GfK
MARKETING
INTELLIGENCE
REVIEW

FROM ACADEMIC RESEARCH TO PRACTICAL USE

Marketing AND DATA SCIENCE

Big Data Challenges

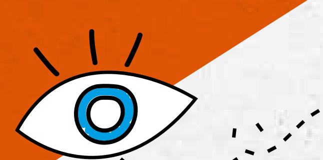
Correlation and Causality

Location data

New skills

Online Advertising

Passive Measurement



Vol. 8 / No. 2 / 2016

Editorial

Data scientists and marketing technologists are among the hottest jobs today and their vocabulary contains such strange words as SQL, MongoDB, Hadoop, API, R, Python, GitHub, A/B Testing, and supervised and unsupervised learning. But why are these jobs and terms so hot?

Well, the amount of available data has literally exploded, capacity to store and transfer data has grown almost infinitely large, and tools to analyze this data keep popping up, some of them even free of charge. These developments create huge opportunities in marketing, and jobs in marketing will attract talented people with a very strong background in data science, statistics, and econometrics. These quants will recognize that the data available in marketing is often much richer and more interesting than data in finance and economics. These people will impact marketing as strongly as they impacted finance during the last two centuries.

We created this special issue to help you embrace this development because superb marketing and industry knowledge is key to realizing the opportunities that more data provides. Data scientists alone might look at the wrong questions as long as marketers do not share their knowledge with them. They might identify what happened but only deep industry and marketing knowledge will enable data scientists to figure out how to add customer value. Marketers will learn that their experience is very valuable but will be even better if it is backed up by data. ''Do you think, or do you know?'' That question will be answered more and more often by analyzing data and running experiments to get the appropriate answers. I hope that this special issue helps you to follow the advice: ''In God we trust, all others bring data.''

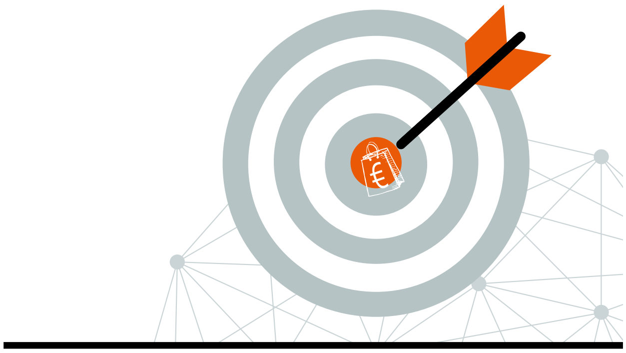
Yours,
Bernd Skiera

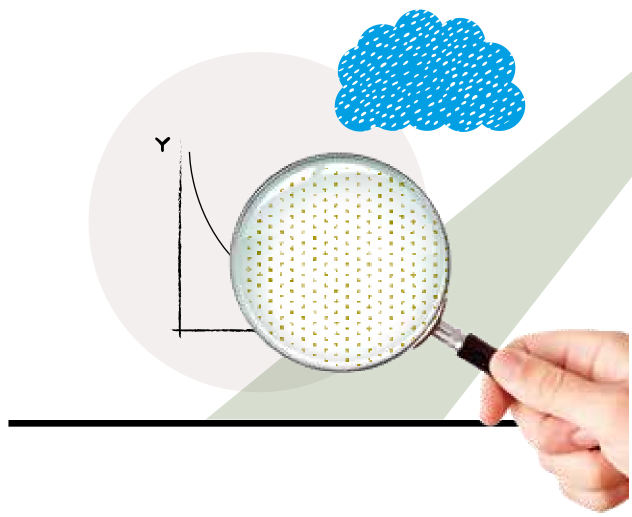
Editor

Bernd Skiera

Frankfurt, August 2016

Contents




Y

Executive Summaries

Data, Data and Even More Data: Harvesting Insights From the Data Jungle

Bernd Skiera

Increasing global digitalization brings huge amounts of data. Finding a successful way to handle all this data and to transform it into real insights will be a critical success factor in the future. The biggest challenge for data science applications in marketing is that many marketing people studied marketing because they no longer wanted to do mathematics. A good marketing campaign will still have to be creative, touch emotions and build a brand, but more and more marketing managers will also need technical and analytical skills. It will more than ever be necessary to determine causal effects to pull the right levers. Consumer insights have always been considered a major driver for growth, but in the digital world, successful growth can also come from improved technical and analytical Implementation and skillful application of new tools and methods.

page 10

Marketing and Data Science: Together the Future Is Ours

Pradeep Chintagunta, Dominique M. Hanssens and John R. Hauser

The synergistic use of computer science and marketing science techniques offers the best avenue for knowledge development and improved applications. A broad area of complementarity between the typical focus in statistics and computer science and that in marketing offers great potential. The former fields tend to focus on pattern recognition, control and prediction. Many marketing analyses embrace these directions, but also contribute by modeling structure and exploring causal relationships. Marketing has successfully combined foci from management science with foci from psychology and economics. These fields complement each other because they enable a broad spectrum of scientific approaches. Combined, they provide both understanding and practical solutions to important and relevant managerial marketing problems, and marketing science is already very successful at obtaining unique insights from big data.

page 18

On Storks and Babies: Correlation, Causality and Field Experiments

Anja Lambrecht and Catherine E. Tucker

The explosion of available data has created much excitement among marketing practitioners about their ability to better understand the impact of marketing investments. Big data allows for detecting patterns and often it seems plausible to interpret them as causal. While it is quite obvious that storks do not bring babies, marketing relationships are usually less clear. Apparent "causalities" often fail to hold up under examination. If marketers want to be sure not to walk into a causality trap, they need to conduct field experiments to detect true causal relationships. In the present digital environment, experiments are easier than ever to execute. However, they need to be prepared and interpreted with great care in order to deliver meaningful and genuinely causal results that help improve marketing decisions.

page 24

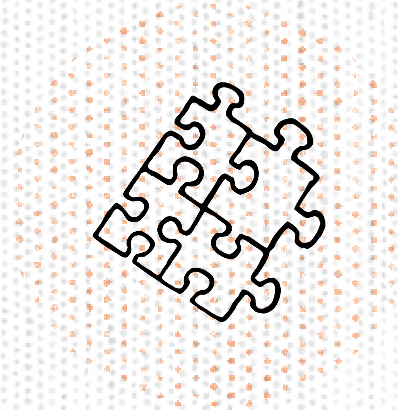
Tell Me Where You Are and I'll Tell You What You Want: Using Location Data to Improve Marketing Decisions

Martin Spann, Dominik Molitor and Stephan Daurer

Location data has become more and more accessible. Smartphone applications such as location-based services collect location data on a large scale. Up to now, most approaches have relied on past data, but new developments in machine learning and artificial intelligence will soon enable more dynamic real-time use of location data. Companies that embrace these technologies will be able to create competitive advantages.

Location data offers great potential to improve a variety of marketing decisions such as targeted pricing and advertising, store locations and in-store layout. Location based advertising is currently the most common application. It allows targeting all customers within a certain distance of a store. Besides advertising, location data can be used for dynamic pricing decisions. Customers close to competitor's locations can be charged a lower price for particular products via discounts in order to reduce switching costs. Indoor tracking can help to optimize store design or the positioning of categories and brands. Granular location data about consumers' movements hence further allows for minimizing potential offline transaction costs based on the distances to stores.

page 30



Using Big Data for Online Advertising Without Wastage: Wishful Dream, Nightmare or Reality?

Mark Grether

Big data contains lots of information about consumers and allows companies real-time and data-assisted decision making to gain significant competitive advantages. Digital advertising is an important application for tailoring services to individual needs. Customized advertising is expected to be more effective, less costly, and better received by society. But what looks deceptively simple when it succeeds is frequently quite difficult to implement in practice. It is difficult to judge and validate the quality of automatically generated data.
And besides quality, there are other aspects that make it tricky to determine the value of data. A reasonable price for data depends on the context of its application and the potential cost savings it generates. And not only the price per impression is unclear. The number of contacts is also less obvious than it seems at first glance.
Primarily third party data providers often incur problems with the monetization of big data and many are struggling to survive. They depend on the fairness of the data buyer and a successful business model has yet to be developed.

page 38

The Art of Creating Attractive Consumer Experiences at the Right Time: Skills Marketers Will Need to Survive and Thrive

Katherine N. Lemon

New technologies have made today's marketing faster, more mobile, more location-based, more digital, more virtual, and more automatized than ever. In this new world, marketers need to be "real-time relevant" – to gain awareness, to change perceptions and to spur action. They need to have their content in the right channel, format, time and context – from a consumer's perspective. Only then do they at least have a chance of the consumer attending to the information and being influenced by it. In such an environment new skills and competences are required.

The amount of available data has virtually exploded. To gain any perspective or apparent "control" in these environments, successful managers must embrace the complexity and learn to analyze, integrate and interpret all this data. A critical skill for marketers will be to identify the metrics that best reflect the desired outcomes of the organization and that sufficiently reflect specific indicators of critical processes. Furthermore, insights from other disciplines such as architecture, design, information-processing, biology or engineering will be important for creating customer experiences. The marketer of the future will need to be supremely curious and creative and to balance and integrate different worlds. It will all come down to delivering memorable and lasting experiences in a constantly and fast changing environment.

page 44

Data Analysis Trumps Specialist Advice: How Direct Banks Function

Interview

While many traditional retail banks are struggling with a business downturn, the direct bank market is enjoying steady and respectable growth despite a challenging environment. Dr. Schmidberger, Fully Authorized Representative at ING-DiBa Germany, offers us a glimpse behind the curtains of this direct bank. We learn how data technology is used so that bank customers are (more) satisfied without bank advisors.

page 50

Big Data in Market Research: Why More Data Does Not Automatically Mean Better Information

Volker Bosch

Big data will change market research at its core in the long term because consumption of products and media can be better logged electronically, making it measurable on a large scale. Unfortunately, big data datasets are rarely representative, even if they are huge. Smart algorithms are needed to achieve high precision and prediction quality for digital and non-representative approaches. Also, big data can only be processed with complex and therefore error-prone software, which leads to measurement errors that need to be corrected.

Another challenge is posed by missing, but critical variables. The amount of data can indeed be overwhelming, but it often lacks important information. The missing observations can only be filled in by using statistical data imputation. This requires an additional data source with additional variables, for example a panel. Linear imputation is a statistical procedure that is anything but trivial. It is an instrument to "transport information," and the higher the observed data correlates with the data to be imputed, the better it works. It makes structures visible even if the depth of the data is limited.

page 56

Data, Data and Even More Data: Harvesting Insights From the Data Jungle

Bernd Skiera

KEYWORDS

Data Science, Big Data, Analytics, Growth Hacking, Online Advertising

THE AUTHOR

Bernd Skiera,
Chair of Electronic Commerce,
Faculty of Business and Economics,
Goethe-University,
Frankfurt am Main, Germany
skiera@wiwi.uni-frankfurt.de.

From data wasteland to data jungle /// Increasing global digitalization brings huge and ever-growing amounts of data. It all began with the invention of the browser that made access to the Internet via a desktop computer so much easier and faster. More and more consumers started to enjoy direct online interaction with each other and with companies… and started to leave their traces. The cost of observing these interactions fell to marginal costs that were very close to zero. It was, for example, possible for the first time to observe on a large scale not only that an advertisement was shown to a consumer, but also how the user reacted to that ad. So marketers were able to measure whether the consumer clicked and even purchased after clicking on the ad. Previously, a comparable measurement of advertising success was only possible for direct marketing activities but the cost of doing so was much higher and the quality of measurement much lower. For example, direct marketers could not even observe whether the consumer opened the letter they sent. Just compare this opportunity to the ones that email marketers have today.

The next major step forward came with the availability of affordable and powerful smartphones and mobile data plans. They now enable companies to target consumers everywhere, add location-based information to consumers' actions, and record consumers' reactions at the location and the time where the reaction occurs. Thus, instead of interacting with

FIGURE 1:
A flashlight on the data science jungle

consumers during the few hours per day that they use their desktop, companies can nowadays interact with consumers essentially 24/7. The availability of data exploded and Hal Varian, chief economist at Google and previously a well-known researcher in microeconomics, became famous for saying around 2005 that "the sexiest job in the next 10 years will be statisticians." So, instead of a data wasteland we seem to be living in a data jungle full of ripe fruit. Can marketers simply pick it up now? Is all of it wholesome? Or is harvesting insights from a data jungle a more challenging task than anticipated and one that requires new skills?

What companies can gain from big data analysis

> ***Insights from academic research*** /// Many companies are convinced that the fruit of the data jungle is wholesome. Insights that arise from big data analyses are in high demand. In contrast to ten years ago, the number of company jobs for PhDs in marketing and economics is growing and growing. Well executed academic studies attract huge interest among companies. Managers are, for instance, willing to use the insight of a study by the researchers Blake, Nosko and Tardelis in 2015 on the unprofitability of Google AdWords for branded keywords to reallocate millions of dollars of advertising budget to other advertising media. The Wharton Customer Analytics Initiative (WCAI) successfully draws the interest of Fortune 500 companies to sponsor competitions that attract academics to analyze the data of those companies and share the respective insights. The list of marketing problems that are analyzed on the analytics platform Kaggle is constantly increasing, and online classes on "machine learning" are among the most popular online courses.

> ***Improving marketing decisions*** /// Marketing can be much more effective if more and better information is available. With this special issue, we intend to help com-

{Box 1}

STRIKING A BLOW FOR MATH IN MARKETING

Loosely speaking, the biggest challenge for data science applications in marketing is that many people in marketing decided to work in marketing or study marketing because they no longer wanted to do mathematics. A good marketing campaign is doubtless creative, touches emotions, builds a brand and encourages consumers to talk about the product. Math alone does not allow for doing so but my prediction is that managers without good technical skills will not even get a chance to be responsible for marketing campaigns. To illustrate my point, let me share a very recent experience with you:

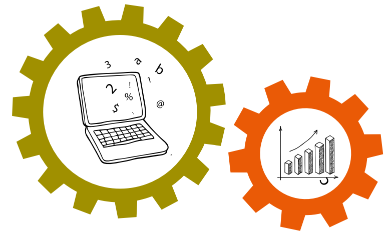
SENSITIZING STUDENTS FOR A DATA-LADEN MARKETING WORLD

A few weeks ago, I had the chance to run an undergraduate seminar together with a superb Business Development Manager at Amadeus, Sandro Cuzzolin, who is deeply involved in its online travel company Travelaudiance. He put together four data sets for this five day seminar. They ranged in size from 770,000 to more than 9,000,000 observations; sizes too large for popular spreadsheet programs. Our marketing students had to analyze the data sets to answer simple questions such as identifying the regions where the company makes largest profit and should spend more on advertising. Not an easy task, indeed:

» The number one problem that most students had was to use a database engine such as SQLite to condense the data in such a way that they could further analyze it in a spreadsheet environment to calculate a return on investment.
» Students stopped complaining about the size of the data sets after Sandro Cuzzolin reminded them that these data sets did not even cover a week.
» He further asked the students to imagine how long such an analysis would probably take in reality if they always had to ask the IT department to properly select and condense the data.
» Needless to say, the students' analysis did not come close to separate causality from correlation.

You might think that my student requirements and my predictions are exaggerated? Maybe, but keep in mind that the largest provider of advertising inventory in most Western countries, Google, started as a high tech company. Its competitive advantage is also due to using superb technology and analytics to provide free-of-charge insights into the success of clients' advertising campaigns. In addition, other companies such as Facebook or Instagram do not have their roots in marketing.

»
The two fields, marketing and information systems, are moving closer together.
«



panies participate in the gains in efficiencies. I am glad that some of the most prestigious researchers in marketing took their time to contribute and highlight how marketing decisions can be improved in our data-intensive environment. Pradeep Chintagunta, Mike Hanssens and John Hauser illustrate how data can be transformed into useful information for various marketing applications (pp. 18) Recently, they served as co-editors of a Special Issue on big data of the flagship journal Marketing Science, and they share some highlights with us. In their article they also predict a much stronger collaboration of marketers with data scientists and computer scientists and stress that marketers can not only benefit but also contribute by modeling structure and exploring causal relationships. Martin Spann and his coauthors show how to get location data, how to analyze it and how to use it to make customized marketing decisions (pp. 30). Including the situational variables that smartphone applications generate helps companies to design their offers so that consumers perceive them to be much more relevant. Martin Spann has a strong background in marketing and now conducts much of his research into information systems. Other researchers came the opposite way. These developments in research interests are fine examples of how the two fields, marketing and information systems, are moving closer together.
Martin Schmidberger looks at the data jungle from a managerial perspective. For many years he has been the head of data and marketing at ING DiBa. Remarkably, this is the only bank whose market share in the German retail banking market has grown substantially over the last ten years. In my interview with him (pp. 50), he confirms that systematic use of data and machine learning techniques leads to a better understanding of customer behavior. Their systems generate customized recommendations with a better response than traditional marketing achieved and thus highlight again the fact that data science and marketing are a winning team.

Marketing challenges for harvesting insights from the data jungle /// However, fruit-picking in a data jungle is not automatically rewarding. It can be very tricky, sometimes even painful to harvest true insights. Mathematical skills are a precondition (see Box 1), but marketers need to react to diverse challenges and our authors deal with other critical issues as well.

> ***Drawing correct conclusions*** /// Two of the strongest women in our field, Anja Lambrecht and Catherine Tucker, highlight the important difference between correlation and causality (pp. 24). On the one hand, this topic is an old one and every student hopefully learns that a regression analysis indicates correlation but not necessarily causation. On the other hand, the ability in a digital world to target individual consumers makes this problem so much more important.
In Box 2 you find an example that illustrates how tempting it is to draw wrong conclusions. Hopefully you will see that only experiments like the ones Anja Lambrecht and Catherine Tucker describe are able to observe the causal effect of marketing actions and that only those causal effects should guide marketing allocation decisions.

> ***Data quality and data pricing*** /// Marc Grether, COO of Xaxis, focuses on the potential of big data use in the online advertising industry (pp. 38). The challenge here is to reduce wastage and target exactly those consumers that form a specific target group. While AdWords builds on search phrases and consumers' self-selection, successful display advertising relies on data about consumers (see Box 3). It is still quite difficult to provide such data in good quality and to find a reasonable price for the data, as Marc Grether outlines nicely. The data industry has yet to develop business models to successfully handle these challenges.
Data quality, or – to be more precise – missing data is also the biggest challenge for using big data in market research. Volker Bosch of GfK notes that it is fairly challenging to technically and methodologically handle "big data" (pp. 56). But the future of big data in marketing research is bright. One major reason is that new technologies allow for passively measuring consumers' attitudes and preferences. He also highlights the opportunities that data imputation provides for solving the problem of limited depth of data and calls for closer collaboration between data science and marketing science.

{*Box 2*}

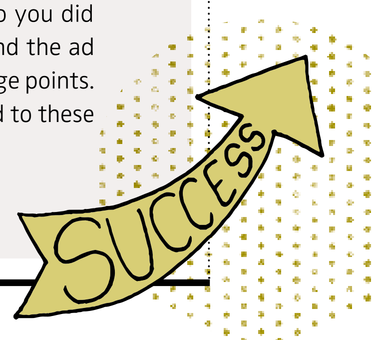
THE REAL RETURN ON MARKETING

•

In Google AdWords or other forms of targeted advertising, you typically present advertising messages to consumers that have already expressed interest in your offering. In Google AdWords, for example, consumers search for a product that you offer. Let us assume that this group of consumers has a purchase probability of 90%. Does 90% represent the impact of your advertising? Most likely not. The reason is that you don't know whether the consumers' purchases were triggered by your ad or if they would have occurred anyway.

Comparing the purchase probability of 90% to the response rate of a group of consumers who have not seen the advertisement won't help neither. Even if such a group displays a response rate of only 5%, you can't attribute the difference of 85% to advertising alone. The reason is that you targeted consumers who are simply much more likely to purchase. Thus these consumers would have more likely purchased even if you did not send them any advertising message. The challenge is that you need to disentangle the observed difference (85%) into the causal effect of advertising and the systematic difference between the two groups of consumers: the ones who already expressed interest in your offering and received an ad and the others who did not express interest before and therefore did not receive an ad. The latter, however, can only be observed if you run an experiment in which you stop advertising for randomly selected consumers from the group that expressed an interest in your offer.

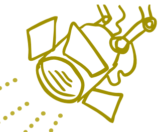
Let us assume that the experiment shows that you still observe an 80% purchase probability for consumers who expressed interest in your offering (e.g., searched on Google) but who you did not target with an ad. Then you can conclude that 80% would have bought anyway and the ad increases the purchase probability to 90%. Thus the causal effect of the ad is 10 percentage points. Consequently, in return of investment calculations, you should compare the cost of the ad to these 10 percentage points.



> ***New skills*** /// One of the key challenges will also be to hire specialists and to form teams that find their way around in the data jungle. Kay Lemon elaborates on the seven skills necessary to survive and thrive in the new market conditions (pp. 44). She currently serves as director of the Marketing Science Institute (MSI) and is therefore one of the key players at the interface between academic marketing theory and business practice. Marketing teams need diverse talents that will work closely together to deliver great experiences at the moments that matter. As a result, marketing stars of tomorrow will be the ones who understand the customer's decision journey and the critical moments in that journey – in real time, and in context.

The future of the data jungle: perish – or thrive?

> ***Even more to gain*** /// Will we observe even more data in the future? It is very hard to believe that this will not be the case. Devices such as watches, glasses, cameras, technologies like face recognition, thermal imaging, WiFi tracking or beacon communities like WhatsApp, WeChat, and Snapchat will generate even more data. Almost every area will be impacted but, given that consumers produce a huge amount of that data, marketing is the area in business that will experience the strongest impact from the availability of data. So the jungle will be even denser with even more to gain for those who have the necessary skills to harvest data and transform it into real insights. We now know so



{Box 3}

A SPOTLIGHT ON ONLINE ADVERTISING

The online advertising industry was essentially turned upside down when Google began in 2002 to use auctions to sell its AdWords. Remarkably, Yahoo! owned a patent for doing so and granted Google a license to use this patent. What happened is that search engine marketing became the most prominent online marketing instrument in most countries, with Google essentially having a monopoly in many Western countries.

This change came along with a disruption of how online advertising was sold. Historically, most online advertisements were sold by sales representatives who usually negotiated a long-term contract with a uniform price for all ads. In Google AdWords, Bing and Yahoo! and the prominent search engines in China and Russia, Baidu and Yandex, each ad is sold in a real-time auction. We have an individual price for each ad that is sold on the spot. I am not aware of another industry whose pricing mechanism changed so quickly.

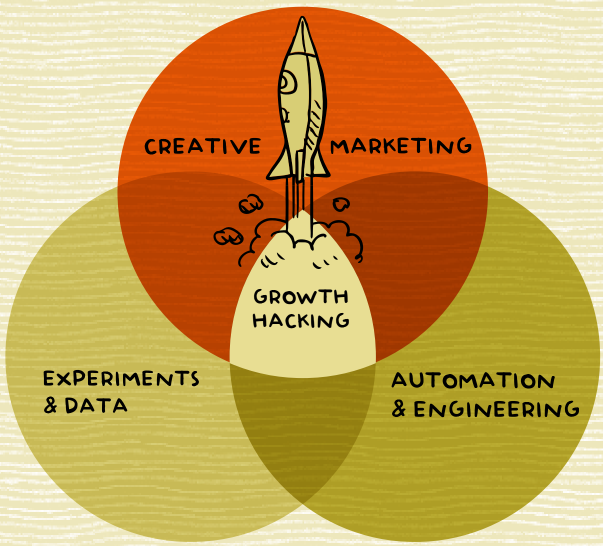
Currently we are observing a similar change for display advertisements, which are more and more often sold via real-time auctions, also known as real-time bidding. The remarkable difference between display advertisement and search engine marketing is the information that is available about a consumer. In search engine marketing, the search phrase essentially contains all the information. In display advertisement, however, data is collected about consumers' interests and preferences, frequently via third parties and data aggregators. In that sense, the kind of data to select customers is much closer to the kind of data that is used in TV advertising, with the important difference that we now have data about each individual consumer.

much better what consumers do, where they are, what they think, or how they react to the companies' messages. No one predicted 25 years ago how much information we would have available today and what opportunities this data provides.

> ***Growth opportunities from technical expertise*** /// Consumer insights have always been considered a major driver for growth, but in the digital world successful growth can come from new angles. "Growth hacking" is a term not too popular in marketing but is likely to be at the core of how marketing will develop in the future. Andrew Chen, an influential Silicon Valley-based manager and blogger, describes a growth hacker in a 2012 blog post as a hybrid of marketer and coder who looks at the traditional question of how to acquire customers and answer them with A/B tests, landing pages, viral factors, email deliverability, and Open Graph. He uses Airbnb and its remarkable integration into the classified advertisement website Craigslist as an example to demonstrate how technical expertise allows to better spread an offering. He concludes by stating: "Let's be honest, a traditional marketer would not even be close to imagining the integration – there's too many technical details needed for it to happen". Even if there is an abundance of ripe fruit, the harvesting and processing equipment needs to be sophisticated and tuned for purpose to insure a successful harvest.

> ***Prepare the field for talent to grow*** /// So technical and methodological skills will be the key to success in a data-laden marketing environment. I am convinced that tomorrow's marketing curricula will include much more data science related topics. Students will learn how to collect data via crawlers or APIs and how to handle such data with sophisticated database software. They will be able to handle unstructured data with text-mining approaches and to analyze data with econometric and machine learning techniques. Students will know how to separate correlation and causation via properly conducting experiments and will understand techniques to visualize relations between many objects and how to automatically react to customers in real time.

FIGURE 2:
Growth Hacking: Skillful technical implementation of marketing ideas to generate growth



With such experts, companies will be able to distinguish between wholesome and perished data fruit and harvest real insights that improve decisions and enable growth. But despite all my enthusiasm, let me conclude by stressing that one thing will never change. Companies, and in particular marketing, need to provide value for the customers. So data science in marketing is just a means to an end. Probably, however, a very powerful one.

/.

FURTHER READING

Blake, Thomas; Chris Nosko; Steven Tadelis (2015):
"Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment," Econometrica, Vol. 83 (1), pp. 155 – 174.

Chen, Andrew (2012):
"Growth Hacker is the new VP Marketing," http://andrewchen.co/how-to-be-a-growth-hacker-an-airbnbcraigslist-case-study/.


One future
DATA
SCIENCE

Marketing and Data Science: Together the Future is Ours

Pradeep Chintagunta, Dominique M. Hanssens and John R. Hauser

KEYWORDS

Data Science, Marketing Science, Computer Science, Big Data, Quantitative Analysis, Modeling, Machine Learning

THE AUTHORS

Pradeep Chintagunta,
Joseph T. and Bernice S. Lewis
Distinguished Service Professor of Marketing,
Booth School of Business,
University of Chicago, Chicago, USA
pradeep.chintagunta@chicagobooth.edu

Dominique M. Hanssens,
Distinguished Research Professor of Marketing,
UCLA Anderson School of Management,
University of California Los Angeles, USA.
dominique.hanssens@anderson.ucla.edu

John R. Hauser,
Kirin Professor of Marketing,
MIT Sloan School of Management,
Massachusetts Institute of Technology,
Cambridge, MA, USA
hauser@mit.edu.

From rules of thumb to formal knowledge /// Nearly half a century ago, young marketing faculty in business schools started to embrace sophisticated research methods, including mathematical optimization, multivariate statistics, and econometrics, to study marketing problems. At the time, the marketing discipline was considered relevant for job prospects, but teaching and practice were focused on institutional knowledge and rules of thumb. These faculty started a revolution that continues to this day to provide formal knowledge, structures for teaching and practice, and the excitement to draw the highest-caliber students to its curricula. Early efforts resulted in pioneering contributions, notably in market segmentation, media mix optimization, data-based planning, and consumer preference modeling that set in motion a new discipline, now called marketing science. New professional organizations and scholarly journals were created, specialized conferences drew progressively larger audiences, and marketing as an area of study became increasingly quantitative in nature.

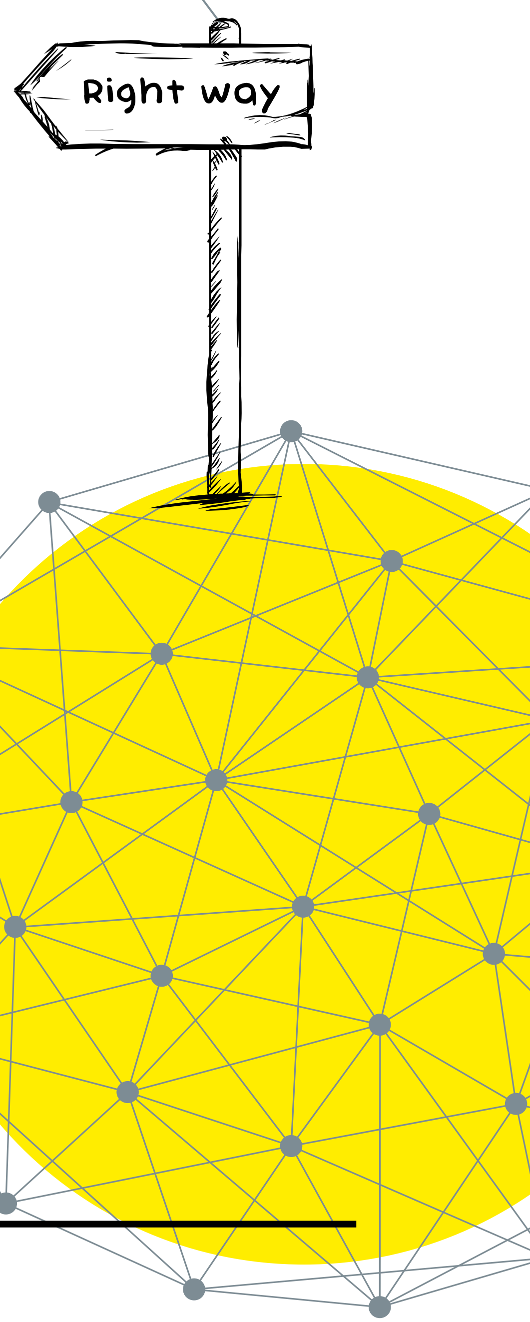
New insights from the constant flow of new data, new methods and new ideas /// The revolution was driven by new data, new methods, and new ideas. While early models relied on warehouse-withdrawal data, the 1980s brought scanner data. While simple regression was once king, advanced econometrics, discrete choice models, Bayesian methods, and improved optimization enabled researchers to tackle bigger and more relevant issues. During this development, quantitative researchers drew increasingly from insights developed in economics, engineering, psychology and sociology, but contributed back with a deep understanding of consumers and markets. Increasingly, data was collected and aggregated online, but often in focused studies. And then the Big Data revolution happened. "Big" Data is

FIGURE 1:
The evolution of data quality: From warehouse-withdrawal data via scanner data to big data

often distinguished from regular "Data" by the three Vs, volume, velocity and variety. Indeed, with remarkable speed, companies and specialized data providers are able to assemble unprecedentedly large digital databases (volume) in real time or near-real-time (velocity) and with a large variety of data characteristics, including numerical, text, sound and video files (variety). Computer science contributed to the academic discipline by providing new methods to structure and store large data, new approaches to process large data, and new techniques for using large data. Analyses of consumers and markets that were once unthinkable because of their complexity, scale, and dynamics were now possible. Much of this development was ad hoc and focused on computation rather than building on decades of consumer and market knowledge.

Embracing the new opportunities /// At the same time, these Big Data technological developments raised new opportunities for marketing intelligence development, and this is where marketing science comes in. Indeed, marketing science is ready for the Big Data revolution. Peer-reviewed solutions for marketing's major challenges are already in place and can only get better with the advent of more and more varied data. In particular, marketing science had already made substantial advances in such fundamental areas as consumer choice modeling, customer lifetime value modeling, new-

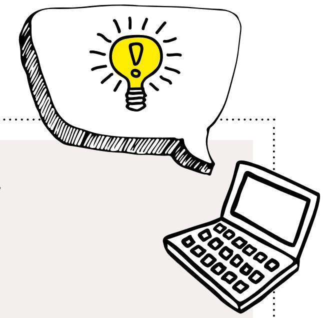
»
The synergistic use of computer science and marketing science techniques offers the best avenue for knowledge development and improved applications.
«



product demand prediction, marketing impact assessment, customized communication and promotion, and brand valuation. We refer the interested reader to an excellent overview of these contributions in The History of Marketing Science, edited by Russell Winer and Scott Neslin.

Going forward, the synergistic use of computer science and marketing science techniques offers the best avenue for knowledge development and improved applications. We have put this perspective to a test by serving as co-editors of a special issue on Big Data of the flagship journal Marketing Science. We received over 50 submissions addressing new issues with new methods. We winnowed this set to approxi-

{ *Box 1* }



GETTING A FLAVOR OF QUESTIONS THE LATEST MARKETING-DATA-SCIENCE CAN ANSWER*

•

Profiling the most promising customers

The internet age offers unparalleled opportunity for brands to target their advertising to consumers who are most likely to respond. But what is the best way to do this targeting, or "profiling" of potential customers? Those who are able to read the tracks can resort to web surfing behavior. Web surfing can provide reliable clues of individual consumers' propensity to purchase. In their article "Crumbs of the Cookie: User Profiling in Customer-Base Analysis and Behavioral Targeting," Michael Trusov, Liye Ma and Zainab Jamal develop and implement a targeting algorithm based on consumers' online surfing data. Their approach is superior to existing methods, both in identifying the best consumers to target with digital advertising, and in avoiding wasted exposures to uninterested consumers.

Identifying relevant choice alternatives from a consumer´s perspective

Some high-technology product categories, for example television sets and digital cameras, offer a bewildering number of choice alternatives for consumers. What's more, these offerings are subject to continuous technological innovation. How do manufacturers know which competitive products are perceived as similar – and therefore competitive - to theirs and how they should identify and target lucrative submarkets for their new offerings? In their article "Visualizing Asymmetric Competition among more than 1,000 Products Using Big Search Data," Daniel Ringel and Bernd Skiera develop innovative mapping methods on search data at price comparison websites to obtain effective visualization of these complex market structures. Their approach offers a fast, easy to understand, yet comprehensive view of how new technological offerings compete with each other, as perceived by the buying public.

Filling individual shopping baskets through relevant product recommendations

In recommendation systems, in automated marketing and in customized targeting, practitioners would like to be able to use a consumer's purchase history to predict the next product the consumer will buy. In their paper "Product Recommendations Based on Latent Purchase Motivations," Bruno Jacobs, Bas Donkers and Dennis Fok apply a method that is often used in text processing to identify, from the consumer's perspective, sets of products that tend to be purchased together. The authors' analysis with latent Dirichlet allocation (LDA) performs better than typical collaborative filters and other model benchmarks. In doing so, it holds promise for a variety of new recommendation systems to build upon the improved predictive ability.

Knowing how consumers truly perceive your brand

Consumer perceptions of a brand are important for the management of the brand. Consumers readily express their opinions about brand attributes such as eco-friendliness, nutrition, and luxury via social media. The article "How #Green is Your Brand? Mining Cause-Related Brand Associations on Twitter," by Arun Culotta and Jennifer Cutler provides a fully-automated method to monitor brand related messages in social media (Twitter). They track these perceptions by mining a brand's social connections and demonstrate the method by monitoring 200 brands for these attributes. Their approach allows managers to react quickly and effectively to both opportunities and challenges in consumer perceptions of their brands.

*The details on methods and procedures are published in the original articles.
They can all be found in Marketing Science, Vol. 35, 3 (May – June 2016).

mately a dozen articles by rigorous peer review. In the box on page 22 we provide high-level summaries of selected papers. These papers provide a flavor of the unique insights that are obtained from the combination of big data and marketing science.

Benefits from complementing disciplines /// One broad area of complementarity between the typical focus in statistics and computer science and the typical focus in marketing is the following. The former fields tend to focus on pattern recognition, control and prediction. Many marketing analyses embrace these directions, but also contribute by modeling structure and exploring causal relationships. Marketing has successfully combined foci from management science with foci from psychology and economics. These fields complement each other because they enable a broad spectrum of scientific approaches. Combined they provide both understanding and practical solutions to important and relevant managerial marketing problems.

We endorse, with enthusiasm, the premise that marketing should embrace data science and machine learning. We also endorse the complementarity. Data scientists and computer scientists will improve their focus and research by taking advantage of the wealth of insights provided by marketing science. Marketing managers will not only benefit from new data science flavors but will be able to choose from ever increasing sophisticated research menus. A well-balanced selection from several disciplines will be able to answer glaring questions that could not be answered before and will improve marketing decision making substantially. Together the future is ours.

/.

» Analyses of consumers and markets that were once unthinkable because of their complexity, scale, and dynamics became possible. «

FURTHER READING

Winer, Russell S.; Neslin, Scott A., (2014):
"The History of Marketing Science,"
World Scientific Publishing Co. Pte. Ltd.

Marketing Science, Special Issue on "Big Data: Integrating Marketing, Statistics, and Computer Science," Vol. 35 (3), 2016.

On Storks and Babies: Correlation, Causality and Field Experiments

Anja Lambrecht and Catherine E. Tucker

KEYWORDS

Correlation, Causality, Field Experiments, Field Tests, Causal Inference

•

THE AUTHORS

Anja Lambrecht,
Associate Professor of Marketing,
London Business School, United Kingdom
alambrecht@london.edu

Catherine Tucker,
Professor of Management Science,
MIT Sloan School of Management, Cambridge, USA
cetucker@mit.edu

Correlation is not causality /// The explosion of available data has created much excitement among marketing practitioners about their ability to better understand the impact of marketing investments. Big data allows for detecting patterns and often it seems plausible to interpret them as being causal. While it is quite obvious that storks do not bring babies, marketing relationships are usually less clear. If marketers want to be sure they are not walking into a causality trap, they need to conduct field experiments to detect true causal relationships. In the present digital environment, experiments are easier than ever to undertake, but they need to be prepared and interpreted with great care in order to deliver meaningful and genuinely causal results that help improve marketing decisions.

Apparent causalities often fail to hold up under examination /// The online marketing world is full of examples of organizations or journalists being tempted to make causal inferences from purely correlational data. For example, Twitter on its website reports the information displayed below in Figure 1. In the original headline it stated that engagement with promoted tweets translates to higher brand favorability and purchase intent and suggests that 'this study result highlights the value of an engagement on Twitter.'

In reality, it is difficult to interpret this data as causal. It more likely illustrates that a consumer who views a brand more favorably is also more likely to engage with a promoted tweet by this brand. Similarly, a consumer who intends to purchase a certain brand is more likely to engage with a message promoting this brand. Indeed, the causality could also be reversed. Note that this does not mean the ad is ineffective, but since the data presented is purely correlational it is impossible to judge whether the ad was effective or not.

FIGURE 1:

Correlation interpreted causally:
Does engagement with promoted tweets really translate to higher brand favorability and purchase intent?

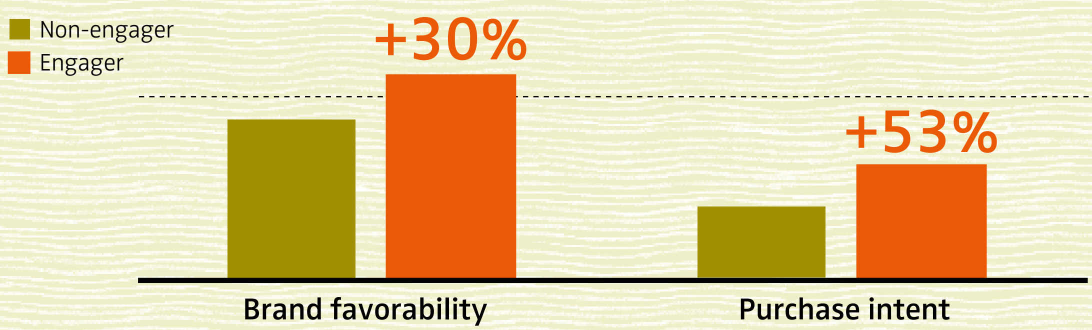
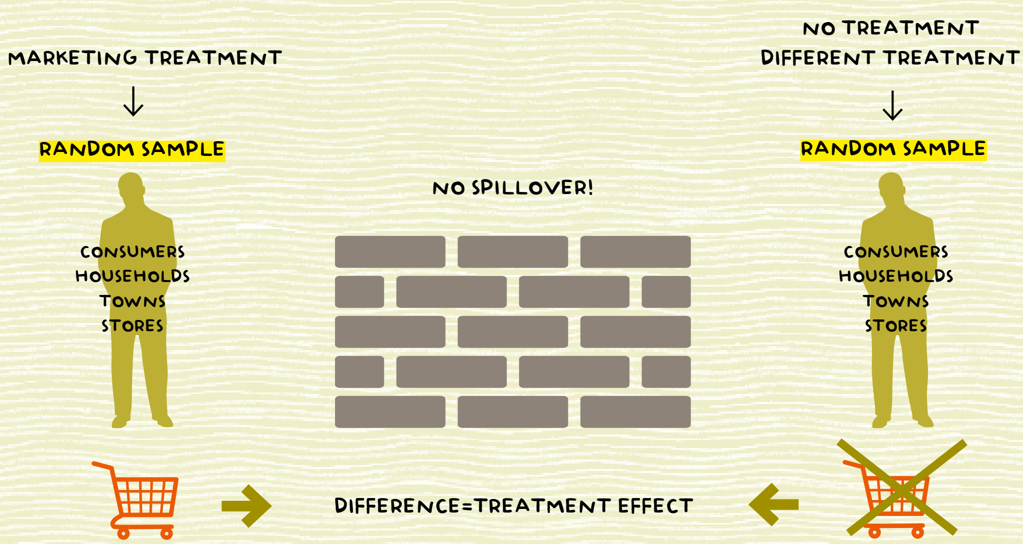


Figure taken from the Twitter website:
https://blog.twitter.com/2013/nielsen-brand-effect-for-twitter-how-promoted-tweets-impact-brand-metrics

FIGURE 2:

How field experiments permit causal inferences



{ Box 1 }

ONLINE ADVERTISING IS SUCCESSFUL – OR ISN'T IT?

Users who viewed more ads bought more often

Imagine a toy retailer that has implemented a particular form of online advertising: retargeting. Its systems identify users who have looked at its website but did not purchase. As these users continue to browse the web, the toy retailer targets them through ads for the online store. The toy retailer collects detailed user-level data on website visits, ad views, subsequent purchases and non-purchases. The marketing team then evaluates this data. In its analysis it finds that users who viewed more ads were more likely to eventually make a purchase.

Does this mean that the ads were effective in converting users into buyers?

No. The data merely illustrates that users who browsed the web more and, as a consequence, were exposed to more online ads were also more likely to purchase. That's just a correlation. To clarify why such data cannot be interpreted as causal, imagine two users, Emma and Anna. Both Emma and Anna visited the toy retailer's website. In the following weeks, Emma is very busy at work and unable to further attend to Christmas shopping and also unable to browse the internet more broadly. Anna, however, is already on holiday and spends a great deal of time exploring many different gift options online. This means that Emma does not purchase, but merely because she is busy at work and for the same reason she does not look at any online ads. In contrast, Anna has plenty of free time, which leads her to spend a great deal of time on the internet. As a result she is exposed to ads and ultimately buys the product. From the data at hand it is impossible to tell whether Anna's exposure to ads in any way influenced her decision to buy.

But what could the toy retailer do to determine the effectiveness of the ads?

The solution would be a field test as described in Figure 2: It randomly assigns every user who has visited the website to a test group and a control group. The users in the test group will be shown the toy retailer's ads while the users in the control group will be exposed to a replacement ad, such as an ad for a charity. Since, on average, the users in the test group and the users in the control group are the same, any difference in purchase behavior can be attributed to advertising exposure. Reverting to our example, the two groups composed randomly would each include the same number of Annas and Emmas, eliminating the effect of their different behavior.

Field experiments permit causal inferences /// In the social sciences, the gold standard for making causal inference is a field experiment, sometimes referred to as an A/B test. In a field experiment, individual consumers or users are, unbeknown to them, assigned to different groups. One group is then exposed to a marketing treatment, say online advertising, whereas the other group is not exposed to it (see Figure 2).

As long as the company randomly assigns a sufficiently large number of users to each experimental condition, the difference in outcome variable between the two groups of users can be attributed to the marketing treatment. Any researcher interested in field experiment techniques should be aware of the potential need for a large sample when conducting a field experiment, especially when the tested effect is hard to predict or assumed to be small. In general, though, it is difficult to give practical advice on sample size beyond aiming for as large a sample and data collection effort as possible. Box 2 highlights the critical decisions necessary to plan and interpret field experiments.

Further applications of field tests to improve marketing decisions /// When the 5 steps described in Box 2 are executed carefully, applications are numerous and we describe some more below.

> ***Comparing the effectivness of generic and personalized ad content*** /// In this study we compared personalized and generic ads for a travel site. Both groups were shown an ad but in one instance users were exposed to a generic brand ad for the site whereas in the other instance the ad

{ *Box 2* }

IMPLEMENTING FIELD EXPERIMENTS SUCCESSFULLY

Step 1: Decide on the unit of randomization
Randomization could happen, for example, at the level of the individual, household, town, website, store, or company. While finely-grained units of observation, like single individuals, tend to provide higher statistical power, their setup is often more expensive and difficult to implement. Also, the risk of potential for spillovers and crossovers is higher.

Step 2: Minimize spillovers and crossovers between experimental treatments
Suppose a company randomly selects an individual to receive a free mobile phone. Potentially his or her adoption of a mobile phone could affect the adoption outcomes of relatives and friends even if the relatives and friends were supposedly not treated. If such spillovers are a large concern, one way of addressing them would be to randomize at the level of plausibly isolated social networks such as a community, rather than randomizing at the level of the individual.
A crossover occurs when an individual who was supposed to be assigned to one treatment is accidentally exposed to another. Suppose, for example, a canned soup company is testing different advertising messages in different cable markets, and individuals are exposed to a different advertising message from that of their home market because they are traveling. This could potentially lead to mismeasurement of the treatment, especially if there were systematic patterns in travel that led to such crossovers not simply being random noise.

Step 3: Decide on complete or stratified randomization
The experimenter then needs to decide whether to conduct stratified or complete randomization. In complete randomization, individuals (or the relevant unit of randomization) are simply allocated at random into a treatment. In stratified randomization, individuals are first divided into more homogenous subsamples. Then each individual in each of these subsets is randomized to a treatment. This stratified technique is useful if some variables are strongly correlated with an outcome. For example, household income may be strongly correlated with purchase behavior toward private label brands. Therefore, it may make sense, if the researcher has access to household-level data, to stratify the sample prior to randomization to ensure sufficient randomization occurs within, for example, the high-income category.

Step 4: Ensure that appropriate data is collected
Researchers also need to carefully consider what type of data they need for their later analysis and to ensure that the practical set-up allows them to collect this data. This is especially important in digital environments where different parties have access to different types of data and it is not always obvious how these can be collected and linked. For example, advertising networks have access to ad exposure data but may require additional steps to ensure that they likewise capture purchase data and can link those to ad exposures.

Step 5: Interpret results from a field experiment carefully
In theory, interpretation of field experimental data should be straightforward, but in practice there are numerous issues to consider when interpreting the statistical results. The key issue is to understand exactly the difference between the groups and to be careful about how to generalize this difference. Also, the duration of the field experiment is critical and will affect the interpretation of results. For example, the researcher needs to have access to a long enough period to understand whether any treatment they measure is stable, dissipates or increases in its effect over time. However, for many field experiments it is hard to measure long-term effects because experiments are limited in time. Therefore, in most settings researchers should carefully consider whether the causal effect they establish truly reflects the long-term treatment effect.

reflected the specific hotels the user had previously looked at on the company's website. We compared the performance of the different ads and found that on average the generic brand ad was more likely to convert a user to purchase. Only when a consumer's browsing history indicated that they had reached a stage where they were actively comparing attributes of different hotels, did the personalized ads became equally effective.

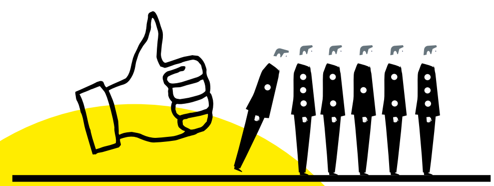
> ***Testing website design*** /// Companies may also wish to compare which of two different designs of their home page is more effective in getting a user to browse products in detail. In this case a company may randomly direct a user to either of the home page versions. The company could then compare the number of users who went on browsing specific products, and later purchased, across the two experimental conditions. Provided that users were randomly assigned to the experimental conditions, the difference in the likelihood to browse or to purchase can be attributed to the difference in the design of the home page.

> ***Optimizing pricing policy*** /// In this article we have mostly focused on marketing communications, but other types of marketing decisions can likewise benefit from insights that come from field experiments. Imagine a company that wishes to estimate how shipping fees affect purchases from their online store. Marketing could set up two different checkout pages where in the first instance the checkout page charges the usual shipping fee and in the second instance the shipping fee is discounted or entirely removed. They could then compare the number of consumers who do not complete their purchase upon reaching the checkout page across conditions and adjust their pricing accordingly.

For companies that want to make sure that they do not invest in storks to get more babies, field experiments represent a very useful avenue in which to obtain truly causal data. When planned and interpreted with care, the results can help to guide a wide range of marketing decisions.

/.

»
In the present digital environment, experiments are easier than ever to undertake.
«



FURTHER READING

Lambrecht, A.; Tucker C. (2013):
"When Does Retargeting Work? Information Specificity in Online Advertising," Journal of Marketing Research, Vol. 50 (5), pp. 561 – 576.

Lambrecht, A.; Tucker C. (2013):
"When Personalized Ads Really Work.," https://hbr.org/2013/06/marketers-serve-no-ad-before-i

Lambrecht, A.; Tucker C. (2015):
"Field Experiments in Marketing," working paper. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2630209

Lewis, R. A.; Rao, J. M. (2015):
"The Unfavorable Economics of Measuring the Returns to Advertising," Quarterly Journal of Economics, Vol. 130 (4), pp. 1941 – 1973.


BERLIN
10:45 a.m.

Tell Me Where You Are and I'll Tell You What You Want: Using Location Data to Improve Marketing Decisions

Martin Spann, Dominik Molitor and Stephan Daurer

KEYWORDS

Location Data, Location Intelligence, Decision Support Systems, Mobile Targeting, Mobile Analytics

THE AUTHORS

Martin Spann,
Professor of Electronic Commerce and Digital Markets,
Ludwig-Maximilians-University of Munich, Germany,
spann@spann.de

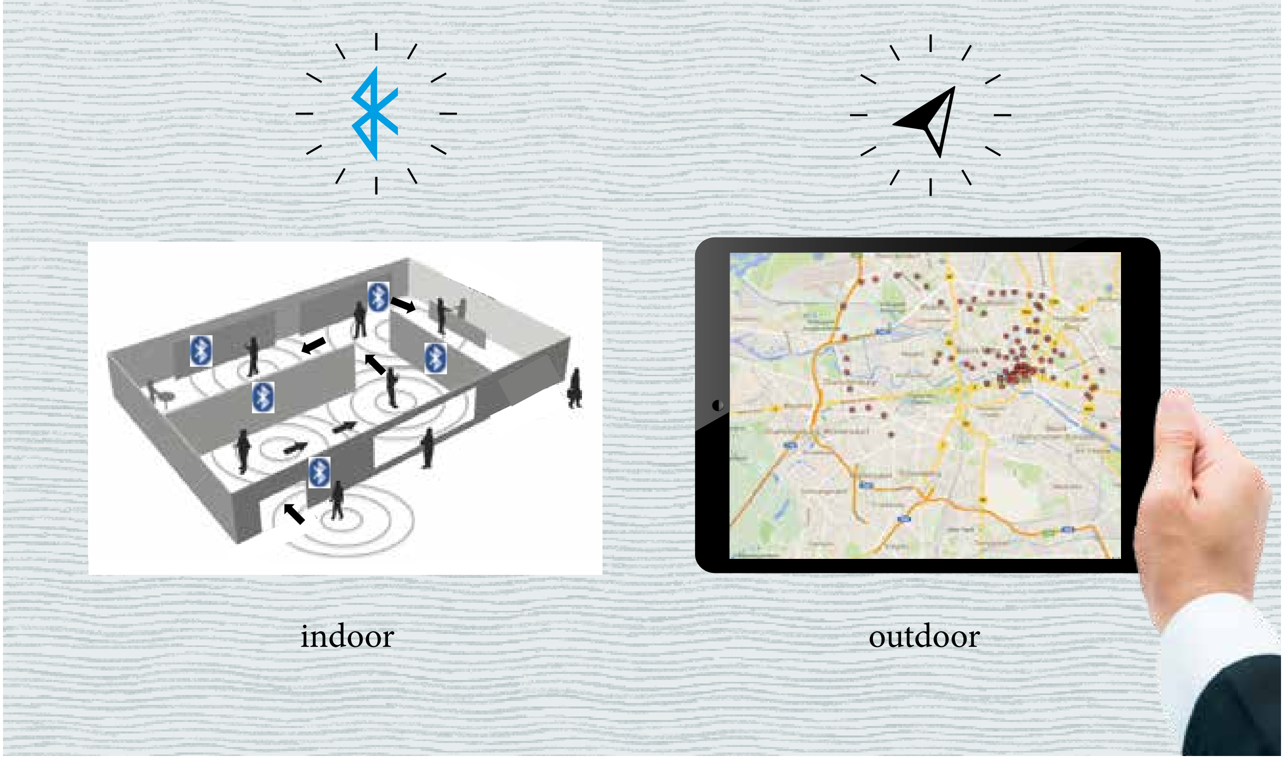
Dominik Molitor,
Assistant Professor, Gabelli School of Business,
Fordham University, New York, USA,
dominik.molitor@gmail.com

Stephan Daurer,
Professor of Business Information Systems
at the Baden-Wuerttemberg Cooperative
State University, Ravensburg, Germany,
daurer@daurer.net

Context matters /// Consumers choose what they like, but their preferences depend on their specific situation. Situational variables like time, day of the week, season, local weather or social situation and, of course, specific locations form the context for all decisions. A person might feel and act differently on a workday than on the weekend, in summer or winter, on a sunny or rainy day, whether alone or with family and friends, out in the countryside or in a shopping mall. The digitalization of business processes has been tremendously helpful in identifying those preferences and leaving a broad and rich data trail. This development has even intensified by the increased popularity of smartphones. Smartphone applications generate situational variables, including path data such as movement sequences through supermarket aisles or the movement patterns within different parts of a city. Figure 1 visualizes movement patterns based on the physical location of smartphones or other devices.

Location data – the new cookie? /// Location data become more and more accessible. Location-based advertising is one application that takes advantage of those context factors. Mobile ads might be tailored and targeted for consumers who happen to be in a certain area or even in or within a small radius of a seller's store at a certain time. The total value of real-time location-based advertising is supposed to grow to about $15 billion in 2018 or approximately 40 percent of total mobile advertising, according to a research report from the Swedish Berg Insight market research company. Not surprisingly, location is sometimes referred to as the new cookie.

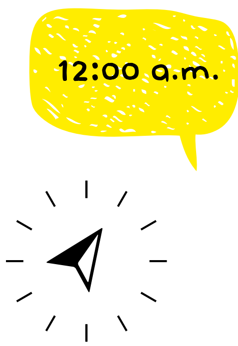
FIGURE 1:
Exemplary movement patterns



While desktop cookies allow identifying a browser's activities over time, a consumer's physical location is an indicator of his or her preferences in the "real" world. Therefore location is more relevant for offline shopping patterns. It is hence tremendously important for companies to understand how this type of data can be leveraged to improve marketing decisions regarding promotions, pricing, the assortment of products and the choice of store locations.

How to obtain location data /// Smartphone applications such as location-based services already collect location data on a large scale. To do so, smartphones use a combination of sensors and determine the current location of the device. The most prevalent method is based on the Global Positioning System (GPS). As GPS has some limitations in areas with tall buildings and as it does not work well indoors, other methods are available for these settings. A database of known locations of cell towers and WiFi networks makes it possible to determine the smartphones' locations using lateration and triangulation. For indoor settings there is Bluetooth Low Energy (BLE). This system is based on short-range radio signals. Active components or senders are placed in specific places. When a smartphone comes into reach of one or more senders the position can be determined. An example for this technology is Apple's proprietary iBeacon protocol.

How to analyze location data /// Analyses of location data can be conducted either in retrospect or in real time. Dependent on the type of observation, we further distinguish between cross-sectional designs with one data point per consumer or longitudinal designs with several measurements over time. Table 1 gives an overview of the different types of analyses and what they show.
Retrospective analyses of cross-sectional or pooled location data can be used to generate snapshots of location-specific preferences based on clustering analyses. In addition, retrospective analyses of longitudinal location data can be conducted by applying vector autoregressive models. Such applications have already been used in the context of online browsing and path analyses between different websites.

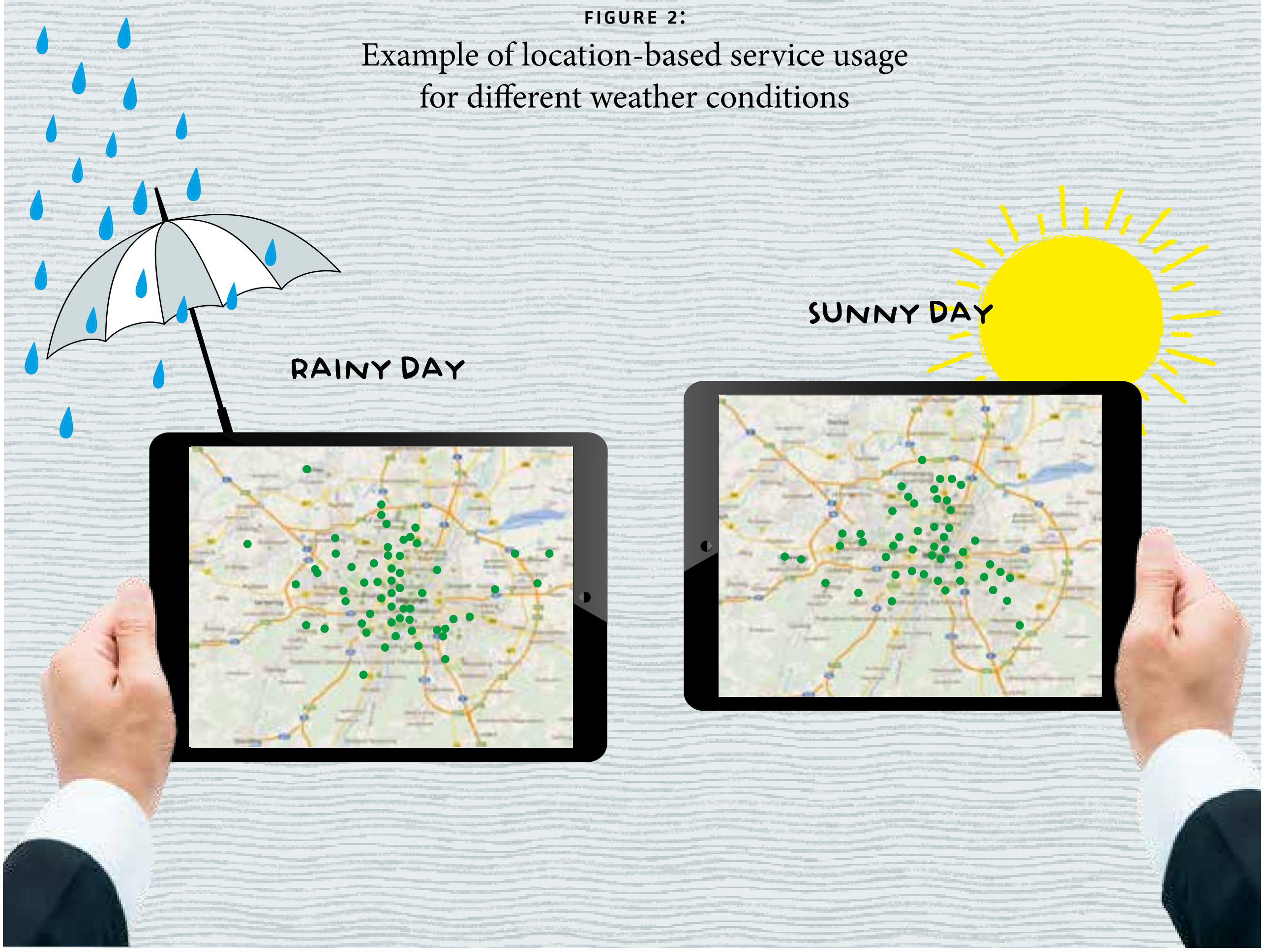


» Mobile ads might be tailored for consumers who happen to be within a small radius of a seller's store at a certain time. «

TABLE 1:
Summary of analyses and observation types and their application

Type of analysis		Observation: Cross-sectional	Observation: Longitudinal
Type of analysis	Retrospect	Objective: Spatial pattern identification Analysis type: Cluster analysis Application: Customer segmentation	Objective: Spatial time series analysis Analysis type: Spatial/vector autoregressive models Application: Trajectory-based segmentation of customers
Type of analysis	Real-time	Objective: Similarity/personalization Analysis type: Collaborative filtering (via machine learning) Application: Distance-based real time discounts (location-based coupons)	Objective: Location prediction/trajectory Analysis type: Clustering and collaborative filtering (via machine learning) Application: Trajectory-based real time discounts (location-based coupons)

FIGURE 2:
Example of location-based service usage for different weather conditions



They allow marketers to segment customers based on their geo-location or their movement paths or trajectories.
On the contrary, real-time analyses are automated approaches in the backend of smartphone applications. An example of real-time analyses of cross-sectional location data is model-based collaborative filtering algorithms. These applications suggest coupons to consumers based on the activities of other customers in the same location, and well-established machine learning techniques already exist in the context of recommendation systems. In addition, machine learning techniques can also be applied to predict preferences based on the location and trajectory of consumers in real time as well as to dynamically adapt the discount depth on personalized coupons. For example, deeper discounts might be necessary for more distant customers than for customers that are closer and already moving towards the store.

Location data might be able to provide even more interesting insights when combined with other data sources such as demographic or transaction information, the weather, social network/co-location, or survey data. For example, the combination of location, co-location and transaction data can be used predict coupon choice and personalize offers based on the weighted information of previous purchases and similar individuals. In addition, survey data like psychographics can also enhance location information by delivering insights into the intentions and motivations of certain location-based activities, which can again be used to improve the prediction of locations or trajectories. Figure 2 shows how a combination of weather and location information can deliver insights into the use of a location-based service: higher intensity in malls and shopping areas on a rainy compared to a sunny day.

TABLE 2:
Overview of marketing decisions that might be improved using location data

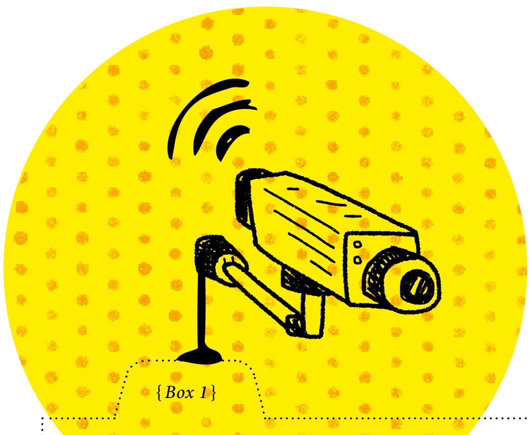
MARKETING DECISION	DATA SOURCE	ANALYSES (EXAMPLES)	PURPOSE
ADVERTISING	PURCHASE HISTORY OWN APP MOBILE AD PROVIDERS	LOCATION PROXIMITY MOVEMENT/TRAJECTORY TEMPORAL PROXIMITY	LOCATION TARGETING REAL-TIME TARGETING
PRICING	COMPETITORS' STORE LOCATIONS CONSUMERS' TRAVEL HISTORY	LOCATION PROXIMITY MOVEMENT PATHS	DYNAMIC PRICING PRICE DISCRIMINATION
STORE ASSORTMENT/ LAYOUT	CONSUMERS' IN-STORE MOVEMENTS (RFID, BLE BEACONS) OWN APP	IN-STORE PATH ANALYSES SHOPPING-BASKET ANALYSES	POSITIONING OF PRODUCT CATEGORIES AND BRANDS IN-STORE ADVERTISING
STORE LOCATION	CONSUMERS' OUTDOOR MOVEMENTS OWN APP MOBILE AD PROVIDERS	LOCATION PROXIMITY MOVEMENT PATHS	STORE LOCATION CHOICE COMPETITIVE ADVERTISING

Improving Marketing Decisions with Location Data

> ***Location-based advertising*** /// Location-based advertising is a natural choice for marketers. It can be utilized either by developing/using an own application or by joining existing mobile ad providers. Location- specific ads change online and offline advertising by providing more effective methods, such as location targeting. Location-based push advertising seems to be particularly suitable for matching retailers and consumers in real time. It is similar to display advertising and allows targeting consumers dependent on their behavior and/or the situational context. Both, spatial and temporal proximity can significantly increase the effectiveness of SMS-based targeting strategies.

> ***Location-based pricing*** /// Besides advertising, location data can also be used for dynamic pricing decisions. For example, companies may conduct location-based price discrimination between own and competitor's customers. More precisely, customers close to competitor's locations can be charged a lower price for particular products via discounts in order to reduce switching costs. The combination of the consumers' whereabouts based on location data and competitors' store locations makes these pricing decisions possible.

> ***Optimization of store layout*** /// Location data from indoor tracking technologies such as Bluetooth Low Energy traces or WiFi networks are able to provide valu-

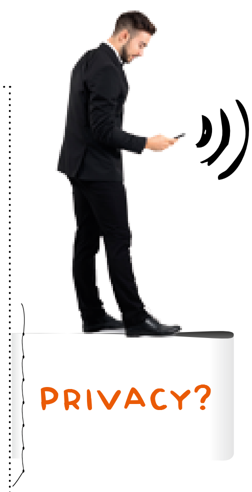


{*Box 1*}

… AND WHAT ABOUT PRIVACY?

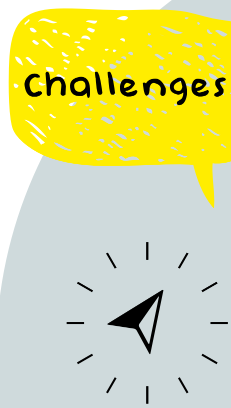
While all forms of targeting raise privacy concerns, tracking tools that record an individual's whereabouts are particularly sensitive. Consumers, especially in Europe and in the United States, are increasingly uncomfortable about the privacy implications of location data.

One critical aspect of the use of location data is its potential to reveal consumer identities – based on path data – even if the raw data is anonymous. This trade-off between more detailed user data leading to highly targeted ads and privacy concerns is currently approached differently. Google, for example, strongly pushes into data-driven personalization algorithms, such as mapping and directions, via Google Now on Android. Apple, on the contrary, is approaching similar personalization features via Siri, with the difference that most automated tasks will be conducted directly on the phone, without being uploaded into the cloud. Previous research has also suggested new privacy-friendly targeting mechanisms without compromising the benefits of location data.



able insights about consumers' in-store movements. Previous research has already analyzed consumers' movements using path data from radio-frequency based RFID shopping carts and identified different clusters of in-store travel activities. The same can be done with mobile devices. Smartphones will also be able to cover the paths of consumers that haven't picked a shopping cart, provided that they have downloaded a retailer's or an ad provider's mobile app. Information on customers in-store movements and product choices allows retailers to optimize their store layout as well as the positioning of product categories and brands within the store.

> ***Choice of store locations*** /// Furthermore, information about consumers' outdoor movements, for example between home and work locations, gives valuable insights into their location-specific preferences. This type of data might help retailers to decide where to open new stores. Granular location data about consumers' movements allows for minimizing potential offline transaction costs based on distances to stores, which are known to be important drivers of store choice. This information may, for example, enable a retailer to pick a side street location for a store, thereby saving rent, with location-based advertising



inducing pedestrians from a nearby main street to visit the store. Table 2 summarizes some major applications of mobile data analysis.

Outlook – the future of location data /// Location data offers great potential to improve a variety of marketing decisions such as targeted pricing and advertising, store locations and in-store layout. Companies have been experimenting with some of these new opportunities, but up to now most approaches are still static, relying on past data. New developments in machine learning and artificial intelligence will, however, soon enable more dynamic real-time use of location data and thus create competitive advantages for companies that embrace these technologies. And while we are just starting to understand how to leverage the power of location data, technological progress already generates new streams of (big) data. Sensor data related to the Internet of Things is one example. It is spanning even further, including data on personal health, smart homes, cars or industrial machines. These novel data sources will also create tremendous new challenges and opportunities for consumers, companies and researchers.

/.

FURTHER READING

Bell, David R. (2014):
"Location Is (Still) Everything: The Surprising Influence of the Real World on How We Search, Shop, and Sell in the Virtual One," New Harvest, Boston.

Daurer, Stephan; Molitor, Dominik; Spann, Martin; Manchanda, Puneet (2015):
"Consumer Search Behavior on the Mobile Internet: An Empirical Analysis (July 2015)," Ross School of Business, Paper No. 1275. http://ssrn.com/abstract=2603242

Luo, Xueming; Andrews, Michelle; Fang, Zheng; Phang, Chee Wei (2014):
"Mobile Targeting," Management Science, Vol. 60 (7), pp. 1738 – 1756.

Molitor, Dominik; Reichhart, Philipp; Spann, Martin; Ghose, Anindya (2016):
"Measuring the Effectiveness of Location-based Advertising: A Randomized Field Experiment." http://ssrn.com/abstract=2645281


?
SALE

Using Big Data for Online Advertising Without Wastage: Wishful Dream, Nightmare or Reality?

Mark Grether

KEYWORDS

Online Advertising, Big Data, Third-party Data Provider, Personalization, Wastage, Retargeting

THE AUTHOR

Mark Grether,
Global Chief Operating Officer,
Xaxis, New York City , US,
mark.grether@gmail.com

Big Data and Online Advertising: High Expectations /// Everyone right now is talking about big data, which is viewed as the next great innovation challenge for marketers. The digitalization of the entire advertising industry is generating ever increasing amounts of data that must be collected, analyzed and interpreted. Using real-time and comprehensive data-assisted decision making, companies are hoping for significant competitive advantages by improving processes and creating more options for tailoring and personalizing services. Digital advertising is an important application for this personalization. Customized advertising will be more effective, cost less, and be better received by society. Companies like Google and Facebook are playing a vanguard role here, while their stock values demonstrate the economic potential that can be realized through big data.

A New Data Market Emerges /// But advertisers are not the only ones who have high expectations for the possibilities of big data. Progress in database and analytics systems has opened the doors to this new business opportunity to more and more smaller companies. Figure 1 shows some of the players who are active in this new big data market. Particularly in online advertising, many companies are trying to develop their own business ideas and claim a piece of the growing online advertising pie by using big data tools. In digital jargon, these companies are called third-party data providers. They are transferring the profitable data business, which large market research institutes like GfK, TNS, Nielsen and Comscore have established in the non-digital environ-

FIGURE 1:
The Big Data Vendor Landscape

Source: http://www.mammothdb.com/the-big-data-landscape-by-mammothdb/

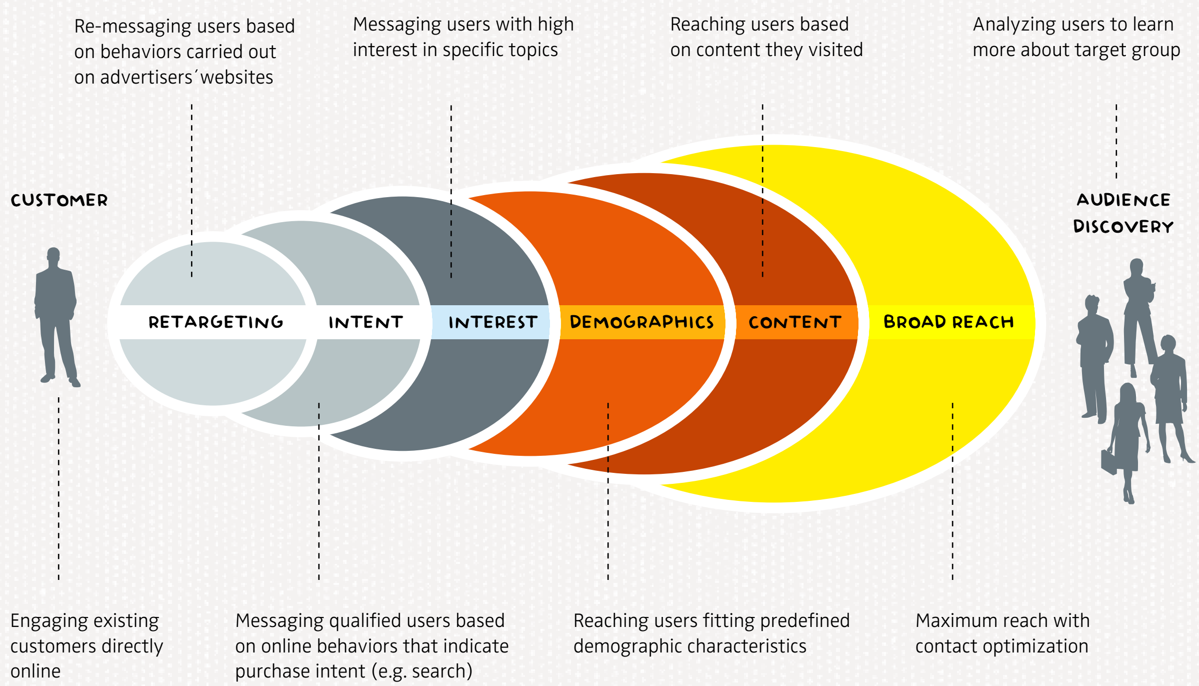
ment, to the digital world – user data that has been acquired through the analysis of user behavior on the internet, is sold and licensed. But the undeniable opportunities are offset by substantial risks. The profitable use of big data is not without pitfalls, and some of the business models on the fringes of big data and online advertising simply do not work. These fledgling companies rarely succeed in achieving a competitive advantage in the market. Many are fighting for economic survival.

Using Big Data to Optimize Advertising /// The big promise of big data to advertising is improved accuracy of communication. Advertising is expected to become more relevant and less expensive as a result of less wastage. Different data is necessary depending on individual advertising goals. Basically, advertising activities can either be performance-related or support the brand image. The greater the focus on immediate sales success, the more data is needed to promote individual customer contact and re-targeting. But if increasing brand recognition is what matters, the focus will be more on general interest data and nonspecific messages (see Figure 2).

> ***Big Data and Performance Marketing*** /// With performance marketing, advertising is billed solely on the basis of the performance of an agreed action by a targeted online user. In simplest terms, this action is a click on the advertisement. Now, with big data, it is possible to investigate which variables influence this click. Predictive re-targeting methods are frequently used to this end. These are based on data on users who were already very close to taking the desired action. Data mining tools are used to search for other users who have matching behavioral profiles. Searching for these statistical twins involves enormous amounts of data with diverse quality in different sets, often exhibiting large gaps. Furthermore, the data must be evaluated within milliseconds. But if a sufficient number of such twins can be found, reach and click rates can be increased significantly. For example, if a campaign with a cost-per-thousand price of €1 achieves a factor of five, the value of the underlying algorithm is €5 (per thousand). The value of the advertising space therefore increases fivefold.

FIGURE 2:

Types of available data and their application in online advertising



> ***Big Data and Branding*** /// Branding campaigns frequently aim to improve brand image or recognition. This is traditionally a domain of TV advertising. Therefore, the online advertising world has adopted indicators like net reach or gross rating points from TV advertising. The success of a branding campaign is judged by maximum contact with a given target group. In many cases, sociodemographics like age and gender determine the relevant segments. Data mining is used to make a valid prediction of these characteristics for as many online users as possible. Usually, the greater the reach, the less precise the forecasting of characteristics, and this is a trade-off that must be considered. Provided that the data is valid, an advertiser can significantly reduce its media costs this way. Advertising is delivered only to its target group, driving down wastage significantly. A good example of this type of data usage is Facebook. With the login data of its users, Facebook has access to well-validated age and gender information, and it achieves enormous reach via various devices. The underlying data is ideal for the precise delivery of advertising to the target group. Particularly with video advertising, which is primarily used to increase brand recognition, age and gender are well-suited criteria for targeting.

The Pitfalls of Monetizing Big Data in Advertising /// What looks deceptively simple through its success is frequently quite difficult to implement in practice. This holds particularly true for the aforementioned third-party data providers, who unlike Facebook or Google do not use their own data, but live on the sale of such data. Aside from data quality, the biggest problem lies in determining a reasonable price for the data. The box on the side describes why it is difficult to determine the value and the quality of data.

{Box 1}

THE PROBLEM WITH VALUING DATA FOR ONLINE ADVERTISING

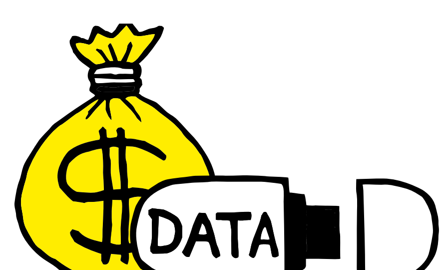
The Value of Data Varies with Different Applications
Specific information about online users, for instance whether they are male or female, interested in finance or sports, or live in New York or Los Angeles, is valuable if it leads to lower wastage in an online advertising campaign. The costs of the advertising space and the targeting effectiveness determine its precise value.
If, for example, the goal is to reach only men but the gender of the recipient is unknown, it becomes statistically necessary to display two ads. But if gender is known in advance, one ad has the same effect. At costs of €1 (per thousand impressions) for a classic banner ad, savings of €1 (per thousand impressions) can be realized. In this case, the value of the information about gender amounts to €1 multiplied by the planned reach. If the information is used for video advertising costing €20 (per thousand contacts), the same information suddenly has a value of €20, twenty times the amount.

Data Quality is Hard to Quantify
The validity of the automatically generated user data is another fundamental problem. Verifying whether the cookies of online users describe them accurately is a service provided by third-party companies like Nielsen or Comscore. However, these companies use proprietary metrics, meaning that their measurements are not always consistent with one another. Tests in the USA and UK have shown that different validating companies assign different genders to one and the same online user. Some may categorize a user as male, while another categorizes the same person as female. As a result, even the data provider cannot be sure of the actual quality of the data in the lead up to an advertising campaign. The same applies to the data user. As long as there is no validation standard, the user cannot know which provider supplies good data, which means that they are taking a risk.

The following challenges are common in setting a price for data.

> ***A Suitable Price for Data is Hard to Determine*** /// As a rule, the data provider will not know for what type of advertising its customers will use the data and is therefore unable to set an optimum price. As described in the box, the value of the data depends, for example, on whether they are used for display or video advertising. Accordingly, is it difficult to decide whether the data provider should demand a cost-per-thousand price of €1 or €20 for the "gender" characteristic? A possible solution would be a pricing model in which the provider participates in the cost savings of the data user. The price of the data could be set as x% of the costs for the saved advertising space. But because the data provider is not familiar with the costs of the advertising space and they are also not

»
Aside from data quality, the biggest problem lies in determining a reasonable price for the data.
«



known in advance for real-time bidding approaches, this pricing model has not been established in practice. Instead, billing is normally done according to a cost-per-thousand price. As a result, data used for display advertising is usually too expensive, while data for video advertising tends to be underpriced.

> ***The Customer Determines the Number of Contacts*** /// Not only the price per impression is unclear. The number of contacts is also less obvious than it seems at first glance. In online advertising campaigns, it has become established practice for both the media seller (the publisher) and the media buyer (media agency, advertiser) to count how frequently advertising is displayed. Billing is done according to the cost-per-thousand price based on counts performed by both parties. Thus, if both parties count 1,000 delivered advertisements, the advertiser will pay the previously agreed price. Only in case of major discrepancies between the two measurements they would technically investigate the matter. To date, however, there is no standardized measurement of data usage for online advertisement. Measurements are usually taken by the advertiser, so that the data provider must rely on the accuracy of the information. Therefore the data provider is dependent on the data user's honesty.

> ***Price Markdowns for Questionable Data Quality*** /// The problems with data validation described in the box can lead to diminished trust in the quality of the data to be purchased. It is thus not uncommon for buyers to demand risk-related markdowns due to the lack of assurance regarding the validity of the data. Whether such markdowns are justified is difficult to tell.

> ***Cost Structure Drives Price Pressure*** /// When digital information has been collected once, e.g. that an online user is male, it can be sold as often as necessary. The selling costs are marginal and therefore the contribution margin is already positive, even at low prices. In a competitive environment, this situation leads to a downward price spiral. Data buyers can frequently negotiate significantly lower prices, because they know that the provider will still generate a positive contribution margin.

> ***Possible Data Theft*** /// Another problem is that data users must incorporate the data into their own system in order to use it. Once incorporated, they can continue to use the data without paying for it. The data provider cannot verify this or can do so only with considerable technical effort, which leads to additional dependency. The commercial success of the data provider is based on trust in the buyers and their proper use of the data.

From Wishful Dream to Nightmare to Reality? /// The challenges described here have led many data providers to disappear from the market or be acquired by major technology companies like Oracle, Salesforce or Adobe. The use of big data, which is already a reality at Google, Facebook and other global players, has become a nightmare for them.

But are there solutions that could enable small data service providers to be successful? The most promising approach appears to be bundling data with advertising technology and advertising space. If these bundles come from one source and are sold as a combi-product, most of the problems involving price setting are eliminated. Big data can be better used to increase the effectiveness and inherent value of media, to achieve a margin from the cost savings of a highly targeted media selection and reduced wastage, and to leverage its economic potential completely.

Once the data quality problems are also solved, then the profitable use of big data is no longer a wishful dream for the data providers or their customers, but rather a reality.

/.

FURTHER READING

https://bigdatablog.de/big-data-3/

http://blogs.wsj.com/cmo/2015/02/23/marketers-question-quality-of-ad-targeting-data-providers/

The Art of Creating Attractive Consumer Experiences at the Right Time: Skills Marketers Will Need to Survive and Thrive

Katherine N. Lemon

KEYWORDS

Marketing Skills, Customer Experience, Consumer Experience, Big Data, Key Metrics, Causal Models, Real-Time, Deep Knowledge, Creativity

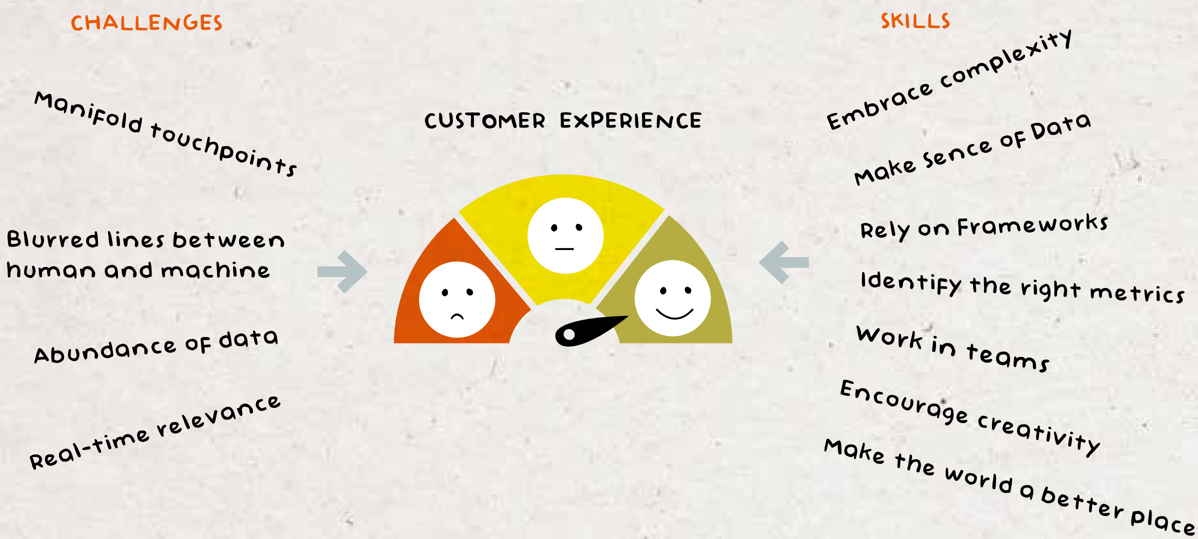
THE AUTHOR

Katherine N. Lemon,
Accenture Professor and Professor of Marketing,
Carroll School of Management,
Boston College, USA, and Executive Director,
Marketing Science Institute, USA
kay.lemon@bc.edu

Technology, data and the fragmentation of attention
/// Delivering a strong, positive and enduring customer experience is still a critical challenge that most companies face, but the way people experience almost everything is being completely transformed. Technological opportunities like mobile, location-based, digital making, virtual reality, drones, neuroscience, automation, machine-to-machine interactions through the internet of things are changing consumer behavior, changing the way companies are organized, changing the role of "humans" in the marketplace. The lines between human and machine are becoming blurred.

The manifold experiences and new technology result in massive amounts of data across all touch points in the entire customer journey, across channels, offerings, platforms and devices, incorporating collaboration with other customers and partners, and across time and space as well. And given all those different touch points, consumer attention has become fragmented. Consumers have gained more control over what media they consume or channels they use. The ability to push messages and ideas to consumers that marketers used to have has greatly diminished. Now, all marketing is pull, there is no more push. In this new world, marketers need to be "real-time relevant" – to gain awareness, to change perceptions and to spur action. They need to have the content in the channel, format, time and context that the consumer wants – to stand at least a chance of the consumer attending to the information and being influenced by it. In this new consumer world, marketing works best when the consumers initiate the conversation – or the relationship – with the company because what they find is relevant at that very instant.

FIGURE 1:
Skills to create convincing customer experiences in complex markets



But how can marketers handle this complexity and this pace? What are the skills necessary to identify the right technologies, to design all the different experiences and to appeal to consumers at the right moment? Essentially, seven skills are required to survive and thrive in the new market conditions (Figure 1).

» The ability to push messages and ideas to consumers that marketers used to have has greatly diminished. Now, all marketing is pull, there is no more push. «

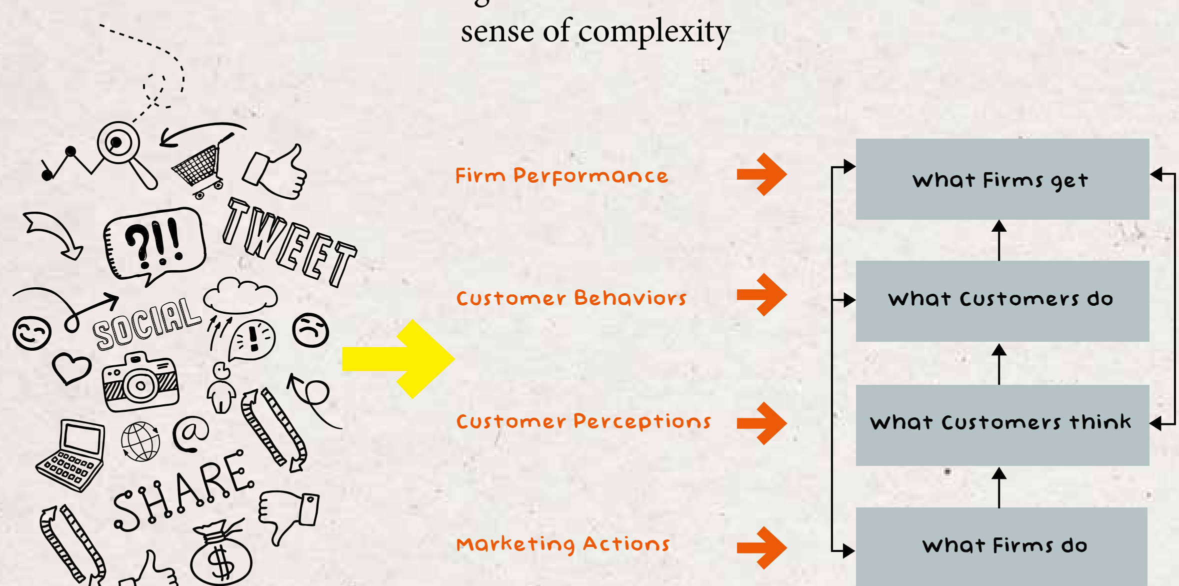
Seven skills marketers will need to survive and thrive

1. Embracing complexity /// There is no simple any more. Marketers need to embrace the messiness and unwieldiness that is the path to purchase for their customers – a dynamic process that happens across a myriad touch points in multiple channels on multiple screens involving multiple partners in uncontrollable contexts with significant social influence. To gain any perspective or apparent "control" in these environments, successful managers have to embrace the complexity and develop models and approaches that enable them to consider and incorporate multiple dimensions – the old 2x2 matrix just isn't going to be enough. Consider just one element in the overall customer journey – digital media. Making sense of how distinct elements of the digital media landscape may influence a consumer's path to purchase is a daunting task. A myriad companies are working to develop multi-touch attribution models that enable them to understand the extent to which each digital element may influence the path to purchase. These models offer much promise but, at the moment, are somewhat scattershot in their effectiveness. That said, it is crucial to embrace the complexity anyway – and to begin to develop these capabilities and models.

2. Making sense of all the data /// Complexity and technology have led to an explosion in data. There is a critical need for new models and approaches to analyze, visualize, integrate, interpret, and present this huge amount of data. We need ways that lead to usable insights and result in improved value for both customers and companies. Marketers will need better causal models, better predictive analytics, more scalable methods, simpler ways to gain insights from unstructured data, easily understandable and updatable data visualization techniques, and easily implementable monitoring approaches, metrics and dashboards. What makes this even more challenging is that these new approaches need to be dynamic and to provide real-time insights, enabling real-time design, access, personalization, and response. With the explosion of data, channels and the digitalization of everything, almost every action is traceable and causal inference models are becoming more ubiquitous. Marketers of the future will need to have the capability to build and test these causal models, to gain insights from them, and to change course when necessary. At the Marketing Science Institute, we have just completed our 2016-2018 Research Priorities, outlining the most pressing topics companies face in the next 2-5 years. Making sense of the data arose as a key priority. As one company noted, we need "to evolve traditional methods to faster and agile consumer/customer 'pulse reads' that can lead to better and faster insights; to read marketing levers in real time, including emerging touch points; and the agility/ability to change plan given real-time learning."

3. Relying on frameworks /// This may sound quite "old school," but I firmly believe that marketing frameworks – maybe even new ones not yet developed – will be a critical resource for the marketer of the future. Frameworks are a way to put order into a complex environment and the more messy and confusing the scene gets, the more important it will be to find structure for moving around successfully. Understanding the critical paths for the customer journey, or the key aspects of a successful innovation and new product introduction, or the key elements of brand engagement, or the keys to a successful content strategy will be necessary to gain insights and make decisions. Frameworks don't have to be unduly complex. For example, in the one pictured on the right side of Figure 2, Gupta and Zeithaml provide a simple framework within which companies can link their own actions with consumers' thoughts and actions and the respective outcomes. What is the value of these frameworks? Keeping everyone on track, and providing a common language for checking progress and moving forward.

FIGURE 2:
Using frameworks to make sense of complexity



»
Perhaps more difficult will be not to pay attention to metrics that don't matter. Traditional metrics that no longer predict key outcomes need to be demoted.
«

4. Identifying the metrics that matter and paying attention to those /// Decision makers can often get bogged down with metrics – and often measure the wrong things. A critical skill for marketers will be to identify the metrics that best reflect the desired outcomes of the organization and that sufficiently reflect specific indicators of critical processes. As causal inference models become better over time, marketers will begin to get a sense of which metrics truly act as leading indicators and are good proxies for desired outcomes. Recent evidence suggests that traditional approaches, such as surveying customers to gauge their satisfaction or likelihood to recommend, may be less effective moving forward. Instead, more in situ measures throughout the customer journey may become better metrics.

We also need to pay more attention to how the metrics we collect may actually influence our customers. Much research suggests that merely asking customers about their future intentions can influence their future behavior – the "mere measurement effect." Recent research my colleagues and I have conducted suggests that these effects may be more pervasive than previously thought. We find that asking customers to focus on what they liked about the experience (via an open-ended prompt at the beginning of a traditional post-purchase survey) led to significant increases in future purchase behavior in a consumer retailing context and to significant increases in conversion from trial to purchase in a business-to-business context. Assuring that our metrics are accurately measuring what we want them to measure is critical. Perhaps more difficult will be the ability to not pay attention to metrics that don't matter. Traditional metrics that no longer predict key outcomes need to be demoted.

5. Create agile teams with deep and broad skills /// Marketers will need to have deep expertise in big data analytics and also have the skills of an anthropologist to deeply and quickly grasp what is going on in the minds, hearts, voices, and hands of consumers. Marketers will also need significant integrative skills. The customer insights group within an organization may be faced with insights from a huge machine learning attribution model and at the same time with insights from qualitative social media data. Also, there are significant opportunities to integrate and leverage insights from design science, architecture, information systems, biology, engineering, and other fields to enhance the processes by which we design experiences.

Successful companies will need individuals and approaches that enable them to compare and to integrate the insights from such disparate approaches. The marketing team will also need to be skilled in the art of developing a compelling narrative – to tell their story up the chain of command, across all groups in the organization that might influence or "touch" the customer. Communicating a brand's value proposition in captivating ways to a heterogeneous and barely attentive market will be crucial. Finally, marketers will also need to respond and react in real time. Much like a sailor navigating an upwind course, marketers will need to both "stay the course" – keeping key objectives in mind – and make quick, necessary "tacking" adjustments in real time given the continuous feedback being received from the market.

6. Encouraging creativity and curiosity /// The marketer of the future will need to be supremely curious and creative. Understanding the role of emotions in experience design and the customer journey will be a significant challenge. Also, how to help consumers and customers get their "jobs done" smoothly, with less friction, and maybe even with more joy. Disruption and innovation are coming from unlikely places – unlikely start-ups are gaining trust and stealing significant share from very well-known and entrenched market leaders. Consumers are empowered, attention is fragmented, and barriers to entry are eliminated. It is much easier to identify market opportunities, to solve customer problems in more efficient and often cheaper ways. Thus, encouraging disruptive thinking, and fanning creative sparks into flames will be critical. Companies will need to hire curious, creative thinkers who can connect disparate things in new ways.

7. Keeping an eye on making the world a better place ///
Last but not least, marketers have a real opportunity to develop products, services and relationships with customers that focus on improving well-being and establishing optimal social contracts. Many companies will need to gain deep insights into the needs of developing markets and into how to serve people at the base of the pyramid. Consumer reliance on technology also creates opportunities for brands to be embedded in the lives of their customers like never before, and with these deeper relationships come greater responsibilities. Recent advances in behavioral decision making and behavioral economics offer great opportunities for companies to work with their customers to ensure that value is created for consumers as well as for companies. Encouraging customers to make better decisions will also be a critical skill for marketers of the future.

Delivering great experiences at the moments that matter. The main objective of all these skills is to still create strong and enduring experiences for customers under ever more challenging conditions. Customer experiences are dynamic, complex, social and embedded in the larger ecosystems that customers inhabit. It does not appear that the focus on designing and delivering great customer experiences is going to diminish anytime soon. Clearly, understanding your customer's decision journey and the critical moments in that journey – in real time, and in context – will be critical. But those marketers who are able to take insights further to design and deliver consistent, seamless experiences will be the marketing stars of tomorrow – and will make their companies thrive.

/.

»
A critical skill for marketers will be to identify the metrics that best reflect the desired outcomes of the organization and that sufficiently reflect specific indicators of critical processes.
«

FURTHER READING

Bone, Sterling A.; Lemon, Katherine N.; Voorhees, Clay; Liljenquist, Katie A.; Fombelle, Paul W.; Money, R. Bruce; DeTienne, Kristen B. (2015): "Mere Measurement 'Plus': How Solicitation of Open-ended Positive Feedback Influences Customer Purchase Behavior," Journal of Marketing Research, forthcoming.

Gupta, Sunil; Zeithaml, Valarie (2006), "Customer Metrics and Their Impact on Financial Performance," Marketing Science, Vol. 25 (6), pp. 718-738.

Marketing Science Institute (2016), Research Priorities 2016 – 2018, Cambridge, MA

ABOUT ING-DiBa

ING-DiBa AG is a direct bank based in Frankfurt am Main. It is a subsidiary of ING Groep, which is based in the Netherlands and operates internationally. With more than eight million customers, it is one of the largest retail banks in Germany. Its core business areas in retail banking are savings deposits, securities trading, construction financing, consumer loans and checking accounts. The bank does not operate branches, but instead gains insights into customer behavior and needs primarily through the systematic analysis of customer data. In Germany's largest bank test conducted by the business magazine "€uro", ING-DiBa was selected as Germany's most popular bank in 2016 for the tenth time in a row in a survey of more than 180,000 bank customers. More than 3,700 employees work for ING-DiBa at sites in Frankfurt, Hanover, Nuremberg and Berlin.

ABOUT MARTIN SCHMIDBERGER

Dr. Martin Schmidberger is a Fully Authorized Representative at ING-DiBa, where he is responsible for product and target group management. Since 1999 he has built up the Customer Analytics department, which focuses on the areas of data warehousing, campaign management and forecasting models. He is currently working with his team on matters relating to "real-time marketing" and the use of innovative algorithms to forecast customer behavior and "machine learning processes". Martin Schmidberger earned his doctorate in 1997 and has taught at the Goethe University Frankfurt am Main as an instructor and lecturer since 2012.

THE INTERVIEWER

Professor Bernd Skiera conducted the interview in April 2016.

Data Analysis Trumps Specialist Advice: How Direct Banks Function

MIR interview with Martin Schmidberger,
Fully Authorized Representative at ING-DiBa Germany
and responsible for product and target group management

Low interest rates and sluggish economic growth are not exactly ideal conditions for the financial services industry. Almost daily we are confronted with reports of inadequate capital bases, declining earnings, and layoffs at banks. But while many traditional retail banks are struggling with a business downturn, the direct bank market is enjoying steady and respectable growth despite a challenging environment. Dr. Schmidberger, Fully Authorized Representative at ING-DiBa Germany, offers us a glimpse behind the curtains of this direct bank. We will learn how data technology is used so that bank customers are (more) satisfied.

MIR: Today ING-DiBa is the third largest private bank in the highly competitive German retail banking market, despite not having any branches. How is this possible?

MARTIN SCHMIDBERGER: In a market as competitive as the German retail banking market, approximately 2,000 banks are competing for the customers' favor. To be successful you have to understand the requirements and desires of your customers perfectly and ideally reflect them in your products. Along with our focus on simple and transparent products, we became involved very early in the analysis of customer data. We recognized that the systematic use of data can lead to a better understanding of customer behavior and customer needs and we regard this knowledge as one of our core assets.

MIR: How do you get to know your customers at a direct bank? What kind of information helps you understand your customers?

MARTIN SCHMIDBERGER: Many of our analytical questions concern our customers' experience with us as a bank when they use our products: What channels do they employ, what products do they use? What options are used frequently or less frequently? How do customers get information about other products and services? What procedures do they complain about? Which customers recommend us to others? We can learn a lot from the results of the analyses and optimize products accordingly.

MIR: Also, sales management is surely different at a direct bank. How do you sell additional services to customers who are never seen in person?

MARTIN SCHMIDBERGER: Of course, numerous analyses revolve around the topic of sales. Here, we evaluate customer behavior in order to determine purchasing probabilities using predictive modeling. Based on these probabilities, we manage and optimize our sales, both in the acquisition of new customers and in cross-selling. In the process, we rely increasingly on digital channels. Daily we have approximately 400,000 logins to our internet banking service and almost 200,000 accesses from mobile devices – that contrasts with about 20,000 incoming calls. Well over 90% of our customer contacts already occur digitally today.

MIR: How has increasing digitalization changed your marketing?

MARTIN SCHMIDBERGER: The digitalization of customer contacts brings two kinds of challenges. On the one hand, we must achieve the transformation from traditional marketing, like from classical advertising mailings, to digital channels such as online banking. Therefore, we need digital forms of advertising whose response rates, for example, do not fall short of the rates for traditional offline media. On the other hand, the digitalization of customer communication entails a significant acceleration of customer interaction. This means that the coordination of advertising campaigns via multiple channels must be quicker and more automated than before. Online and offline advertising messages must be selected simultaneously and blended into a consistent, customized, multi-channel approach.

MIR: Earlier you addressed the optimization and management of sales. What are you doing specifically?

MARTIN SCHMIDBERGER: Our goal is to determine the most appropriate message for each customer individually and in real-time and to display it on the customer's digital device, for example for internet banking or on the app. We have developed a modern big data system to adapt our selection procedures to digital channels and to deliver service and sales messages to customers. Along with traditional data from customer relationship management, we also consider current data, such as a customer's immediate financial situation or specific user entry. This enables messages with very high relevance. For example, we can instantly offer a customer whose checking account has just gone into the red a more favorable global credit facility as an alternative to a relatively expensive overdraft facility.

» Once the customer logs into online banking, we have no more than half a second to select the suitable sales or service messages. «

MIR: So your systems transform data into customized recommendations. What does the technology behind such applications look like?

MARTIN SCHMIDBERGER: High standards for the speed of data processing require powerful, modern big data technologies. Once the customer logs into online banking, we have no more than 500 milliseconds to select the suitable sales or service messages. In this half-second, a series of filters and complex selections are executed in real time to deliver the appropriate sales message. A rule engine manages the algorithms for determining the service and sales messages. Subject area specialists can flexibly optimize and expand them at any time without IT support. This flexibility is important to adapt the system to different applications so that future ideas can also be implemented quickly and efficiently.

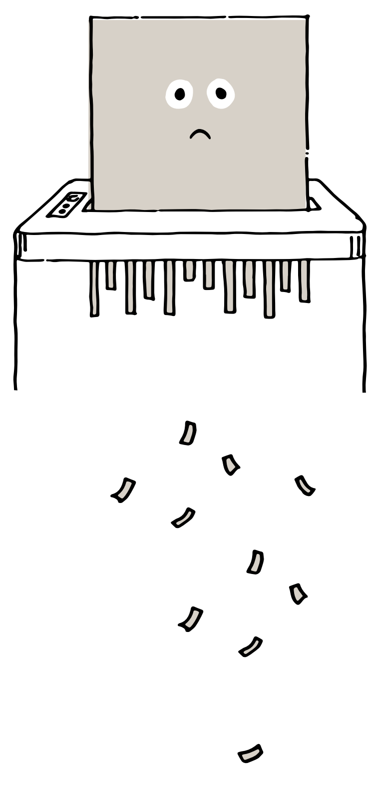
MIR: Big data technologies are a natural match with major investments. Can you determine their advantages for sales?

MARTIN SCHMIDBERGER: Yes, we can measure that precisely. Our digital sales have become much more efficient and less costly. By increasing digital sales contacts with our customers, we have been able to reduce the circulation of postal mailings by approximately 80%. Cross-selling rates for existing customers increased in this period by more than 20%. This shows that our digital methods help us selling more successfully than in the earlier "paper-world".

» Our digital methods help us sell more successfully than in the earlier "paper-world". «

MIR: Are investments in real-time technologies also paying off?

MARTIN SCHMIDBERGER: With real-time advertising, we are much more successful in reaching customers at the right moment and with relevant topics. These highly individualized digital forms of advertising result in similar and sometimes even higher response rates compared to personalized mailings. To quote just one example: within a few days, a purely digital campaign to sell our "direct custody account" product had a result comparable to a €120,000 printed mailing. The investment for the digital campaign consisted almost exclusively of creating a corresponding banner and setting up a rule designed for the target group.



MIR: So digitalization has radically transformed advertising and sales. Surely the entire company with all its processes had to change as well?

MARTIN SCHMIDBERGER: Yes, we scrutinized many established processes and replaced them with modern, digital implementations. Most importantly we overcame the pre-existing ''sales channel silos'' through a consistent system of multichannel management. Previously the planning and selection of postal and digital campaigns was certainly coordinated, but done in separate departments. With our new focus, we now design campaigns across all available channels in an integrated manner. For example, we only send traditional mailings, if a digital and costless contact had not been successful. Instead of mailing lists, we now create selection criteria and real-time rules for our digital channels. We have also restructured our target group management to streamline it for the new opportunities of increasing digital communication.

MIR: Has digitalization had an impact on the methodology of your data analysis or has everything in this area remained as it was before?

MARTIN SCHMIDBERGER: Analytical concerns in the digital world are entirely different. In the traditional ''offline'' world, such as mailing, we faced an outbound situation. Using statistical models, we selected those customers who best fit a given product offer. This basic framework changes completely because digital sales entail an ''inbound situation''. We as a bank do not initiate contact with the customer. Instead, the customer decides whether and when a contact is made. Therefore, our task is no longer to select the best customers for a given product, but to find the most suitable product for each customer that makes contact with us. In this next-best-offer approach, traditional response models must be adjusted methodically.

MIR: So everything is new, even the methodical approaches to customer analysis?

MARTIN SCHMIDBERGER: Of course, many established analytical and forecasting methods continue to be applicable, but they are becoming more complex. Frequently, the choice of the ''best'' product offer requires a combination of several regression or response models. In addition, in a digital real-time environment, we can take advantage of additional data, such as contact frequency, time, or placement of advertising space that was not available with traditional offline media. This means that we can now link established analytic methods, such as regression analysis, with real-time decisions about contacting customers online. Typical applications are re-targeting or frequency capping for deciding what message should be used and the maximum number a customer should be contacted. In the future, I expect a further blurring of the boundaries between the proven, traditional techniques of the ''offline'' world and the situational and real-time possibilities of the ''online'' world.

MIR: How can you achieve these methodological blends?

MARTIN SCHMIDBERGER: We work on a series of innovative developments that fuse old and new techniques. For instance, as already mentioned, regression analysis still plays an important role. Its primary strength lies in the robustness and good interpretability of its results. Naturally, we also use the new processes of machine learning, for example with random forest algorithms. In various applications, these processes have shown very promising results and a slightly superior forecasting power to regression models. However, it is currently still difficult to integrate such computationally-intensive algorithms into a real-time context.

MIR: But whoever has the best algorithm wins?

MARTIN SCHMIDBERGER: More important than the choice of algorithm is our ability to create and use an entirely new category of input variables generated by real-time applications and web data. For example, we see that the response rate depends on the time of day. Accordingly, variables like time of day, the amount of time needed to process applications, the terminal device used or navigation behavior are completely new data that was previously not available. My belief is that the systematic exploitation of such data has a greater impact on response optimization than the selection of the algorithms to be used.

»

Our task is no longer to select the best customers for a given product, but to find the most suitable product for each customer that makes contact with us.

«

MIR: In conclusion, perhaps a look at the bigger picture. You have provided an impressive explanation of how a direct bank benefits from data analysis. In general, how do you assess the outlook for "analytics"?

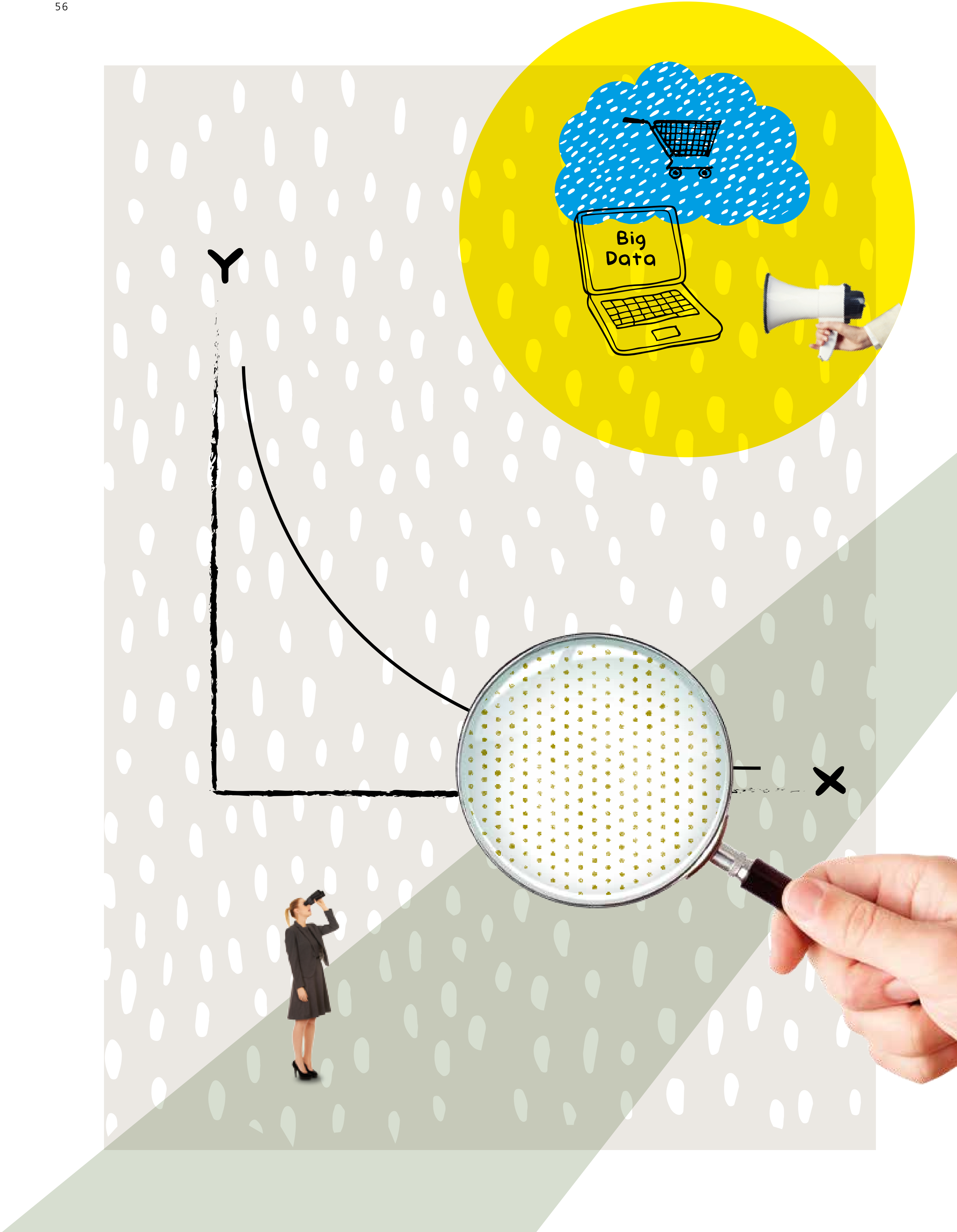
MARTIN SCHMIDBERGER: The view that marketing is increasingly data-driven is gaining more and more acceptance in everyday business activities. I believe that in a few years, analytics will be a natural part of sales and will be represented in senior management more prominently than to date.

MIR: Does this mean that in future, career opportunities will depend on big data knowledge and analytic skills?

MARTIN SCHMIDBERGER: Analytics today is still a specialized discipline of a few experts. However, I expect that in many professional fields such as product management or web design, analytic knowledge will become much more relevant. In the midst of the current debate about "big data", great expectations have been raised for an explosive increase in market efficiency, but these expectations are likely to be only partially met. It would seem that analytics still have some maturing to do, including managing these expectations. Still, I believe more than ever that the analytic expertise of companies will play a decisive role in their competitive success.

MIR: Thank you very much for your assessment and informative insights into the data landscape of Ing-DiBa. We wish you continued success in shaping your digital customer relationships.

/.


Y
Big
Data
X

Big Data in Market Research: Why More Data Does Not Automatically Mean Better Information

Volker Bosch

KEYWORDS

Big Data, Market Research, Information, Representativeness, Data Integration, Data Imputation

THE AUTHOR

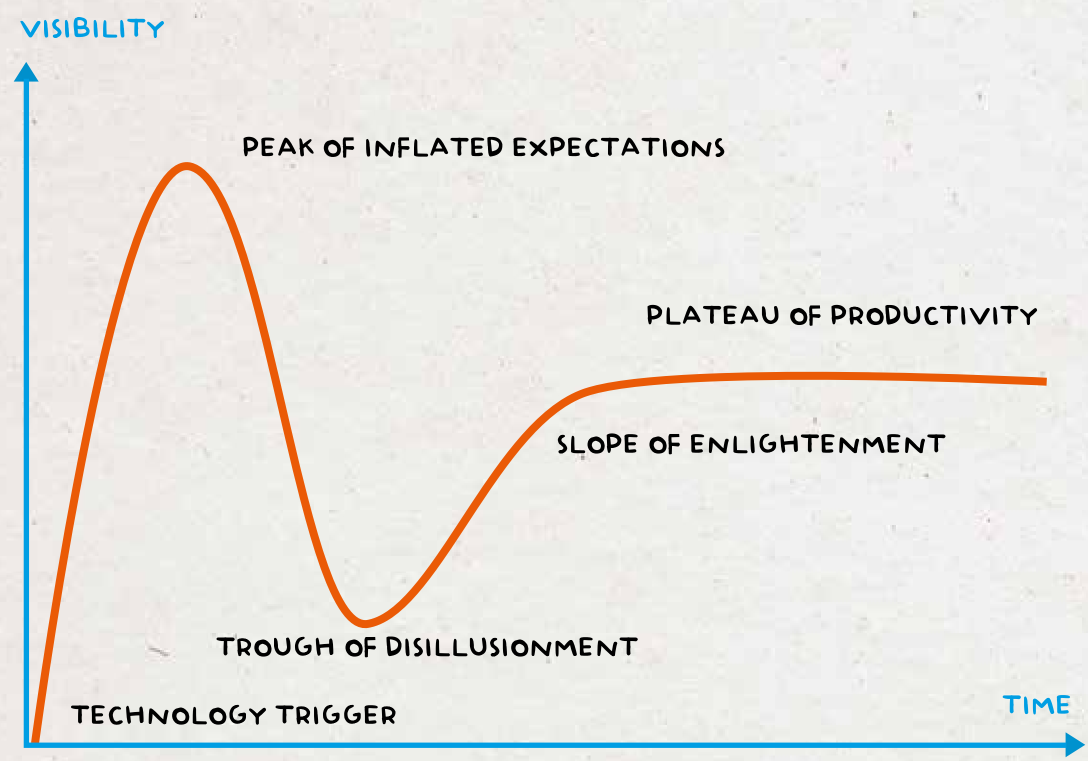
Volker Bosch,
Head of Marketing & Data Sciences,
GfK SE, Nuremberg, Germany;
volker.bosch@gfk.com

Everyone is Talking about Big Data /// "Big data" is a megatrend, although not everyone means the same thing when they talk about it. Generally speaking, it involves enormous amounts of data, which are generated almost automatically with the help of the latest technological developments, and examines how this data can be converted into useful information. Big data has raised big expectations, but the profitable monetization of data usually turns out to be much more complex than anticipated. That new technological developments are no guarantee for a start-to-finish victory is nothing new, though. Over time, many conform to a typical pattern which was first described by the consulting firm Gartner in 1995 and has since been referenced in numerous publications as the "hype cycle" (Figure 1).

At this time, big data has probably passed the "peak of inflated expectations" and is still yet to cross the "trough of disillusionment" before reaching the "plateau of productivity". But what are realistic expectations for big data? And what does big data mean for the market research industry?

Big Data Conquers Market Research /// Big data will change market research at its core in the long term. While other trends such as neuromarketing have not been able to gain a substantial foothold, big data business models will assume a central role in the value chain. The consumption of products and media can be logged electronically more and more, making it measurable on a large scale. In some areas of market research, big data is already established today, with social media analytics and the use of cookie data to measure internet coverage being two prominent examples. The use of panels for the passive measurement of media consumption through the internet, television, and radio also falls under

FIGURE 1:
Hype cycle according to Gartner



big data. But what additional benefits does big data provide compared to traditional market research data?

Big Data = Passive Measurement /// The 4V definition describes the core characteristics of big data: volume, velocity, variability (of the data structure) and (questionable) veracity. However, 4V doesn't tell the whole story, because the origin of the data is especially decisive. New sensor technologies and processing architectures enable entirely new possibilities for gathering and processing information. We are dealing with a fundamental paradigm shift – in traditional market research, data is collected actively, e.g. through human interaction or interviews. With big data, by contrast, the information no longer needs to be processed by slow, limited-capacity, mistake-prone and emotional human brains for a dataset to be created. As a result, passive measurement is the actual driver of the efficiency of big data in market research. It creates economies of scale that were formerly the stuff of dreams.

»
Passive measurement is the actual driver of the efficiency of big data in market research. It creates economies of scale that were formerly the stuff of dreams.
«

Defining big data based as passive measurement does not necessarily mean massive amounts of data. Equipping just a few measured units with sensors can already create datasets that are hard to manage. Examples include the software-based measurement of internet behavior in a panel or equipping shopping carts with RFID technology to transmit the precise location and purchase history of a customer in a supermarket. Perhaps it would be more appropriate to speak of "new data" than "big data."

Is Twice as Much Data Worth Twice as Much? /// The size of a typical big data dataset leads to the false assumption that it provides a correspondingly large amount of information. From an organizational perspective that is absolutely correct, but from a statistical perspective, it is wrong, because "information" in statistical analysis is defined as the reduction of uncertainty. But twice the data does not mean twice the accuracy, only an improvement by a factor of 1.4, as measured by the confidence interval of a sample. Marginal utility declines significantly as data amounts increase. Ignoring declining marginal utility will almost certainly result in overestimating the value of big data for market research and overlooking its actual benefits.

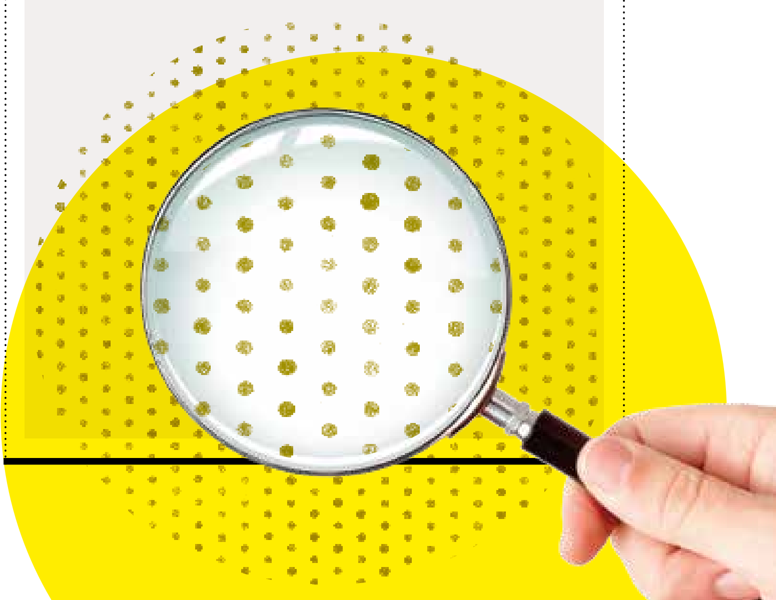
» Big data can be used like a microscope to see structures that would appear blurred with conventional market research or even be entirely unrecognizable. «

In visual terms, a larger amount of data results in greater statistical resolution, enabling structures of finer granularity to be described with statistical validity. Examples are smaller target groups, websites in the "long tail" of the internet, or rare events. Big data can be used like a microscope to see structures that would appear blurred with conventional

{*Box 1*}

MORE DATA ≠ BETTER DATA: THE LITERARY DIGEST FARCE

In 1936, the leading news magazine in the 1930s, the Literary Digest, delivered a very clear prediction for the outcome of the presidential elections. It was based on an extensive mail and telephone survey using sources available at the time, the telephone book and a list of car owners. 2.4 million citizens participated and a clear victory for Alf Landon, the challenger to the incumbent F.D. Roosevelt, was predicted. For the standards of the day, that was certainly big data, even though there were no methods for passive measurement in 1936. But based on a much smaller sample of about 50,000 persons, George Gallup predicted the exact opposite. After analyzing the non-representative subsamples of telephone and car owners, he predicted that the Literary Digest forecast would be wrong. He was vindicated, and shortly after this glaringly wrong prediction, the Literary Digest had to cease publication.



market research or even be entirely unrecognizable. In other words, the declining marginal utility is mitigated by the fact that with big data, structures of the finest granularity can be defined in the midst of the statistical noise.

Big Data Must Be Scientifically Evaluated in Market Research /// In direct marketing, with CRM systems, or in intelligence agencies, it mainly comes down to describing individual characteristics. In market research, by contrast, the aim is to find valid, generalizable statements based on scientific standards. When analyzing the use of products and media by populations and their segments, it must also be possible to describe statistical errors. This has a decisive impact on the applied algorithms and processes. Statistical methods, data integration, weighting, variable transformations, and issues of data protection present much larger challenges than was previously the case with conventional datasets. In particular, the three following challenges must be overcome.

> ***Challenge 1: Big Data is (Almost Always) not Representative*** /// Massive amounts of data do not necessarily result in good data and more does not automatically mean better. Big data can easily tempt us to fall into the same "more is better" trap that the Literary Digest fell into back in 1936 (see box 1). It is a truism of market research that sample bias cannot be reduced by more of the same and that the representativeness problem remains. Consequently, traditional topics of sampling theory like stratification and weighting are highly topical in the age of big data, and must be reinterpreted. Rarely is it possible to measure all units of interest and avoid bias. For example, the scope of interpretation for social media analysis is restricted because the silent majority usually cannot be observed. This may explain why social media data often behaves unexpectedly in predictive models. By using smart algorithms, however, it is possible to achieve astounding precision with non-representative digital approaches, as for example in the elections in the United States in 2012 and in Great Britain in 2015.

> ***Challenge 2: Big Data is (Almost Always) Flawed*** /// The passive measurement of behavior along with its proxies and the high level of technology being used may lead to the false assumption that practically no measurement error exists and that data can be further processed without hesitation. But that is very rarely the case. Such technologies are highly complex and often not designed for use in market research. Big data must be processed with very complex and therefore error-prone software and many measurement errors arise. In addition, the internet ecosystem is subject to constant updates (in the best case) or to changes in technology (in the worst case): Internet Explorer yields to Edge, HTML5 replaces the old HTML4, http pages turn into https, or Flash is no longer supported. In measuring internet behavior in the GfK Cross-Media-Link panel, we observed how browser updates, technological upgrades, changes in website behavior, and end-of-life systems can lead to a measurement failure. If updates occur unannounced and unexpectedly, emerging measurement gaps might even be noticed (too) late.

Things become even more difficult with systems that were originally constructed for another purpose, for example, if mobile internet use is measured by a mobile network operator and not by the market researcher directly. This is referred to as network-centric measurement, in contrast to user-centric measurement in a panel or site-centric measurement using cookies. Data processing capacities in such systems primarily serve to maintain telephone or internet service and for billing. Market research requirements were not even a factor in the original design at all. Therefore, so-called "probes" must be laboriously installed in order to retrieve the relevant information. Control over data quality is limited. Undetected data blackouts frequently occur because the primary tasks of the system take priority and no error routines have been installed for other requirements. GfK discovered this the hard way in its "Mobile Insights" project.

> ***Challenge 3: Big Data (Almost Always) Lacks Important Variables*** /// The biggest market research challenge from a methodological point of view is the limited data depth of big data. Despite the sometimes overwhelming amount of data in terms of observed units, the number of measured variables is low or critical variables are missing. By contrast,

{ *Box 2* }

DATA IMPUTATION

In a traditional data matrix, the columns represent variables and the rows represent observation units like people or households. Variables observed in the census are available for all units, while other variables, for example sociodemographic characteristics, can only be collected in a subset, e.g. the panel.
In image data, gray values represent the measured values of a variable (Figure 2). In the example, 75% of the data or image points are missing. Only a few randomly selected rows (panel members=donors) and columns (census data=common variables) are fully observed. In order to ensure that an algorithm cannot use pure image information (the physical proximity of image points), rows and columns are randomly sorted. As a result, the image behaves like a market research dataset and the data can be processed accordingly.
The missing values are filled in by imputation. Many algorithms exist and all work with different assumptions regarding the statistical properties of data. As a common feature they learn from donors how the common variables relate to the specific variables to be transferred and they fill the data gap for the recipients using this knowledge. In the big data context, imputation is particularly difficult because large quantities of data must be processed and finding the optimal model is usually too costly. In addition, the data rarely follows a multivariate normal distribution or other well-defined distributions.
This is why the Marketing & Data Sciences department at GfK developed the "linear imputation" process. It requires a minimum of theoretical assumptions and delivers good results, even for highly non-linear data structures (as in the image) by using local regression models.
In the image example, the quality of the imputation can be judged immediately if the matrix is sorted back to its original order (Figure 3, middle picture).

FIGURE 2:

Data imputation using an image file with random sorting of rows and columns

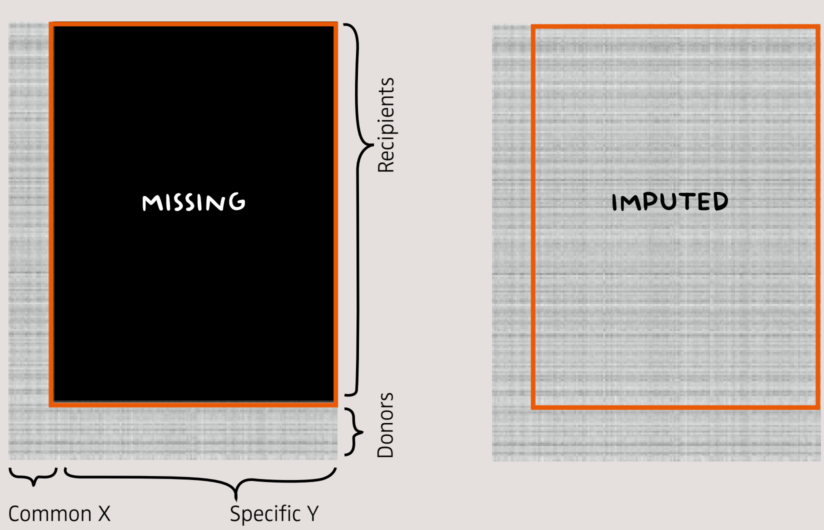
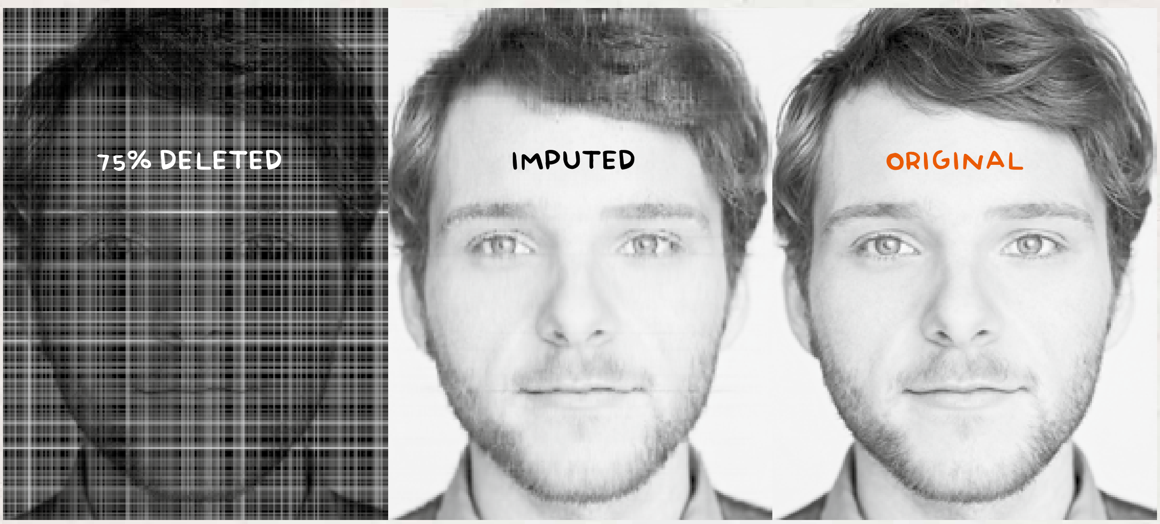


FIGURE 3:
The original sequence of rows and columns was restored following imputation



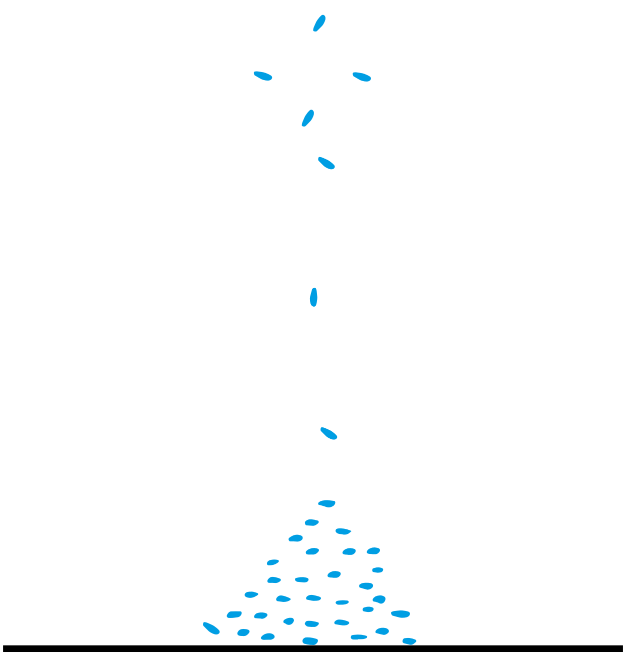
in traditional opinion research the relevant variables are optimized for a specific subject and can be very extensive. Internet coverage research based on cookies or network-centric data illustrates this. Even if almost the whole population is reached, like in a census, critical information is missing, such as sociodemographic data. Therefore the value of the collected data is limited and important evaluations such as target group or segment-specific analyses cannot be conducted. The missing information can only be filled in using statistical data imputation. This requires an additional data source with the additional variables, for example a panel. The source must also contain the variables of the big data dataset. Imputation is a statistical procedure that is anything but trivial. Box 2 describes the underlying logic using an image dataset, which is handled like market research data.

However, imputation is not a tool that lets information be gathered by magic. Statistics do not create information, only observation does. Statistics makes structures visible. Therefore, imputation is an instrument to "transport information" and the higher the observed data correlates with the data to be imputed, the better it works.

More Value through More Data: Also in Market Research

/// Big data poses special challenges for market research. It is by no means sufficient to master the technologies for processing large amounts of data, or to engage in pure "data science". It is also necessary to develop in-house market research algorithms, which can be applied to the new data and successfully address the three challenges of representativeness, measurement errors, and statistical data integra-

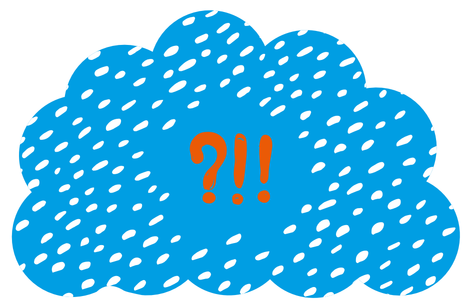
»
Despite the sometimes overwhelming amount of data in terms of observed units, the number of measured variables is low or critical variables are missing.
«



tion. Therefore, the young discipline of "data science" must be fused with the classic field of "marketing science" to help market research expand its core business successfully.

And at least as far as applications in market research are concerned, big data is well on its way to the plateau of productivity in the hype cycle of new technologies.

/.



FURTHER READING

Fenn, Jackie (1995):
The Microsoft System Software Hype Cycle Strikes Again

Gaffert P., Bosch V., Meinfelder, F. (2016):
"Interactions and squares. Don't transform, just impute!," Conference Paper, Joint Statistical Meetings, Chicago

http://www.ibmbigdatahub.com/infographic/four-vs-big-data

http://fivethirtyeight.blogs.nytimes.com/2012/11/10/which-polls-fared-best-and-worst-in-the-2012-presidential-race/?_r=0

Editors

EDITOR

Bernd Skiera,
Chair of Electronic Commerce,
Faculty of Business and Economics, Goethe-University, Frankfurt am Main, Germany
www.ecommerce.wiwi.uni-frankfurt.de
skiera@wiwi.uni-frankfurt.de

MANAGING EDITOR

Christine Kittinger-Rosanelli,
GfK Marketing Intelligence Review
christine.kittinger@gfk-verein.org

ABOUT BERND SKIERA

Bernd Skiera took over the first chair of Electronic Commerce in Germany at the Goethe-University Frankfurt in 1999. In addition, he is a founder and member of the Board of the E-Finance Lab, a Board member of the German Marketing Association and a Board member of the Schmalenbach Association, as well as an Advisory Council member of the INFORMS Society of Marketing Science (ISMS). His main research focus is on online marketing, marketing analytics, electronic commerce, customer management, and pricing. The German business journal "Handelsblatt" currently ranks him the most productive researcher in Germany, Austria and Switzerland. His publications have appeared in highly prestigious journals such as Management Science, Marketing Science, Journal of Marketing Research, Journal of Marketing, Journal of Product Management, Journal of Management Information Systems, and International Journal of Research in Marketing. He co-founded two companies. Bonpago GmbH (www.bonpago.de) helps companies to optimize their digital process. Marini Media GmbH (www.marini-media.de) develops solutions that support companies to better sell online.

Advisory Board

Imprint

PUBLISHER
GfK Verein
Nordwestring 101
90419 Nuremberg
Germany
Tel +49 911 395 22 31
Fax +49 911 395 27 15
Email: hello@gfk-verein.org
www.gfk-verein.org

•

The GfK Verein was established in 1934, primarily to "allow the voice of the consumer to be heard." Consequently, the purpose of the Verein since its inception has been to promote market research.

Today, the GfK Verein is a market research think tank acknowledged by those in both academic circles and engaged in practical application. Its remit as a not-for-profit organization is to create and pass on knowledge.

The GfK Marketing Intelligence Review© (GfK MIR) is a Marketing magazine directed at managers and market research professionals who are interested in new insight and methods of academic marketing research. It is published biannually.

•

DESIGN & ART DIRECTION
Scheufele Hesse Eigler
Kommunikationsagentur GmbH

•

PRINT
Druckerei Eugen Seubert GmbH,
Nuremberg

•

SUBSCRIPTIONS
75 € per annum

•

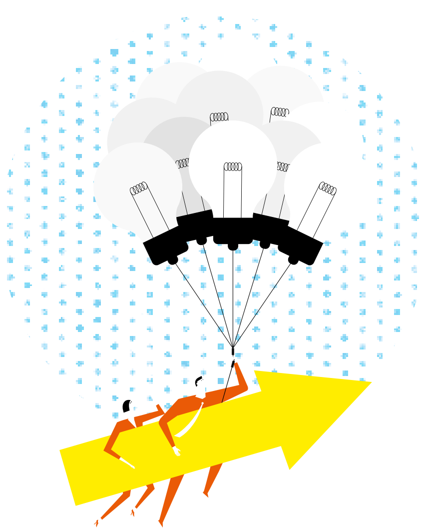
ISSN 1865-5866

•

ONLINE VERSIONS
www.gfkmir.com

Next Issue Preview

Digital Transformation



Digital Transformation: New Ways to Success
Werner Reinartz and Srinivas Reddy

Business Model Innovation –
How to Create Value in a Digital World
Christoph Zott and Raphael Amit

Platform Power
Marshall W. Van Alstyne

Creating Effective Display Advertising
Bruce Norris

Digitally Transforming Retailing
Werner Reinartz

Smart Cities - Livable Cities
Anil Menon

Crowdfunding
Srinivas Reddy

Too Successful to Digitalize? How to
Overcome Innovation Blindness and
Check Your Business Model's Assumptions
Andreas Neus

GfK MIR Vol. 9, No. 1, 2017 is forthcoming in May 2017

GfK Verein

MARKET RESEARCH THINK TANK