the media-buying process. This part of fragmentation bias strongly affects deterministic data segments, which must be matched (=onboarded) to identifiers of websites (typically web cookies) to be able to execute targeted online ad campaigns. On the other hand, it is also possible that fragmentation bias affects providers of probabilistic data segments. Users with multiple devices or browsers may generate mixed or incomplete signals if a data vendor does not obtain a unified view of a single person (Trusov et al., 2016). As a result of missing critical signals of individuals or incorrectly assuming signals belong to the same person, a data vendor applying probabilistic methods to customers' online browsing behavior may make wrong inferences about a profile attribute.

Second, we show that being an ITDM depends on the digital footprint users leave and their online behavior. For example, some people use browsers that don't allow tracking, which strongly affects the performance of probabilistic data segments. We also find that browsing online or shopping online daily correlates with the likelihood of being an ITDM. This findings may be partly due to being an ITDM behavioral characteristic and partly due to providing more online signals that allow identifying or modeling their behavior or background more accurately.¹⁹

Third, we find low match rates for many pieces of information from the deterministic data segments (in particular, location and job function) when validating these with other well-known databases based on first-party data (e.g. LinkedIn). This finding suggests that the provided contact details may be outdated or incorrectly associated. Incorrect user-specific information primarily affects providers of deterministic data segments.

Our findings on the most effective and efficient digital targeting tactics for reaching ITDM have important implications for marketing practice. Mozilla Firefox and Apple's Safari have already made policy changes that disable third-party cookies by default. Google has announced to phase out third-party tracking in the Chrome browser in 2024 (Gonzales, 2022). Thus, many current targeting solutions that rely on third-party cookies are likely not feasible options in the future. Using deterministic data segments may still be possible with individual platforms, but is also controversial in terms of privacy considerations, can be expensive and is still subject to several issues, such as incorrect information or fragmentation bias. In contrast, we show that content-interest (contextual) targeting performs well, but is also relatively less intrusive. Demographic targeting based on first-party data and collected with consent may be also able to be more compliant with existing privacy-regimes that are based on consent. Given our results on varying performance of onboarding platforms, marketers are also advised to perform due diligence about the exact matching methods and to carry out their own tests.

Our study has limitations. First, we investigate targeting tactics to reach IT decision-makers through digital advertising. It would be interesting to see how other very specific segments fare. Second, we use self-reported data from surveywalls. We acknowledge that this data may suffer from inaccuracies and be subject to typical survey response biases that are often unavoidable. Given the nature of our research

¹⁹ Interaction effects of increased online behavior and probabilistic data segments do not reach statistical significance.

question and the complexity of B2B purchase processes, there is no obvious viable single metric that can be used as a source of ground truth. Moreover, robustness checks and our different complementary research methods suggest that our findings and conclusions are unlikely to be explained by survey response bias. Third, our tests for targeting alternatives mimicked a targeting procedure in the field by filtering and selecting respective profiles within the publisher network of our campaign. While the respective users were online and could have been served an actual ad, the publisher network did not offer either of the two tested targeting tactics when our study was conducted.²⁰ However, the data stems from users of a digital network of popular content providers and publishers, which could easily collect such data to build and sell the described segments using our approach. Consumers may provide the required data with consent to their publishers of interest in exchange for access to content. Content-interest targeting, the most efficient tactic in our tests, is also often referred to as contextual targeting. Publishers can easily create such segments from the browsing history of the first-party cookies or signed-in users on their platform. We can see publishers adopting this strategy.

For example, the New York Times or News Corp, report that they have successfully created their own audience segments using their first-party data (including collected survey data) or by building a partner network (NewsCorp Press Release, 2019; NYT Open Team, 2020).

One of the most crucial findings in our study relates to the impact of onboarders' performance when third- or second-party data needs to be matched with different identifiers of media sellers. This process step seems prone to errors due to fragmentation bias. In contrast, audience information generated by publishers (= first-party data segments) is less likely to be subject to some errors based on linking profiles to the wrong person as the identifiers to create and leverage any audience information for media buying are the same. In other words, as long as advertisers rely on first-party data from the publisher [network] where they run their targeted campaigns, additional onboarding/ identifier matching between different networks is not necessary.

Appendix

A: Questionnaire

Q1: Please select your gender [single choice]:

- Female
- Male

Q2: Please select your age [single choice]:

- 13-17 yrs [excluded]
- 18-24 yrs

²⁰ In fact, we want to stress that the online measurement company does not use or sell any data they collect in the publisher network for media buys or targeting.

- 25-34 yrs
- 35-44 yrs
- 45-54 yrs
- 55-64 yrs
- 65+ yrs

Q3: Which of the following most closely matches your job title? [single choice]

- Intern/ casual/ part-time
- Entry Level
- Analyst/Associate
- Manager/ Senior Manager
- Director/ VP
- C-suite/ owner
- Student [SKIP Q4, Q5, Q7 and Q8]
- Not currently employed [SKIP Q4, Q5, Q7 and Q8]

Q4: What best describes your role with regards to IT (Information Technology)? Please indicate what applies to your role with regards to IT (Information Technology).

Likert scale [single choice per item]: This is my main responsibility/ significant portion of my role/ half of my role/ minor portion of my role/ not part of my role.

- I identify the needs for new IT products in my company/ department.
- I select/ shortlist vendors for IT purchases in my company.
- I sign contracts/ make financial decisions for IT purchases in my company

Q5: What best describes your area of work/ department [single choice]:

- Operations
- Administration/ Legal/ Compliance
- Human Resources
- Information Technology (IT)
- Accounting/ Finance
- Marketing/ Sales
- Other

Q6: Which of the following types of content do you read about online each week? Select all that apply. [multiple choice]

- Business & Finance
- Technology
- Travel
- Home & Garden
- Food & Entertainment
- Health & Fitness/ Sports
- Computer & Games

Q7: Please estimate the number of employees in your company at all locations, nationally and internationally [single choice]:

- 1 to 99 employees

- 100 to 249 employees
- 250 to 499 employees
- 500 to 999 employees
- 1,000 to 4,999 employees
- 5,000 or more employees

Q8 : Which of the following best describes your situation at work? [multiple choice]

- I have a work laptop
- I use my private laptop for work
- I have a work mobile
- I use my private phone for work
- I do not need to be much online for work

Q9: Please indicate how frequently you do each of these activities: [single choice per item]. Options: Multiple times a day, roughly once a day, A few times a week, Less than once a week, Never

- Social networking (e.g. Facebook, LinkedIn, Twitter, etc.)
- Browsing the internet
- Shop online
- Playing mobile games

B: Additional descriptive statistics

Table 8 Descriptive statistics of the sample - proportion of people (%)

| Audience | A | B | C | D | E | Baseline |
|---------------------------------|------|------|------|------|------|----------|
| Content read each week | | | | | | |
| Travel | 12.8 | 10.4 | 11.2 | 13.3 | 11.3 | 12.5 |
| Home & Garden | 14.0 | 20.0 | 12.4 | 13.8 | 12.7 | 12.3 |
| Health & Fitness | 15.8 | 15.7 | 12.4 | 15.8 | 13.2 | 16.4 |
| Computers & Games | 12.9 | 11.3 | 18.0 | 10.7 | 14.0 | 11.6 |
| Food & Entertainment | 21.7 | 29.6 | 22.5 | 24.5 | 23.1 | 21.7 |
| Sports | 13.6 | 11.3 | 14.6 | 13.3 | 13.5 | 13.4 |
| Other | 9.2 | 1.7 | 9.0 | 8.7 | 12.3 | 12.3 |
| Department/ area of work | | | | | | |
| Accounting/Finance | 8.0 | 8.3 | 0.0 | 3.5 | 14.9 | 7.4 |
| Administration/Legal/Compliance | 9.6 | 0.0 | 7.7 | 12.3 | 9.1 | 4.5 |
| Human resources | 5.6 | 16.7 | 15.4 | 1.8 | 0.8 | 5.3 |
| Information Technology (IT) | 9.9 | 8.3 | 0.0 | 5.2 | 6.6 | 6.1 |
| Marketing/Sales | 9.2 | 12.5 | 0.0 | 8.8 | 11.6 | 12.7 |
| Other | 40.2 | 29.2 | 61.6 | 47.4 | 36.3 | 38.9 |
| Operations | 17.5 | 25.0 | 15.4 | 21.1 | 20.7 | 25.0 |

Table 9 Proportions of consumer characteristics for ITDM/ no ITDM across segment types

| Audience | ITDM proportions | | Senior executive | | Manager | | Entry-level employee | | Intern/Casual/Part-time | |
|---------------|------------------|---------|---------------------|---------|--------------|---------|----------------------|---------|-------------------------|---------|
| | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM |
| Baseline | 16.0 | 84.0 | 15.6 | 1.6 | 47.9 | 10.1 | 11.5 | 4.2 | 9.4 | 4.0 |
| Deterministic | 18.0 | 82.0 | 31.0 | 2.3 | 34.5 | 8.9 | 7.8 | 5.3 | 5.2 | 4.5 |
| Probabilistic | 10.4 | 89.6 | 31.7 | 1.9 | 38.1 | 7.2 | 3.2 | 3.5 | 7.9 | 3.7 |
| | Works in IT | | Works in Operations | | Small firm | | Mid-sized firm | | Large firm | |
| Audience | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM |
| Baseline | 10.4 | 1.0 | 24.0 | 7.5 | 40.6 | 8.5 | 28.1 | 10.3 | 31.3 | 10.5 |
| Deterministic | 14.7 | 1.9 | 25.0 | 4.0 | 44.0 | 10.0 | 18.1 | 9.4 | 37.9 | 10.6 |
| Probabilistic | 7.9 | 1.1 | 12.7 | 5.7 | 57.1 | 8.7 | 15.9 | 7.2 | 27.0 | 7.8 |
| | Male | | Age 18 to 24 | | Age 25 to 34 | | Age 35 to 44 | | Age 45 to 54 | |
| Audience | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM |
| Baseline | 57.3 | 47.0 | 8.3 | 11.3 | 15.6 | 19.8 | 22.9 | 17.3 | 19.8 | 14.9 |
| Deterministic | 51.7 | 37.9 | 4.3 | 5.3 | 11.2 | 8.3 | 12.1 | 8.7 | 19.0 | 13.4 |
| Probabilistic | 54.0 | 41.1 | 0.0 | 2.6 | 6.3 | 5.4 | 4.8 | 5.9 | 20.6 | 11.1 |

Table 9 continued

| Audience | ITDM proportions | | Senior executive | | Manager | | Entry-level employee | | Interm/Casual/Part-time | |
|---------------|------------------------|---------|---------------------------|---------|---------------------|---------|--------------------------|---------|-------------------------|---------|
| | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM |
| Audience | Age 55 to 64 | | Age 65 and over | | Daily internet use | | Daily social media use | | Shop online daily | |
| Baseline | 20.8 | 15.7 | 12.5 | 21.0 | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM |
| Deterministic | 31.0 | 24.9 | 22.4 | 39.4 | 67.7 | 55.2 | 60.4 | 53.4 | 31.3 | 25.2 |
| Probabilistic | 34.9 | 28.5 | 33.3 | 46.5 | 77.6 | 59.4 | 65.5 | 59.8 | 37.1 | 23.2 |
| Audience | Mobile gaming daily | | Anti-tracking browser use | | Work laptop | | Private laptop for work | | Work phone | |
| Baseline | 37.5 | 39.5 | 46.9 | 50.4 | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM |
| Deterministic | 49.1 | 43.8 | 15.5 | 15.5 | 45.8 | 9.1 | 28.1 | 3.6 | 19.8 | 3.2 |
| Probabilistic | 49.2 | 41.3 | 1.6 | 18.5 | 44.0 | 9.6 | 27.6 | 4.2 | 24.1 | 1.9 |
| Audience | Private phone for work | | Two laptops for work | | Two phones for work | | Reads technology content | | Reads business content | |
| Baseline | 46.9 | 50.4 | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM | ITDM | No ITDM |
| Deterministic | 15.5 | 15.5 | 12.5 | 1.2 | 5.2 | 0.6 | 36.5 | 10.9 | 30.2 | 9.5 |
| Probabilistic | 1.6 | 18.5 | 6.9 | 0.9 | 3.4 | 0.6 | 37.1 | 18.9 | 33.6 | 12.3 |
| | | | 6.3 | 0.2 | 4.8 | 0.2 | 28.6 | 14.6 | 28.6 | 10.4 |

C: Original outcome variables

Table 10 Chi-Square and Fisher's exact tests for original ITDM responsibility binary outcomes

| ITDM Responsibility | Comparison | Proportion 1 (%) | Proportion 2 (%) | Chi-square statistic | Chi-square p-value | Odds Ratio | Fisher's exact p-value |
|------------------------|--|------------------|------------------|----------------------|--------------------|------------|------------------------|
| IT need identification | All off-the-shelf audiences vs. baseline | 0.131 | 0.135 | 0.05 | 0.8263 | 0.97 | 0.8263 |
| IT need identification | Deterministic vs. probabilistic data segment | 0.164 | 0.096 | 12.61 | 0.0004 | 1.84 | 0.0004 |
| IT need identification | Deterministic data segments vs. baseline | 0.164 | 0.135 | 2.06 | 0.1509 | 1.26 | 0.1540 |
| IT need identification | Probabilistic data segments vs. baseline | 0.0962 | 0.135 | 4.43 | 0.0352 | 0.68 | 0.0380 |
| IT vendor selection | All off-the-shelf segments vs. baseline | 0.097 | 0.092 | 0.13 | 0.7207 | 1.06 | 0.7996 |
| IT vendor selection | Deterministic vs. probabilistic data segment | 0.118 | 0.075 | 6.60 | 0.0102 | 1.65 | 0.01257 |
| IT vendor selection | Deterministic data segment vs. baseline | 0.118 | 0.092 | 2.23 | 0.1352 | 1.32 | 0.1402 |
| IT vendor selection | Probabilistic data segments vs. baseline | 0.075 | 0.092 | 1.15 | 0.2844 | 0.80 | 0.2978 |
| Sign IT contracts | All off-the-shelf segments vs. baseline | 0.092 | 0.105 | 0.88 | 0.3476 | 0.86 | 0.3538 |
| Sign IT contracts | Deterministic vs. probabilistic data segment | 0.111 | 0.070 | 6.57 | 0.0104 | 1.67 | 0.0106 |
| Sign IT contracts | Deterministic data segment vs. baseline | 0.111 | 0.105 | 0.13 | 0.7141 | 1.07 | 0.7845 |
| Sign IT contracts | Probabilistic data segments vs. baseline | 0.070 | 0.105 | 4.71 | 0.02986 | 0.64 | 0.0320 |

D: Feature analysis for original outcome variables

Below we present the results for logit models where we regress the likelihood of having at least some responsibility for the originally three measured ITDM responsibilities (identifying needs, selecting vendors, signing contracts) on the covariates from the 1,849 consumer responses. We again split the analyses by demographic/ firmographic features (Table 11) and digital behavior characteristics (Table 12).

We also explored whether we can use proportional log-odds models that consider the ordinal nature of the response scale. However, the data was too sparse to examine a meaningful number of features (Brant tests suggested that many covariates violate the proportional odds assumption). We therefore collapsed the categories into a binary outcome of “Having at least some responsibility” vs. not.

Table 11 Firmographic and demographic features of three ITDM responsibilities

| Dependent variable: Question: | Identify needs | | Select vendors | | Sign contracts | |
|----------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Intercept | -2.378*** (0.319) | -3.110*** (0.377) | -2.870*** (0.383) | -3.672*** (0.437) | -2.378*** (0.319) | -3.638*** (0.423) |
| Senior Executive | | 2.688*** (0.274) | | 2.521*** (0.285) | | 3.091*** (0.293) |
| Manager | | 1.906*** (0.206) | | 1.690*** (0.234) | | 1.632*** (0.243) |
| Works in IT | | 2.575*** (0.340) | | 1.897*** (0.345) | | 1.722*** (0.360) |
| Works in operations | | 0.593** (0.235) | | 0.155 (0.264) | | 0.225 (0.272) |
| Age 25 to 34 | 0.323 (0.368) | -0.052 (0.432) | 0.394 (0.437) | -0.007 (0.492) | 0.323 (0.368) | -0.337 (0.488) |
| Age 35 to 44 | 0.517 (0.361) | -0.208 (0.426) | 0.336 (0.440) | -0.502 (0.499) | 0.517 (0.361) | -0.542 (0.481) |
| Age 45 to 54 | 0.572* (0.348) | -0.434 (0.414) | 0.583 (0.415) | -0.509 (0.474) | 0.572* (0.348) | -0.775* (0.464) |
| Age 55 to 64 | 0.549* (0.333) | -0.108 (0.394) | 0.752* (0.395) | 0.057 (0.447) | 0.549* (0.333) | -0.373 (0.438) |
| Age 65 and over | -0.150 (0.338) | -0.255 (0.399) | -0.039 (0.403) | -0.278 (0.458) | -0.150 (0.338) | -0.504 (0.446) |
| Male | 0.433*** (0.139) | 0.070 (0.171) | 0.503*** (0.161) | 0.242 (0.192) | 0.433*** (0.139) | 0.350* (0.201) |
| Mid-sized firm | | 0.765*** (0.239) | | 0.986*** (0.275) | | 0.961*** (0.289) |
| Small firm | | 1.173*** (0.212) | | 1.547*** (0.237) | | 1.787*** (0.242) |

Table 11 continued

| Dependent variable: Question: | Identify needs | | Select vendors | | Sign contracts | |
|----------------------------------|----------------|-----------|----------------|----------|----------------|----------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Observations | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 | 1,849 |
| Log Likelihood | -707.899 | -503.458 | -568.229 | -408.708 | -707.899 | -377.953 |
| AIC | 1,429.799 | 1,032.915 | 1,150.457 | 843.416 | 1,429.799 | 781.905 |

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 12 Digital behavior characteristics of three ITDM responsibilities

| Dependent variable: Question: | Identify needs (1) | Select vendors (2) | Sign contracts (3) |
|---|-----------------------|-----------------------|-----------------------|
| Intercept | -2.800*** (0.185) | -3.021*** (0.207) | -2.793*** (0.198) |
| Anti-tracking browser | -0.004 (0.206) | -0.134 (0.238) | -0.443* (0.244) |
| Probabilistic data segments | -0.074 (0.194) | 0.045 (0.215) | -0.138 (0.218) |
| Anti-tracking browser X probabilistic data segments | -2.598** (1.085) | -2.031* (1.071) | -1.708 (1.072) |
| Work laptop | 1.802*** (0.197) | 1.289*** (0.231) | 1.091*** (0.237) |
| Private laptop | 2.291*** (0.283) | 1.953*** (0.309) | 2.236*** (0.304) |
| Work phone | 1.300*** (0.288) | 1.807*** (0.296) | 1.889*** (0.299) |
| Private phone | 0.559** (0.255) | 0.640** (0.281) | 0.623** (0.284) |
| Two phones | -0.491 (0.690) | -1.433** (0.679) | -1.135* (0.684) |
| Two laptops | -1.931*** (0.505) | -1.354** (0.528) | -1.181** (0.526) |
| Daily social-media user | -0.467** (0.221) | -0.677*** (0.251) | -0.224 (0.252) |
| Daily online browsing | 0.191 (0.228) | 0.092 (0.258) | -0.441* (0.262) |
| Daily online shopping | 0.518** (0.229) | 0.833*** (0.264) | 1.127*** (0.270) |
| Daily mobile gaming | -0.185 | -0.212 | -0.418 |

Table 12 continued

| Dependent variable: Question: | Identify needs (1) | Select vendors (2) | Sign contracts (3) |
|----------------------------------|-----------------------|-----------------------|-----------------------|
| | (0.224) | (0.260) | (0.265) |
| Travel content | 0.236 (0.205) | 0.085 (0.232) | 0.197 (0.235) |
| Home & Garden | 0.373* (0.207) | 0.157 (0.239) | 0.112 (0.243) |
| Health & Fitness | -0.295 (0.209) | -0.383 (0.239) | -0.234 (0.240) |
| Computers & Games | -0.276 (0.224) | -0.451* (0.256) | -0.393 (0.264) |
| Food & Entertainment | -0.131 (0.195) | 0.056 (0.220) | -0.102 (0.225) |
| Business | 0.328 (0.212) | 0.506** (0.232) | 0.604** (0.236) |
| Technology | 0.641*** (0.213) | 0.697*** (0.237) | 0.271 (0.247) |
| Sports | 0.310 (0.193) | 0.264 (0.216) | 0.389* (0.218) |
| Observations | 1,849 | 1,849 | 1,849 |
| Log Likelihood | -545.288 | -450.777 | -448.408 |
| AIC | 1,134.576 | 945.554 | 940.817 |

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

E: Feature analysis for sample of employed respondents

Below we present the logit model results for our analysis of ITDM/ MITDM characteristics when we only use the sample proportion of respondents who were employed ($n = 710$). The model specifications are identical, with one exception: For this analysis, we also include the variable ‘not much online for work’, which only applies to respondents who are employed. The effect is negative as expected, while all our other main findings are not affected (Tables 13 and 14).

F: Robustness check: impact of prior ad exposures on reaching ITDM

We examine whether the number of prior ad exposures before our validation campaign is correlated with the likelihood of being an ITDM or MITDM. Table 15 summarizes the mean and median impressions of prior ads per use for each of our five vendors.

We then carry out logit models in which we add the number of impressions to the model in addition to either demographics (age, gender) and firmographics (firm size)

Table 13 Firmographic and demographic features of IT decision-makers

| Dependent variable: | Likelihood of online user being an ITDM | | | Likelihood of online user being an MITDM | | | | | | |
|---------------------|---|----------------------|----------------------|--|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Intercept | -0.571*** (0.092) | -1.048*** (0.120) | -1.014*** (0.356) | -0.506*** (0.133) | -1.244*** (0.392) | -1.480*** (0.114) | -2.143*** (0.171) | -2.145*** (0.468) | -1.846*** (0.187) | -2.614*** (0.519) |
| Works in IT | 0.992*** (0.296) | | | | 1.247*** (0.318) | 0.730** (0.316) | | | | 0.971*** (0.349) |
| Works in operations | 0.166 (0.191) | | | | 0.168 (0.211) | -0.178 (0.250) | | | | -0.275 (0.279) |
| Senior Executive | | 1.909*** (0.249) | | | 1.872*** (0.266) | | 2.044*** (0.263) | | | 1.857*** (0.288) |
| Manager | | 0.828*** (0.175) | | | 0.871*** (0.186) | | 0.747*** (0.234) | | | 0.781*** (0.250) |
| Age 25 to 34 | | | -0.037 (0.401) | | -0.096 (0.423) | | | -0.131 (0.536) | | -0.202 (0.562) |
| Age 35 to 44 | | | 0.230 (0.395) | | -0.020 (0.422) | | | -0.175 (0.535) | | -0.513 (0.567) |
| Age 45 to 54 | | | 0.224 (0.380) | | -0.187 (0.407) | | | 0.104 (0.500) | | -0.445 (0.534) |

Table 13 continued

| Dependent variable: | Likelihood of online user being an ITDM | | | | Likelihood of online user being an MITDM | | | | | |
|---------------------|---|----------|---------------------|--------------------|--|----------|----------|---------------------|---------------------|--------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Age 55 to 64 | | | 0.425 (0.371) | | 0.049 (0.396) | | | 0.721 (0.478) | | 0.260 (0.507) |
| Age 65 and over | | | 0.907** (0.390) | | 0.445 (0.421) | | | 0.981** (0.494) | | 0.253 (0.531) |
| Male | | | 0.424*** (0.158) | | 0.200 (0.173) | | | 0.590*** (0.199) | | 0.474** (0.217) |
| Mid-sized firm | | | | -0.382* (0.205) | | | | | -0.141 (0.287) | -0.082 (0.303) |
| Small firm | | | | 0.380** (0.180) | | | | | 0.950*** (0.231) | 0.626** (0.253) |
| Observations | 710 | 710 | 710 | 710 | 710 | 710 | 710 | 710 | 710 | 710 |
| Log Likelihood | -468.097 | -438.491 | -463.841 | -466.231 | -423.392 | -342.306 | -314.513 | -331.512 | -331.617 | -296.848 |
| AIC | 942.194 | 882.981 | 941.682 | 938.463 | 872.783 | 690.612 | 635.026 | 677.024 | 669.234 | 619.696 |

Note: *p < 0.1; **p < 0.05; ***p < 0.01

Table 14 Digital behavior characteristics of IT decision-makers

| Dependent variable: | Likelihood of online user being an ITDM | | | | Likelihood of online user being an MITDM | | | | | |
|---|---|----------------------|----------------------|----------------------|--|----------------------|----------------------|----------------------|----------------------|----------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Intercept | -0.417*** (0.106) | -0.966*** (0.123) | -0.345** (0.159) | -0.731*** (0.119) | -1.385*** (0.261) | -1.386*** (0.129) | -1.760*** (0.153) | -1.107*** (0.187) | -1.683*** (0.150) | -1.385*** (0.261) |
| Anti-tracking browser | 0.165 (0.198) | | | | -0.247 (0.281) | -0.223 (0.258) | | | | -0.247 (0.281) |
| Probabilistic data segments | -0.091 (0.192) | | | | 0.263 (0.249) | 0.074 (0.230) | | | | 0.263 (0.249) |
| Anti-tracking browser X probabilistic data segments | -2.877*** (1.051) | | | | -1.845* (1.093) | -1.684 (1.069) | | | | -1.845* (1.093) |
| Work laptop | | 0.472*** (0.182) | | | -0.316 (0.267) | | -0.141 (0.258) | | | -0.316 (0.267) |
| Private laptop | | 1.288*** (0.261) | | | 0.542 (0.330) | | 0.731** (0.292) | | | 0.542 (0.330) |
| Work phone | | 1.006*** (0.261) | | | 0.885*** (0.314) | | 0.828*** (0.291) | | | 0.885*** (0.314) |
| Private phone | | -0.017 (0.221) | | | 0.064 (0.290) | | 0.216 (0.262) | | | 0.064 (0.290) |
| Two phones | | -0.464 (0.618) | | | -0.211 (0.688) | | -0.370 (0.629) | | | -0.211 (0.688) |
| Two laptops | | -0.468 (0.491) | | | 0.648 (0.558) | | 0.483 (0.511) | | | 0.648 (0.558) |
| Not much online for work | | | -0.739*** (0.268) | | -0.099 (0.372) | | | -0.449 (0.335) | | -0.099 (0.372) |
| Daily social-media user | | | -0.499*** (0.209) | | -0.763*** (0.281) | | | -0.876*** (0.266) | | -0.763*** (0.281) |
| Daily online browsing | | | 0.318 | | -0.250 | | | 0.156 | | -0.250 |

Table 14 continued

| Dependent variable: | Likelihood of online user being an ITDM | | | | Likelihood of online user being an MITDM | | | | | |
|-----------------------|---|----------|----------|----------|--|----------|----------|----------|----------|----------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Daily online shopping | | | (0.214) | | (0.289) | | | (0.263) | | (0.289) |
| | | | 0.539*** | | 1.157*** | | | 1.121*** | | 1.157*** |
| | | | (0.209) | | (0.300) | | | (0.277) | | (0.300) |
| Daily mobile gaming | | | -0.221 | | -0.417 | | | -0.534* | | -0.417 |
| | | | (0.207) | | (0.299) | | | (0.278) | | (0.299) |
| Travel content | | | | 0.248 | 0.133 | | | | 0.154 | 0.133 |
| | | | | (0.202) | (0.261) | | | | (0.245) | (0.261) |
| Home & Garden | | | | 0.469** | 0.077 | | | | 0.107 | 0.077 |
| | | | | (0.211) | (0.283) | | | | (0.263) | (0.283) |
| Health & Fitness | | | | -0.452** | -0.351 | | | | -0.305 | -0.351 |
| | | | | (0.203) | (0.267) | | | | (0.250) | (0.267) |
| Computers & Games | | | | -0.252 | -0.481 | | | | -0.484* | -0.481 |
| | | | | (0.219) | (0.295) | | | | (0.275) | (0.295) |
| Food & Entertainment | | | | -0.272 | -0.132 | | | | -0.293 | -0.132 |
| | | | | (0.192) | (0.254) | | | | (0.240) | (0.254) |
| Business | | | | 0.584*** | 0.834*** | | | | 0.932*** | 0.834*** |
| | | | | (0.204) | (0.256) | | | | (0.232) | (0.256) |
| Technology | | | | 0.795*** | 0.401 | | | | 0.527** | 0.401 |
| | | | | (0.215) | (0.271) | | | | (0.254) | (0.271) |
| Sports | | | | 0.201 | 0.246 | | | | 0.246 | 0.246 |
| | | | | (0.188) | (0.241) | | | | (0.225) | (0.241) |
| Observations | 710 | 710 | 710 | 710 | 710 | 710 | 710 | 710 | 710 | 710 |
| Log Likelihood | -464.114 | -445.779 | -463.689 | -452.027 | -303.375 | -342.035 | -330.681 | -330.964 | -329.444 | -303.375 |
| AIC | 936.228 | 905.559 | 939.378 | 922.055 | 652.750 | 692.069 | 675.362 | 673.927 | 676.888 | 652.750 |

Note: *p<0.1; **p<0.05; ***p<0.01

Table 15 Number of ad impressions per user prior to our calibration test

| Audience | Mean | Median |
|----------|------|--------|
| A | 11.6 | 5 |
| B | 16.5 | 8 |
| C | 11.1 | 3 |
| D | 9.92 | 3 |
| E | 5.42 | 4 |

or behavioral data (content regular browsed, social-media/shopping/gaming activity, number of laptops and phones). For this analysis, we only use the 1249 respondents from the five vendors as the baseline group has no prior impressions data that we can access (there were no tagged users which were tracked across campaigns). The results of these 10 models are shown in Table 16 below.

We find that our conclusions do not change and that the number of prior ad exposures (measured in impressions per user) results in no significant effect on our outcome measures.

G: Robustness check: impact of customer heterogeneity on reaching ITDM

Our analyses in Study 1 only compared the frequencies of ITDM/ MITDM across our different audience segments. Here we examine whether possible customer heterogeneity in our measured covariates may affect our study conclusions. We carry out logit and linear probability models ([LPM], using ordinary least squares regression) in which we compare the impact of deterministic and probabilistic data (in comparison to the baseline, our reference category) for the likelihood of reaching an ITDM or MITDM. In particular, we add consumer characteristics as further control variables to the models to assess whether our effects could be based on slightly different user populations across each of the samples of the five ITDM vendors. We include either demographics (age, gender) and firmographics (firm size) or behavioral data (content regular browsed, social-media/shopping/gaming activity, number of laptops and phones). The results of the regressions are shown for the logit models in Table 17 and for the LPM in Table 18 below.

We can see that our key finding of no significant differences between probabilistic or deterministic data in comparison to the baseline (random user selection) is not affected. If we include demographics and firmographics as control variables (columns 2 and 4) or no covariates (column 1), we even find statistically significant negative differences for probabilistic data segments in the case of ITDM reach (in comparison to the baseline).

Thus, our further parametric analyses confirm the relevance of online behavior signals and that customer heterogeneity does not change our main finding of poor performance of probabilistic and deterministic data in reaching IT-decisionmakers.

Table 16 Logit models examining impact of audience type and number of prior expressions on being an ITDM

| Dependent variable: | Likelihood of being an ITDM | | | | Likelihood of being an MITDM | | | | | |
|--------------------------------|-----------------------------|----------------------|----------------------|--------------------|------------------------------|---------------------|---------------------|---------------------|-------------------|-------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Probabilistic data segment | -0.629*** (0.168) | -0.637*** (0.170) | -0.650*** (0.191) | -0.333* (0.201) | -0.395* (0.214) | -0.441** (0.220) | -0.468** (0.223) | -0.514** (0.247) | -0.072 (0.256) | -0.248 (0.275) |
| Prior ad exposures | | -0.001 (0.005) | -0.003 (0.006) | 0.002 (0.006) | -0.0003 (0.006) | | -0.006 (0.008) | -0.008 (0.008) | -0.003 (0.008) | -0.005 (0.009) |
| Demo- & firmographics included | | | Yes | Yes | Yes | | | Yes | Yes | Yes |
| Behavioral covariates included | | | | Yes | Yes | | | | Yes | Yes |
| Log Likelihood | -506.0 | -505.9 | -411.9 | -370.8 | -342.2 | -331.5 | -331.2 | -260.9 | -250.5 | -222.6 |
| AIC | 1,016.0 | 1,017.9 | 845.8 | 783.6 | 742.5 | 667.0 | 668.3 | 543.8 | 543.0 | 503.3 |

Note: Reference category is deterministic data. * p<0.1; ** p<0.05; *** p<0.01

Table 17 Logit models examining impact of audience type and covariates on being an ITDM

| Dependent variable: Variable: | Likelihood of being an ITDM | | | | Likelihood of being an MITDM | | | |
|----------------------------------|-----------------------------|---------------------|-------------------|--------------------|------------------------------|-------------------|------------------|-------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Deterministic data segment | 0.139 (0.151) | 0.173 (0.175) | 0.013 (0.182) | 0.015 (0.195) | 0.296 (0.212) | 0.228 (0.238) | 0.286 (0.238) | 0.226 (0.255) |
| Probabilistic data segment | -0.490*** (0.174) | -0.459** (0.201) | -0.327 (0.202) | -0.384* (0.223) | -0.144 (0.236) | -0.227 (0.268) | 0.162 (0.261) | -0.045 (0.290) |
| Demo-& firmographics included | | Yes | | Yes | | Yes | | Yes |
| Behavioral covariates included | | | Yes | Yes | | | Yes | Yes |
| Log Likelihood | -769.8 | -629.0 | -575.9 | -529.5 | -481.1 | -381.8 | -386.0 | -337.8 |
| AIC | 1,545.6 | 1,280.1 | 1,193.8 | 1,117.0 | 968.1 | 785.5 | 814.0 | 733.6 |

Note: Reference category for segment data is baseline group (random user selection). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 18 Linear probability models examining impact of audience type and covariates on being an ITDM

| Dependent variable: Variable: | Likelihood of being an ITDM | | | | Likelihood of being an MITDM | | | |
|----------------------------------|-----------------------------|-------------------|-----------------|------------------|------------------------------|-----------------|----------------|------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Deterministic data segment | 0.02 (0.02) | 0.02 (0.02) | 0.003 (0.02) | -0.002 (0.02) | 0.02 (0.01) | 0.01 (0.01) | 0.02 (0.01) | 0.01 (0.01) |
| Probabilistic data segment | -0.1*** (0.02) | -0.04** (0.02) | -0.03 (0.02) | -0.03* (0.02) | -0.01 (0.01) | -0.01 (0.01) | 0.01 (0.01) | -0.001 (0.01) |
| Demo-& firmographics included | | Yes | | Yes | | Yes | | Yes |
| Behavioral covariates included | | | Yes | Yes | | | Yes | Yes |
| Adjusted R ² | 0.01 | 0.2 | 0.3 | 0.3 | 0.001 | 0.1 | 0.1 | 0.2 |

Note: Reference category for segment data is baseline group (random user selection). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Acknowledgements We are grateful for generous support of this study from an anonymous measurement company, HP and their media agency. We wish to specifically thank Freddie Liversidge and Ian Mundorff for their outstanding help with this project. Thank you to NSF CAREER Award 6923256 and the MBS Internal Research Grant Program for financial support.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflicts of interest Catherine Tucker has a disclosure statement at <https://mitmgmtfaculty.mit.edu/ctucker/disclosure/>, which lists various companies she has consulted for.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Andrews, M., Luo, X., Fang, Z., & Ghose, A. (2016). Mobile ad effectiveness: Hyper-contextual targeting with crowdedness. *Marketing Science*, 35(2), 218–233.
- Ansari, A., & Mela, C. (2003). E-customization. *Journal of Marketing Research*, 40(2), 131–145.
- Bergemann, D., & Bonatti, A. (2015). *Selling Cookies*. *American Economic Journal: Microeconomics*, 7(3), 259–294.
- Bucklin, R. E., & Sismeiro, C. (2003). A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, 40(3), 249–267.
- Chen, Y., Li, X., & Sun, M. (2017). Competitive mobile geo targeting. *Marketing Science*, 36(5), 666–682.
- Chen, Y., Narasimhan, C., & Zhang, Z. J. (2001). Individual marketing with imperfect targetability. *Marketing Science*, 20(1), 23–41.
- Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4), 404–413.
- Coey, D., & Bailey, M. (2016). People and cookies: Imperfect treatment assignment in online experiments. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, (pp. 1103–1111). ACM Press. New York, New York, USA
- Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1), 92–112.
- De Bruyn, A., & Otter, T. (2022). Bayesian consumer profiling: How to estimate consumer characteristics from aggregate data. *Journal of Marketing Research*.
- Goldfarb, A. (2014). What is Different About Online Advertising? *Review of Industrial Organization*, 44(2), 115–129.
- Goldfarb, A., & Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30, 389–404.
- Gonzales, O. (2022). Google Pushes Back Third-Party Cookie Blocking in Chrome to 2024. <https://www.cnet.com/news/privacy/google-pushes-back-third-party-cookie-blocking-inchrome-to-2024/>, retrieved 12-Dec-2022.
- Google (2021a). An updated timeline for Privacy Sandbox milestones. <https://blog.google/products/chrome/updated-timeline-privacy-sandbox-milestones/>, retrieved 15-Oct-2021.
- Google (2021b). Charting a course towards a more privacy-first web. <https://blog.google/products/ads-commerce/a-more-privacy-first-web/>, retrieved 25-Mar-2021.
- Hauser, J. R., Urban, G. L., Liberali, G., & Braun, M. (2009). Website morphing. *Marketing Science*, 28(2), 202–223.
- Johnson, G. A., Lewis, R. A., & Nubbemeyer, E. I. (2017). Ghost Ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, 54(6), 867–884.
- Johnson, G. A., Shriver, S. K., & Du, S. (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science*, 39(1), 33–51.
- Lambrecht, A., & Tucker, C. (2013). When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research*, 50(5), 561–576.
- Lambrecht, A., & Tucker, C. E. (2019). Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of stem career ads. *Management Science*, 65(7), 2966–2981.
- Lebow, S. (2021). US B2B tech ad spending increased by 50% in 2020. *eMarketer*.
- Li, H., Kannan, P. K., Viswanathan, S., & Pani, A. (2016). Attribution strategies and return on keyword investment in paid search advertising. *Marketing Science*, 35(6), 831–848.
- Lin, T., & Misra, S. (2022). Frontiers: the identity fragmentation bias. *Marketing Science*, 41(3), 433–440
- Luo, X., Andrews, M., Fang, Z., & Phang, C. W. (2014). Mobile targeting. *Management Science*, 60(7), 1738–1756.
- Manchanda, P., Dube, J.-P., Goh, K. Y., & Chintagunta, P. K. (2006). The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1), 98–108.
- Miller, K. M., & Skiera, B. (2017). Economic loss of cookie lifetime restrictions. Available at SSRN 3046254.

- Murthi, B., & Sarkar, S. (2003). The role of the management sciences in research on personalization. *Management Science*, *49*(10), 1344–132.
- Nagpal, K. (2021). Identity resolution misses the mark without systemic measurement and attribution. <https://www.adexchanger.com/content-studio/identity-resolution-misses-the-mark-without-systemic-measurement-and-attribution/>, retrieved 26-Oct-2021.
- Narayanan, S., & Manchanda, P. (2006). Heterogeneous learning and the targeting of marketing communication for new products. *Ssrn*, *28*(3), 424–441.
- Neumann, N., Tucker, C. E., & Whitfield, T. (2019). Frontiers: How effective is third-party consumer profiling? evidence from field studies. *Marketing Science*, *38*(6), 918–926.
- Newscorp Press Release (2019). <https://www.newscorpaustralia.com/news-corp-australia-takesnews-connect-to-new-levels/>, retrieved 10-Feb-2021.
- Nylen, E. (2018). Comparing value of audience versus context using attention and brand metrics. *The Drum*.
- NYT Open Team (2020). <https://open.nytimes.com/to-serve-better-ads-we-built-our-own-dataprogram-c5e039bf247b/>, retrieved 10-Feb-2021.
- Park, Y.-H., & Fader, P. S. (2004). Modeling browsing behavior at multiple websites. *Marketing Science*, *23*(3), 280–303.
- Ryan, J. (2020). eMarketer Report - US B2B Digital Advertising 2020.
- Summers, C. A., Smith, R. W., & Reczek, R. W. (2016). An audience of one: Behaviorally targeted ads as implied social labels. *Journal of Consumer Research*, *43*(1), 156–178.
- Trusov, M., Ma, L., & Jamal, Z. (2016). Crumbs of the cookie: User profiling in customer-base analysis and behavioral targeting. *Marketing Science*, *35*(3), 405–426.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Nico Neumann¹  · Catherine E. Tucker² · Kumar Subramanyam³ · John Marshall³

Catherine E. Tucker
cetucker@mit.edu

Kumar Subramanyam
kumar.subramanyam@hp.com

John Marshall
john.marshall@hp.com

- ¹ Melbourne Business School, University of Melbourne, Melbourne, Australia
- ² MIT Sloan School of Management, Cambridge, MA and Research Associate at the NBER, Cambridge, MA, USA
- ³ HP Inc., Palo Alto, CA, USA