

Arnold, René; Marcus, J. Scott; Petropoulos, Georgios; Schneider, Anna

Conference Paper

Is data the new oil? Diminishing returns to scale

29th European Regional Conference of the International Telecommunications Society (ITS):
"Towards a Digital Future: Turning Technology into Markets?", Trento, Italy, 1st - 4th August,
2018

Provided in Cooperation with:

International Telecommunications Society (ITS)

Suggested Citation: Arnold, René; Marcus, J. Scott; Petropoulos, Georgios; Schneider, Anna (2018) :
Is data the new oil? Diminishing returns to scale, 29th European Regional Conference of the
International Telecommunications Society (ITS): "Towards a Digital Future: Turning Technology into
Markets?", Trento, Italy, 1st - 4th August, 2018, International Telecommunications Society (ITS),
Calgary

This Version is available at:

<https://hdl.handle.net/10419/184927>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen
Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle
Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich
machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen
(insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten,
gelten abweichend von diesen Nutzungsbedingungen die in der dort
genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

*Documents in EconStor may be saved and copied for your personal
and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to
exhibit the documents publicly, to make them publicly available on the
internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content
Licence (especially Creative Commons Licences), you may exercise
further usage rights as specified in the indicated licence.*

Is data the new oil?

Diminishing returns to scale

René Arnold¹, J. Scott Marcus¹, Georgios Petropoulos¹, Anna Schneider²

Abstract

A key advantage of online advertising over offline is that online advertising can, with sufficient data, be far more accurately targeted than traditional advertising. But how much data is enough? The empirical literature tends to suggest that there are indeed economies of scale in using data for market targeting, but that these benefits are subject to diminishing returns in a static perspective. Is there a plateau, and is it perhaps very large?

It is clear that a certain amount of data is necessary to identify meaningful consumer segments and to offer targeted advertising space as part of an advertising campaign; however, a simple correlation between the volume of data gathered by an advertiser and the return on investment of an advertising campaign neglects the complexity of advertising effectiveness.

We provide a general assessment of key elements of the literature on economies of scale in the use of data for online advertising, and then seek to link these to the general literature on market targeting in order to provide insights as to the factors that limit effectiveness in using big data for market targeting.

1 Introduction

Data is very different from oil in almost all of its characteristics.³ However, there is one parallel to oil: While 100 years ago, the companies with the highest market valuation at the stock exchange built their success on oil, today the most valuable companies appear to depend on data just as much.⁴

In part, their success can be attributed to their superior access to larger volume of data, which these companies use to improve their services. One of these services is providing advertising space and targeting the advertising content that is being displayed to a specific audience. Beyond competition related issues, the alleged correlation between the sheer volume of data and success holds implications for consumers' privacy online as well as for the potential manipulation of consumers' preferences and behaviour.

³ For a discussion on the characteristics of (big) data and its value see Hildebrandt and Arnold (2016).

⁴ Notably, the most highly valued company in the world that is not listed on the stock exchange is Aramco – a Saudi Arabian oil company - with an alleged value of US\$2 trillion.

Thus, the correlation between data and success has entered policymakers' focus of attention. However, such a straightforward correlation neglects the fundamental characteristics of data and the complexity of advertising as it relates to consumer behaviour. Our paper seeks to shine a light on these two issues drawing on a multi-disciplinary literature review.

The remainder of the paper is structured along these themes. First, we discuss the economics of data focusing on the question of whether there is a generally diminishing return to scale for data used to target consumers. Second, we draw on studies investigating the role of data in targeting and the implications for advertising effectiveness as well as market actor behaviour. By and large, these studies stem from the realm of marketing and economics. We then introduce some more fundamental issues relating to targeted advertising effectiveness, drawing on consumer behaviour literature from both positivist and constructivist / interpretivist schools of thought. We close the paper with a short conclusion.

2 Data and economies of scale in the literature

In terms of advertising targeting, is more data always better? How much better? What might we expect?

Data-based market targeting can be viewed as a form of *predictive modelling* – we are trying to infer how an individual is likely to respond to a given advertisement, and more generally the individual's likely predisposition to purchase a product or service, based on aspects of the individual's known behaviour.

To put some rigour into the discussion, we introduce a few definitions from Junqué de Fortuny, Martens, and Provost (2013): "Predictive modeling is based on one or more data *instances* for which we want to predict the value of a *target variable*. Data-driven predictive modeling generally induces a model from *training data*, for which the value of the target (the label) is known. These instances typically are described by a vector of *features* from which the predictions will be made." [emphasis added]

We are all familiar with prediction and estimation exercises from everyday life. We all recognise that, if we flip an unbiased coin, the more flips, the closer the fraction of "heads" is likely to be to 0.50 (an example of the *law of large numbers*). This is a trivial example of trying to estimate the central tendency of a distribution (typically the arithmetic average or *mean*) based on a *sample* drawn from a larger *population*. Often we also want to know the standard deviation of the distribution, which is a measure of dispersion from the mean.

The sample mean has its own standard deviation (due mainly to variation in the samples that could be chosen). The *standard error of the mean* is known to be equal to σ/\sqrt{n} , where σ is the (unknown) standard deviation of the population, and n is the number of observations in the sample. What this simple formula tells us is that, as the sample size increases, the sample is likely to provide better and better estimates of the true population mean; however, since the term in the denominator is based on the square root of the number of observations rather than the number itself, the improvement with increased sample size is *less than linear*.

It is for this reason that political polls are often conducted with between 1,000 and 2,000 randomly selected respondents. On the one hand, the sample size must be large enough to ensure that the answers sufficiently approximate the real distribution of sentiment in the electorate, which is to say that the standard error of the mean must be sufficiently small. On the other hand, sampling costs money. At some point the *law of diminishing returns* kicks in based especially on the less than linear improvement with sample size.

This means that at some point, further expanding the sample size is not cost-justified. Expressed in economic terms, the marginal increase in the economic value does not exceed the marginal cost of increasing the sample size.

Given that estimation of the population mean based on a sample is an example of predictive modelling, the natural intuition is that the accuracy and economics of prediction for market targeting using big data might follow roughly the same rules. A natural set of (unproven at this point) hypotheses is thus:

- Increasing the number of instances in a predictive model is likely to always increase prediction accuracy.
- The improvement in prediction accuracy can be expected to be less than linear in the number of instances.
- If the cost per instance of increasing the number of instances is greater than zero, there will always be some point at which the marginal utility of having more instances no longer exceeds the cost of obtaining and maintaining them. At that point, a rational player would no longer invest in expanding the number of instances.

Intuition is all well and good, but it is not an altogether reliable guide. What do we know based on actual empirical results? Further, is there any number of instances for which marginal benefits are zero or negative?

Banko and Brill (2001) provide an excellent early assessment. These Microsoft researchers were looking for ways to improve the effectiveness of natural language processing. They had observed that most natural machine learning approaches to language processing at the time were based on relatively small training *corpora*. “While the amount of available online text has been increasing at a dramatic rate, the size of training corpora typically used for learning has not.” This was largely a consequence of the “potentially large cost of annotating data for those learning methods that rely on labeled text”. They wanted to explore the use of much larger training corpora, but foregoing human annotation. A key driver for their approach was the observation that a simply enormous volume of (English language) text had become available on the web, and that it was available to researchers at no cost. In particular, they were trying to understand whether improved grammar checking could be achieved with greater cost-efficiency through better algorithms, versus by simply expanding the volume of text (i.e. the corpora) used to drive the machine learning exercise.

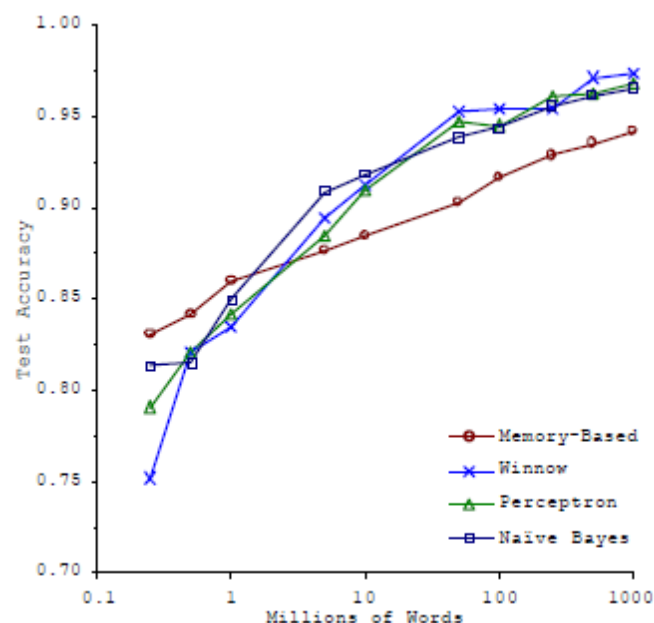
Banko and Brill (2001) applied four different algorithms to the natural language problem of *confusion set disambiguation*. “Confusion set disambiguation is the problem of choosing the correct use of a word, given a set of words with which it is commonly confused. Example confusion sets include: [principle, principal], [then, than], [to, two, too], and [weather,

whether].” Their choice of this particular natural language problem for the study was motivated in large part by the observation that “labeled training data is essentially free, since the correct answer is surface apparent in any collection of reasonably well-edited text.”

They obtained a striking and much cited result (see Figure 1). For corpora sizes up to a billion words, every increase in corpus size was associated with an increase in prediction accuracy; however, a law of diminishing returns is manifest in their results (note the downward curve, implying a negative second derivative). There was, however, no tendency for the benefits of large corpora to “top out” for any of the algorithms at the corpus sizes used (which are, to be sure, tiny by today’s standards).

The differences among the algorithms were far less important for this problem than the increase in the size of the corpora. Among the algorithms, the memory-based algorithm was however something of an outlier – it presumably benefitted somewhat less from large corpora because it “used only the word before and word after as features”, while the other algorithms used “the set of words within a window of the target word, and collocations containing words and/or parts of speech” (Banko & Brill, 2001). One might infer that the other three algorithms were better able to take advantage of large corpora.

Figure 1. Test accuracy versus size of corpora for confusion set disambiguation.



Source: Banko and Brill (2001)

Subsequent experience has confirmed that, in many contexts (but not all), the benefits of large training corpora tend to be more important than the benefits to be obtained by improving algorithms. As the Google researchers Halevy, Norvig, and Pereira (2009) put it, “The biggest successes in natural-language-related machine learning have been statistical speech recognition and statistical machine translation. The reason for these successes is not that these tasks are easier than other tasks; they are in fact much harder than tasks such as document classification that extract just a few bits of information from each document. The

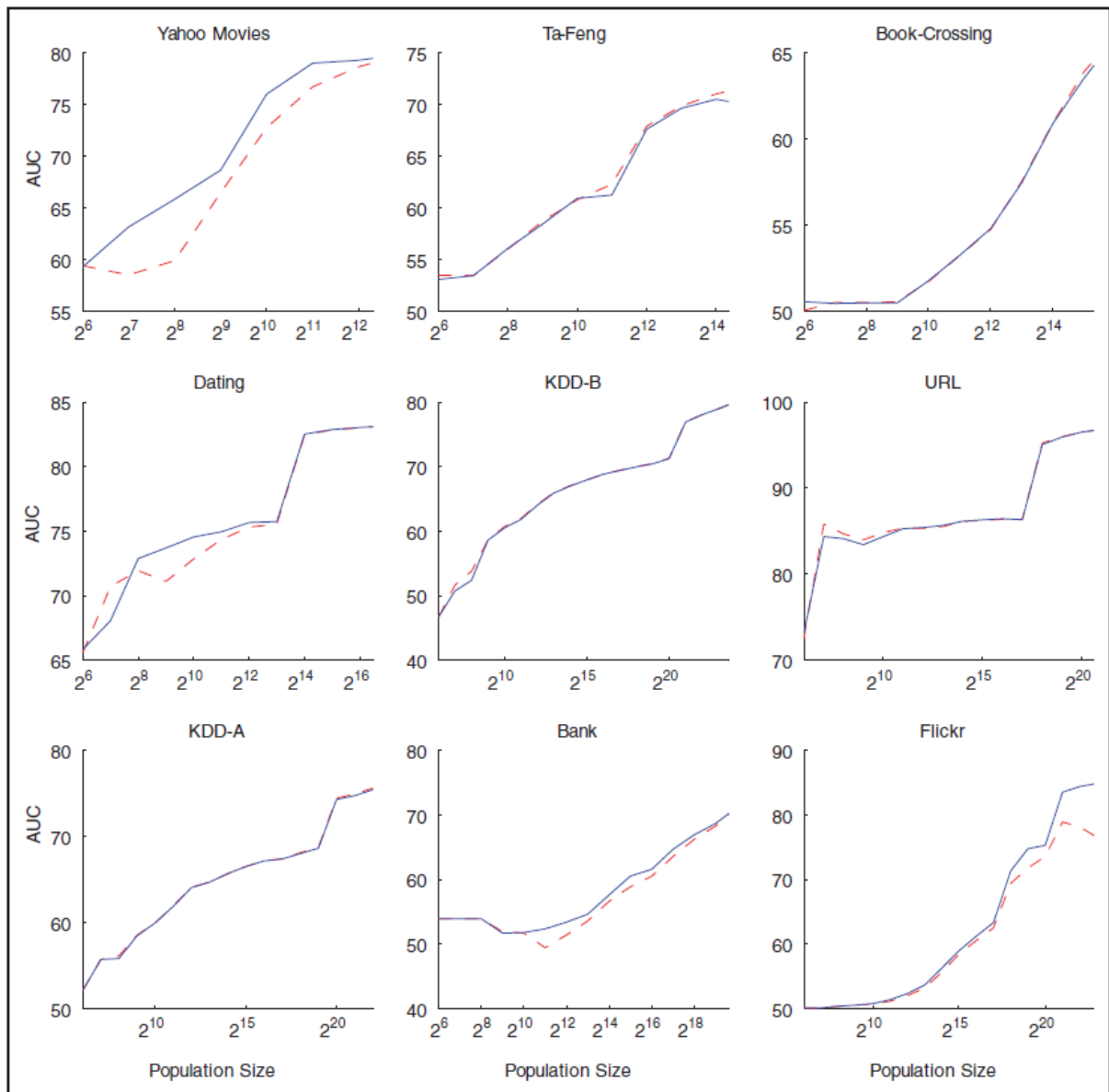
reason is that translation is a natural task routinely done every day for a real human need (think of the operations of the European Union or of news agencies). The same is true of speech transcription (think of closed-caption broadcasts)."

Junqué de Fortuny et al. (2013) obtained results similar to those of Bank and Brill for nine different predictions based on nine different large corpora, and using two different algorithms. In Figure 2, AUC is "area under the curve" (i.e. accuracy), while the multivariate algorithm is depicted as solid, blue and the multinomial Naïve Bayes as dashed red. In each case, a general improvement in prediction is visible with increased corpus size, even at very large sizes. That the curves appear to increase monotonically suggests that, if an asymptote exists, it has not yet been reached.

Diminishing returns to scale are however evident, which is, as the authors note, "typical for the relationship between the amount of data and predictive performance, especially for linear models. The marginal increase in generalization accuracy decreases with more data for several reasons. First, there simply is a maximum possible predictive performance (the 'Bayes rate') due to the inherent randomness in the data and the fact that accuracy can never be better than perfect. Second, modelling (especially linear modeling) tends to find the larger generalities first; modeling with larger datasets usually helps to work out nuances, 'small disjuncts,' and other nonlinearities that are difficult or impossible to capture from smaller datasets." (Junqué de Fortuny et al., 2013)

A common characteristic among each of these nine corpora is that each represents "sparse, fine-grained feature data, such as that derived from the observation of the behaviours of individuals. For example, data on individuals' visits to massive numbers of specific web pages are used in predictive analytics for targeting online display advertisements." (Junqué de Fortuny et al., 2013) For this kind of data, for any given instance (typically an individual), hardly any of the features contain values other than null or zero. This is to say that there are an enormous number of potential locations or potential websites that most users have never visited.

Figure 2. Learning curves (AUC) for two different algorithms with nine different large scale corpora.



Source: Junqué de Fortuny et al. (2013)

Is more data always better? Clearly not. Suffice it to say that there are numerous instances in the literature where more data either produces negligible gain, or in some cases negative gain. Indeed, the use of data volumes “beyond what can currently be handled well on typical modern platforms” seems not to be particularly advantageous for traditional market analytics applications where the data are more dense, such as banking and telecommunications, credit scoring, and attrition management. (Junqué de Fortuny et al., 2013)

This observation brings us to a discussion of the number of the number of *features*, as distinct from the number of *instances*. A key observation in Junqué de Fortuny et al. (2013) is that, if too many features are included that are totally irrelevant to the target variable to be predicted, their inclusion can lead to *overfitting* that actually reduces the accuracy of the model. Conversely, if the features make a tiny but positive contribution to prediction, than they can be expected to improve the accuracy of prediction. In the former case, it may be

appropriate to reduce the number of features used in the model (but determination of which features to drop is a complicated problem). In the latter case, working with a huge number of features may be in order, both to improve model accuracy, and because some long tail rare but important behaviours may be visible only in a sufficiently large corpus. Junqué de Fortuny et al. (2013) demonstrate a case where there are returns to scale even with millions of observations and 400,000 features.

If data are biased, gathering more data might be of limited value to the extent that the new data are just as biased as the old; however, correction for bias will presumably also require substantial volumes of data.

This addresses the *benefits* of large corpora of data. Much less appears in the literature as regards the *cost* of obtaining and maintaining large corpora of data. Recall that the cost is key because a rational actor will be motivated to increase the size of the corpus until it finds the marginal cost of doing so exceeds the marginal benefit of a larger corpus.

Here it is necessary to distinguish between the costs of *storing and processing* the data versus the cost of getting it in the first place, which is often a matter of *persuading the consumer to provide it*.

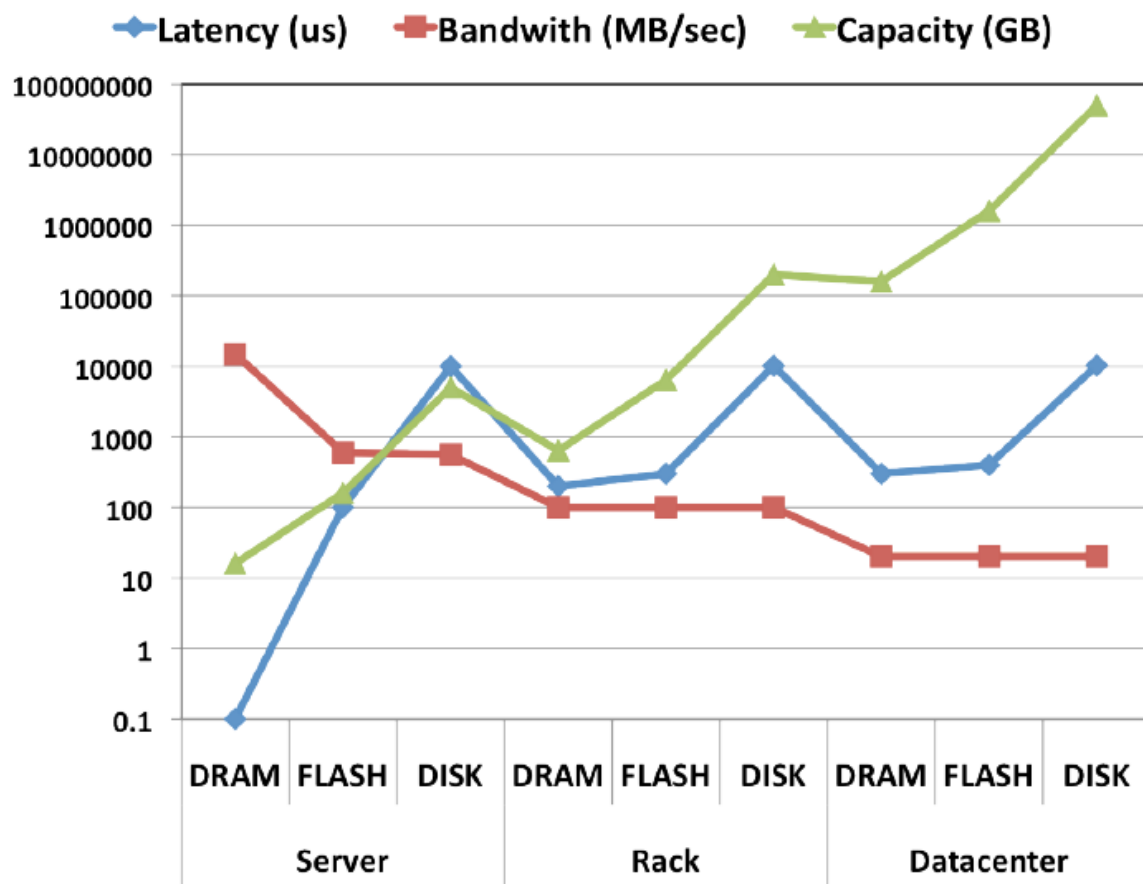
Even in the early work in Banko and Brill (2001), it is already evident that “gains in accuracy do not come for free”. They thought that compression might be needed in order to reduce storage costs of a large corpus.

Unit costs of data processing, storage and transmission are vastly lower today than they were in 2001, thanks to (1) Moore’s Law effects on the unit costs of processing and storage,⁵ (2) progressive improvements in electronic communications, especially in the form of Dense Wave Division Multiplexing, and (3) improved big data technology.

Peter Norvig (2011) of Google has noted that there are trade-offs as a data corpus grows. The amount of data that can be stored in a data centre or in a cloud of servers is vastly greater than the amount of data that can be stored in a single server, or in a single rack of servers. Distributed functionality increases robustness and total capacity, but it comes at a price. In Figure 3, He implicitly claims that while scaling up can achieve a huge gain in storage capacity, it does so at the cost of a huge increase in latency and a huge decrease in the speed of accessing the data.

⁵ This trend may not continue in the future as it has in the past. There are numerous indications that the long-standing predictions that Moore’s Law would sooner or later reach the limit of what could be achieved are now becoming reality.

Figure 3. The impact of storage hierarchy on latency, bandwidth, and capacity



Source: Norvig (2011)

Much less is known about the effective costs of persuading users to make their data available, and its relationship to the marginal benefits of using the data.

Taking Alphabet (Google) as the archetypical example, we observe (1) that Google invests many billions of euros per year in continuing development and maintenance of their search capabilities, the Android ecosystem, and a range of supporting tools such as Google maps and Google translate; (2) that their profits come primarily from advertisers based on the data that they have obtained and their ability to display advertising to consumers, not from sale of the software that they have developed; and (3) that their continued investment suggests that they do not believe that they have reached the limits of their business model; and (4) that the investors who have rewarded them with a market cap well in excess of their annual earnings likewise appear to believe that there is more room to grow the business.

On the basis of those observations, we conjecture that Google finds that more data is indeed advantageous, even with probable diminishing returns to scale. We also conjecture that, if there is an upper limit to the amount of data that is useful, Google does not believe that they have hit it.⁶

⁶ Note that this is not necessarily equivalent to a claim that a firm like Google has an insuperable advantage due to the volume of data that they hold. Thanks to diminishing returns, the curve becomes nearly flat as one

Clearly, these costs are not one-time costs. The preferences and interests of a consumer change rapidly over time, so it must surely be necessary to continuously acquire new data.

Given the apparent importance of data for the success of advertising-based online business models like the one of Google, we examine the role played by data in online advertising more closely in the following section.

3 Data and targeting in online advertising

3.1 Data, targeting and advertising effectiveness in online advertising

The fundamental premise of advertising is to promote ideas, goods or services to a target audience in order to accomplish a communication task (Kotler & Armstrong, 2008). Intuitively, accomplishing the communication task is easier the more positively the target group addressed is predisposed towards the communication task. Many current practices regarding targeting in online advertising reflect this idea, as they seek to match the advertisements served to the audience with their (previous and/or current) behaviour online (Goldfarb, 2014; Lee & John, 2001; Manchanda, Dubé, Goh, & Chintagunta, 2006). While the fundamental functioning across online advertising formats is similar, the required data inputs differ depending on the specific format and the online targeting method employed. We begin by reviewing the literature relating to the role played by data for two major types of online advertising: search and display advertising.⁷

Search (keyword) advertising is the most popular type of online advertising in most countries. Users of an online search engine such as Bing, DuckDuckGo and Google Search enter a search term and receive matching advertising results together with the so-called organic search results. Typically, advertisers pay for each click on their advertisement.

Blake, Nosko, and Tadells (2015) investigate the substitution between paid and unpaid traffic arriving on eBay through search engines. They find that this substitution is almost complete for keywords that include the brand name of the advertiser's site. While the data that consumers enter (i.e. the search term) enables a very precise targeting of consumers with an apparent intention to purchase an item on eBay's platform, the apparent effectiveness of paying for keyword advertisements for these consumers masks the fact that they probably would have arrived on eBay's platform, and would have been just as likely to make a purchase there, without the paid click. For non-brand searches, the authors find no overall significant effect on sales; however, they observe that these search advertisements can increase the number of newly registered users and infrequent users of eBay's platform.

moves to the right. The implication would be that a smaller competitor has some disadvantage, but the disadvantage for a competitor with a smaller but still substantial volume of data might perhaps be manageable. Further research seems to be needed in this area. See also Schaefer et al. 2018.

⁷ According to eMarketer Display and Search advertising formats together represent around 95% of the total digital advertising spending in the US in 2018. The remaining formats Classifieds and directories, Lead generation, Email and mobile messaging account for roughly 5%. (eMarketer 2018)

Johnson, Lewis, and Reiley (2016) find on the other hand that with increasing proximity, the effectiveness of search advertisements increases. They attribute this difference to the fact that the particular advertiser whose advertisements they tested does not advertise continuously on search engines.

Online display advertising consists of various formats such as banner, expandables (rich media ads), interstitial, overlays or video advertisements. Each of them can be integrated into a website in various shapes and sizes. (Junqué de Fortuny et al., 2013) These variations can significantly influence the effectiveness of the individual advertisement. (Chatterjee, 2008; Li & Lo, 2015; Moore, Stammerjohan, & Coulter, 2005; Namin, Hamilton, & Rohm, 2017; Rodgers, Cameron, & Brill, 2005; Tutaj & van Reijmersdal, 2012)

Goldfarb and Tucker (2011) indicate that display advertising's effectiveness can be increased by contextual targeting of the banner as well as by making it more visible (obtrusive). Interestingly, display advertisements that combined both methods performed significantly worse than those using either contextual targeting or high visibility. The authors explain this effect as a reflection of privacy concerns of consumers. This early finding is broadly in line with the effects of online behavioural advertising (OBA) noted by Boerman, Kruikemeier, and Zuiderveen Borgesius (2017) in their recent comprehensive review of the OBA-related literature. Consumers may find personalized online advertising intrusive. This effect can severely limit the effectiveness of this kind of online advertising.

Lambrecht and Tucker (2013) investigate the effect of (dynamic) retargeting of display advertisements. They find that dynamically targeted advertisements (i.e. advertisements that show the specific product a consumer had clicked on earlier) are effective only when the consumer already had a specific idea of the purchase ahead. When this is not the case, showing a generic appeal to revisit the site performs better. Their results are in line with the Elaboration Likelihood Model (ELM) (Petty & Cacioppo, 1996; Petty, Cacioppo, & Schumann, 1983).⁸ Hoban and Bucklin (2015) extend these insights by means of their analysis of cookie-level data to investigate the effectiveness of display advertisements along the purchase funnel. They differentiate four stages: (1) non-visitor, (2) visitor, (3) authenticated user and (4) converted customer. They find that display advertisements play different roles along the purchase funnel. They build awareness for non-visitors or serve to remind authenticated and converted users of the services on offer. For visitors who decline to provide identifying information on the advertiser's website, there is no discernible effect of display advertising.

Bleier and Eisenbeiss (2015) employ two large-scale field experiments and two lab experiments in their study. Their results further underscore the findings of Lambrecht and

⁸ The ELM is a model proposed to explain/predict how persuasive stimuli (e.g. advertising messages or packaging designs) lead to attitude and/or behavioural change (e.g. purchase intention). Specifically, it proposes that the types of stimuli that are most likely to be persuasive depend on the likelihood of 'thought elaboration' on the part of the consumer, that is, whether he / she is motivated and able to devote cognitive effort to process the persuasive stimuli in a careful way or not. If the consumer is motivated, he / she is proposed to follow a central information processing route, in which messages requiring extensive thought will be most likely to persuade. In particular, such messages are considered to be ones with high quality arguments that can convince the recipient in the face of careful consideration. If the consumer is not motivated, he / she is proposed to follow a peripheral information processing route, where messages that do not require extensive effort are more likely to persuade. In particular, these are considered to be ones, which offer simple, easy cues such as the relationship of the stimuli to attractive, credible celebrity endorsers.

Tucker (2013) and Hoban and Bucklin (2015) inasmuch as they show that personalised advertising can enhance effectiveness, but only do so when the targeted consumer has recently visited the advertiser's online presence. Less personalised display advertisements become more effective as time elapses between visiting the online store and seeing the advertisement.

Taken together, the studies reviewed here indicate that data's role in targeting is mostly linked to identifying those consumers who are most likely to click on an advertisement. This is in line with the incentives of the entity serving the advertisement to consumers, be it a search engine or an advertising network, because they commonly get paid for a click but not for merely showing the advertisement. Although the causal link between the last click and the conversion (e.g. a purchase on a website) can be easily measured, there is substantial noise in these measurements, and the ability to actually attribute a specific sale to a specific promotional interaction with the consumer is limited. Independent of the specific accuracy of attribution of advertising effects, the majority of the studies reviewed question whether these conversions of consumer who are predisposed to a purchase or an idea in any case actually create added value for the advertiser, since a significant portion of the paid clicks and sales might have happened anyway without additional promotion. In essence, they call into question whether advertising to consumers who are already positively predisposed actually pays off. The studies further suggest that the incentives for the entity distributing the advertisements may not always be aligned with the best outcome for either advertisers or consumers.

Researchers approaching online advertising from an economics perspective have shed some light on these two questions, as we describe in the following section.

3.2 The impact of targeting on prices for online advertising

From an economics perspective, it is particularly interesting to explore how market actors in online advertising behave in light of the decreasing cost of capturing and storing data and the consequently increasing targetability of consumers. As we explained in the previous section, increased targetability reflects the ability of the entity delivering the advertisement to match promotional goals to those consumers who are positively predisposed. However, as the specificity of this target group increases (i.e. as the individuals in the targeted audience become more and more homogenously predisposed to a specific promotional goal), they likewise become homogenously less positively disposed to competing promotional goals.

In line with this thought, Levin and Milgrom (2010) argue that ever thinner segments of the market attract fewer and fewer potential bidders for the offered advertising space. The price for a click correspondingly decreases as the targeted audience becomes thinner. In their investigation of a large-scale dataset in mobile advertising, Rafieian and Yoganarasimhan (2017) arrive at a similar conclusion. Concretely, they identify a clear incentive for the entity distributing the advertisements to limit the number of features and thus the thinness of the targetable audience in order to yield the greatest profits from advertisers.

More generally, Chen, Narasimhan, and Zhang (2001) model the competition between two firms with identical targetability. They show that the initial impact of improving targetability is to increase the profits of both firms; however, when targetability becomes too high, profits of both firms decrease.

3.3 Limitations to what data and targeting can achieve

Beyond the immediate effects of targeting on advertising effectiveness and competition around advertising space, we would also like to position online behavioural advertising in the broader traditions of marketing and consumer research. This may explain some of the results of the studies reviewed in the previous sections. Several different schools of thought are visible, with significantly different implications as to the effectiveness of online behavioural advertising (OBA).

OBA rests on the idea that observed past behaviour can be informative for current and future behaviour. Attribution is typically measured by clicks or some other predefined outcome (e.g. a purchase or a subscription to a newsletter). Thus, OBA assumes a strict and predictable consumer reaction to specific stimuli that can be inferred from previous behaviour. Essentially, the concept follows a simplistic stimulus-response (S-R) or stimulus-organism-response (S-O-R) paradigm.

Drawing on the basic assumptions of rational action, various models of consumer decision-making processes have been developed. Three commonly cited examples are the models of Engel, Kollat, and Blackwell (1968), Howard and Sheth (1969) and Nicosia (1966). The objective of these models was to depict the complete consumer decision-making process from recognising that there is a need that may be satisfied by for example a purchase, through the purchase itself, up to post-purchase evaluation. The data rich environment of online advertising and online shopping may overcome some of the earlier criticisms of these models – criticisms that were based on their empirically untestable excessive complexity, poor specification and an apparent lack of methods to measure the implied concepts. (Ehrenberg, 1988; Olshavsky & Granbois, 1979) However, fundamental issues with OBA remain unsolved, inasmuch as the S-R / S-O-R paradigm neglects the reflexivity of communication as well as the existence of cultural codes, symbols and autonomous (sometimes irrational) actions by consumers who quite possibly interpret the messages that they receive in ways that are quite different than those that were intended. (Abel, 1948; Hirschman, 1983; Marsden, 1993)

Insights on consumer behaviour accumulated in studies adhering to the competing constructivism / interpretivism (CI) school of thought indicate that first and foremost, “Human experiencing involves continuous active agency.” (Mahoney, 2003, p. 5) This active agency is also reflected in human organisational processes, which are “emotional, tacit and categorical [...] and [...] the essence of meaning making.” (ibid.). When applied to the self, organisational processes shape one’s identity. Social-symbolic processes continuously shape all of these processes. Everyone is entangled in “living webs of relationships, many of which are mediated by language and symbol systems.” (ibid.). Throughout their lives, humans are not only exposed to dynamic dialectical developments, but are actors in those

developments. “Complex flows among essential tensions (contrasts) are reflected in patterns and cycles of experiencing that can lead to episodes of disorder (disorganization) and, under some circumstances, the reorganization (transformation) of core patterns of activity, including meaning making and both self- and social relationships.” (Mahoney, 2003, p. 5) With all of this in mind, one might question whether the data used in OBA can in fact reflect the complexity of consumer behaviour. Interestingly, this subtlety of consumer behaviour is also reflected in the self-image of those advertising agencies who see their task as being closer to art than to a mechanical matching exercise. (Viale, Gendron, & Suddaby, 2017)

4 Conclusions

The internet and the tracking technology associated with it have opened up a rich pool of data on consumer behaviour. As data storage and processing capabilities have increased and their cost has decreased, granular exploration of consumer behaviour has become more and more cost-effective. Advertising appears to benefit disproportionately from this new availability of consumer and consumption related data, as the data and the ability to analyse it enables ever more precise targeting of specific audiences. The resulting alleged increase in advertising effectiveness is often cited as the key success factor for Google, Facebook and other online businesses largely based on advertising revenues.

Indeed, advertisers’ steeply increasing spending on online advertising indicates that they perceive a superior effectiveness of online advertising compared to other advertising formats like TV, radio or newspaper advertising.

The literature reviewed in the present paper suggests, however, that there may be some notable limitations to the unambiguously positive impact both of additional data and of the associated increased targetability.

First, in terms of ability to conduct the predictive analytics that drive data-based market targeting, empirical studies suggest that there are diminishing returns to scale for data in general. Moreover, while enormous data volumes are useful for predictive analytics (and thus for market targeting) on sparse, fine-grained feature data, there are numerous traditional market targeting applications that do not appear to benefit from enormous data volumes.

Second, a number of studies cast doubt on the actual return on investment for advertisers who employ highly targeted advertisements. Advertisements are most likely to lead to completing a sale when consumers are already positively predisposed to the promotional goal. The concern is that narrow targeting primarily reaches consumers who would have been likely to purchase the product or service in any case. There is little hard data on this to date.

Third, there are additional doubts as to the overall effectiveness of market targeting to the extent that simplistic S-R / S-O-R tradition in marketing research neglects for the most part the agency and reflexivity of consumers. There is reason to doubt that more data will ultimately result in (almost) perfect predictions of consumers’ intentions.

Finally, economic models on the impact of increasing targeting suggest that as the consumer segments that can be targeted become thinner and thinner, fewer and fewer businesses are interested in bidding to place an advertisement for these thin segments. This will tend to lower the price that an advertiser is willing to pay. Improving the accuracy of targeting may be useful for more reliably slotting consumers into large segments, but may be of limited commercial value in slotting consumers into increasing narrow slots.

References

- Abel, T. (1948). The Operation Called *Verstehen*. *The American Journal of Sociology*, 54(3), 211-218.
- Banko, M., & Brill, E. (2001). *Scaling to Very Very Large Corpora for Natural Language Disambiguation*. Paper presented at the Proceedings of the 39th Annual Meeting on Association for Computational Linguistics.
- Blake, T., Nosko, C., & Tadells, S. (2015). Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment. *Econometrica*, 83, 155-174.
- Bleier, A., & Eisenbeiss, M. (2015). Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where. *Marketing Science*, 34(5), 669-688.
- Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online Behavioral Advertising: A Literature Review and Research Agenda. *Journal of Advertising*, 46(3), 363-376.
- Chatterjee, P. (2008). Are Unclicked Ads Wasted? Enduring effects of banner and pop-up ad exposures on brand memory and attitudes. *Journal of electronic commerce Research*, 9(1), 51-61.
- Chen, Y., Narasimhan, C., & Zhang, Z. J. (2001). Individual marketing with imperfect targetability. *Marketing Science*, 20(1), 23-41.
- Ehrenberg, A. (1988). *Repeat Buying: Theory and Applications* (2 ed.). London: Griffin.
- Engel, J. F., Kollat, D. T., & Blackwell, R. D. (1968). *Consumer Behavior*. New York: Holt, Rinehart, and Winston.
- 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(3), 389-404.
- Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8-12.
- Hildebrandt, C., & Arnold, R. (2016). Big Data und OTT-Geschäftsmodelle sowie daraus resultierende Wettbewerbsprobleme und Herausforderungen bei Datenschutz und Verbraucherschutz. Bad Honnef: Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK).
- Hirschman, E. C. (1983). Aesthetics, Ideologies and the Limits of the Marketing Concept. *Journal of Marketing*, 47(3), 45-55.
- Hoban, P. R., & Bucklin, R. E. (2015). Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment. *Journal of Marketing Research*, 52(3), 375-393.
- Howard, J. A., & Sheth, J. N. (1969). *The Theory of Buyer Behavior*. New York: Wiley.
- Johnson, G. A., Lewis, R. A., & Reiley, D. H. (2016). Location, Location, Location: Repetition and Proximity Increase Advertising Effectiveness.
- Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). Predictive Modeling with Big Data: Is Bigger Really Better? *Big Data*, 1(4), 215-226.
- Kotler, P., & Armstrong, G. (2008). *Principles of marketing* (12th ed. ed.). Harlow: Pearson/Prentice Hall.
- Lambrecht, A., & Tucker, C. E. (2013). When Does Retargeting Work? Information Specificity in Online Advertising. *Journal of Marketing Research*, 50(5), 561-576.
- Lee, S., & John, D. (2001). Banner advertising: Measuring effectiveness and optimizing placement. *Journal of Interactive Marketing*, 15(2), 60-64.
- Levin, J., & Milgrom, P. (2010). Online advertising: Heterogeneity and conflation in market design. *American Economic Review*, 100(2), 603-607.
- Li, H., & Lo, H.-Y. (2015). Do you recognize its brand? The effectiveness of online in-stream video advertisements. *Journal of Advertising*, 44(3), 208-218.
- Mahoney, M. J. (2003). *Constructive psychotherapy: a practical guide*. New York ; London: The Guilford Press.
- Manchanda, P., Dubé, 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.

- Marsden, R. (1993). The Politics of Organizational Analysis. *Organization Studies*, 14(1), 93-124. doi: 10.1177/017084069301400107
- Moore, R. S., Stammerjohan, C. A., & Coulter, R. A. (2005). Banner Advertiser-Web Site Context Congruity and Color Effects on Attention and Attitudes. *Journal of Advertising*, 34(2), 71-84.
- Namin, A., Hamilton, M. L., & Rohm, A. J. (2017). Impact of message design on banner advertising involvement and effectiveness: An empirical investigation. *Journal of Marketing Communications*, 1-15.
- Nicosia, F. M. (1966). *Consumer Decision Processes; Marketing and Advertising Implications*. Englewood Cliffs, N.J.: Prentice-Hall.
- Norvig, P. (2011). *Internet scale data analysis*. Paper presented at the KDD.
- Olshavsky, R. W., & Granbois, D. H. (1979). Consumer Decision Making—Fact or Fiction? *Journal of Consumer Research*, 6(2), 93-100. doi: 10.1086/208753
- Petty, R. E., & Cacioppo, J. T. (1996). *Attitudes and persuasion : classic and contemporary approaches*. Boulder, Colo.: Westview Press.
- Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement. *The Journal of Consumer Research*, 10(2), 135-146.
- Rafieian, O., & Yoganarasimhan, H. (2017). The value of information in mobile ad targeting: Working paper.
- Rodgers, S., Cameron, G. T., & Brill, A. M. (2005). Ad placement in e-newspapers affects memory, attitude. *Newspaper Research Journal*, 26(1), 16-27.
- Schaefer, M., Sapi, G., & Lorincz, S. (2018), The effect of big data on recommendation quality: The example of internet search, DICE Discussion Paper, No. 284.
- Tutaj, K., & van Reijmersdal, E. A. (2012). Effects of online advertising format and persuasion knowledge on audience reactions. *Journal of Marketing Communications*, 18(1), 5-18. doi: 10.1080/13527266.2011.620765
- Viale, T., Gendron, Y., & Suddaby, R. (2017). From “mad men” to “math men” The rise of expertise in digital measurement and the shaping of online consumer freedom. *Accounting, Auditing & Accountability Journal*, 30(2), 270-305.