

Hans-Wilhelm Eckert

Storytelling with data

Gaining insights, developing
strategy and taking corporate
communications to a new level



Springer Gabler

Storytelling with data

Hans-Wilhelm Eckert

Storytelling With Data

Gaining Insights, Developing Strategy
And Taking Corporate Communications
To A New Level



Springer Gabler

Hans-Wilhelm Eckert
Momentum Communication
Munich, Germany

ISBN 978-3-658-38554-5

ISBN 978-3-658-38555-2 (eBook)

<https://doi.org/10.1007/978-3-658-38555-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive licence to Springer Fachmedien Wiesbaden GmbH, part of Springer Nature 2022

This book is a translation of the original German edition „Storytelling mit Daten“ by Eckert, Hans-Wilhelm, published by Springer Fachmedien Wiesbaden GmbH in 2021. The translation was done with the help of artificial intelligence (machine translation by the service DeepL.com). A subsequent human revision was done primarily in terms of content, so that the book will read stylistically differently from a conventional translation. Springer Nature works continuously to further the development of tools for the production of books and on the related technologies to support the authors.

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer Gabler imprint is published by the registered company Springer Fachmedien Wiesbaden GmbH, part of Springer Nature.

The registered company address is: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Foreword

Dealing with data is one of the core competencies of our digital age. And it has long since ceased to be the domain of computer science, but needs to build bridges to other disciplines. In his book, Hans-Wilhelm Eckert shows the role that data plays in communication and marketing and how it becomes an important source of storytelling.

The story is not in the data, but in our heads: this is the core message of the book and an appeal to our own power of judgement. Data do not have meaning in themselves, but only gain it through the context in which we place them. In this way, the author also formulates the claim to all those who deal with data: Don't hide behind data, but reflect and defend your point of view.

Hans-Wilhelm Eckert calls this cognitive process the oracle principle: a 3000-year-old cultural technique that has not lost its fascination even in the age of digitalization. On the contrary, it seems to be more effective than ever. The only difference is that today's oracles are no longer located in Delphi, but in Mountain View or Shenzhen. And that the source of wisdom is no longer to be found in divine inspiration, but in data.

What matters are the questions we ask of this data. This has not changed since Delphi. Hans-Wilhelm Eckert shows this with concrete examples from the everyday life of communication and marketing managers such as the positioning of brands, the analysis of target groups and the identification of relevant topics. In addition to exciting insights into

narratives of artificial intelligence and the fight against epidemics, the book provides solid foundations, useful tools and examples around storytelling with data.

And it has another good message in store: It has never been easier to emancipate yourself from the big oracles from China and the USA. Data is much more readily available today. Communication and marketing managers are well advised to find their role in the use of data in the company, to look for allies and to see the interdisciplinary interaction as an opportunity. Digitalization gives them an important asset: access to the customer.

With this approach, the book provides an important impetus to take an interdisciplinary approach to the challenges of digital transformation. This is also a concern of the Bavarian Research Institute for Digital Transformation (bidt), which is committed to advancing digitization in Germany for the benefit of the people. This means turning digitization, which for a long time seemed to be a purely technical matter, into a matter for everyone, because it affects everyone.

Bavarian Research Institute for Digital
Transformation (BIDT), Board
of Directors ISF
Munich, Germany
November 2020

Andreas Boes

Preface

The idea for this book came to me during a project report by a colleague. At a meeting of communications managers, he told me about his goal of predicting his clients' issues. The company he works for covers a wide range of consulting services. Clients expect high-quality input. It takes a lot of time and effort to create such content. But what if the elaborately produced content doesn't meet the client's interest at all? Disappointed authors, lost time and wasted money. That's why this colleague set out to identify the topics that mattered to his clients. More specifically, the goal was to predict which topics would be relevant to them in the future. After all, content production takes time. Big data and artificial intelligence should help to support this process. Data from social media channels and journalistic publications were used to identify relevant topic clusters and predict their development.

This question opened the view on a wide field of new searching movements around the telling of stories. Data plays a central role in this. In this case, data should provide information about customers' interests and needs. They were not the content, but the context of the message.

But data is often also the content of messages. Here, it is a matter of their preparation, usually in visual form – this is the classic field of data storytelling. In both cases, data plays an important role in gaining insights and developing messages.

They help us gain insights into people's needs, desires and interests. More and more even believe they can predict the future with data. That's a big promise. The process is similar to that of an oracle, where prophets and priests have dedicated themselves to foreseeing the future.

A second impulse came from the side of economists: stories have an influence on our behaviour and our economy. Robert Shiller expressed this thesis in his book *Narrative Economics*. The economist was one of the first in his profession to analyse the economic influence of grand narratives – of the gold standard, of the American dream, of the Great Depression. I was particularly fascinated by his approach of understanding the spread of stories as epidemics and applying mathematical modelling methods to them.

Then came the real pandemic, and we all witnessed the importance given to data in explaining the spread, and how this gave rise to a multitude of readings that competed for the prerogative of explanation and constantly evolved. Corona triggered two epidemics: the one caused by the virus itself and the epidemic of narratives about the virus.

That was a good breeding ground for my work. A book project is always a leap into the unknown. Of course, at the beginning, I had a specific idea of what it should be about and why. I used this to convince my publisher. But then new insights emerged during the research, each time combined with the question of whether they were important enough to change the original plan. The external developments caused by the pandemic have once again shifted some accents. To put it briefly, they have confirmed and sharpened my attitude: We are all looking at new infections, R-values, incidences, mortality rates and many other key figures. But more than the numbers, we hang on the lips of the interpreters of these data. What's really exciting about storytelling with data are the prophets and priests who interpret it, and thus move others to action. Thanks to digital media, we can analyse more than ever the networks that emerge through the interaction of these stories and in which we negotiate our view of the world.

In my writing, I have always noticed how much the historian in me comes through. So, I have often given in to the temptation to trace the emergence of developments such as data journalism, artificial intelligence and data ethics, in order to understand what the protagonists' motives

were and the reasons for their success. A fundamental principle of our discipline also shapes the approach in this book: the principle of source criticism. Data, in this sense, are sources. And with sources, a historian always asks himself a series of critical questions, above all: why does this source exist? What is its quality? What is its intention? This has proven to be a very useful attitude when looking at data sources.

Of course, the book is primarily shaped by my experience in the communications industry over the past 25 years: as a journalist, as a marketing manager, as a press officer, as an investor relations manager and now as a communications consultant. Perhaps these different roles already make it clear that I don't think too much of thinking in strict boundaries of disciplines such as advertising, marketing and PR. The digitalization of communication has done a lot to dissolve these boundaries and open up new fields of activity. Much more important than the boundaries is therefore the question of what we can learn from the others. This could be designers, data analysts or journalists, for example. To this end, it helps to have an understanding of the respective roles and to focus on synergies. With this in mind, the book is not written for any particular communications discipline, but rather to encourage us to look beyond our own discipline and learn from others. In view of the rapid, also technologically driven, change, I consider this indispensable.

I particularly enjoyed the excursions into other disciplines – such as computer science, epidemiology and psychology. Of course, I am not an expert in these subjects, and the experts from these disciplines may forgive me for that and for my mistakes. Rather, I was also concerned here with the question of what impulses we can take up for communication from these subjects.

The book is aimed at practitioners from the communication disciplines. Due to the large number of topics addressed, it does not provide concrete instructions for implementation, but gives impulses to deal with interdisciplinary questions and should ideally open up new perspectives for one's own field of activity.

Such a book is not possible without the experts who supported me with their expertise and gave me valuable feedback. My thanks go to these people who have supported me in word and deed while writing the book. These are in particular:

- Katharina Brunner, data journalist
- Lutz Klaus, Marketing ROI Experts
- Martin Szugat, data thinker

A special thank you goes to my family: With my sons I was able to deepen some technological questions. They have also supported me in collecting, or rather scraping, many a data series. Above all, however, I would like to thank my wife, who accompanied me throughout the entire process of writing this book and constantly encouraged me to sharpen the theses of my book.

I would like to express my sincere thanks to Imke Sander. As an editor at Springer Gabler Verlag, she supervised the editing of the work and brought it to completion.

For reasons of better readability, I have refrained from using the language forms masculine, feminine and diverse simultaneously in this book. All references to persons apply equally to all genders.

I hope you find my book stimulating and full of new insights.

Munich, Germany
November 2020

Hans-Wilhelm Eckert

Interactive Graphics

Graphics are an important component of storytelling with data. Digital channels offer additional features such as data download, mouseover effects, sorting options and linking. To help you take advantage of these, I've put together the appropriate graphics on a website:

www.data-storyteller.de



I indicate in the captions when an interactive version of the graphic is available.

» *pharmakon*, Greek: poison, drug, medicine

» “Sugar Man, you are the answer, that made my questions disappear” Rodriguez, “Sugar Man” from the album “Cold Fact”.

» “Data wins arguments” Tim Campos, Facebook CIO

Contents

1	The Pitch: What Is Storytelling with Data?	1
1.1	The Oracle Principle	1
1.2	Data Is the New Smoke	2
1.3	Prophets and Priests: Who Predicts the Future?	3
1.4	What Makes Storytelling with Data Special?	8
	References	12
2	Storytelling in the Digital Age	13
2.1	From Stone Age Caves to Echo Chambers	13
2.2	Stories, Narratives and Epidemics	18
	References	25
3	From the Question to the Data	27
3.1	The Brand: Why Am I Relevant?	30
3.1.1	Adidas and the Quantification Bias	30
3.1.2	Volkswagen: Brand Communication with Big Data	32
3.2	The Target Group: For Whom Am I Relevant?	34
3.2.1	An Online Shop Sharpens Its Customer Profile	35
3.2.2	A Fashion Retailer Reinvents Itself	37

3.3	The Topics: What Is Attracting?	40
3.3.1	What Advertising Does the Customer Want to See?	40
3.3.2	Listening to What Moves the User	41
3.3.3	A Viral History of Artificial Intelligence	47
3.3.4	Context and Change of Meanings	56
3.3.5	Setting Tomorrow's Topics Today	59
3.4	Where Silence Is Golden: Uplift Modelling	61
3.5	Towards a Data Strategy	64
	References	66
4	From Data to Story	69
4.1	Data Journalism: With Wikileaks to the Breakthrough	69
4.2	Visualization: Basics, Tools and Best Practice	75
4.3	Finding the Hero in the Data	86
4.4	Saving Lives with Data: From Cholera to Corona	92
4.5	Data Storytelling Is Teamwork	98
	References	102
5	Fair Play: What Counts in Data Stories	105
5.1	False Certainties and Deliberate Manipulation	106
5.2	Machine Bias	118
5.3	Ethical Issues of the Data Oracle	122
5.3.1	The Protection of Privacy	124
5.3.2	Algorithmic Accountability	127
5.3.3	Voluntary Commitments	128
5.3.4	Data Literacy	131
5.4	Less Is More: An Opportunity for Storytelling	133
	References	136



1

The Pitch: What Is Storytelling with Data?

Abstract Man is a seeker of meaning. With his stories he organizes the world, networks with others and constructs connections. Storytelling is so effective because it is deeply embedded in the structures of our brain. More and more of these stories today feed on data. Whether data is the new oil, gold, or even plutonium: Ultimately, it is we humans who ascribe these meanings to data. In storytelling, data has a dual role: it is the content and context of the message.

1.1 The Oracle Principle

“If you cross the Halys, you will destroy a great empire.” Such was the saying of the oracle of Delphi to Croesus, king of Lydia. In his day it was customary to consult an oracle before making great decisions. Previously, Croesus himself had tested seven oracles in a kind of benchmark, and only Pythia, the divining priestess of Delphi, had reliably provided the correct answers. So it was natural to trust Pythia’s answer in this case as well. Croesus then crossed the river Halys in 546 B.C. and attacked the Persian king Cyrus, who had moved ever closer to his kingdom. And in doing so, he actually destroyed a great empire: his own.

Delphi was considered the center of the world at that time. The oracle was the most important institution of the Hellenic culture. For more than a thousand years, the powerful sought advice from the oracle before making important decisions. The oracle was dedicated to the god Apollo, the god of light, purity, the arts and divination.

In direct contact with Apollo was the prophetess of the oracle, called Pythia. Only the priests had contact with her. Pythia sat in the holy of holies of the temple on a tripod above an opening in the floor, through which she could enter into contact with Apollo according to the view of that time. The sentences uttered by the prophetess Pythia were recorded, processed, interpreted and then given to the questioner as a prophecy (about the oracle of Delphi see Maaß 1996).

The oracle was known to provide answers that had multiple readings. The intention of the questioner was directed into the future. The answers, in turn, were often only revealed in retrospect, and until then they often misled the questioner. They sent the hero on a journey in the course of which he had to win battles and at the end of which, in the best case, he recognized himself. This was not always pleasant. Neither with Croesus nor with Pyrrhus or Oedipus. For that, the heroes' journey provided the dramas that make up Greek mythology. And not only them. The invention and telling of stories is a basic principle of our human existence. And the heroes' journey became the much-cited formula of storytelling.

The oracle principle forms the blueprint of this book: the art is not only to ask the right questions to an appropriate addressee, but rather to then interpret the answers received correctly. The span between the formulation of the question, i.e. the inner departure of a hero, and the insight, i.e. the end, is the drama of the journey that is at the core of what we now call storytelling.

1.2 Data Is the New Smoke

What has not already been attributed to data? "Data is the new oil" was the headline of the Economist in 2017, referring to the new driving force of the fourth industrial revolution (Economist 2017). "Data is the new plutonium," said Jim Balsillie, founder of Blackberry, referring to the

toxic content of the material from which all modern dystopias are fed (Balsillie 2018). “Data is the new gold,” says a study by Accenture, which believes that as democratization continues, data will be the central driver of growth for every company (Accenture 2020).

The thesis of this book is: data is the new smoke. The source of wisdom today is no longer Apollo, but it is data. And at the heart of the cult of data today are no longer prophets and priests, but people who know how to distill information from data and those who use it to make predictions and assert their claim to interpretation.

Data is the smoke that pours from the crevices of our oracles. Everything is contained in this smoke. Whether it is fuel, poison or gold, whether it harms or helps us, soobs or intoxicates us, that is up to us. Only the Pythia and the priests make predictions from it that give direction to our actions. They decide signal or noise. They weave the stories from it, they create what we call meaning from the data. And we all, as oracle receivers, have the freedom to draw our own conclusions from these oracle sayings and base our actions on them. Even though today we are surrendering more and more of this interpretive power to so-called intelligent machines, it is still us who have decided to do so. These machines have as much power over us as we are willing to give them. It is in our hands to do so. In that sense, this book is also an appeal to our own power of judgment.

1.3 Prophets and Priests: Who Predicts the Future?

I worked in the financial industry for many years, including several years in investment banking. During this time I was able to observe the system of companies, stock exchanges, banks and investors from the inside. A lot of money was made in the capital market for a long time and some still make it in these times. This led to the development of a division of labor around the question of how money will be worth more in the future, consisting of issuers, brokers, asset managers, analysts, private and institutional investors and the media, who fuel the system with their stories.

For the capital market, the data oracle is constitutive: CFOs, together with their investor relations departments, develop the story of their company in order to achieve the best possible price for their company shares on the markets. To interpret these stories, there are analysts on the sellers' and buyers' sides who formulate their own readings of these stories, at the end of which the oracle says: "Target price 435 euros". The negotiation of these stories takes place at the stock exchanges (and of course at many other trading places), where we can follow the outcome of the negotiation every second.¹ The big data reservoirs here are Bloomberg and Reuters, which pay dearly for access to the most precious data and therefore secure an information advantage for those who can afford it.

Similar structures can also be found in other industries. Anyone looking for current technology trends will not be able to avoid the analyst firm Gartner and its hype cycle (Fig. 1.1). The hype cycle depicts the phases of public attention through which a new technology passes: from

Gartner Hype Cycle

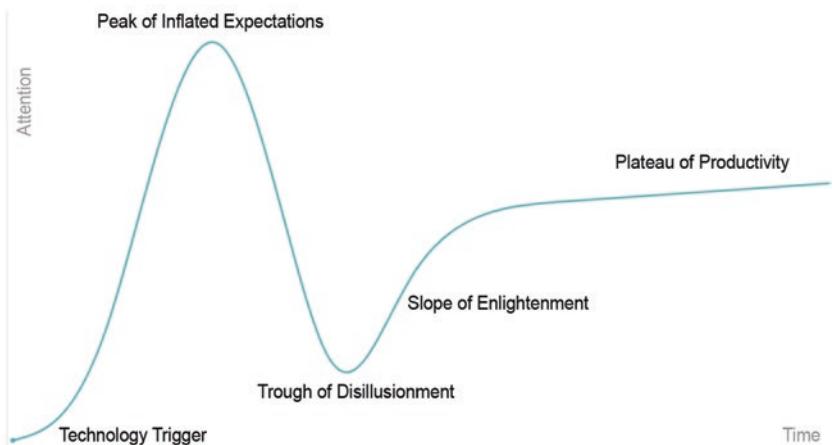


Fig. 1.1 The Gartner Hype Cycle: every year the market researcher updates its cycle of current technology trends. (Source: Gartner)

¹ Here I am only talking about the visible area. The flash boys of high-frequency trading are another issue that falls more into the realm of piracy.

its technology trigger, through the peak of inflated expectations, to the trough of disillusionment and finally to the plateau of productivity. The image of the hype cycle was coined by Gartner consultant Jackie Fenn and still forms the company's brand essence today. Every year, new assessments of current technologies are eagerly awaited: Where are new hypotheses emerging, which technology is on its way to the trough of disillusionment, and where are real productivity gains? A marketing machine driven by annual updates feeds off the basic narrative embodied by the curved curve of the hype cycle. It's hard to build a business model on a story better than that.

What Bloomberg and Reuters are to the capital markets, market research institutes such as Nielsen, Ipsos, IQVIA and Kantar are to the economy. They deliver relevant studies for almost every question. In the field of communications, the major market media studies should be mentioned here, which conduct representative research into media use and consumer behavior on the basis of large case numbers. Market researchers such as the Sinus Institute access this data material directly and offer individual counts and analyses. Evaluations are possible, for example, according to various markets such as nutrition, finance or travel, and offer the possibility of evaluating them according to socio-demographic characteristics and also according to the so-called Sinus Milieus.

The days of these oracles and many other business models based on this pattern are numbered. The sources of our narratives are and will remain data. But the way data is collected and distributed is about to undergo a massive change. The business model of analyst firms and market researchers is based on the fact that they have exclusive access to data. They have a pool of market data, company data, supplemented by primary and secondary surveys. The collection, processing and preparation of the data is costly, requires in-depth methodological knowledge and for many years secured the companies an interpretative edge on which their business models are based.

But the business model of these data giants is dissolving. The quality and quantity of the available data are completely different today. We have long since tapped into data sources that let us make decisions faster and find new insights. The internet, social media and the cloud make data available in real time. Company figures, for example, can be

automatically extracted from press releases, investor presentations and annual reports at the time of publication. Today, hardly any humans are needed for this. A well-programmed algorithm works faster, more precisely and around the clock.

Even more significant than the Internet itself is a piece of hardware: the first iPhone. In 2007, Apple launched the mother of all smartphones. At the time, perhaps not even Steve Jobs had any idea how much it would change our entire lives. Things we previously owned materially disappeared into it: first our address books, then our camera, music collection, wallet. Not only did they disappear into it, but they miraculously combined to create new things and generated new data and metadata, data for organizing and classifying data. WhatsApp doesn't know what we say, but it knows when, where and with whom we speak. The photos on our smartphone store time and place, which then reappear in our Time and Travel line as memories, as well as on Facebook, Instagram and Co. where we posted them. These and many other useful or even just nice features make the smartphone so powerful that we have adjusted our behavior to it and are busily generating even more data and metadata.

Social media and the movement data from our mobile devices are constantly providing new amounts of data. Taught by us humans – for example when “liking” a post or commenting on a picture – machines are now increasingly capable of understanding what we say.

In addition to the consumer market, industrial applications are flooding much more data onto the market or into the cloud: sensors in cars, robots and entire industrial plants generate many times more data and feed cloud systems in which it is provided and processed. Estimates from the industry analyst IDC from 2019 assume an almost unimaginable amount of over 40 billion networked devices in 2025 (IDC 2019).

If data is the source of our divination, then these devices provide the raw material that feeds our data oracles, their prophets and priests. They drive all modern business models, they create new oligopolies and make Facebook, Google, Amazon and Co. so valuable. What this means for the oracle market is that a new Delphi League owns and uses the knowledge of humanity: These are the GAFAs (Google, Amazon, Facebook and Apple) in the Western world and the BATXs (Baidu, Alibaba, Tencent and Xiaomi) in the Eastern world.

But data has long since ceased to be generated solely in the system of the major platforms. Therefore, companies are well advised to build up their own data sovereignty in order to develop a model and control communication activities from it. Market researcher Forrester predicts that brands that own their data will outperform those that merely obtain their data from third parties. To achieve this, it is important to make oneself independent of the large data networks and to feed knowledge about customers from one's own sources. Access to data has become increasingly easy, allowing any company to build its own oracle. Many companies have taken this step and created technological structures and built departments dedicated to developing questions, collecting, processing and interpreting data. In this way, they gain their own access to a valuable raw material and secure an edge over the competition.² Business Intelligence is the name of the discipline that deals with all forms of data analysis. And where it is already more about predicting the future, it is also referred to as business analytics.³

However, data in itself is not relevant. They only become valuable when we give them meaning. Today, we often do this in the belief that data is the source of objective truth. Wherever data are the sources of wisdom, there needs to be a Pythia who opens up access to these sources and a priest who interprets the words of the Pythia. In today's corporate world, multiple functions vie for this interpretive authority. Communications and marketing managers are not always ahead of the game. But their access to the customer gives them an advantage that they should use profitably. The better companies understand customer behavior and draw the right conclusions from it, the more loyal customers will be to the company in the long run. Therefore, they should have good access to the oracle.

² For example, brick-and-mortar retail has recognized how important a direct line to the customer is for business success. The Otto Group also makes its concentrated knowledge of 25 million CRM data in 150 segments available to third parties, see Zimmer (2019).

³ A good overview of these and related terms such as predictive analytics and predictive maintenance is provided by Mauerer (2017).

1.4 What Makes Storytelling with Data Special?

Storytelling is so effective because it is deeply embedded in the structures of our brain. It is based on our tendency to make decisions emotionally in order to justify them rationally afterwards. The psychologist Daniel Kahneman described the phenomenon in his book “Thinking, Fast and Slow” with many examples (Kahneman 2011). Storytelling approaches use this way of working of our brain and rely on the emotional appeal of the counterpart. They are based on the pattern of the hero’s journey, the drama and its resolution. This describes very well the formal side of the phenomenon. However, this approach focuses only on a few formats, namely reportage, commercials and other subjective stories and accounts in which such personalization can take place. But what about analyses, studies, commentaries news, infographics or posts? They too take up themes and explicitly or implicitly refer to stories and their heroes.

It therefore needs a broader view of the storytelling environment. The focus of storytelling lies on the second component of the word and explores the question of *how* we tell stories. Much has been written since Aristotle about the construction of stories and their structure, i.e. the formal side of the subject. I go into this in more detail in Sect. 1.2. The *what*, the content and its context, its development, transformation and interaction with other stories, is rarely addressed in storytelling.

The term “narrative”, borrowed from English, is now also used in German in a theory-heavy or at least meaning-heavy way, for example when it comes to the big topics such as nationalism, globalisation, the welfare state or growth. A buzzword, perhaps, but one with the claim to dock with postmodern narrative theory. More on this in Sect. 1.2.

For the purposes of our question and the rest of this book, I take narrative to mean the context of the story being told. This context is essential for our understanding of the world and the story we have just told. For man is a seeker of meaning. With his stories he organizes the world, networks with others and constructs contexts. Through stories he communicates knowledge and forms networks. These stories are interwoven, through them we interact. People with similar backgrounds, attitudes,

and desires create the primary resonant spaces of these stories. Depending on the mix, the stories are constantly changing and can go viral far beyond these resonant spaces. With such a network concept, the relevance and spread of stories can be mapped and interactions can be tracked. Via digital channels, their infection paths become traceable.

So when we speak of stories here, we mean the specifically told story. Narratives,⁴ on the other hand, are the “larger” stories that create meaning and order and are agreed upon and shared by social groups. They provide the context of the specific story.

Based on this consideration of narratives (the macro-level with the context of the story told) and storytelling (the micro-level with the formal structure of the story told), the vertical dimension of the playing field emerges. On the left is the context level, on the right the actual narrative level.

The second, horizontal dimension is provided by the data themselves: The way it is prepared differs in whether it has an explanatory or exploratory function. If the data is prepared in such a way that patterns become apparent, but no solution is presented yet, the task of exploring and drawing conclusions lies with the viewer. Dashboards are built this way. They provide patterns from which experts can derive explanations. Visual presentations of complex processes, the visualization of user flows or the color clustering of medical images are also based on these methods. Exploratory processing is the preliminary stage for the explanatory processing of data. In this book, I will use the examples of cholera and covid to show how exploratory and narrative processing differ (see Sect. 1.4).

When data provides explanations, it has a narrative structure. This is the real playing field of data storytelling – outlined in orange here.

- **Process of cognition:** Explaining and exploring differ in who does the work – sender or receiver. In exploring, the receiver does the work of

⁴The term “narratives” may have come into German usage through Lyotard from the French “recit” via English, cf. Heine (2016): “The Oxford English Dictionary explicitly names Lyotard as the originator of the latest English meaning of narrative, defining it: ‘a narrative or account used to explain or justify a society or historical period’. One is therefore probably not wrong in assuming that the conjuncture of the noun narrative in German goes back to the influence of English.”

interpreting and making sense, in explaining, the sender provides the insight.

- **Narrative level:** On which level does the narrative take place? Does it provide context or is it itself the story? In terms of data: Does the data provide insights into which stories are relevant or is it itself the subject?
- **Data provides the narrative:** by segmenting customers, identifying needs and themes, developing brand stories. This forms the strategic basis for all stories that are told. Data provides the context for drafting and spreading stories. It provides us with deep insights into the world of our fellow human beings, their needs, expectations and desires.
- **Data delivers the story:** Data is the content of the story. This is the domain of data journalism and all journalistically data-inspired formats such as content marketing and social media campaigns, but also analyst reports. Visualizations are the guiding discipline of these stories.
- **Data provides the insight:** Humans are masters of pattern recognition. Visualizations help to recognize and analyze correlations. Data provides the patterns, the human or/and an AI interprets them and derives an insight. This can form the basis for a narrative and a story.
- **Data is the issue:** recognizing anomalies, monitoring systems: visually prepared data enables the control of processes, workflows and states. Dashboards, i.e. overviews of data categories considered relevant, are structured according to this principle.

But as soon as data provide explanations, they have a narrative structure. This is the real playing field of data storytelling (Fig. 1.2). With the processing, analysis, interpretation and embedding in the context of existing narratives, we humans are able to tell relevant, activating stories from them and to develop contexts of meaning.

The journey in this book begins with a look at the literature on storytelling and narratives. The focus here is on the question of what digitalization has actually changed and what remains as communication patterns (Chap. 2).

The topic of Chap. 3 “From the Question to the Data” uses examples to show the contribution that data makes to the development of narratives: Since we are talking about corporate communication, the journey starts with the brand. The brand embodies the personality of a company

Storytelling with Data: the Playing Field

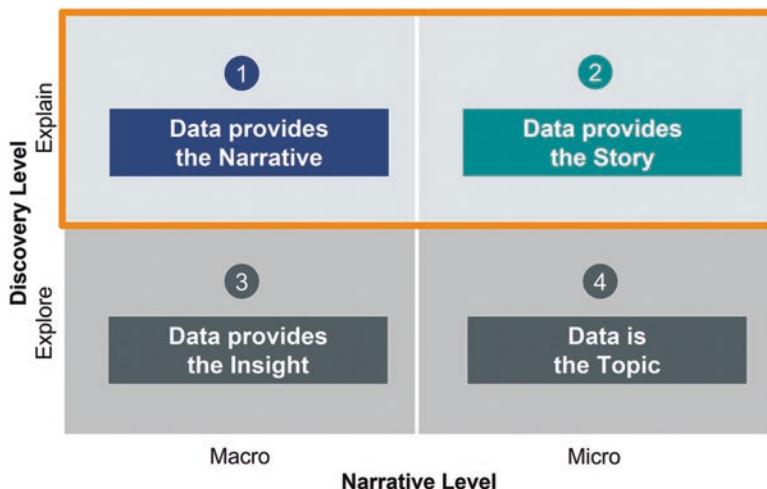


Fig. 1.2 Data storytelling – the playing field. (Source: Own illustration)

or product and ideally answers the “why?”. As there is no brand without a market, in the next step we look at how customers can be classified (“For whom?”) and with which topics they can be addressed (“How?”).

Once the context is clarified, it is a matter of developing the specific stories. This is the focus of Chap. 4 “From Data to Story”. What can we learn from data journalism, how can stories be visualized and what are the roles in the oracle team?

Every game needs rules. Chap. 5 “Fair Play: What Counts in Data Stories” deals with the ethical framework and the sovereign handling of data: Around false certainties, deliberate manipulations, and what data literacy means for storytelling.

The more the raw material of our forecasts consists of data, the more important it is that we not only ask the right questions, but that we are also aware of what assumptions are contained in our question. Otherwise, in the frenzy of oversupply, we fall into the illusion that the answer is already in the data. The Croesus case shows how prone to error we humans are when it comes to interpreting information. And this doesn’t just affect those who base their actions on the oracle. Even the oracle itself can err if it is not aware of the nature of its sources.

References

- Accenture (2020) The human impact of data literacy. Accenture. https://www.accenture.com/_acnmedia/PDF-115/Accenture-Human-Impact-Data-Literacy-Latest.pdf. Accessed 11 Juni 2020
- Balsillie J (2018) Data is not the new oil – it's the new plutonium. Financial Post, 18. Mai. <https://business.financialpost.com/technology/jim-balsillie-data-is-not-the-new-oil-its-the-new-plutonium>. Accessed 22 Jan. 2020
- Economist (2017) The world's most valuable resource is no longer oil, but data (06.05.17). <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Accessed 26 Nov. 2019
- Heine M (2016) Hinz und Kunz schwafeln heutzutage vom „Narrativ“. Die Welt, 13. November. <https://www.welt.de/debatte/kommentare/article159450529/Hinz-und-Kunz-schwafeln-heutzutage-vom-Narrativ.html>. Accessed 4 Juli 2020
- IDC (2019) The Growth in Connected IoT Devices Is Expected to Generate 79.4ZB of Data in 2025, in IDC 18.06.2019. <https://www.idc.com/getdoc.jsp?containerId=prUS45213219>. Accessed 2. Mai 2020
- Kahneman D (2011) Thinking, fast and slow. Penguin, London
- Maaß M (1996) Delphi – Orakel am Nabel der Welt. Thorbecke, Sigmaringen
- Mauerer J (2017) Was ist was bei predictive analytics? Computerwoche, 11. Dezember. <https://www.computerwoche.de/a/was-ist-was-bei-predictive-analytics,3098583,5>. Accessed 28 Mai 2020
- Zimmer F (2019) Daten, zu mir! – Wie Marken sich datentechnisch emanzipieren Dmexco, 8. November. <https://dmexco.com/de/stories/daten-zu-mir-wie-marken-sich-datentechnisch-emanzipieren/>. Accessed 25 Jan. 2020



2

Storytelling in the Digital Age

Abstract Stories shape our view of the world. Innovations such as printing, newspapers, radio, television and now digital channels have made it possible to spread stories and connect humanity on a global scale. Increasingly today, we are putting control over the selection and spreading stories into the hands of machines whose actions are fed by data. We are also unlocking new content through data: Through it, we are able to venture into regions where us humans lack an awareness. We recognize connections and patterns that we would not have seen without this data. By evaluating and contextualizing this data through language, we expand our world.

2.1 From Stone Age Caves to Echo Chambers

Somewhat human, somewhat animal: Six creatures with ropes and spears are chasing a huge bovine. With their dark red pigments on the bare wall, the paintings look a bit like graffiti. They are the oldest hunting scenes in the world in the cave Leang Bulu Sipong 4 in the southwest of the Indonesian island of Sulawesi. They tell an ancient story of hunting and

killing. Comparable depictions were also created somewhat later in Europe in the cave Cueva de El Castillo near Puente Viesgo in Cantabria, Spain.

What makes these paintings so significant is the artists' ability to create things that do not exist (Callaway 2019). Whoever immortalized himself or herself there: He or she left behind one of the first records of human creativity. The paintings in the caves are the first known sign that our ancestors began to develop ideas and communicate about them. Even if today we do not know what our ancestors told each other about it. Researchers assume that language developed at the same time as painting (Kuke 2012).

Storytelling became a need like only food, sleep, sex, or companionship. "Storytelling, however, was not a pastime, it was training. It allowed people to theorize about what the other might be up to, who was good friends with whom, or who might just be putting on an act." (Siefer 2015).

Storytelling is as old as humanity. In a community it was conducive to be cooperative and helpful, to be able to read emotions and react to them. This helped people to stand their ground and stay on top of things. The means to do this was through stories. When our ancestors began to talk around the campfire about things that existed only in their imaginations, they began to network and organize.

Storytelling has been deeply inscribed in the structures of our brain in the course of evolution. Neuroscientists have studied how stories work in our brains and have come to impressive conclusions. Uri Hasson vividly explained the effects of stories on our brain in a TED Talk (Hasson 2016).

- Neurophysiologically, stories are "cinema in the head": they activate different areas of our brain. In addition to the language centre, which is located in Broca's and Wernicke's areas in the cerebral cortex, areas in the insular cortex (cortex insula) are also activated, an area that is responsible for feelings of empathy, pain and pleasure.
- Stories stimulate our brain to release hormones that are important for feeling stress or empathy, for example.
- With storytelling, stories synchronize the brains of the teller and the listener. This is how we appropriate stories and take them for our own experiences.

- Because of the powerful effects on our brains, stories can be remembered up to 22 times better than pure facts, says Jerome Bruner, Harvard professor and co-founder of cognitive psychology in his work “Actual Minds, Possible Worlds” (Bruner 1986).

This is how patterns developed around the campfires of our ancestors and shaped our narrative structures – from the first cave pictures to posts on Instagram. Images were likely the foundation of oral narratives. Subsequently, it was always new media that people used to extend the reach of their stories. After the introduction of writing, it was the invention of the printing press that made the book the first mass medium. In his country, Martin Luther helped the technology to achieve a breakthrough by translating the New Testament into German in 1521, giving a large audience immediate access to the Scriptures.

On the basis of language, all the great narratives that define our lives have emerged: Religions, philosophies, ideologies and political theories. Even in the so-called exact sciences, the social consensus of a group of scientists determines which view prevails. In his 1968 analysis of the structure of scientific revolutions, Thomas Kuhn showed how in science the wrong is not replaced by the right, but only one agreement is replaced by another, and coined the term paradigm shift for this.

After all, all the institutions that shape our world have also developed from these narratives. States, churches, and schools, for example, have emerged from stories that people share, negotiate, agree upon, and on the basis of which they eventually build the organizational structures. In this way, we have created our powerful systems that define our world. Whether it is science, religion, politics or economics, in the end it is all about asserting the interpretive sovereignty of social groups and thus exercising power. And where power is concerned, manipulation is not far away.

Media play a central role in this. Innovations such as printing, newspapers, radio, television and now digital channels have contributed decisively to the emergence of systems of order and the networking of humanity. The more powerful a medium became, the stronger the possibility of manipulation.

One certainly significant difference from analogue times is that today we have put the control for selecting and distributing stories into the

Echo Chamber - Filter Bubble

Percentage of search terms (n-gram) (*1E9) in Google Ngram Corpus German 2019

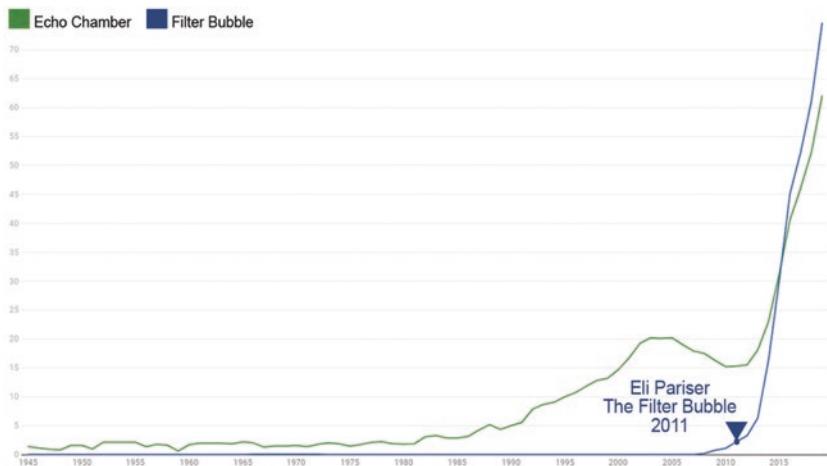


Fig. 2.1 The term filter bubble was popularized by internet activist Eli Pariser (2011) and surpassed the term echo chamber in late 2015. Interactive graphic at www.data-storyteller.de. (Source: Google Ngram Corpus German 2019)

hands of machines. Algorithms decide what we see, bots spread stories. A shadow industry floods the world with fake news¹ and so-called deep-fakes make it harder to identify fakes today. But most importantly, much of the channel is in the hands of a very small number of companies. This is worrying.

Regarding the growing power of Google and Facebook, the Internet activist Eli Pariser (2011) expressed his criticism with the image of the filter bubble (Fig. 2.1). The large Internet corporations would exploit knowledge about their users to bring together those who have similar interests. A filter effect controlled by algorithms leads to my Facebook timeline, for example, only showing me posts that confirm my own opinion. In this way, Pariser's allegation goes, the algorithms of internet

¹ See, for example, the great report by Samanth Subramanian on the city of Veles in Macedonia, which has become the epicenter of the fake news industry in the Trump election campaign. On the Russian disinformation industry, see the website of the EU East StratCom Task Force with a weekly newsletter on the latest fake news at <https://euvdisinfo.eu/>

corporations further reinforce opinions in so-called filter bubbles and thus further fuel extreme positions.² In the discussion triggered by it, his thesis was in particular attacked on two points: On the one hand, manipulation was not as pronounced as Pariser implied; on the other hand, given the flood of information, a filtering function of all media was unavoidable.³ An internal Facebook investigation, recently reported by the Wall Street Journal, proves that Facebook was well aware of how much its algorithms contributed to the polarization of its users. The findings were inconsequential, however, and the study ended up in the company's archives (Horvitz and Seetharaman 2020).

Furthermore, the accusation remains that the user himself has no influence on what is fed to him by the algorithm and has no insight into how this decision was made.

The related theory of echo chambers is also based on the assumption that interaction with like-minded people in social networks leads to a fragmentation and narrowing of the world view. However, it is broader in scope and does not focus solely on the influence of algorithms. It is based on a bias we call confirmation bias. In psychology, this describes the tendency to select and interpret information in such a way that it corresponds to one's own wishes.⁴

The confirmation error, by the way, also sealed the fate of Croesus mentioned at the beginning. The message of the oracle was: "You will destroy a great empire." In his wishful thinking Croesus perceived only one possibility of interpretation, namely, that it must be the empire of his adversary. That it might be his own empire he did not consider. Yet the oracle itself warned of the effects of his sayings. At the entrance of the temple there is said to have been an inscription, "Know thyself." Perhaps Croesus should have thought about this inscription before he crossed the Halys.

The multiplicity of digital channels allows us to avoid self-awareness and surround ourselves with the people (or machines, it is not always

² Here I simply assume that Facebook is also a medium, even if Marc Zuckerberg stubbornly denies this. On the filter bubble, see Pariser (2011).

³ See Seemann (2018) for the current debate.

⁴ The theory for this comes from Peter Wason and Philip Johnson-Laird (1968).

certain) who share our opinions. In this sense, the Internet has only multiplied the caves of our ancestors. The basic pattern of communities forming by inventing and spreading stories hasn't changed for 40,000 years. It's just that today we have the freedom to change caves if we don't like the narrative that prevails there.

2.2 Stories, Narratives and Epidemics

The literature around storytelling focuses on a recurring pattern: the hero's journey. It is explored in countless facets.

In essence, they can be reduced to a few essential elements:

The Hero: A single character is the protagonist of the story. This personalization provides the reader with the opportunity to identify with him. From there, the connection to a larger narrative develops. The journey of the hero, who leaves his familiar surroundings and sets out into a new, unknown world. Here, various stations await him: adventures, challenges, and tests of endurance that he grows from – physically or even as a personality. Until he finally has to pass the decisive test. Once he has successfully completed this, the reward awaits him. From there he returns home as a new person, is celebrated and reports on the change (Campbell 2008, pp. 151–206).

The essence of any good story is based on such pattern. But a hero rarely comes alone. It takes personnel to animate a story. This is where the antagonist comes in: the anti-hero, the villain, or however the role is fleshed out. The stronger the antagonists, the greater the fall, the more drama the story offers. The Beauty and the Beast, Dr. Jekyll and Mr. Hyde, Faust and Mephisto. All the other variations can be derived from this basic constellation: Allies and traitors, converts and renegades. The most important ingredient of a good story is the resolution of a conflict.

This has been described extensively in the literature. Anyone who wants to sharpen their pen as a screenwriter, for example, will not be able to avoid the standard works by Joseph Campbell and Robert McKee. Campbell spans millennia of cultures and discovers recurring narrative

patterns (what he calls monomyths) that can be found in Greek mythology, Siegmund Freud's interpretation of dreams, and Carl Gustav Jung's teaching of archetypes, among others. George Lucas was inspired by Campbell's work to develop his Star Wars saga. Robert McKee also focuses on the big themes and introduces the art of screenwriting for the Hollywood stuff. For a heroic epic, such an approach certainly works well. To captivate my audience for 21 h and 48 min (that's how long all Star Wars episodes combined last), I can explore many facets of my hero's struggle against evil. This is rarely the case in corporate communications, where the issues are usually not big enough to dock onto the basic questions of life. As readable and inspiring as these works are: they are of limited use as guides to our concerns. Other works are more helpful for communicators: "Storytelling" by Petra Sammer and "Think Content" by Miriam Löffler provide very good basics on storytelling for companies (Sammer 2017; Löffler 2014).

That a picture is worth a thousand words is a common stereotype. Our eye is a master of pattern recognition. This already helped our ancestors, for example when they distinguished edible from poisonous mushrooms or discovered the outline of a sabre-toothed tiger in the savannah in time. Our optic nerve forms the shortest and fastest path of all sensory organs to the brain. That's why visualizations are so effective. And so it is not surprising that a whole branch of storytelling deals with it. For example, the book "Visual Storytelling" by Pia Kleine Wieskamp, of which, however, only a small part of the book also deals with infographics, i.e. the data-based part. Nancy Duarte uses data storytelling in the title of her new book. Unfortunately, the result is only a guide on how to create trenchant presentations for the busy CEO (Kleine Wieskamp 2019; Duarte 2019). A systematic treatment of storytelling with data is provided by Brent Dykes in his 2020 work Effective Data Storytelling. He precisely analyzes the specifics of data stories and makes clear in which form the narrative structures of a hero's journey can be implemented here as well. He sensitizes to perceptual distortions and shows which visualizations are suitable for data stories. It is the most profound work on the subject, which is why we will encounter it again and again in this book. At this point, it is also worth looking at developments in data journalism, which has taken off with the processing of big data from WikiLeaks and

has not only set standards for the analysis and processing of big data, but can also serve as a template for collaboration between interdisciplinary teams. More on this in Chap. 3.

Humans are pattern seekers, says Daniel Kahneman in his book quoted earlier. We believe in a coherent world in which all things make sense and events connect causally. People recognize patterns even in things that are statistically random (Kahneman 2011, pp. 114–118). Storytelling describes the formal side of this pattern recognition. By inventing and telling stories, we are able to structure and interpret our world and neutrally couple ourselves with other people. In a sense, stories are the data packets in which the myriad of impressions pelting us are filtered, compressed, and connected to meaning. Thanks to powerful media, we can use these stories to synchronize very large numbers of people in their actions over very large distances. According to this understanding, storytelling is nothing more than data processing for people.

But the approach still falls short for the following consideration. It needs a look at the narratives, the content and context of the many stories. For only those who have sovereignty over the interpretation of the stories also have power over the people.

The oracle provides the paradigm here. Today there is no longer a clear centre like the one at Delphi, which provides us with a binding interpretation and an organising principle for all stories. The world appears fragmented. In our post-industrial society, the certainties of modernity have also been destroyed. The idea of the Enlightenment, with its goal of the self-liberation of the individual, has proven to be a mistake. This is how Jean-Francois Lyotard, a French philosopher and literary theorist, put it in his 1979 work *La Condition Postmoderne*. The narratives of modernity, he argues, began with the claim to possess a scientific legitimacy. Lyotard declares this narrative to have failed. Our postmodern world is characterized by the coexistence of different narratives, which Lyotard calls discourses. These discourses follow their own rules as isolated language games between which there is no understanding and no interaction (Lyotard 1979).

Still believing in the triumph of liberalism as the dominant interpretation of the world, the American political scientist Francis Fukuyama argued in his 1992 book *The End of History* that the world had a predetermined path and that economic and political liberalism had won the

day. With the fall of the Berlin Wall and the collapse of the Soviet Union, the last grand narrative competing with the Western world – Marxism – had effectively failed. Western democracy and liberalism had prevailed and thus achieved the goal of history (Fukuyama 1992). This reading proved to be a fatal error with the 2001 attacks on the World Trade Center and the Pentagon.

The American military already saw things differently in the 1990s. The acronym VUCA was used to describe a world that had lost its old certainties and was characterised by volatility, uncertainty, complexity and ambiguity. The concept was developed at the United States Army War College (USAWC) and initially served to describe the multilateral world after the end of the Cold War and to develop strategies for asymmetric warfare. Today, the term plays an important role in the teachings of strategic leadership and organizational theory, particularly in the context of digitalization and agile leadership (for its origins, see U.S. Army Heritage and Education Centre 2019). Most importantly, the concept is another expression of a fragmented mishmash of narratives in which many oracles compete for interpretive authority.

Stories do not even need to be true in order to have an effect. What is decisive is that a community shares and spreads these stories (and thus also conceals other stories). It is by no means new that this involves deliberate falsification in order to put things in the right light. In the face of some excited fake news, it helps to remember that stories have always been invented to achieve goals. Fakes were also deliberately commissioned in the past, for example to prove ownership claims or to legitimise ruling houses. The Privilegium Maius of 1358/1359, for example, is one of the most skilful document forgeries of the Middle Ages. With its help, the Habsburgs legitimized the state right to dominate the Austrian lands. These forgeries and manipulations have left their mark on human history. Historian Eric Hobsbawm, in an edited volume, has called these narratives the “Invention of Tradition,” noting that many so-called traditions that we think have endured for a long time have been recently invented (Hobsbawm and Ranger 1983). And another historian, Yuval Noah Harari, refers to man as a post-factual being who has only been driven, forming his networks, and gaining power over the world by believing in fictions (Harari 2018, pp. 231–244).

There are no longer any binding narratives in our society. Our knowledge consists of a multiplicity of narratives that coexist. To this extent, postmodern Lyotard's ideas seem to coincide with the masterminds of the American military academy and the media theorists with their echo chambers. But rather than persisting in isolated resonant spaces, some scholars view narratives as large, interactive systems. Stories would interact with each other as if on a large playing field – reinforcing each other, fighting each other, and constantly modifying each other. That's the model economist Robert Shiller develops in his book *Narrative Economics*. He shows that these stories actually also have real influence on our actions and thus on economic development – an observation neglected by economists for far too long. Shiller demonstrates the effect of various popular narratives on economic development in the US.

Shiller has described their enforcement in terms of epidemics: The triumphant march of a story behaves much like the course of an Ebola infection (The book is written before Corona). The infection curve initially rises as more people get infected than recover or die. This process reverses once the epidemic has passed its peak, i.e. the number of new infections decreases relative to those who die or recover. Overall, the hump-shaped course typical of epidemics is thus formed. According to Shiller, the same applies to the rate of infection of narratives. Here, similar to viruses, there are narratives that are more infectious than others. "Similarly, with narrative epidemics there may be two different narratives, one with some minor story details that make it more contagious than the other. The minor story details make the first narrative, and not the second, into an epidemic" (Shiller 2019, pp. 18–21, quoting p. 21). Shiller examines the impact of these stories using common economic narratives – such as the gold standard, the Great Depression and Bitcoin.

Its truthfulness plays no role in the enforcement of the story: "Ultimately, a story's contagion rate is unaffected by its underlying truth. A contagious story is one that quickly grabs the attention of and makes an impression on another person, whether that story is true or not" (Shiller 2019, p. 96). Stories unfold their impact when they appear coherent and help to create meaning by linking up with other stories. In many cases, stories are in circulation long before they become infectious. This is well illustrated by the American Dream narrative, for example. First

circulated in James Truslow Adams' 1931 bestseller *The Epic of America*, Shiller sees the narrative as an example of a slow epidemic that is still relevant and growing today. From Adams's perspective, the American Dream was the dream of a social order in which everyone would find a place according to his or her ability that would offer recognition and livelihood, regardless of birth or social position. Adams thus refers to the American Declaration of Independence and the principle of equality formulated there.

But for the narrative of the American Dream to go viral, it needed more triggers. One of them was Martin Luther King's speech in 1963. Here we see the main three building blocks that helped to significantly increase the infection rate:

- a famous protagonist: in this case in the person of the civil rights activist Martin Luther King
- a large audience: his speech "I have a Dream" on 28 August 1963 in front of the Lincoln Memorial in Washington
- a massive conflict: racial discrimination and the demand for equality of blacks and whites

It is only in this constellation that the story goes viral and provides the impetus for the further infection cycle of the American Dream narrative, which continues into the present. Thus, this narrative is also a very good example of how long infection cycles can last (Shiller 2019, pp. 151–154).

Taking this approach further, one of the most important principles of storytelling is to form topics and messages in such a way that they can dock onto already familiar themes and messages. Ideas, topics and attitudes shared by a group of people promote social cohesion. Thus, narratives form something like the social glue of a community.

In times of the Corona pandemic, however, some communication managers will ask themselves whether it is still an appropriate metaphor to infect people with stories. The asymmetrical communication model with the clear distribution of roles between sender (communication driver) and receiver ((target) customer) has become obsolete at the latest with the triumph of social media platforms. It has long been accepted in communication science that the roles of the receiver of content have

shifted to the creators, distributors and commentators of content. Through digital channels, we now also have the data to trace the emergence of stories, their changes and paths of infection. Such networks emerge when connections are made between social actors (“nodes”) – which can be individuals or organizations. The collections of these connections can be condensed into patterns or network structures that describe the interaction of the system. It also provides a model for the stories that remain in the resonant space of our echo chambers without going viral. These chambers, on the other hand, are not closed systems, but rather incubators where new mutations emerge and then find their way out.

So if a virus now goes world-spirit, will virologists become the new priests of our oracles? That's exactly what Shiller said. And there's a lot to be said for giving them an important place on the interpretive team. After all, to predict the spread of infection, we need to understand how people behave collectively. To do that, you need the skills to understand how infections are transmitted. But it also involves analyzing data and developing mathematical models. This is more of an interdisciplinary task, in which the engineering sciences also have a lot to contribute, for example when it comes to mathematical theories, in particular chaos theory and statistical modelling. Together with other representatives from disciplines such as mathematics and bioinformatics, they can develop the tools to recognise patterns for which humans have no sense of their own and therefore make use of the machine. And it needs the interpreter to place these patterns in a larger context and make a coherent story from them.

If one conclusion can already be drawn from the Corona pandemic, it is that data will play an even greater role in our daily lives in the future:

- The current crisis is giving digital communication a huge boost. This means that much more data is being created in the industry's systems, especially by companies whose business model is based on this form of data collection.
- Data is currently helping to improve our understanding of how the virus is spreading. We have a real-time global pandemic development laboratory that is providing us with an unprecedented volume of data.

- Data is the basis of the surveillance systems that will enable us to control, predict and manage our behaviour even better in the future. Many of these practices already exist and are being used in the current crisis to at least (and hopefully only) temporarily suspend fundamental liberties in order to slow or prevent the spread of the virus.

Thus, the pandemic will further strengthen the role of data in understanding our world. Enormous sums of money are flowing into their collection and processing, new oracles are developing and with them new priest boxes for interpreting all these new volumes of data, which are constantly providing new material for interpretation.

In doing so, we simultaneously broaden and narrow what constitutes our world. The philosopher Ludwig Wittgenstein once said, “The limits of my language are the limits of my world.” The stories woven from language make up this world, with its networks, nodes and echo chambers. By means of data, we are able to advance into regions for which we humans have no sense of our own. We see connections and patterns that we would not have seen without this data. By evaluating and contextualizing this data through language, we expand our world. At the same time, however, we run the risk of narrowing this world to that which exists through data. We lose sight of the things that cannot be measured and quantified. This is quite helpful to keep in mind when we ascribe so much interpretive power to data.

References

- Bruner J (1986) Actual minds, possible worlds. Harvard University Press, Cambridge
- Callaway E (2019) Is this cave painting humanity's oldest story? Nature. <https://www.nature.com/articles/d41586-019-03826-4>. Accessed: 13. Dez. 2019
- Campbell J (2008) The hero with a thousand faces, 3. Aufl. New World Library, Novato
- Duarte N (2019) Data story. Explain data and inspire action through story. IdeaPress, Washington, DC
- Dykes B (2020) Effective data-storytelling. Wiley, Hoboken
- Fukuyama F (1992) The end of history and the last man. Free Press, New York

- Harari YN (2018) 21 lessons for the 21st century. Jonathan Cape, London
- Hasson U (2016) This is your brain on communication, Youtube 03.06.2016.
<https://www.youtube.com/watch?v=FDhlOovaGrI>. Accessed: 25. Nov. 2019
- Hobsbawm E, Ranger T (1983) The invention of tradition. Cambridge University Press, New York
- Horvitz J, Seetharaman D (2020) Facebook executives shut down efforts to make the site less divisive. Wall Street Journal. <https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>. Accessed: 1. Juni 2020
- Kahneman D (2011) Thinking, fast and slow. Penguin, London
- Kleine Wieskamp P (2019) Visual Storytelling im Business: – mit Bildern auf den Punkt kommen. Hanser, München
- Kuke U (2012) Wer schuf die Zeichen von El Castillo? Die Welt, 2. Ocktober.
<https://www.welt.de/kultur/history/article108596340/Wer-schuf-die-Zeichen-von-El-Castillo.html>. Accessed: 25. Nov. 2019
- Löffler M (2014) Think content! Rheinwerk, Bonn
- Lyotard J-F (1979) La Condition Postmoderne. Les Edition de Minuit, Paris
- Pariser E (2011) The filter bubble. What the internet is hiding from you. Penguin, London
- Sammer P (2017) Storytelling. O'Reilly, Heidelberg
- Seemann M (2018) Filterbubbles, ja, nein, doch, gerne, oder – Stand der Debatte PIQD, 6. September. <https://www.piqd.de/technologie-gesellschaft/filterbubbles-ja-nein-doch-gerne-oder-stand-der-debatte>. Accessed: 2. Mai 2020
- Shiller R (2019) Narrative economics, how stories go viral and drive major economic events. Princeton University Press, Princeton
- Siefer W (2015) Wer erzählt, der überlebt Die Zeit, 23. Dezember. <https://www.zeit.de/2015/52/geschichten-maerchen-christentum-mythos-legende-ueberleben>. Accessed: 23. Jan. 2020
- U.S. Army Heritage and Education Centre (2019) Who first originated the term VUCA (volatility, uncertainty, complexity and ambiguity)? USAHEC, 7. Mai. <https://usawc.libanswers.com/faq/84869>. Accessed: 29. Dez. 2019
- Wason P, Johnson-Laird P (1968) Thinking and reasoning. Penguin, Harmondsworth



3

From the Question to the Data

Abstract Data sources flow so abundantly that when starting any project, it's important to ask precise questions. In this chapter, I show approaches to using data to contextualize stories and help define which audience I'm targeting, which themes I'm setting, and why they're relevant in the first place. The examples illustrate the challenges of developing the question and in each case provide specific answers to key communication issues: Why we tell something, who we tell it to, and what we tell.

To enter the land of milk and honey, the aspirant must first eat his way through a wall of dough. Once inside, he finds himself in a place where food and drink abound. Wine flows from overturned jugs straight into his mouth, pancakes fall from the sky, a roasted pig equipped with a knife runs around as a mobile snack. This is how Pieter Brueghel the Elder depicted the popular paradise in the sixteenth century.

The path to data sources is somewhat reminiscent of the farmer's path to the roast. An oversupply of food meets a limited stomach. In Breughel's painting, the peasant, like his fellow sufferers, has capitulated and surrendered to sleep. He is no longer even able to enjoy all the good things.

To avoid a similar fate when processing data, we need a clear goal and a lot of discipline to avoid succumbing to the obvious temptations.

There are so many ways to tap into data sources that a precise question is important at the start of every project. Only on this basis does it make sense to develop your own data strategy for the project.

Companies usually offer a multitude of possible sources that can be tapped. For communication topics, these are first of all the core systems in which customer data is stored. Ideally, it's stored centrally in a customer relationship system (CRM) and is also used and maintained by several units of the company such as communications, sales and service. In addition, all communication and marketing systems such as mailing programs, websites, shop systems and apps are of course relevant for the collection of data. So-called enterprise resource planning (ERP) systems provide valuable sources, for example, about flows of goods and purchasing behavior. In the meantime, more and more companies also have their own so-called business intelligence, in which the data from the various internal and external sources can be stored, analyzed and played back in a central instance.

The advantage of self-collected data is obvious: You have direct insight into its quality and significance, know how it was collected and thus gain access to important contextual information for its interpretation. And you can access it at any time.

For most questions, however, companies are also dependent on external sources. For communication and marketing topics, the channels on which the company is active and in contact with its target groups are ideal. Countless data sources can be tapped from media response analyses, advertising impact studies and search engine evaluations as well as from listening on social media channels, i.e. social listening. They are further supplemented by market media studies and other offerings from market research firms. The interpretation of this data is usually more complex because information about the method of collection, its quality and significance must first be obtained and evaluated.

Regardless of where the data come from: They are not themselves the subject, but in this case they provide the narrative, i.e. the context of the stories we tell (see Fig. 3.1 and the comments on this in Sect. 1.4).

Storytelling with Data - the Context of Stories

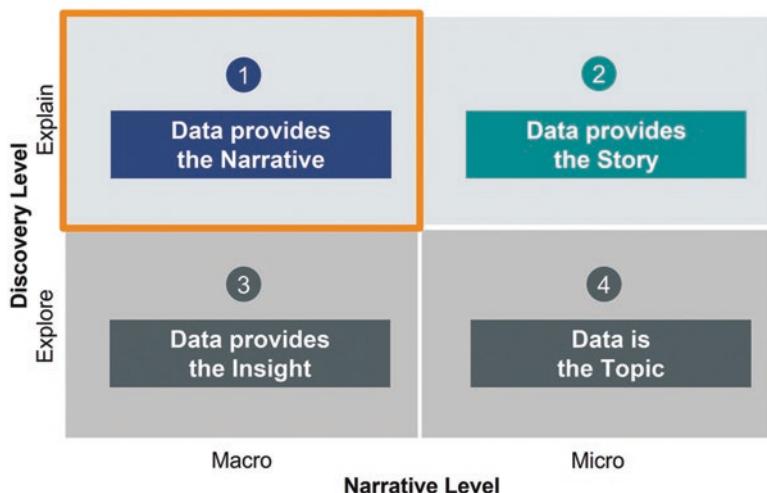


Fig. 3.1 Storytelling with data- the context of stories: In this chapter, data is not itself the subject, but provides the context of the stories we tell. (Source: Own representation)

In this chapter, I present selected examples of approaches that use data to define storytelling frameworks. These examples show the challenges in developing the question and each provide specific answers to the central communication issues:

- Why we tell a story: The brand as an anchor of the narrative can be better grasped in the perception of the target groups and the effect of the communication measures.
- Who do we tell: the target groups, their needs, expectations and wishes can be analysed.
- What we tell: We find new stories in the data. And we discover ways to connect them to the topics of our target groups.

3.1 The Brand: Why Am I Relevant?

3.1.1 Adidas and the Quantification Bias

A failure of Google's advertising server in Latin America in 2017 brought Adidas surprising insights. After all, a large part of the advertising budget there ran via Google AdWords and was intended to boost sales of sneakers and sportswear. So it was natural to assume that when the servers went down, so would sales. But what happened was: nothing. Although Adidas could no longer place paid advertising, both traffic and sales remained constant.

The server failure brought Adidas a salutary insight: the model with which the advertising budget had been distributed until then was wrong. Far too much of it was allocated to so-called performance marketing, i.e. primarily to paid search engine and banner advertising for specific products. The good performance and measurability of these advertising media and the possibility of tracking them in the shop system had led Adidas to place a special emphasis on this in its marketing. The server failure, however, made it clear that the company had overestimated this influence.

Adidas global head of media Simon Peel used the lion's den of his guild, the annual performance marketers' gathering in London, EFF Week 2019, to admit the mistake: "We over-invested in digital advertising". The focus on return on investment (ROI), he said, had led the company to over-invest in performance-based advertising. Many of Adidas' products hit an oversaturated market where they were sold through discounts, making customers increasingly price-sensitive. There was no long-term brand orientation. As a result, of the total budget, only 23% went to the brand, while performance marketing took 77% of the advertising budget.

This budget was directly tied to e-commerce sales. Behind this was the belief that digital ads drive sales. Adidas was keen to drive online sales, as this was the most profitable part of the business. "We assumed it was digital advertising – desktop and mobile – that was driving those sales, and as a result we were over-investing in that area," Peel said. The insight led Adidas to introduce an econometric model that brought together brand, sales and customer structure. Until then, the company had

believed that loyal customers drove sales. But the results from this model revealed that 60% of revenue came from first-time buyers.

Adidas also had to abandon another basic assumption: Namely, the belief that business units only drove their own sales, so soccer advertising, for example, would drive sales of soccer boots. In reality, all advertising paid into the sale of Adidas products, i.e. the soccer boot also paid into the jogging outfit.

After all, Adidas' old model was based on tunnel vision that looked at each channel in isolation. Little was invested in video because it didn't perform particularly well in last-click metrics. But a closer look showed that TV, out-of-home and cinema were also paying into e-commerce. The new model highlighted that brand activity was responsible for 65% of wholesale, retail and e-commerce sales, while performance marketing conversely also drove wholesale and retail sales.

So it was time to shift the allocation of the advertising budget significantly towards the brand. Adidas used the four years following the server outage to develop a new model and approach for it. With its marketing playbook – called 'Creating the New' – Adidas introduced a new campaign framework in 2019 that was focused on emotional brand activation. Three to four major campaigns per year are expected to recharge the brand in the future (Peel 2019).

This shift from marketing efficiency to marketing effectiveness made waves. The industry discussed little else in the aftermath. Peel's speech was a kind of wake-up call, in the wake of which other advertisers announced that they would rethink their performance share of their marketing mix, including Booking Holdings ([Booking.com](#), Priceline, Kayak), Tripadvisor or the Old Navy brand from the Gap Group (Rentz 2019).

Adidas is a good example of the quantification bias that many companies succumb to. It's just too tempting to surrender to tunnel vision and focus on digital advertising efforts and here especially the last click. This data is readily available in large numbers. It can be easily tracked, segmented and analysed, making it easy for the communications manager to make a numbers-driven controller and perhaps many a CFO happy by explaining to them how to derive impact relationships, ROIs and the budgets required. It is more difficult to get a view of long-term

developments of a brand and cross-channel effects. Such a model is costly to develop and needs a common understanding of the entire management on the impact correlations. As a rule, the figures for this have to be collected first.

But it's worth taking this route and not falling into the quantification trap. By investing their brand(s), companies build a web of narratives that helps them stabilize their sales (see Adidas) and also builds them a buffer for times of crisis. General Electric, for example, lost around 60% of its share value during its downturn from 2017 to 2019, but only 40% of its brand value ([Economist 2020](#)). The fact that it can also go the other way and a decision can destroy brand capital that has been painstakingly acquired over decades is again shown by the example of Adidas. With the decision not to pay rents for the stores during the Corona crisis, Kasper Rorsted incurred the wrath of many loyal fans of the brand. Even though Adidas reversed this decision after the storm of indignation: the damage was there and can probably only be repaired in the very long term. Adidas is now at the forefront of companies that only look out for the good of shareholders and shirk their responsibility to society, especially in times that call for solidarity. This is not a good story for a brand that wants to connect people through sport.

3.1.2 Volkswagen: Brand Communication with Big Data

The question of why a brand exists can only be answered by the company itself. The power it needs for this positioning comes from the culture, the shared experience and the ideas about future development. On the other hand, to balance the internal and external views, it needs sound market research. Amazon founder Jeff Bezos summed it up as follows: "A brand is what people say about you when you're not in the room." Only with a clear picture of how a brand is perceived by its target groups can it be determined what steps are needed to bring the target closer to the actual. So in Bezos' words, matching what people say about me when I'm not in the room with how I want to be seen as a company. That forms the basis of all brand management.

Data is essential when it comes to the perception of a brand. The example of Volkswagen shows how big data can support this. In the time before the diesel scandal, brand perception at Volkswagen was characterized by values such as discipline, precision and reliability. A sober, objective, one could say engineer-driven positioning with a penchant for clearance. The brand's reputation has suffered badly as a result of the diesel scandal. Its credibility was heavily affected after the interventions in the exhaust gas purification systems. In order to reposition itself among the relevant target groups, Volkswagen has taken an unusual route.

The company began with a new accentuation of the core values. Out of the field of factual values around the concept of discipline and into the field of emotional values around the topics of experience and joy. Here, the focus was now on virtues of being innovative, valuable, responsible, credible, fair and reliable. In order to enliven them with a new story, Volkswagen specifically sought insights into the expectations and attitudes of the target groups. This is usually done by interviewing focus groups and online panels. Here, Volkswagen gave the country managers creative freedom. In Poland, for example, Volkswagen relied on the "Flash AI" Big Data platform from NeuroFlash. This is a kind of semantic database that makes it possible to discover and weight associations of words. In this way, attitudes of target groups can be determined without interviewing them. To this end, NeuroFlash has compiled articles from online sources such as Wikipedia, mass media and social media channels and evaluated them using machine learning methods to identify and weight word meanings and associations.

Volkswagen used the platform to identify characters that fit the brand particularly well. Here, two top occupational groups were found through semantic analysis: the doctor and the teacher. As an intersection of these two characters, the two core ideas from which the campaign ideas were developed emerged: "schoolchildren" and "first aid". The "Little Heroes" campaign focused on teaching first aid to school children and their families (see Fig. 3.2).

The result: a more than fourfold higher interaction rate compared to the successful T-Roc/Tuareg campaigns on Facebook. Brand-relevant indicators such as brand consideration were also clearly addressed. In addition, the campaign still paid into Volkswagen's CSR perception (Mall 2020).

Repositioning the Brand with Big Data and AI

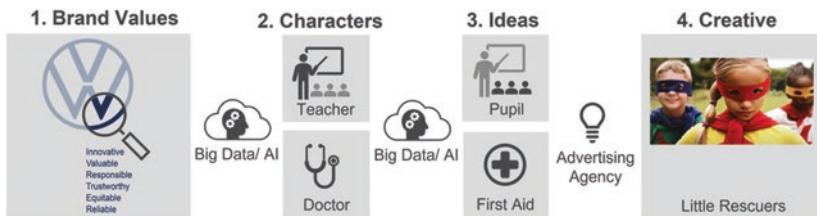


Fig. 3.2 Using a semantic database, characters were identified for VW that fit the brand particularly well. From this, the two core ideas crystallized, from which the campaigns were developed

The example shows: When developing a brand message, the internal view is not a good guide. It is important to translate the “Why?” into the language of the target groups and to connect with their narratives. New methods such as the semantic analysis of words and the identification of association fields on the basis of big data provide important impulses here, can be implemented faster and with less effort than classic market research and will continue to gain relevance.

3.2 The Target Group: For Whom Am I Relevant?

KYC is not a new fast food chain, but stands for Know Your Customer. In the data age, knowledge about the wishes, needs and situation of customers is increasingly becoming the most important success factor for companies. This is because we have very powerful sources through digital channels that provide valuable insights into customer attitudes, expectations and desires. Those who use them correctly are ahead of the game. This can be done in very different ways. Bottom-up or top-down, based on available data or based on a strategic decision. But best of all based on a combination of both.

3.2.1 An Online Shop Sharpens Its Customer Profile

The shoe retailer Seven Feet Apart from England has chosen a bottom-up approach to optimize its target group appeal (www.sevenfeetapart.com). The company sells high-quality shoes exclusively through its online shop. Two years after its launch in 2016, the company had collected a lot of data that was suitable for analyzing and clustering its customers and allowed it to review and develop its original premises. The company used data from website analytics, in this case Google Analytics, data from its Shopify storefront system, and data from its Dotmailer email program. This was supplemented by customer demographic data and – very important for the mail order business – return rates. On this basis, the company developed a data platform that helped it understand and segment its customers and offer them the right shoes at the right time.

Using the clusters of Experian, a leading market research company from England, Seven Feet Apart segmented its own customer groups. In this way, customer personas could be developed via the Experian tool. Personas are personalized, typical representatives of a particular segment. They can be used to gain deeper insights into attitudes, expectations, desires, values and, in particular, purchasing behavior.

With the help of the personas, the company was then able to model the target customers as a projection of the existing customers. The basis for this was the identification of the most valuable customers, which were determined on the basis of purchase volume, frequency and the time of the last purchase. To expand this customer group, Seven Feet Apart focused on finding and targeting statistical twins. This was done by analyzing the engagement of the corresponding customers on the channels and their preferred devices. This allowed the questions to be answered:

- How do most valuable customers differ from everyone else?
- How are the most valuable customers acquired and retained?
- How do these customers shop?
- What triggers a purchase?

Among other things, the company found out that the age of this core target group is not between 35 and 45, but between 45 and 60. This led to some changes in the approach: the selection of models in the campaigns, the product range and its presentation as well as a stronger focus on values in general and on sustainability in particular. For example, the product presentation of the shoes was supplemented by story elements, i.e. alternative entrances combined via colours, themes or occasions. With the 100 Square Feet project, the company also addresses the sustainability issue: with every purchase of a shoe, Seven Feet Apart donates to the World Land Trust in order to preserve 100 square feet of rainforest.

Another important insight was that the most valuable customers were also the wealthiest. This led the company to evaluate its pricing strategy. The fact that price had no great impact on conversion rates and that customers were willing to accept higher prices confirmed this move.

The analysis and classification of customers has brought more profitability to Seven Feet Apart. Above all, it created the prerequisites to properly address this profitable customer group, to select the appropriate products and to set the right topics (Aaron 2019).

Communication and marketing units always have a potential hero, the customer in all his facets. The persona concept provides an important approach to identifying these customers, their needs and desires. For the development of stories on the basis of personas, data is the foundation: It flows in from the various contact points – the purchase history, the visit to the website, the opening behavior of newsletters, the contacts with sales and service employees, some of them possibly bundled in a CRM or BI system. In this way, it is possible to address age groups with particularly high purchasing power, their topics and wishes. These can be supplemented by external sources that provide information about the interests and relevant topics of these personas.

On this basis, the dramaturgy can then be developed for each individual persona. The journey of this persona, the customer journey, is a narrative concept that is fed by data. This dramaturgy can be orchestrated along the so-called touchpoints. Data helps to understand at which point this persona is and which impulse he needs next. In fact, he should analyze exactly which typical journeys there are for his offer and at which points he loses particularly many travelers.

Such customer journey analyses quickly become very complex. Here, a core competence of our brain can be used: We are masters of pattern recognition. And the fastest way to recognize patterns is through our sense of sight. This helps us to analyze complex relationships. For example, in the question of what the customer's journey to the company looks like. Such an analysis brings together an enormous amount of data – from a wide variety of sources, such as ad tracking, the website, social media channels, the shop system, the email system. By visualizing this data, it is possible to see where most customers come from, through which channels they are reached and which path they take to purchase and beyond. Especially in such complex contexts, humans are sometimes superior to machines.¹

3.2.2 A Fashion Retailer Reinvents Itself

If there is more than one distribution channel, it becomes much more complex. Multichannel is the term retailers use to describe their claim to meet customers on the exact channel they are using. Whether they are browsing through the catalog or scrolling through their Instagram timeline, it is important to set exactly the right impulse to lead them in the right direction on their journey until they finally place a product from the retailer in their shopping cart and order it. However, the change from a goods economy to customer-oriented thinking also requires the consistent realignment of all processes in the company, a good data strategy on the basis of which the paths of the customer can be tracked.

Multichannel is a lofty claim that is best summed up this way: The easier the purchase is for the customer, the more complex the sale is for the company. After all, it's not just about managing the many channels so that they address the right person at the right time with the message tailored to them. In most cases, the data behind them is in different silos:

¹A good example of this is the analysis of the customer journey of applicants of a personnel service provider. With the visualization methods of the process mining tools, it was not possible to understand the behavior patterns of the applicants. This was finally achieved by visualizing detailed processes. On this basis, not only the data specialists, but also many employees from the most diverse departments were then able to analyze the processes and created the basis for a broad and fruitful discussion about optimization potentials: Münster (o. J.).

knowledge about the goods (usually in the ERP systems) and knowledge about the customers (usually in the CRM systems) are not networked with each other; moreover, data from online and offline channels is often held in different systems. A change of the customer from one medium to another can currently still be tracked well with cookies on online channels. But how do I bring together the activities from the online shop with the retail stores or the catalogue? How do I manage to bring together a uniform view of goods and customers in such a way that I can draw the right conclusions from them and give appropriate impulses in real time?

It is clear that retail companies, as soon as they have more than one distribution channel, face particular challenges here due to the abundance of data and systems involved. The example of fashion retailer Atelier Goldner Schnitt shows how the right course can be set. For the longest time, the company had only one sales channel: That was a high-quality catalog of fabric samples that customers had to return with their order in a specially designed envelope. The company's brand promise: fashion in a perfect fit for an older target group. The average customer was 83 years old, and the business model was very stable for a long time.

But over time, the expectations and experiences of this target group also changed, and new competitors came along who used digital sales channels and put pressure on earnings. This made a reorientation necessary.

This reorientation was essentially based on a rejuvenation of the target group and the introduction of two sharply defined customer profiles. In addition to the traditional regular customer, a further, somewhat younger customer profile was developed. In addition to the age of the target customers (83 vs. 68 years), these profiles also differ in terms of their fashion tastes, their online affinity (9% vs. 50% online use) and the cancellation rate (low vs. higher).

These customer profiles were not only mapped in the brand world, but in the entire organization – from purchasing to the assortment to the customer approach. The new customer group also brought other contact points into play: above all the web shop and e-mail communication. An important point of the entire reorientation was the data organization.

This was realized by a data model centered on the customer: All contacts online and offline are stored with the customer and thus provide a holistic view of the relationship. This is a great challenge, especially for companies from the stationary trade. This is because the web shop is usually developed at some point as a second channel alongside the stationary activities. With a separate customer and inventory management and a completely separate approach, you then have two silos and nothing gained.

Atelier Goldner Schnitt avoided this mistake by focusing on a strict customer centricity from the very beginning of the reorientation. The CRM system quickly reached its limits. The challenge was to map the increased complexity and to draw the right conclusions from it as automatically as possible: to integrate all information from the contact points, including outbound measures such as calls, mailings, newsletters, and orders from the catalog and the shop, and to derive customer loyalty activities and impulses for customer development (first to second purchase) from this.

This was realized through an interaction of a so-called scoring engine (Gpredictive), a marketing automation tool (Cross-Engage) and a business intelligence platform (incuda). The BI platform provides the aggregated data about customers and products, the scoring engine calculates with the use of machine learning continuously and automatically what the future sales of the customers look like. The marketing automation solution in turn controls the corresponding actions in the appropriate channels.

The basic decision was of utmost importance: the entire data budget was aligned with the female customers. With the introduction of two distinct profiles, the company is able to gather valuable information about the behavioral patterns of its female customers. This means that the company is now in a position to develop a precise approach for the two groups of female customers – across all channels. On this basis, two fashion worlds could be created and transported on the corresponding communication channels (Anton 2019).

3.3 The Topics: What Is Attracting?

3.3.1 What Advertising Does the Customer Want to See?

John Wanamaker, the so-called department store king and father of modern advertising, is reported to have once said, “I know half my advertising is money thrown out. I just don’t know which half.” Since then, an entire discipline has been devoted to the study of these two halves: so-called advertising pretesting. It aims to improve the ratio in favor of effective advertising. Because ad space costs a lot of money. And it’s better to invest some of it in testing procedures so that you can then advertise more effectively. As a rule, this involves testing an advertisement or campaign in a kind of laboratory situation before it is placed, to determine whether it can meet the requirements placed on it or whether and in what form it can be optimised (for more on this, see Trommsdorff and Becker 2009).

But in the meantime, there are also automated processes that determine the probability of success of an advertising spot on the basis of specially developed metrics. The soft drinks giant Coca-Cola, for example, increases advertising effectiveness using its own “One Number Score”. This is a metric that can be used to indicate whether creations are achieving strong business results and appealing to consumers. Spots have to land in the top quarter of market researcher Kantar’s creativity database to ultimately make it to market.

This was originally a big shock to the system – to the ad agencies and the marketers who weren’t used to testing. It was also important to educate the internal brand managers and their agencies about how the pre-testing system works and to get them involved so that they have an indicator of how the spot might work before production.

Now, however, Coca-Cola is also automatically assessing the likelihood of success of new spots on the basis of a large database with over 30,000 videos. The technology relies on facial recognition methods and uses neuroscientific approaches to screen out the best spots. In this way, the company shifts pretesting to the machine and makes the results from one country transferable to other countries (see Kantar 2020).

Other methods are also being developed here. For example, the company Aiconix AI wants to recognize at which point a viewer of a film breaks off and automatically cuts the films together so that the viewer continues watching. This type of optimization should lead to films having a higher acceptance and effectiveness on social media channels.²

3.3.2 Listening to What Moves the User

Where the most questions are asked, that's where the most knowledge arises. The priestesses of the Oracle of Delphi already knew that. Success fueled the business model, which at some point profited from its own fame and provided the best answers thanks to the many questions asked from all sides. Today's search giants, led by Google and Amazon, are no different. While Google "only" monetizes knowledge in ads, Amazon also has the opportunity to let its users generate ideas for new products and add them to its portfolio. So the platforms themselves benefit the most. But companies can also use this knowledge via various tools to keep an eye on relevant topics and to know what the audience wants to read.

Even in the more consulting-intensive B2B business, digital channels such as Internet searches and videos account for almost 60% of information procurement. This was revealed by a survey of 2745 sales managers conducted by Google and the consulting firm Roland Berger in 2015 (Roland Berger 2015, p. 6). Given the growth of digital channels, it can be assumed that this figure has, if anything, risen even further in recent years. Here, it is a matter of attracting attention in the early phases of the customer journey in order to become relevant to the potential buyer in the first place.

Since in the early stages there is usually still little knowledge about a concrete solution or product, search movements are more focused on advice, problem solving, product comparisons and evaluations.

Take the purchase of a mattress, for example. This happens relatively rarely in most households. Hardly anyone will therefore type the name of a manufacturer into the search engine line at the beginning of their

²The use case is still in development, talk Eugen Groß, AI in Marketing, IHK Munich 28.01.2020 and <https://www.aiconix.ai/anwendungen/>

search. A strategy based on a strong brand will not be relevant here for the acquisition of new customers, but rather for the retention of existing customers and their activation as referrers. Any provider looking to expand their reach beyond their existing visitor and customer base will therefore be on the lookout for topics that may be relevant to a potential customer. The aim here will be to provide answers to more general questions, such as: How low in toxins is a mattress? How can a mattress be cleaned? How often should it be turned? Do I need two mattresses for a double bed? But a mattress can also solve a specific problem: e.g. How does the mattress relieve my back pain? How does it support a peaceful sleep?

Search queries provide an insight into what moves the potential customer. In order to approach the right topics here, the data that users enter into the search engines helps. Many tools have emerged around the major search engine operators that help to understand the interests and language of users. Keyword databases help find the right search terms – they provide related terms and longer phrases. They can be used to discover trends and cycles. They provide information on how certain topics develop over time – whether they fluctuate seasonally, decrease or increase. Although search volumes are often highly rounded and search queries never provide a complete picture of all possible queries, they are a valuable and powerful tool for identifying topics and trends.

In addition to the potential customer, you also need to look at the competition. Which competitor is doing a particularly good job? This can be seen from two things:

- Backlinks
- SEO relevance

Backlinks: Indicate how many links point to your own website or that of a competitor. They provide an insight into the quality and relevance of the competition and of course also information about which pages are worth linking to. Link databases provide the material to analyze the link strategies of competitors. The providers of these databases use them to build their own web index. This has a certain subjectivity, which results from two things: On the one hand, the link data are not always complete

and do not provide a complete or up-to-date picture of the situation. On the other hand, they are only ever approximations of the search engine operators' algorithms. Nobody knows how Bing and Google really count.

SEO relevance: Once your own topics and keywords have been identified, they should also be analysed in comparison to your competitors. The tools provide an important indication of how visible one's own website is compared to competitors. Specifically, they answer the questions:

- Where does a competitor rank in front?
- How do these positions per keyword develop over time?

This allows conclusions to be drawn about content strategies and also the “quality” of content (from the point of view of the search engines and their algorithms, not from the point of view of the people who read the content!) Even if the tools do not capture all rankings and are rather difficult in niche markets, they provide important indicators for the visibility of one's own content strategy in the competitive environment.

If you have identified the relevant backlinks and developed your own strategy, where your own backlinks are desired, a **link building** offers itself. This is an important part of the so-called content seeding. The aim of the procedure is to intercept potential customers where they already are by spreading links to one's own content. Tools help to target websites for specific topics that want to link to or mention a piece of content. Here, there are a number of tools that help identify and categorize relevant search terms (see Fig. 3.3).

Which tool is the right one depends above all on the specific needs of the user – and of course also on the price he is willing to pay. There is a whole range here – for the beginner and for the professional, for the website of a start-up and that of a corporation, free of charge or with a three-digit monthly fee: free tools such as Seobility, Seorch and Ubersuggest have a limited range of functions, others are specialized in certain topics: Depending on whether keyword rankings, backlink or competitor research, the choice will fall on one of the tools. Since the keyword indices are always based on the quality of the provider data, they also differ according to topic, industry and region.

Search Engine Tools



Fig. 3.3 Search queries provide an insight into the interests of potential customers. The search engine operators' data can be evaluated according to different criteria using various tools. Interactive graphic at www.data-storyteller.de. (Source: Own representation)

Almost all of them make it possible to monitor competitors and thus determine at any time who is ahead with which search term. This provides valuable information about where there is room for improvement in the search term strategy.

In addition, some tools also offer semantic methods that can be used to identify not only search terms, but entire topic areas. In this way, content managers receive important information about which topics interact with other topics and where synergies may arise.

Of course, not only search queries are decisive in determining the topics of the target groups. Everything that is written, filmed, commented on and redistributed plays a role here. Television, radio, print media, the Internet and social media form a huge pool of sources for these types of analysis. Here, press and blog articles, radio, TV broadcasts and video contributions, social media posts, likes, comments and shared content are included in the evaluation.

This gives communication and marketing managers valuable ideas for discovering trends, identifying media representatives and influencers, and finding suitable channels for disseminating their own topics (see Fig. 3.4).



Fig. 3.4 Content marketing tools can be used to discover trends, identify influencers and find suitable channels for disseminating one's own topics. Interactive graphic at www.data-storyteller.de. (Source: Own representation)

Among other things, the tools help to discover topic areas, identify channels and find multipliers such as bloggers, influencers and press representatives through whom content can be further disseminated. They make it possible to compare oneself with competitors, to gauge the mood for topics, to plan one's own content and to distribute it on social media channels. Many also have integrated monitoring, which can be used to monitor key figures such as click rates, engagement and follower growth.

There are major differences in the depth and breadth with which the sources are integrated. As the overview shows, there is a focus on social media in all tools. Other channels such as print and radio/TV are only integrated by a few providers. This is also a cost issue, as the preparation of these sources is costly. Here, companies like Argus, Echobot and Unicepta, which come from media resonance analysis, are ahead. Google's Ngram Viewer occupies a special position in the list. It covers a much longer period than all the other tools and is based on the so-called Ngram Corpus, digitized books from five centuries up to the year 2019.

In general, the same applies to all tools: It is worth taking a very close look and asking exactly which sources are tapped in the categories mentioned. Here, the tools differ considerably. Also, not all social media is the same: There are differences in the channels monitored. And within the channels, the range is enormous. No tool offers comprehensive access. This is not even technically possible, as the interfaces do not allow this. So only selected, publicly accessible posts can be extracted. In order to be able to evaluate these, the providers have to provide appropriate storage. Storage space is cheap, but the amount of data is enormous. Therefore, all providers only have selected content over a certain period of time. What these are depends primarily on how much money a provider puts into the storage solutions and who the main users of the system are. So the users' willingness to pay and their areas of interest have a big impact on the sources available. Companies should take a close look at which focal points are stored there and what options there are for incorporating individual requirements.

Here, too, the following applies: There is a wide range and it is worthwhile to think about the specific use case in advance before you start looking for the right tool. Because it makes a difference whether I want to address a B2B or B2C market, search for the right influencer or

primarily produce content myself, which channels I use to address my customers, how I plan, produce and play out my content and according to which key figures my communication is controlled and the success is measured.

3.3.3 A Viral History of Artificial Intelligence

When it comes to processing data, methods from the field of artificial intelligence (AI) are being used more and more frequently. It has long been impossible to imagine our everyday life without these technologies. The recognition of language and images, for example, is not possible without the use of AI. But the methods are still used in many other applications and are currently being tested in new fields. In this book, we encounter these technologies in several places and see examples of how AI supports the discovery, telling, and interpretation of stories.

At this point, however, I will show how the topic area around AI itself has developed as a narrative and gone viral. For this purpose, Google provides the Ngram Viewer and Trends, two tools for a practical and freely accessible evaluation. With Trends, Google presents an aggregated overview of its search queries. There, various search terms and topic areas can be entered, combined and displayed in a graphical progression. The tool thus provides valuable indicators for discovering relevant trends – for the period from 2004 to today.

A much longer period is covered by the Ngram Viewer, which also offers a visualization of keywords. These are based on the so-called Ngram Corpus, which are digitized books from five centuries in eight languages, representing 6% of all books ever published, according to Google (Lin et al. 2012). The time frame ranges from the year 1500 to 2019, making it a good tool to observe longer trends. Unlike search queries, mentions on Ngram show the result of engagement with the topics covered in the books. They are thus slightly later indicators than a search query, but that doesn't matter as much in the long run. Over a 15-year period – between 2004, when trends began to be recorded, and 2019, when Ngram's record ends – the two tools overlap. Unfortunately, you look in vain for absolute numbers from both Ngram and Trends. They only provide percentage

values, so that no statement can be made about how large the actual amount of keywords or search terms is.

In order to enter the matching terms, Ngram requires some prior knowledge. Thus, it is useful to identify word fields, associations and related terms in order to find matching tracks. This can also be done using tools that evaluate association fields and semantics. However, none of them has a data history going back to the previous century. In this case, we would miss out on the term “cybernetics” (CY), for example, with which the US mathematician Norbert Wiener laid the foundations of AI in the 1940s.

If you search the English corpus of Ngram for the terms “Artificial Intelligence” (AI), “Cybernetics” (CY), “Machine Learning” (ML) and “Neural Network” (NN), you will see the picture in Fig. 3.5.³ We see here four different progressions of an infection, to speak with Shiller. These four progressions can be used to tell the story of the development of artificial intelligence.

A first, flat curve is formed by the term CY, much steeper curves by the two terms AI with a peak in 1989 and NN, the latter offset by about ten years in 1997. ML, on the other hand, slowly but steadily establishes itself from about the 1980s in the slipstream of the two steep curves and overtakes the term CY in terms of frequency of mentions at the beginning of the new millennium.

The term CY goes back – as already mentioned – to Norbert Wiener. Since 1943, the US-American mathematician – spurred on by the entry of the USA into the war – thought about how the behaviour of fighter pilots could be predicted in order to be able to shoot them down better. His 1948 work “Cybernetics – Or Control and Communication in the Animal and the Machine” laid many of the foundations of AI. Most importantly, he formulated the idea that there was no essential difference between humans and machines. This is a breakthrough for the development of the later idea that the human brain is nothing but “wetware” and thus follows the same rules as a computer with hardware and software. In his words, “In fact, the whole mechanist-vitalist controversy has been

³ “Big Data” doesn’t yield relevant results in Ngram until 2007, while “Algorithm” yields too many that have nothing to do with AI.

Artificial Intelligence: On the Way to a New Hype?

Percentage of search terms (n-gram) (*1E7) in Google Ngram Corpus English 2019

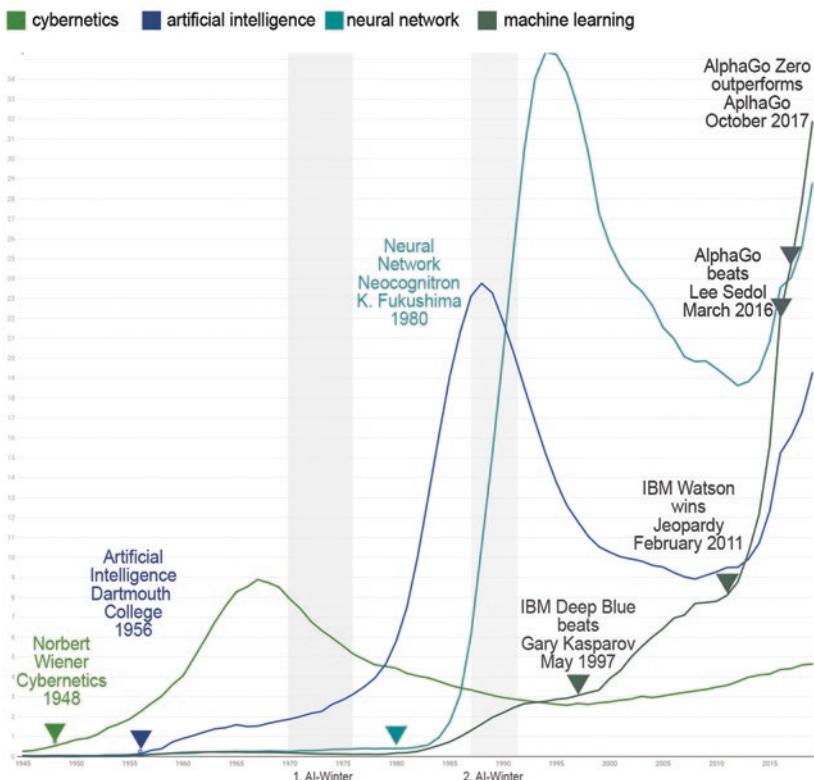


Fig. 3.5 The field of artificial intelligence is developing in several waves. Since 2011, it has started another hype cycle, Interactive graphics at www.data-storyteller.de. (Source: Google Ngram Corpus English 2019)

relegated to the limbo of badly posed questions" (Wiener 1948, p. 44). But if the human brain and a computer function in the same way, the precondition is also created for seeing man and machine in a fruitful and, in some eyes, fearful synthesis, as reflected, for example, in the ideas of transhumanism (see, for example, Lordick 2016).

Through the musician John Cage, Wiener came to John Brockman, the initiator of the Serpentine Marathons in London, a dinner series in

which he gathered the leading minds of his time (Kreye 2018, p. 23). Today he would be called an influencer. Brockman made Wiener and his ideas famous, as can be seen in the curve that peaked after Wiener's death in 1964.

The term AI first appeared in the mid-1950s. It was invented by John McCarthy, a US logician. At the time, he needed a powerful term for a grant application to the Rockefeller Foundation for a conference in Dartmouth the following year (McCarthy et al. 1955). Wiener had already laid the groundwork for the man-machine narrative, but the neologism AI condensed this narrative into a much more powerful term compared to CY. The concept won over the Rockefeller Foundation, and the grant application was approved. Thus, in the summer of 1956, Dartmouth College became the birthplace of AI. The term took off for an initial success story in the 1970s, then finally took off steeply in the 1980s. Wiener was not invited, his cybernetic concept was ignored by the initiators of the conference.

In the years from 1970 to 1975, public and private funding for AI research declined, investments were cut and start-ups in the field received less support. This phase is referred to as the first so-called AI winter. Strong, exaggerated expectations led to a "trough of disillusionment" of the hype cycle in those years (see on Gartner and the hype cycle Sect. 1.2). Interestingly, the first so-called AI winter only led to a flattening of the CY curve, while the term AI continued to be popular in the literature – even if it had not yet reached the level of CY. Research and publications on the topic of AI, on the other hand, continued to grow.

Then, in the mid-1970s, a practical use returned to research with expert systems and also gave the term AI a further boost. AI then reached its first peak towards the end of the 1980s. After expectations of expert systems also proved to be overblown, a second AI winter dawned in 1987, which lasted into the nineties.

A new concept gained the attention of researchers in these years: the concept of neural networks, which had already been developed in the 1940s, was decisively advanced by the Japanese computer scientist Kunihiko Fukushima: in 1975, the concept of the cognitron emerged, which Fukushima expanded in 1980 as the neocognitron, thus laying the foundations for a so-called Deep Convolutional Neural Network, which

was used for the recognition of handwriting and other visual patterns (Fukushima and Miyake 1982). With this concept and through the technological development in the 1980s, neural networks got the decisive push. For the first time, it was possible to implement nodes and networks based on the structures of the human brain using powerful computers. This goes hand in hand with Marvin Minsky's concept of distributed systems, which he formulated in his 1986 book "Society of Mind" and which describes how simple building blocks can be used to solve complex problems through interactions. Minsky, incidentally, was a co-initiator of Dartmouth College with McCarthy. Even though powerful computers became available in the mid-1980s, raising hopes for further development of neural networks: once again, data volumes and computing power were not enough to satisfy expectations in the technology. The second AI winter fell in 1987, this time clearly evident in the decline of AI's uses (A good overview of the history of AI is provided by Manhart 2018). But this winter also set the stage for the growth of something new: the concept of the NN took off here, then replaced the notion of AI in the hit list in the early 1990s, peaking in the last third of the 1990s.

As a subfield of Artificial Intelligence, the term ML takes on a very different trajectory. The rise has been slow but steady since the mid-1950s. In the early 1960s, it was roughly on par with the term NN, but the latter was subsequently used much more frequently. The rise of ML, on the other hand, continued to be steady but slow until the late 1970s, when mentions of it also began to rise significantly. ML is based on research in pattern recognition, which uses mathematical and statistical models to learn from data sets. Neural networks form the basis of ML, in that the success of one term pays into the rise of the other.

The development of AI remained a topic for experts for a long time. But shortly before the turn of the millennium, events staged at great expense were to give the subject area more resonance among the public. IBM kicked things off with the series of chess matches between the chess computer Deep Blue and the world champion Garry Kasparov. Kasparov won the first match against the chess computer in Philadelphia in 1996. In the second match in New York in 1997, Deep Blue was victorious. This game was the first victory of a machine over a world chess champion. The documentary film *The Man vs. The Machine* picked up on the theme.

The victory divided experts: some described it as a milestone in AI research, others as a dead end, since the chess computer based its superiority on pure computational power, which had nothing to do with real AI (Heßler 2017).

Even though the competition, with its narrative of man's struggle against the machine, generated a great deal of media coverage, it did not pay off in terms of further publications on the subject area. Publications on AI had been in decline since 1988, those on NN had peaked in 1995, and even for ML there was no impetus from the event. For search queries, unfortunately, there is no data on this event yet, even if Google was already online.

IBM itself did not associate Deep Blue with machine learning methods. The company's communication was aimed more at supercomputing, i.e. powerful mainframes that could rival the human brain with sheer computing power. Deep Blue's success was based on the fact that it made its decisions of analyzing several thousand games. With its high computing power, it used the analysis to calculate the next moves. The method is also called Brut Force. It is used when there are no known efficient algorithms that can solve the problem. The most natural and simplest approach to an algorithmic solution to a problem in this case is to try all potential solutions until the right one is found.⁴ Sympathies in this battle between man and machine were clearly on Kasparov's side. His supporters consoled themselves with the fact that at least the computer couldn't gloat over his victory. The supercomputer, even with its brute force approach, played more the role of a muscular but cold and somewhat simple-minded Goliath.

IBM made another attempt to reach a wider public in 2011. Under the name Watson, the company presented a computer program from the field of artificial intelligence. The program was developed as part of the DeepQA research project and was able to provide answers to questions that people asked it in natural language. Watson used speech recognition, a sub-discipline of machine learning, to do this. To prove its capabilities, Watson competed on the quiz show Jeopardy in February 2011 against two human opponents who had previously won record amounts of

⁴ But even on the term "supercomputer" the action did not pay off, as an evaluation on Ngram shows.

money. A prize of one million dollars was offered for the match. The media picked up on the theme, comparing the contest to Deep Blue's duel with Garry Kasparov. Again, the big theme was the battle of man versus machine. After human Jeopardy opponents Ken Jennings and Brad Rutter were tied after the first round, Watson emerged as the clear winner in the following two rounds. IBM focused on the human versus machine theme in both competitions. In communication, this antagonism is dangerous: the sympathies are with the humans in this case. The machine is possibly admired as a dangerous opponent, perhaps even the technical engineering achievement is appreciated. What is decisive here, however, are the ideas of struggle, opposition and technical superiority of the machine that defeats man.

The success of the spectacular Go game series of Google's AlphaGo machine learning system may have played a significant role in the growing public interest. In October 2015, the computer had already defeated a professional gamer. In March 2016, the AlphaGo team finally managed to beat the reigning Go world champion Lee Sedol in the fifth round on move 37, winning a total of 4 games.

Google has not staged the communication of the topic as a battle of man against machine, but as a sporting competition between a programming team and an ingenious Go player. There are people on both sides – that is the crucial difference in communication, which the film about the event "AlphaGo – The Movie" (<https://www.youtube.com/watch?v=WXuK6gekU1Y>) also highlights. This gives the AlphaGo team a chance to gain sympathy. And the decisive contribution is made by Lee Sedol himself by talking about the beauty of the 37th move of AlphaGo. At that moment, he says, he felt like he was playing with someone who had a mind of his own. In this way, he opens the door to a different understanding of artificial intelligence, which in categories such as beauty actually makes people not opponents, but equal (sparring) partners. The human, in this case the AlphaGo team became the creator of the AI. The rise in searching activity in all three fields – ML, AI and NN – suggests that this communication has struck a chord. Google relies on a narrative where humans and machines are not adversaries, but partners in a competition. They are successful when humans become creators.

Google shifted the emphasis somewhat with the next expansion stage of the Go game: In October 2017, Google released AlphaGo Zero, a new, much more powerful version of DeepMind's Go software. What was new about it was that it didn't learn from existing games, only from playing against itself. Within three days, AlphaGo Zero had already surpassed its predecessor's playing strength in defeating Lee Sedol. By training an AI solely by itself and foregoing human expertise, Google broke new ground (Silver and Hassabis 2017). The release was not widely publicized in the media, but it caused a stir beyond the professional community. In search queries, the echo of this resonance can be traced in the form of short search spikes for all three terms – AI, NN and ML. The narrative here is still the same as in the first competition against Lee Sedol, the interplay of man and machine, but in this step the created has already conquered a bit of autonomy from its creator.⁵

Even if the Gartner Hype Cycle suggests otherwise: Ngram's evaluation shows that trends can recur. The constellation around the year 2010 makes this clear. AI sinks to an interim low in 2008, before a trend reversal sets in and the number of mentions begins to rise again. The trend is similar, only slightly delayed, for NN, which reaches its interim low in 2012 before the number of mentions also starts to rise again. The beginning of the new decade marks a turning point: interest in AI and NN returns, while ML begins a steep rise.

At this point, it's worth taking a look at Google Trends. A search intent is to be evaluated differently than a mention in a publication. Therefore, a direct comparison of the two tools is not possible. Nevertheless, Trends provides a good indicator of the development of interest in the topic area (see Fig. 3.6). My hypothesis is that Ngram is more likely to represent engagement in the professional community, while the trend in Google Trends suggests broader interest among the general public. Trends has the

⁵The picture can be completed by a search on Google. Here, "Machine Learning" comes up with 147 million results, "Artificial Intelligence" with around 136 million results, while "Neural Network" is far behind with only 24 million results (search from 16 April 2020. Search settings without further filters, i.e. "any language", "any time" and "all results"). It is not possible to say exactly where the peak in demand for the term "Neural Network" in January 2020 came from. It may be related to Google's launch of Flax. Flax is a Neural Network Library for JAX, a library for high-performance machine learning.

Artificial intelligence: machine learning takes the lead

Google Trends: Search terms 2004 to 2020 worldwide

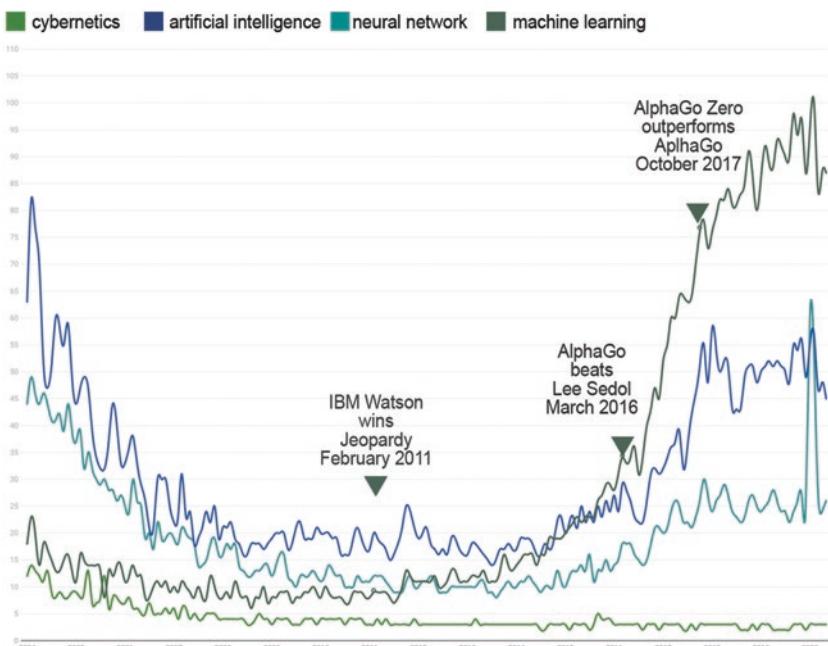


Fig. 3.6 Machine learning is currently the most searched AI term. Interactive graphic at www.data-storyteller.de. (Source: Google Trends)

merit of reflecting real-time reactions, while Ngram documents developments with a time lag determined by the publication process.

For both Ngram and Trends, the numbers increase from 2010/2011 onwards for all terms except CY (here a slow increase is noticeable for the specialist audience, while for the wider public the term is irrelevant). ML has already overtaken AI in the specialist audience in 2013, but in the broader public (trends) this is only the case two years later. Further developments also show that the professional world sets the topics, which are then received by the public with a delay. Mentions of ML begin to increase significantly in Ngram as early as 2012, while in Trends this effect only becomes visible from 2015 onwards.

Looking at these longer cycles around AI as a topic, you see several things:

- Trends can return – as evidenced by the resurgence of interest in AI, ML and NN from 2010.
- The driver of the AI topic is the professional world. The changes in trends are – despite the known delays – first visible on Ngram before they appear in Trends.
- In Ngram, a new peak of interest in the topic of AI is emerging (in the terms AI, ML and NN), while in Trends the curves have already peaked or passed. If the professional community is the driver of the topic, it is likely that we are heading for a new hype.

For those who want to occupy relevant topics and shape an agenda, an analysis of such correlations is helpful. Overall, the topic area shows how developments can be analysed and predictions made for the further success of topics using relatively simple means via tools such as Ngram and Google Trends. It also shows that a story only has a truly far-reaching resonance if it docks onto a successful narrative.

3.3.4 Context and Change of Meanings

To find out keywords that map a topic or at least partial aspects of it, expert knowledge is a good start. But this is not always enough. Word meanings are too diverse, depending on the culture, the social group and the context in which they are used. Tools such as NeuroFlash, presented in Sect. 3.1.2, support the analysis of word usage by evaluating millions of sources. They help to form clusters of related words and to discover association fields.

When placing advertisements, selecting search terms and developing a content strategy, such procedures are important. Google was one of the first commercial users with AdSense. The software made it possible to automatically place an ad that matched the topic of a website. IBM used such a process in Watson's use in Jeopardy in 2011. They also play an important role in voice assistants from Alexa to Siri. Search engine tools as well as content marketing tools offer such methods (Sect. 3.3.2).

Those who analyse developments of key words should be aware that the meanings of these words change. The context plays a decisive role here. This can be seen well in the example of the word “crisis”. In 2008, there was a direct connection to the bankruptcy of the investment bank Lehman Brothers and the financial crisis. When the financial crisis reached the EU economy in 2009, there was then also talk of the euro crisis and economic crisis. In 2015, on the other hand, the term crisis was used particularly frequently in connection with the discussion about refugees. Five years later, in 2020, crisis is again closely associated with the outbreak of the pandemic. Software developer and open data analyst Johannes Filter has analysed this change in context around “crisis”. To do this, he fed 13 million comments from ten years of a German news site into his database. He then used machine learning to evaluate the texts. His study shows how changeable terms are, even over a relatively short period of time, and highlights the importance of contextual information in the analysis (Fanta 2020).

A data team from the Süddeutsche Zeitung has taken on a much longer period of time with the Bundestag minutes: 70 years, 213 million words from over 4200 sessions. Such context analyses are only possible on the basis of algorithms, neural networks and automated processes. This involves translating words into numbers and relating them to neighbouring words. The process is called word embedding. It is a collective term for a number of language modeling techniques in natural language processing in which words are mapped to vectors of real numbers. In this case, Word2vec was used, a technique from Google that builds vector spaces with several hundred dimensions. The trick now is to bring this high-dimensional data back into a two-dimensional format that humans can understand. With the help of these methods, the major lines of discussion in the Bundestag debates can be traced and interpreted. It becomes apparent, for example, that the debate on climate change began relatively late in the mid-nineties. Initially, the focus was on the destruction of the environment, and it was not until later in the 2000s that concrete consequences such as species extinction and resource scarcity came into play. The role of humans (man-made, climate crisis) has only become a stronger focus of the debates in the last two legislative periods since 2013 (Schorries o. J.), see Fig. 3.7.

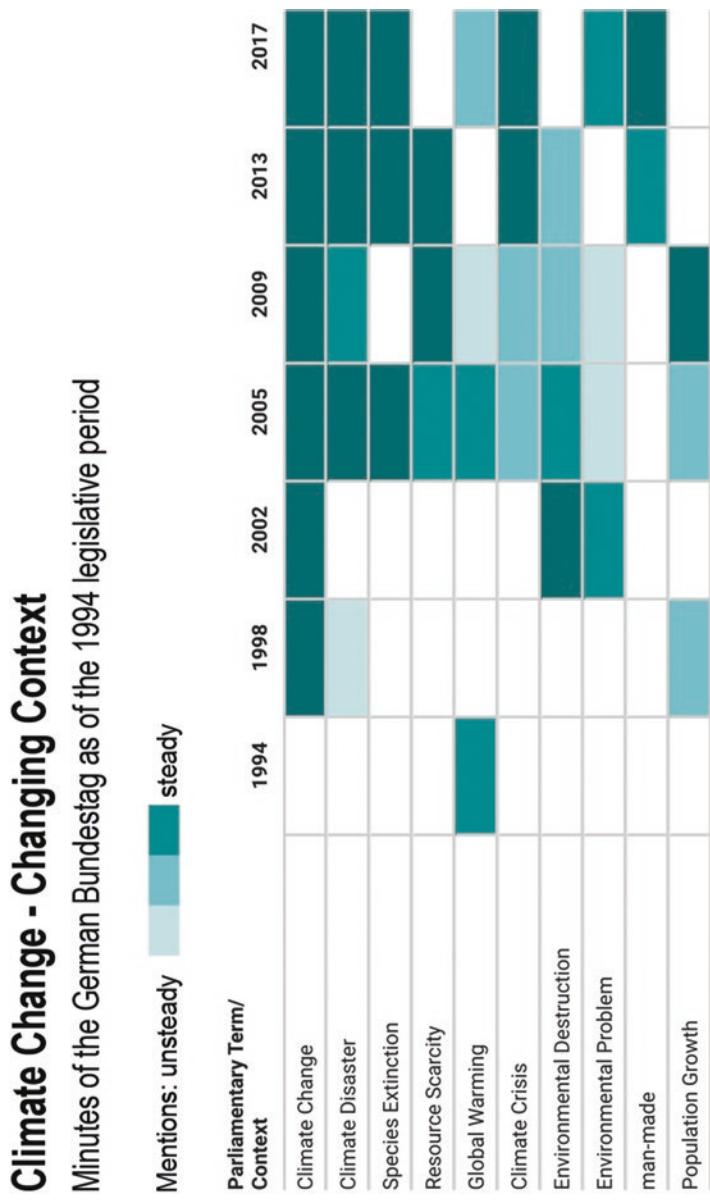


Fig. 3.7 Word meanings change. The Süddeutsche Zeitung has investigated this using the example of climate change debates in the Bundestag. (Source: Schories o. J., own representation)

Such procedures go far beyond the procedure of mere word counts shown in Sect. 3.3.3. Of course, the mere accumulation of a word does not tell us much about the context in which it was used. The authors make this clear with the example of the term “environment”: with the entry of the Greens into the Bundestag in 1983, “environment” was used in connection with nature conservation and biodiversity. In the fifties and sixties, on the other hand, “environment” was understood in the sense of the world around us.

3.3.5 Setting Tomorrow’s Topics Today

We have seen how digital channels can provide insights into topics and trends. And we have seen a number of tools that can be used to analyze this. But how can we specifically identify topics that are relevant to my company’s customers? That is the core question of any company that addresses its target groups via relevant content.

The Frankfurt-based start-up Pythia-ai (<https://www.pythia-ai.de/> – the reference of the name to the High Priestess of Delphi is probably not entirely coincidental) has dedicated itself to the analysis of trends: Whether fitness trends or product range policy at a drugstore – Pythia positions itself as a new oracle that uses artificial intelligence to process enormous amounts of data and derive predictions from it.

The company examines Google and Amazon searches as well as other data sources for new trends. The start-up promises to create trend analyses for upcoming market demands and claims to be much more precise and faster than classic market research. Pythia’s algorithms take over the collection, control and evaluation of data. The drugstore chain Rossmann, for example, uses the tool to optimize its product range. So far, the tool has only been responsible for selected products in the range, including cannabidiol oil (CBD), which Rossmann says has become a sales driver. In addition to Rossmann, other retailers, a shirt manufacturer, financial companies, agencies, and publishers are among the customers who use Pythia to find content and topics (Schobelt 2020).

Those who want to understand exactly what data flows in, how it is collected, combined and interpreted, should consider in-house development. It does involve significantly more work to take the process into

your own hands. However, this is the only way to ensure that the selection of data and its interpretation are also relevant to your own research question. Moreover, with the exception of Pythia, there is no prediction function for future trends and topics.

Of course, every topic should be in demand accordingly. To achieve this, companies should deal with a few questions:

- What topics did we miss?
- How does my target group talk about a topic?
- How are the themes related?
- How relevant is my topic to the target group?
- Is relevance increasing or decreasing?

The content marketing tools in Sect. 3.3.2 are suitable for monitoring. However, they offer little insight into the source situation and analytics. They do not allow you to create a topic map and, above all, they do not have a prediction function.

In order to find relevant trends and topics for their customers, communication and marketing managers best proceed in three steps:

1. Identify trends and themes: Trends and themes can be derived from the available content of the observed channels. However, the available material is so large that it is possible to draw valid conclusions from it using automated evaluations. So you need a suitable technology to analyze the data. Usually, artificial intelligence is used here to identify and cluster topics. This can be done with or without human intervention. These processes are then called supervised or unsupervised. In the latter case, the network finds patterns in the sources independently and without any external input. With the use of unsupervised learning you make sure not to miss any topic due to limitations of your own view. However, to train such networks, you need an enormous amount of data. The AI then finds thematic relationships on its own and can cluster them. In the other case of supervised learning, a human looks at it and provides categories and topic areas. Such methods are more common today and less data intensive. However, they have the disadvantage that certain topics and trends may not be discovered in the first place due to the human focus

Developments in social networks are an important and quick indicator of future success. Relevant topics and trends emerge in social networks such as Twitter, Facebook, Instagram, TikTok, YouTube, and so on. Companies can analyze these developments and use them to make predictions for future topics.

To identify content with broad mass appeal, it makes sense to combine various factors: socio-demographic data, overarching content themes and interactions. With an intelligent combination of these data sets, consumption patterns can be identified and linked to the target groups of one's own company.

2. Allocate topics and trends to the company's focal points: At this point, there is a matching between the topic areas that emerged from the analysis of the sources and the company's topic clusters. The aim here is to determine which topics can play a role for the company. This step requires a deep understanding of the company's topics. A machine would quickly reach its limits here.
3. Predict life cycles of topics: The selected topics are weighted according to future relevance. The question here is: Is the relevance of the topic increasing or decreasing and over what period of time will it develop. This allows the company to prioritize which content should be produced and played out at what time. Mathematical models can be used for the prediction, which make statements about the future relevance on the basis of stochastic procedures using the past development.

3.4 Where Silence Is Golden: Uplift Modelling

Every year just before Christmas I get mail from donation organizations. Especially from those I have donated to before. Thick letters, in which the special meaning of this donation is colorfully described to me. Sometimes there are give-aways like seeds, wooden spoons and other gimmicks. You don't just throw something like that away. But what does such a mailing achieve? From the organization's point of view, it's clear: attention, consternation and a reach for the wallet.

Most of these organizations are masters of tactical storytelling: an individual story, often a child in a life-threatening situation, a rescue intervention through the organization's efforts. And I can make it all happen by making a small contribution from my wallet. Good conscience is as simple as that.

But they don't always understand the situation their addressee is in. For me, this mail led to the fact that I am annoyed that my money does not reach the children, but is used for postage, colorful flyers and gimmicks. An email as a little reminder that now would be the time to donate something again would be perfectly sufficient for me. Am I the only one who thinks like that?

This question is addressed by the so-called uplift modelling (Table 3.1). The procedure is used when I want to know whether the use of a communication measure causes a change in behaviour. This by no means only applies to donation organisations. The procedure is used above all in the retail trade and by providers of telecommunications and electricity. The aim is to understand what effect my communication has and whom it is better not to address. Electricity or telephone customers often use it to discourage customers from switching providers. But it can just as easily lead to a customer receiving the impulse to look into their contract precisely because of the letter and then discovering that there are much cheaper competitors. With this customer the action would have approximately the same effect as the donation mailing with me. He is then gone.

Uplift modeling distinguishes two things:

- What happens when I approach the customer?
- What happens if I don't approach the customer?

Table 3.1 Typical uplift modelling use cases according to Thurber

Use case	Destination	Procedure
Telephone customer	Do not migrate	Upgrade offer
Patient	Get healthy	Treat
Voter	Choose	Message
Donor	Donate	Appeal for donations
Candidate	Accept offer	Switching bonus
End customer	Purchase	Special offer

Source: Thurber (2017)

The two questions result in a constellation of four address possibilities (Fig. 3.8):

- (a) “Do Not Disturb.” These are all those for whom a promotion does the opposite of what it was intended to achieve. So the customer who switches providers when they get an offer to stay.
- (b) “Lost”: Here it doesn’t matter whether the person receives a pulse or not. He will not respond in any case. The mailing campaign is in vain.
- (c) “Safe”: Here it also does not matter whether the person receives an impulse or not. He will react in any case. So here, too, futile actionism.
- (d) “Persuadable”: the only case where the communication activity is worthwhile.

An incremental procedure can now be used to determine the right target group: First, a selection group is created, which is divided into a campaign group and a control group. This is important to get feedback on the different effects that arise without and with action.

Uplift Modeling - Four Categories

		No	A Do-Not-Disturb	B Lost Cause
Response if Treated	No			
	Yes	C Sure Thing	D Persuadable	
		Yes	No	
Response if not Treated				

Fig. 3.8 Uplift modelling helps to answer the question whether the use of a communication measure causes a change in behaviour

In the further procedure, there are several methods. There are essentially three main approaches here: the two-model approach, the class transformation approach and direct modelling (see Gutierrez and Gerardy 2016). They differ mainly in how they compute the probabilities of the behavior of the campaign group and the control group. Crucial in our context is the realization that sometimes I am more successful if I don't tell my story to everyone.

3.5 Towards a Data Strategy

Every company has its specific culture, its market and customers, and its current issues. Data helps answer these questions and make the right decisions. Once the right question has been formulated, a strategy can be developed to tap into the appropriate data. This data strategy must be developed by each company based on its specific constellation of expertise, sales channels, budgets, available data and the company's maturity level in terms of data culture.

Successful data strategies are an integral part of overall business strategy. They establish common and repeatable methods, practices, and processes to control and distribute data across the enterprise. When the entire organization is involved from the beginning, they can advance their data-driven approach.

But it seems that many companies, especially in Germany, are not yet ready. Only 26% of data teams in German companies see themselves in a position to draw the required insights from their data. This is according to the study "Data Strategy and Corporate Culture" by Exasol, which was published in February 2020. The report is based on a survey of more than 2000 data strategy decision-makers in four key markets: the UK, Germany, the US and China. On a global level, this compares to 32%. However, this also highlights that two-thirds of all companies still feel unable to use their data properly or have a precise idea of what data they want to use and for what purposes.

The authors of the study are also surprised that many companies ask the second question before the first: Namely, they are prioritizing where they store data and make it available in their business intelligence

systems – in the cloud or on premise. This is certainly an important step towards an open data culture, where everyone has access to the same data set (Exasol 2020). But the real question of what answers they are looking for in the data does not seem to be on the minds of many data managers either.

Martin Szugat, for example, shows how they can proceed with his company Datentreiber. With data strategy design, Datentreiber helps everyone involved to ask the right questions, to identify important stakeholders in the process, to define goals in an interdisciplinary team and then to develop strategies for collecting and evaluating the data. A free toolset is available for this purpose. Datentreiber's canvases can be used to develop the various fields of a data strategy. Twelve canvases help identify the topic areas and ways to address them. They cover strategic questions about growth horizons and value chains as well as the identification of concrete use cases and the development of a customer contact point analysis, for example. A sample procedure shows how a marketing strategy can be developed on the basis of data.⁶

From one use case, different use cases and their benefits are identified. For example, if I want to optimize my customer journey as a company, the individual customer contact points, such as social media posts, blog entries, and emails, can be analyzed. On this basis, it is possible to determine on a case-by-case basis what data the company already has about these, what potential the contact point offers and with how much effort an optimization can be implemented. This provides the basis for prioritizing the next steps. Only at the end of these steps can the data sources and tools for preparing the data be identified.

The procedure makes it clear how important it is to have a clear and jointly developed focus at the beginning. Without a clear focus, the complexity of the next steps increases rapidly and bears the risk of getting bogged down. In order to avoid this and to successfully implement data projects, a common understanding is needed about which goals are being pursued, which data is needed for this, how it is to be interpreted, and how it contributes to the goals (Klaus 2019) describes which prerequisites are needed for this in marketing and how important people are in this

⁶ <https://www.datentreiber.de/methode/#canvas>

process). A common data culture is therefore needed in the areas involved, on which such questions can flourish. This can only succeed if all those involved from IT and the business departments have agreed on common perspectives and clarified their respective roles. An important step towards data culture is the democratization of data. It gives employees at all levels access to data insights that are relevant to their roles. This enables employees to make better-informed decisions and find new insights. This drives a cultural shift, with every employee contributing to data analytics, embedding the data strategy in the business.

References

- Aaron J (2019) Valuable insights drive growth for online retailer. Station, 10. Februar. <https://www.station10.co.uk/case-studies/valuable-insights-drive-growth-for-online-retailer>. Accessed: 4. Mai 2020
- Anton M (2019) Kundenzentrierung & Künstliche Intelligenz: Erfolgshebel in der Neuausrichtung einer Traditionsmarke. Gpredictive, 2. Dezember. <https://blog.gpredictive.de/kundenzentrierung-k%C3%BCnstliche-intelligenz-erfolgshebel-in-der-neuausrichtung-einer-traditionsmarke>. Accessed: 4. Mai 2020
- Berger R (2015) Die digitale Zukunft des B2B-Vertriebs. München
- Economist (2020) It has never been easier to launch a new brand. Economist, 23. Januar. <https://www.economist.com/business/2020/01/23/it-has-never-been-easier-to-launch-a-new-brand>. Accessed: 4. Mai 2020
- Exasol (2020) Exasol-Studie: Nur 32 Prozent der Daten-Teams können die Erkenntnisse gewinnen, die ihr Unternehmen für eine bessere Entscheidungsfindung braucht Pressemeldung Exasol, 13. Februar. <https://wwwexasol.com/de/company/newsroom/news-and-press/exasol-studie-nur-32-prozent-der-daten-teams/>. Accessed: 4. Mai 2020
- Fanta A (2020) Die Diskursmaschine. [Netzpolitikorg](https://netzpolitik.org/2020/die-diskursmaschine/), 20. Mai. <https://netzpolitik.org/2020/die-diskursmaschine/>. Accessed: 19. Juli 2020
- Fukushima K, Miyake S (1982) Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets. Springer, Berlin, pp S267–S285
- Gutierrez P, Gérard J-Y (2016) Causal inference and uplift modeling. JMLR: workshop and conference proceedings, 67/2016, S 1–13. <https://proceedings.mlr.press/v67/gutierrez17a/gutierrez17a.pdf>. Accessed: 4. Mai 2020

- Heßler M (2017) Der Erfolg der “Dummheit”. NTM 25:1–33. <https://doi.org/10.1007/s00048-017-0167-6>
- Kantar (o. J.) Ad testing. [millwardbrown.com](https://www.millwardbrown.com/mb-global/what-we-do/advertising/ad-testing). <https://www.millwardbrown.com/mb-global/what-we-do/advertising/ad-testing>. Accessed: 4. Mai 2020
- Klaus L (2019) Data-Driven Marketing und der Erfolgsfaktor Mensch. Springer, Gabler, Wiesbaden
- Kreye A (2018) Macht Euch die Maschinen untertan. Süddeutsche Zeitung Edition, München
- Lin Y, Michel J-B, Lieberman Aiden E, Orwant J, Brockman W, Petrov S (2012) Syntactic annotations for the Google books Ngram corpus. In: Proceedings of the 50th annual meeting of the association for computational linguistics, Jeju, Republic of Korea, pp S169–S174
- Lordick M (2016) Transhumanismus: Die Cyborgisierung des Menschen. Zukunftsinstut, September 2016. <https://www.zukunftsinstut.de/artikel/transhumanismus-die-cyborgisierung-des-menschen/>. Accessed: 4. Mai 2020
- Mall JC (2020) Improve your brand positioning like Volkswagen. Neuroflash, 15. Januar. <https://neuro-flash.com/improve-your-brand-positioning-volkswagen-neuro-flash-example/>. Accessed: 4. Mai 2020
- Manhart K (2018) Was Sie über Maschinelles Lernen wissen müssen. Computerwoche, 19. Juni. <https://www.computerwoche.de/a/was-sie-ueber-maschinelles-lernen-wissen-muessen,3329560>. Accessed: 16. Apr. 2020
- McCarthy J, Minsky ML, Rochester N, Shannon CE (1955) A proposal for the Dartmouth summer research project on artificial intelligence, 31. August. <https://web.archive.org/web/20080930164306/>, <http://www.formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>. Accessed: 4. Mai 2020
- Peel S (2019) Walking the walk. Eff Week, 24. Oktober. https://www.youtube.com/watch?time_continue=971&v=rbT8TqBUgOs&feature=emb_logo. Accessed: 3. Sept. 2020
- Rentz I (2019) Die Rückkehr des Gänsehaut-Faktors. Horizont, 19. Dezember. <https://www.horizont.net/marketing/nachrichten/brand-vs.-performance-die-rueckkehr-des-gaensehaut-faktors-179822>. Accessed: 4. Mai 2020
- Schobelt F (2020) Die KI sortiert das Regal: Wie Rossmann Künstliche Intelligenz einsetzt. OneToOne, 14. Januar. <https://www.onetoone.de/artikel/db/915861frs.html>. Accessed: 4. Mai 2020
- Schories M (o. J.) So haben wir den Bundestag ausgerechnet. Süddeutsche Zeitung. <https://projekte.sueddeutsche.de/artikel/politik/so-haben-wir-den-bundestag-ausgerechnet-e893391/>. Accessed: 15. Okt. 2020

- Silver D, Hassabis D (2017) AlphaGo zero: starting from scratch. Deepmindcom, 18. Oktober. <https://deepmind.com/blog/article/alphago-zero-starting-scratch>. Accessed: 12. Juli 2020
- Trommsdorff V, Becker J (2009) Verfahren des Werbemittel-Pretesting. In: Bruhn M, Esch F-R, Langner T (eds) Handbuch Kommunikation. Gabler Verlag, Wiesbaden, pp 923–938
- Wiener N (1948) Cybernetics. MIT Press, Cambridge



4

From Data to Story

Abstract Data is the content of stories. In addition, the proliferation of digital channels is giving rise to a new form of storytelling. More than ever, visual elements are becoming the hook and anchor of stories. Especially in a society flooded with stimuli, strong visuals are gaining weight. Journalism has shown how stories can be developed using data. In the meantime, these practices have also arrived in companies. If you want to turn data into stories, you need a team with very different skills. Communication and marketing managers are therefore well advised to form networks and develop a common data culture.

4.1 Data Journalism: With Wikileaks to the Breakthrough

Anyone who has ever tried to filter out important information from an Excel list can imagine the task the journalists faced when they received a document with 92,201 lines from Wikileaks. On July 25, 2010, Wikileaks published the war diary of the war in Afghanistan and made the complete

documents available in advance to selected journalists from the Guardian, Der Spiegel and the New York Times.

The evaluation of the document, which listed events of the war, happened under great time pressure, as Wikileaks was about to make the entire documents publicly accessible. The journalists' information advantage melted away in a short time. At the same time, they were faced with the task of filtering out militarily relevant information, especially so as not to endanger informants and NATO troops. The majority of the documents contained frontline reports from the years 2004 to 2010, so a thorough analysis was required, because any mistake would put people's lives at risk.

Editing such a huge Excel list was enormously time-consuming, if only because each save operation took a long time. The data was also inconsistently formatted, with many rows simply blank. Scrolling did not provide much insight given the quantity. The journalists were faced with a seemingly impossible task.

But a bit of luck is involved. Harold Frayman, editor at the Guardian, had already gained experience with structuring data with his colleague John Houston and developed an internal database that could process such large amounts of data faster and easier than Excel. Journalists could transfer the data to this and search for keywords and events there.

For example, the journalists filtered out reports on booby traps and evaluated the approximately 7500 reports on explosions from 2004 to 2009. Using this data, the journalists were able to analyse the development of attacks over time, by region and by destructive force. With the help of a developer, they added the coordinates of the attacks and inserted them into a map. A graphic designer helped them to prepare the data for the newspaper (see Rogers 2010a, b).

The Wikileaks revelations may not have been the birth of data journalism, but this was the first time it was able to impressively demonstrate its relevance. Never before had such large amounts of data with even remotely similar political implications come to light. And never before had it been possible to distil stories from such a large amount of data. With this form, journalists opened up access to stories that would not have been visible by conventional means.

In this case, data itself is the subject of the story being told. This characterizes data journalism and also the many other data stories we tell in the corporate context, which form the subject of this chapter (see Fig. 4.1 and the explanations in Sect. 1.4).

The WikiLeaks example vividly illustrates the challenges authors face when filtering stories from large amounts of data:

- Large amounts of data cannot be handled without machine assistance.
- In order to exploit the data, it must be structured and standardised.
- The data (in our case the times of explosions of booby traps) are enriched by further data (geo-coordinates) in order to gain further insights or to be able to present them better.
- The analysis of the data requires know-how among the employees and a corresponding data infrastructure in the company.
- In many cases, the main narrative thread is determined by graphic elements, in our example an interactive map. The preparation in turn requires own skills in the field of data visualization in order to make

Storytelling with Data - Stories based on Data

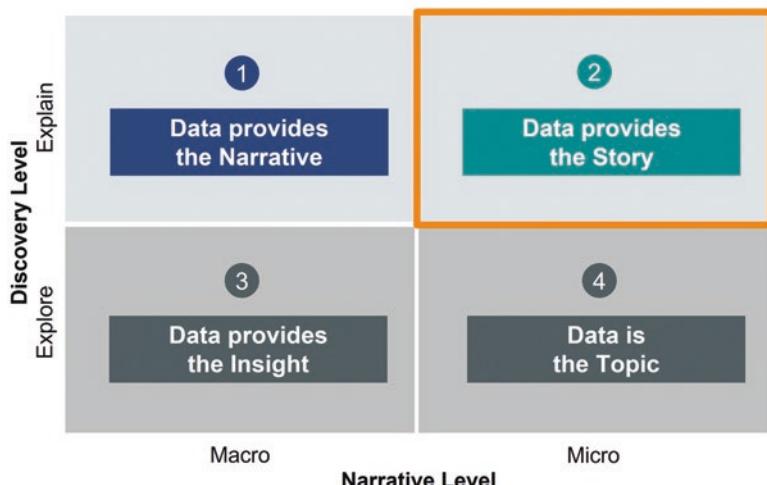


Fig. 4.1 Storytelling with data: in this chapter, data itself is the subject of the story being told. (Source: Own illustration)

the important findings visible and to leave out everything unimportant without exposing oneself to the suspicion of manipulation.

Simon Rogers, then news editor of The Guardian, recalls the Wikileaks period: “It really started with a simple idea: what if we could publish data in a format that would be easier for others to use? … We launched with 200-odd datasets, stored on Google drives because I couldn’t get any resources for a database. That had the weird side-effect of making our work very easy for others to replicate. We were the first blog about data anywhere in the mainstream media. … Through a combination of big stories – the WikiLeaks war records, the 2011 riots coverage and the MPs’ expenses crowdsourcing – it really took off. Suddenly, there was data everywhere, and we explained it and made it more available” (Barr et al. 2019).

The triumph of data journalism came up against classic journalism, whose credibility was questioned at the time. Ten years ago, the representatives of the new discipline were therefore able to compete with the conviction that they could deliver comprehensible and objectifiable results with access to sources and tools.

The successes achieved by Wikileaks spurred the journalists on to open up new sources. Once the technical prerequisites had been created and the journalists had gained initial experience with analysing and processing, they set about collecting further large amounts of data: from offices, authorities, cities, mobile phone providers, weather stations, satellites, etc. On this basis, stories about corruption, the East-West divide, climate change, forest fires, etc. could be developed.¹

Data journalism was fuelled by the growing possibilities to gain access to data. The democratization of data access was advanced by the Open Data movement. The idea emerged as early as the late 1950s and originally served to facilitate the sharing of data for scientific use. An

¹The Guardian has published its data stories in a dedicated collection: <https://www.theguardian.com/data>. Here are a few more selected examples to give you an idea of the breadth of the genre: The topics of the Bundestag over 70 years from 1949 to 2019 have been analysed by the data team of the Süddeutsche Zeitung, identifying focal points and shifts: Schories (2020). The evaluations by Biermann (2011) and Valentino-DeVries et al. (2018), for example, show how movement data and good contextualization can be used to learn a great deal about a person. The spread of Fake News based on Macron’s alleged homosexuality is analyzed by Hamann (2017). Casselman and Dougherty (2019) report on the practices of investors in the US real estate market.

important milestone was the 1995 report of the National Academy of Sciences in the USA, which called for an international exchange of data on changing the world and the environment (National Research Council 1995). Closely related to this is the open government approach, which combines transparent governance with freedom of information and citizen participation. The city of New York, for example, began early on to make its data available to everyone. There, the Mayor's Office of Data Analytics (MODA) and the Department of Information Technology and Telecommunications (DoITT) jointly form the Open Data team. As a hub of analytics in the city, MODA advocates for the use of Open Data in citywide data analytics and in the community. The website provides guidance on how to use data and an overview of current projects (NYC Open Data 2020).

Politicians in Germany have also launched initial initiatives to enable access for citizens and companies. For example, the direct federal administration was obliged to make electronically collected data available as Open Data ("Open by Default") by default from 13 July 2018 at the latest: freely accessible, free of charge and machine-readable. Thus, topics such as health, mobility, air quality, weather, water levels and radiation can be accessed on portals, e.g. at the Federal Statistical Office as well as the statistical offices of the Länder and many other authorities and administrations. But much remains to be done. Valuable data treasures such as traffic data and geodata are not yet publicly accessible.

Open data helps to make administration and government more transparent and to improve citizen participation. It can also become a valuable source for companies to develop new business models based on data. They can use public data to create new value chains as well as to expand existing ones.²

Data journalism benefited from these large, publicly accessible data pools. In the early days, the discipline consisted mainly of the visual preparation of large amounts of data, which testified to the pride of those who compiled and prepared this data. It often lacked a narrative element such

²Cf. on the situation in Germany Bildesheim (2019). A very good overview of publicly accessible data can be found on Github (2020). Proponents of Open Data see data as freely available public property. Since this material is of course also of economic interest, conflict arises here with those who wish to derive benefit from the exploitation of the data.

as a main character to walk you through the story. These were stories that stood out primarily because of the way they were graphically prepared.

Much has changed since then. The joy of the interactive tools soon gave way to the disillusionment that the reader does not use them. This treatment also faced criticism for not giving the reader help in explaining the data. Today's data stories are committed to linear and journalist-led storytelling. Visual editing relies less on charts and more on interactive graphics. The linear narrative form combined with many visualizations in digital formats also earned the genre the name "scrollytelling."³

The appeal of the new thanks to sensational visuals has now worn off. This has certainly led to a more mature approach to the genre. Above all, however, the big data companies have also discovered the topic. The second edition of the "Data Journalism Handbook", for example, is sponsored by Google, among others. And Simon Rogers, who was part of the founding team at the Guardian, is now – after a brief stopover at Twitter – at the Google News Initiative.³

Teams of data journalists are now established in many media houses – from Bayerischer Rundfunk to DIE ZEIT, from the Economist to the New York Times. Interdisciplinary work is a must here. According to the self-image, the journalist is also the one who develops the questions and classifies the answers. But he needs helpers. Whereas the classic journalist falls back on his sources, i.e. consults contacts and tracks down information himself, the data journalist is dependent on a team.

Organizationally, data journalism has now become part of everyday editorial life: Whereas in the early days it was mainly small, specialized teams that worked on topics in isolation, today's data journalists are part of networked teams. When processing very large amounts of data, they increasingly work together with other departments of the medium or – in the case of major revelations such as the Panama Papers – with other media.

But even the most successful visualization does not necessarily tell a story. For data to become relevant to us humans, it needs a context and a protagonist. Analysis, visualization and story belong together. Only by embedding them in narratives and personalizing them do data analyses acquire an emotional component that is crucial for their reception.

³The already available articles of the second version of the Data Journalism Handbook are evidence of this.

Today, the focus is more on the human dimension of the story: “Now we amplify the stories we find in data by collaborating with specialist reporters to put human voices at the center of our stories,” reports Caelainn Barr, data projects editor at the Guardian (Barr et al. 2019). Ben Casselman, editor of the New York Times, describes this approach as follows: “The best stories almost always emerge from talking to people, whether they are experts or just ordinary people affected by the issues we write about. They’re the ones who pose the questions that data can help answer, or who help explain the trends that the data reveals, or who can provide the wrinkles and nuances that the data glosses over. … At the end of the day, data isn’t the story; people are the story” (Casselman 2019). People remain the central actors in data stories.

The authors of the report on the real estate rental market in Germany in the Süddeutsche Zeitung, for example, take this insight to heart. They link the data story with concrete protagonists – here, for example, the Riedel family (Fig. 4.2). The family exemplifies the 44% of people who spend more than 30% of their net household income on rent. This is a larger than average proportion of their income, which can lead to them being in a financially critical situation. Another protagonist, Anna Meier, is representative of a young, well-educated woman who, as an IT consultant, is quite financially able to afford an apartment, but has been living with her mother in Munich for three years because she cannot find one. Thus, from over 57,000 responses to a survey, the abstract result becomes vivid and concretely tangible (Beitzer et al. 2018).

4.2 Visualization: Basics, Tools and Best Practice

The proliferation of digital channels is creating a new form of storytelling. More than ever, visual elements are becoming the hook and anchor of stories. They are easy to produce and easy to distribute via smartphone and computer. Especially in a society flooded by stimuli, in which every story fights for the limited attention of the addressee, strong images gain weight. The construction of our sense of sight also contributes to this: Between eye and brain runs the fastest data connection of all senses in us

The Rental Market in Big Cities

Income and Rent of 30 to 49 Year-olds

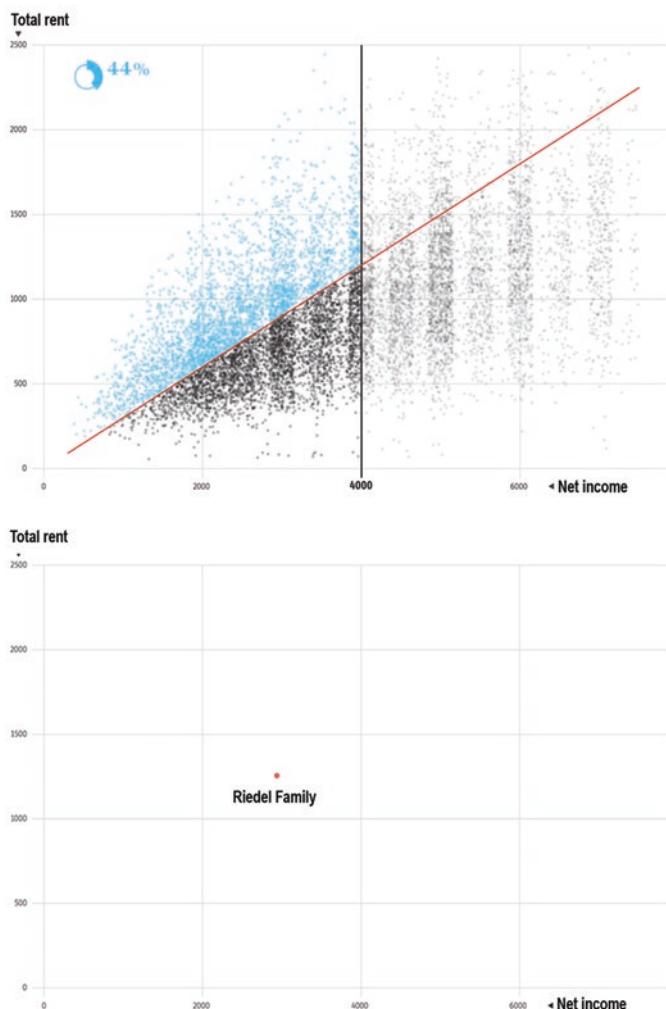


Fig. 4.2 Personal fates make the result of a survey of 57,000 data records vivid and tangible. (Source: Sarah Unterhitzenberger/Süddeutsche Zeitung graphic in Beitzer et al. 2018)

humans. Our brain has been optimally trained by evolution to recognize visual patterns very quickly.

The combination of text and visualization facilitates the reception of stories because it addresses different brain regions. Text is processed in the left hemisphere of the brain, whereas the visual content is processed in the right hemisphere, which is also responsible for emotional impressions. The more brain regions are activated, the more intense the effect of such a story.

The visualization of data is still a fairly young discipline. As recently as the 1970s, Frank Anscombe, an English statistician, was still complaining that his colleagues relied too much on table overviews and paid too little attention to visualizations. At that time, graphs were still considered rough approximations, while number series, on the other hand, were considered accurate. He created four data series whose most important statistical ratios (mean, variance, correlation, linear regression) were identical but differed fundamentally as soon as they were visualized (Anscombe 1973; Fig. 4.3).

These four graphs then became famous as the Anscombe Quartet (Fig. 4.4). In this way, he was able to show that visualizations were an important step in the process of analysis, which could be used both for exploration and for explaining the story. This is where the ability to recognize patterns helps us. Conversely, this means: Once data is visualized in the appropriate form, we process and understand it more quickly and intuitively.

Our fast thinking processes these signals and has recognized the patterns before our slow thinking can intervene. In the psychology of perception, we speak of so-called preattentive perception, in which our brain filters and processes these sensory impressions before we become aware of them. By means of preattentive features (see Fig. 4.5), we can emphasize similarities or highlight differences. Shapes and colours can be used to highlight certain things.

In addition to preattentive traits, Gestalt theory has contributed important insights into perception. As early as the 1920s, psychologists Max Wertheimer, Kurt Koffka, and Wolfgang Köhler sought to understand how people recognize patterns. This resulted in a set of Gestalt principles that help us understand how visualizations of data are perceived (Seel 2012, p. 79).

Anscombe's Quartet

The four data sets

1	2	3	4				
x	y	x	y	x	y	x	y
4.0	4.26	4.0	3.10	4.0	5.39	8.0	5.25
5.0	5.68	5.0	4.74	5.0	5.73	8.0	5.56
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.76
7.0	4.82	7.0	7.26	7.0	6.42	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	6.89
9.0	8.81	9.0	8.77	9.0	7.11	8.0	7.04
10.0	8.04	10.0	9.14	10.0	7.46	8.0	7.71
11.0	8.33	11.0	9.26	11.0	7.81	8.0	7.91
12.0	10.84	12.0	9.13	12.0	8.15	8.0	8.47
13.0	7.58	13.0	8.74	13.0	12.74	8.0	8.84
14.0	9.96	14.0	8.10	14.0	8.84	19.0	12.50

Fig. 4.3 Four data series whose main statistical ratios are identical (mean, variance, correlation, linear regression). (Source: Anscombe 1973)

Data storytelling can take advantage of this by optimizing design for human perception. When data is visualized, the quantitative information is encoded in shapes, color, position, and so on. Viewers must then decode this information. William S. Cleveland and Robert McGill, in their 1984 foundational work Graphical Perception, identified the

Anscombe's Quartet

The four scatter plots

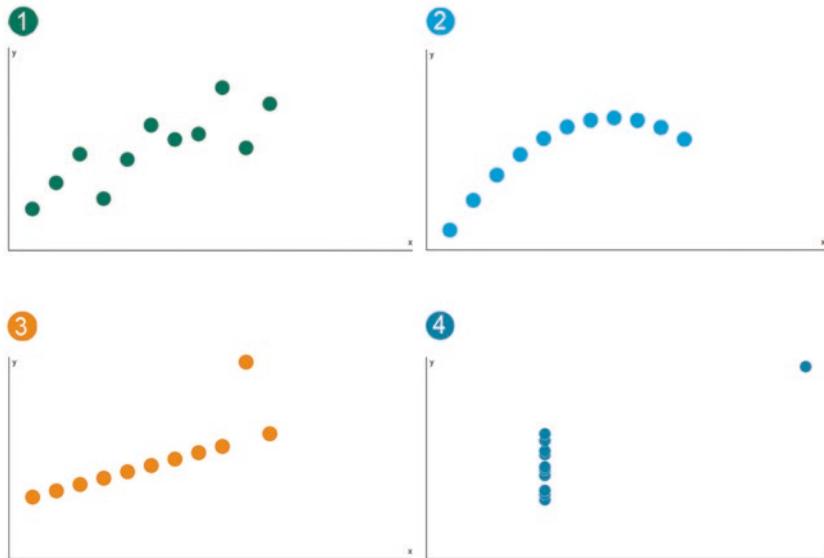


Fig. 4.4 Anscombe illustrates the role of visualizations in the analysis process with the four scatter plots based on data series whose statistical ratios are identical. (Source: Anscombe 1973)

essential processes of human pattern recognition and described the decoding of information contained in graphics. The study deals with a small but important part of the whole process of graphical perception:

- of detection,
- the merging and grouping of the elements (assembly) and
- estimation and comparison (estimation)

Through numerous testing procedures, the authors have determined which forms of graphics are best suited to decode this information and have come up with the following elements:

Preattentive Processing

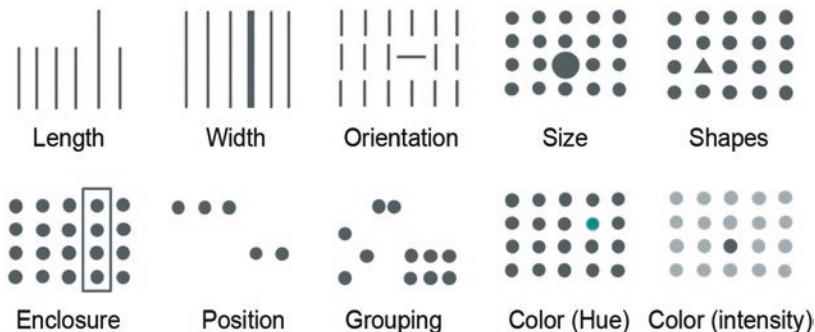


Fig. 4.5 Our brain filters and processes sensory impressions before we become aware of them. The preattentive features help us to recognize patterns quickly. (Source: Funke and Frensch 2006, p. 131)

- Position along a common scale, e.g. scatter plot
- Position on identical but unaligned scales, e.g. multiple scatter plots
- Length e.g. bar chart
- Angle & slope, e.g. pie chart
- Area, e.g. blisters
- Volume, density and color saturation, e.g. heat map
- Hue e.g. Newsmap

Even though the range of visualizations has expanded since then, viewing habits have changed, and interactive preparations in particular have added another dimension, the study provides a solid foundation for telling stories with data (cf. Cleveland and McGill 1984).

Reduction to the essentials is a basic principle of presentation: Edward Tufte, a US information scientist and graphic designer, coined the term “data-ink ratio” for this. All useless components or those that distract from the core message should disappear. Ink should only be used to convey and interpret really significant data.⁴ This is in line with the journalistic demand for clarity, simplicity and unambiguity.

⁴Tufte (1983). Tufte also became known for his criticism of PowerPoint: “PowerPoint is evil”. The presentation tool determines the style of thinking and thus leads to a loss of information, see Tufte (2003).

But which visualization fits the message? There are a number of common types of graphics, which can also be combined with each other depending on the question:

Comparison: Similarities and differences can be shown with bar or column charts. Scaled symbols also work for understanding orders of magnitude.



Trend: Developments can be displayed using line, column and area diagrams.



Composition: In addition to pie and donut charts, grouped bars or even pictograms can be used for this purpose.



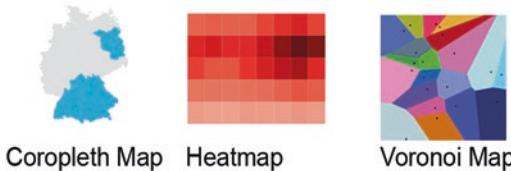
Relationships: Dot and bubble plots help identify outliers and clusters, a tree map shows dependencies, and a flowchart shows movement patterns.



Distributions: Range charts, word clouds, and bubble charts convey orders of magnitude.



Spatial Distributions: Maps, Heatmaps, and Voroni Diagrams⁵ provide information about spatial distributions and clusters.



Today, a variety of tools (Fig. 4.6) make it possible to create infographics. Visualizations in media such as the New York Times, the NZZ, Der Spiegel and the SZ are based on “R”, an open-source programming language for statistical data analysis and visualization, the first version of which was published in 2000. Today, R is one of the most important programming languages for solving statistical tasks. The availability of big data has helped the language gain popularity alongside Python and Scala (Neumann 2018).

But even without programming skills, graphics can now be created. The repertoire of tools includes infographics, scales, geographical maps, concept clouds, heat maps and fever curves. Interactive functions, download options for data and graphics are part of the programs’ range of services, as are a variety of templates on which designs can be created.

Most applications are cloud-based and can be used without further local installations. The results can usually be prepared for online and

⁵Voronoi diagrams are based on a simple but powerful concept: given a set of locations in a space, divide that space into cells – one cell for each location. Each cell contains all the points that are closer to that location than to any other. This form of visualization is useful to many different fields, such as spatial/network analysis, pattern recognition, label placement on maps, and graphs. Voronoi diagrams are now often visualized using a javascript library cf. Rivière (2017).

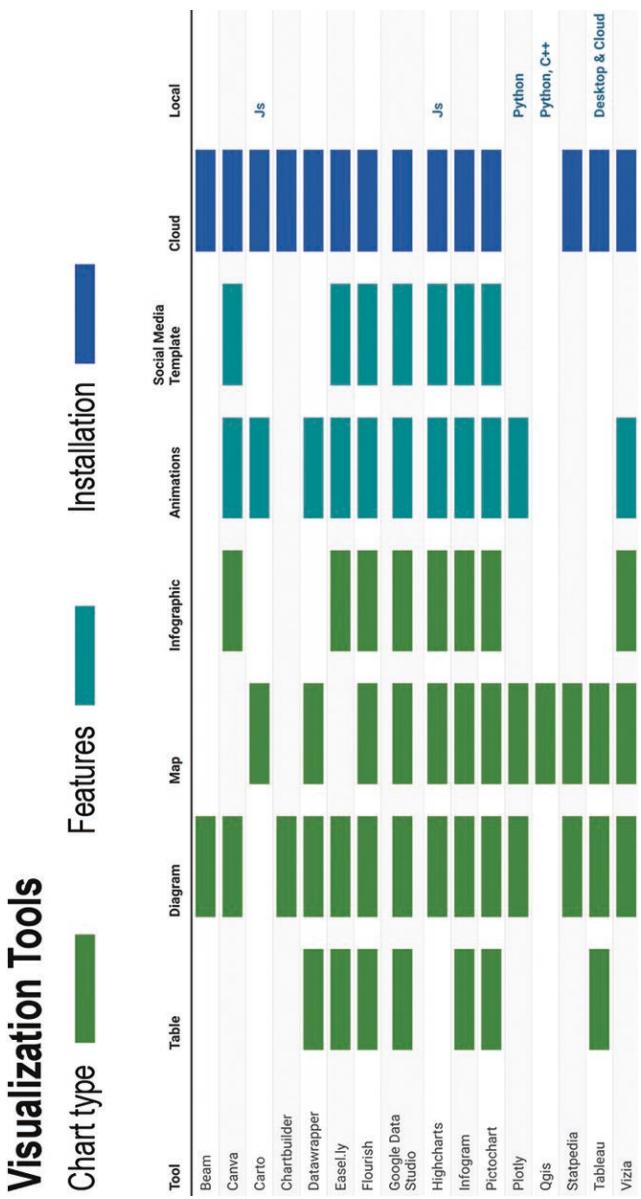


Fig. 4.6 Graphics can be created with the appropriate tools even without programming knowledge. Interactive functions belong to the performance spectrum of the programs as well as a variety of templates on which designs can be created, Interactive graphics at www.data-storyteller.de. (Source: Own representation)

social media channels. For the latter, many of them have their own templates with the optimal formats. There are complete free solutions such as Chartbuilder, Dash, Google Data Studio and Qgis as well as paid versions such as Tableau. In between, numerous tools position themselves with free versions and paid updates. In this way, you can get a good insight into the range of functions, the interfaces, the available templates and the user-friendliness and then decide whether a subscription solution with usually monthly payments comes into question.

To get an overview, it is also a good idea to take a look at the collection of Visualising Advocacy. There, numerous concrete use cases and suitable tools are listed, including:

- Combine and integrate different data sources
- Purge a record
- Build table with text and numbers
- Comparing numbers and counting words
- Making changes visible over time
- Show things on a map
- Visualise network structures (Visualising Advocacy [2020](#))

Through its proliferation of apps on the Android operating system and its web applications, Google has a large and growing influence on data visualization standards. Its Material Design subsidiary has set guidelines for data visualization based on reduction and minimalism. The design is based on map-like surfaces and the flat design approach, which is known for its minimalism. Animations and shadows are used to represent objects as physical objects with appropriate behavior, allowing the user to immediately see which areas contain important information or are interactive and what that interaction will do. The company is continuously updating its web services based on Material Design and also provides interfaces for other developers to implement the design guidelines. In the currently available beta version, the company gives three guiding ideas for the design:

- Accurate: Data should be presented accurately, clearly and with integrity so that information is not distorted.

- Scalable: Visualizations should be adapted to device sizes. User requirements in terms of data depth, complexity and nature should also be taken into account during implementation.
- Helpful: Users should be guided in navigation and encouraged to make their own comparisons and explorations (Material Design [2020](#)).

The first two points would also be signed by data journalists and UX designers. The last design principle makes it clear that Material Design, and thus Google, have a slightly different perspective and allow the user more freedom in their own exploration here than a journalist would.

When implementing infographics, diagrams and maps, the first thing to avoid is mistakes in craftsmanship: The use of certain colors and color combinations leads to the fact that people with red-green weakness or another limitation of color perception have difficulty reading the graphic. Datawrapper has a useful function here, with which you can simulate the common visual impairments on your own graphics right away and thus get an impression of how the implementation affects people with disabilities.

With black-and-white illustrations, on the other hand, representations with multiple elements quickly reach the limits of what the eye can still distinguish. Different hatchings can help here. However, if the lines are too close together, they may create a flickering effect on the viewer, the so-called moiré effect.

But even if you can make everything nice and colorful thanks to a freely selectable color palette, it's not always advisable to use the entire spectrum. Lexie Kane, UX designer at the Nielsen Norman Group, recommends reducing the color spectrum: Rather than a colorful hodge-podge, she advises using a few accent colors that emphasize the core message (Kane [2018](#)).

Not only the colors, but also the other elements of graphical representations lend themselves to standardization. This also includes key figure formats, units, fonts, headings and visualization types. Such recurring elements make it much easier to absorb information. The best way to document the essential building blocks is in a style guide. If such visualizations play a major role in corporate communication, such documentation can also be designed in great detail. The International Business

Communication Standards, which are provided by the non-profit IBCS Association at www.ibcs.com, offer very helpful principles and examples.

Less is more: this also applies to interactive representations. Interaction is used so often in digital media that people are already looking for it. Online visualization offers the possibility for users to ask questions of a tool themselves and thus explore the data pool. For example, the Guardian has developed a pandemic simulator that can be used to identify various factors influencing the spread of epidemics (Evershed and Ball 2020). In this case, this certainly helps to familiarise oneself with the impact of different influences and how they interact. However, not every interaction serves the cause either, but sometimes rather distracts from the actual message. Therefore, you should think carefully about what is necessary and what is better simply left out.

After all, a visualization should always speak for itself. Because it is not always possible to prevent it from reappearing in another context. Because graphics can be easily copied, forwarded in messenger services and posted on social media channels: Always remember that the representation also works without context and that sources are deposited. Then the creator will also benefit if the graphic goes viral.

4.3 Finding the Hero in the Data

Human references are an essential building block of a data story. How does an abstract number become tangible in a personal destiny? In Sect. 2.2 I briefly introduced classical storytelling – the hero's journey with the setting out, the development of a conflict, its dramatic climax and its resolution. Aristotle described the original form of this pattern in his Poetics in 335 B.C. and thus founded Western literary theory (Höffe 2009).

Data provide the raw material of the story. But it is only in the interplay of linguistic and visual interpretation that they achieve their effect. And in more ways than one:

- Memory: Chip Heath, a Stanford professor, concluded in a study that 63% of an experimental group could remember stories, but only 5% could recall a single statistic. It was not the graph but the story that got the group's attention (Heath and Heath 2008, p. 242).

- Persuasion: In another study, researchers tested two versions of a brochure for the charity Save the Children: one was based on the story of Rokia, a seven-year-old child from Mali. The other used infographics to highlight the plight of children in Africa. The version with the story generated twice as much donations per capita as the one with the graphic presentation (Heath and Heath 2008, p. 166).
- Engagement: mathematician John Allen Paulos observed that when listening to stories we tend to suspend our critical minds in order to be entertained, while conversely when processing statistics we are more inclined to find a hook in order not to be deceived (Dykes 2016).

Aristotelian tragedy develops in a kind of triangular movement from the initial situation through the development of a complication, which then culminates in the central conflict and is subsequently resolved. For Aristotle, the whole consists of a beginning, middle and end – the classic three-step of tragedy. The core element in his model is the so-called myth, which can best be described as the plot or sequence of a story.

An extended form of narrative structure is provided by the German writer Gustav Freytag in his “Technique of Drama” of 1863. His insights are based on the analysis of ancient dramas and contemporary tragedies by Shakespeare. Freytag extended the Aristotelian triad to a development in five steps. The central point here is also the climax and turning point of history. Freytag adds to the Aristotelian model one step before and one step after the climax:

The introduction: Here the characters are introduced and the constellation is explained. In addition to the hero, his antagonist also appears here. The introduction is followed by an initial impulse that triggers the course of the story. This intensifies as it progresses, culminating in the climax. This is followed by a phase of deceleration, in the course of which the resolution is prepared as the next turning point.

The resolution ends with either the death of the hero and/or the resolution of the conflicts and the purification of the hero.

The much more comprehensive model developed by Joseph Campbell in his hero analysis “The Hero with a Thousand Faces” in 1948 is more suitable for epic areas in which large stories with many facets, developmental steps and parallel strands can be depicted. The aforementioned

Star Wars episodes and other Hollywood material are better suited for this than stories about companies and products.

These narrative structures have long since found their way into corporate communication. Today, this technique is used in many disciplines and long since not only for films, but also for corporate formats (cf. Dykes 2016, p. 171). The structure of speeches often follows this pattern. Many presentations, workshops, webinars, and even project reports and board papers today, for example, have a three-stage structure of problem, solution, and implementation. Stories now play an important role for strategy and change processes, because there, too, it has been recognized how important shared narratives are for coordinated action.

Such a structure can also be applied to data stories. Here, the basic Aristotelian pattern of the three-stage structure of a drama is the most universal principle, suitable for a wide variety of stories. He required that every drama must have a unified, closed plot with a beginning, middle, and end (unity of action), be set in a single location (unity of place), and not exceed a reasonable duration (unity of time).

Aristotle's drama form can be translated into a three-stage structure of a story:

1. Start: Problem/Conflict
2. Middle: Insight/Climax
3. End: Solution and decision

Exploring, trying out, developing and discarding again are components of the process that leads to the story. But the process of discovery and the story you later make out of it are two different things. The two should remain strictly separate. In exceptions, a few elements of the cognition process can be incorporated into the story later. But only if it plays a role in the internal logic of the story you want to tell and is relevant to the target audience. Otherwise, it has no place in the story. In Sect. 4.4, I will give two examples to show the difference between the explorative and the explanatory approach.

But who is the acting person? With data it is obvious to let numbers speak and to argue with percentages, averages and normal distributions. But our brain looks for the human dimension and thus remembers the story much better. The good thing is that a lot of data can be traced back to human behavior. After all, a car's sensor data doesn't just tell us

something about its current fuel consumption or engine temperature, but also something about the person sitting behind the wheel and pressing the accelerator. Any investigation, no matter how data-heavy, can be made more concrete with insights into people's situations. This is what the authors of the article on the rental market just cited in Sect. 4.1 have done: With their protagonists, they highlight typical examples of a general development and make the concrete effects clear.

In English-language literature, which is much more strongly influenced by narrative approaches, such stories are often developed on the basis of storyboards. The Walt Disney Studios first introduced the concept in the early 1930s for the development of animated films. Today, the storyboard approach provides a practical procedural model for developing stories of all kinds, including data stories.

For the practical implementation, it is good to work with sticky notes that are attached to a magnetic board or a flipchart with the model of the story. In this way, different steps along the story can be tried out and visualised. The sticky notes force brevity and conciseness. And they can be rearranged at any time so that different ways of telling the story can be tried out (cf. Dykes 2016, pp. 170–180; Nussbaumer Knaflic 2020, pp. 20–21).

The pivotal point is the insight/climax of the story. Why is this insight so relevant for the addressees, what are the implications of this insight? It is worth investing time in drawing out this central moment and sharpening the argument well. The storyboard format helps with this. It can be used to try out and discard different points. It should not take more than two sentences to formulate this insight. If it is longer, it should be sharpened.

Let's take the example of the shoe retailer "Seven Feet Apart" from Sect. 3.2 again. The situation – fictitious here, of course – is as follows: Two years after the launch of the online shop, growth is flattening out and sales are shrinking despite the same marketing activities. The managing director wants to know from the marketing manager why the current marketing efforts are no longer really working and what measures he proposes for further sales growth. The marketing manager analyses the data of the website usage, the shop system and the access to the newsletters. After dividing the customers into age groups and analyzing the sales of these groups, he comes to the conclusion: The core target group is older than we originally assumed. The key point of his presentation might look something like this:

Our core target group is older than we thought.

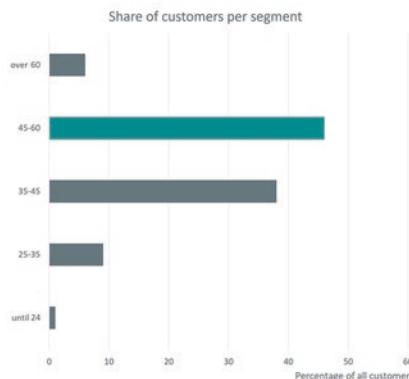
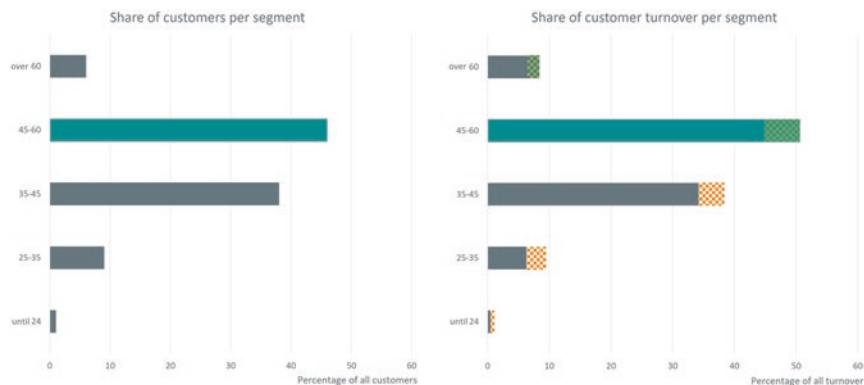


Photo by Logan Weaver on unsplash.com



Step 1: From the Insight to the Initial Situation From the insight, he develops the story backwards for the presentation to management. This way he easily finds the right point. He sharpens the problem and arrives at the following statement: We make too little turnover per customer.

Once the problem is set, the empathetic part follows: whether the story hits depends largely on how it connects to the audience's expectations and experiences. With too little context, the marketer runs the risk of losing his audience because they don't understand the problem

statement. Too much potentially dilutes the flow of the story and bores his audience. Most importantly, he should know the desires, objections, and resistance these people have so he can address them in his argument. In this case, he doesn't need deep research. He knows the management since they started the company together and knows where the crucial points are for the CEO (wants to become market leader in premium online retail for shoes), his CFO (wants stable earnings) and the purchasing manager (is looking for suppliers with the best margins).

The same presentation in front of a circle of marketing colleagues would need an introduction to the specifics of the fashion industry, online retail and perhaps also the shoe market in England. It is worth putting a lot of effort into this and finding out as much as possible about the context and the protagonists. Knowing the desires, attitudes and motives of these people as well as possible provides the basis for a good story.

Step 2: From the Problem to the Insight How many arguments the marketer needs between the two points depends on the story and the audience. Anything that doesn't fit this line, he left out. A few questions help him choose the relevant points:

- What insights help to come to understanding or provide relevant context?
- What questions from the audience can be anticipated?
- Which findings were surprising and unexpected?
- What insights can be left out without the story suffering as a result thereof?

The structure of a storyboard quickly provides the clues as to which arguments fit in well, where some are missing, and which do not fit the story. The marketing manager decides on the following arguments:

- New customer growth is slowing (This worries the CEO).
- The repeat buyer rate is lower than expected (This surprises the buyer).
- Revenue per purchase is below the industry average (the CFO would have asked him for this benchmark anyway).

Step 3: From Insight to Solution After presenting the main point, the marketing manager presents his solution approach. The basis is the realization that a too young customer group was addressed. This realization was possible after he had formed age group clusters and evaluated these groups according to their buying behavior. This then led to the solution: the focus in future will not be on the target group of 35–45 year olds, but on 45–60 year olds. This group has more purchasing power and also offers growth potential because market penetration is still low here. However, this requires not only a reorientation of communication (other channels, new themes), but also a change in the product range (wearable shoes, sustainable production). For this, he needs the CFO and the purchasing manager on his side. Because in the latter's experience, margins are low here, which reduces profits. The core argument at this point is: this customer group is less price-sensitive. So larger margins are possible.

After he has outlined these approaches and the associated growth potential (important for the managing director), he still provides a proposal for the concrete next steps with which the implementation can begin: Develop a new customer approach with new content on the website and a campaign specifically for the age group. Introduction of customer value as a central control variable, development of a new product range policy, etc.

4.4 Saving Lives with Data: From Cholera to Corona

In August 1854, the people of London experienced a devastating cholera outbreak. This was part of a cholera pandemic that was rampant throughout the world from 1846 to 1860. The largest city in Europe at the time, London saw a large influx of people and was the center of the industrialized world with a population of more than two and a half million. The buildings and infrastructure were not up to the task. Where a sewage system existed at all, it transported human and animal excreta as well as industrial waste water directly into the Thames. The then slum area of Soho was particularly badly affected. Within a week, 10% of the district's

residents became infected. Stables, slaughterhouses and grease houses lined the streets, leaving behind animal excrement, rotting fluids and other filth that could not even be discharged into the Thames through sewers.

The physician John Snow had been studying cholera for some time. Six years before the outbreak, he had published his views on the ways of the disease. Until then, the miasma theory had dominated medical doctrine. According to this, cholera spread through the air, which was not entirely far-fetched given the stench in the city. Snow, however, was convinced that the spread was not through the air, but through germs. But at this point he could not trace the path of the disease. The cholera outbreak in Soho gave him the opportunity to test his theory and study the ways in which it was spread. To do this, he plotted all the infections on a map. Every bar a dead person. This map put him on the trail. Through his work, Snow discovered a particular cluster of deaths around a pump in Broad Street (now Broadwick Street, see Fig. 4.7). Only workers at a nearby brewery did not fall ill.

It turned out that the pump's water was contaminated by sewage from a septic tank. The map provided a detailed statistical analysis of the deaths and led him to realize that cholera is transmitted by germs in the water. It also allowed him to explain why the workers at the brewery didn't get sick: They had their own well. Because Snow prevailed with this view and the health authorities and the population drew the right conclusions from it, this was the last cholera outbreak London had seen. Four years later, in 1958, the year of John Snow's death, one of the most significant hygiene measures was put in place with the construction of an efficient new sewerage system. It still forms the backbone of London's wastewater management today.⁶

Snow's map is significant because it enabled the breakthrough of a medical discovery. Without visualization, Snow would not have been able to track the germs in the water, which enabled him to disprove the miasma theory and establish a new interpretation of the infection. It was

⁶For detailed documentation and the map, see the John Snow Archive and Research Companion website at Vinten-Johansen (2020). For more on Snow, see Rogers (2013) and Menden (2020). The concept of this visualization is a so-called Voroni diagram, see Sect. 4.2.

John Snow's Cholera Map of London

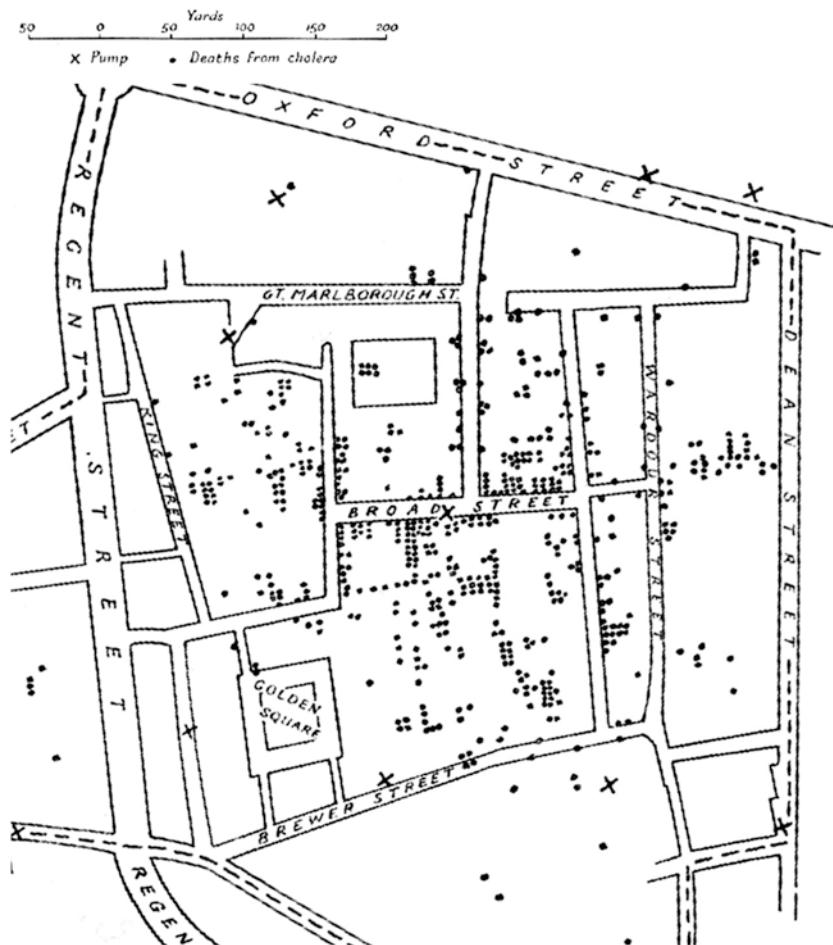


Fig. 4.7 Deaths around a pump in Broad Street during the 1854 cholera outbreak in London. (Source: Wikimedia Commons <https://upload.wikimedia.org/wikipedia/commons/c/c7/Snow-cholera-map.jpg>)

only by recording the exact locations of the dead that he was able to find the source of the epidemic, thus pointing the way to its effective control. The Cholera Map has since been hailed as a major pioneer of data visualization. It highlights how visualizing helps people see patterns and draw

the right conclusions from them. But it is much more than that, it is the document of a scientific breakthrough.

But this visualization does not tell a story. It made it possible to derive a story from it. It is the explorative preliminary stage of a story.

Another example shows how visualizations tell stories and bring them precisely to the point. In view of the spread of Covid-19, Siouxsie Wiles, a microbiologist from New Zealand, developed a graphic together with illustrator Toby Morris and published it on Twitter on 8 March 2020. It went viral under the hashtag #flattenthecurve.

The graph shows two possible progression scenarios for the new Corona pandemic: one with a steep rise and fall and a short time course, and one with a flatter but longer course. A dashed line marks the capacity of the health care system. The graphic idea comes from a 2007 publication by the Centers for Disease Control and Prevention (CDC). That's the U.S. government's disease control agency. The pre-Corona version already showed the two progression scenarios with and without intervention (<https://stacks.cdc.gov/view/cdc/11425>).

An essential element has been added in the Wiles and Morris version (Fig. 4.8): a dashed line. This marks the capacity limit of the health care system. This gives the story a crucial twist: the steep curve clearly exceeds the capacity limit, while the flat curve remains below it. The underlying message: As soon as the healthcare system is no longer able to treat all those who fall ill, the death rate will rise massively. Not only that, but when the system becomes overwhelmed, things will get worse for all of us. The creators of the graphic then backed up the two attitudes with two figures: An indifferent man, who considers the risk to be low, represents the first curve. A woman who urges caution stands for the second curve.

Spurred on by the response to this visualization, Wiles and Morris developed another graphic (Fig. 4.9). It gives concrete action instructions on how everyone can act in the epidemic and went viral on 21 March 2020 under the hashtag #stopthespread. Our brains cannot conceive of exponential progressions, which is especially dangerous in times of epidemics. This is exactly where the animation comes in and makes this clear in a very descriptive way: the graphic with dots and lines shows in the first run how an infection spreads exponentially. A starting point on the left side connects via lines with three more, which in turn connect several

#FlattenTheCurve - Siouxsie Wiles

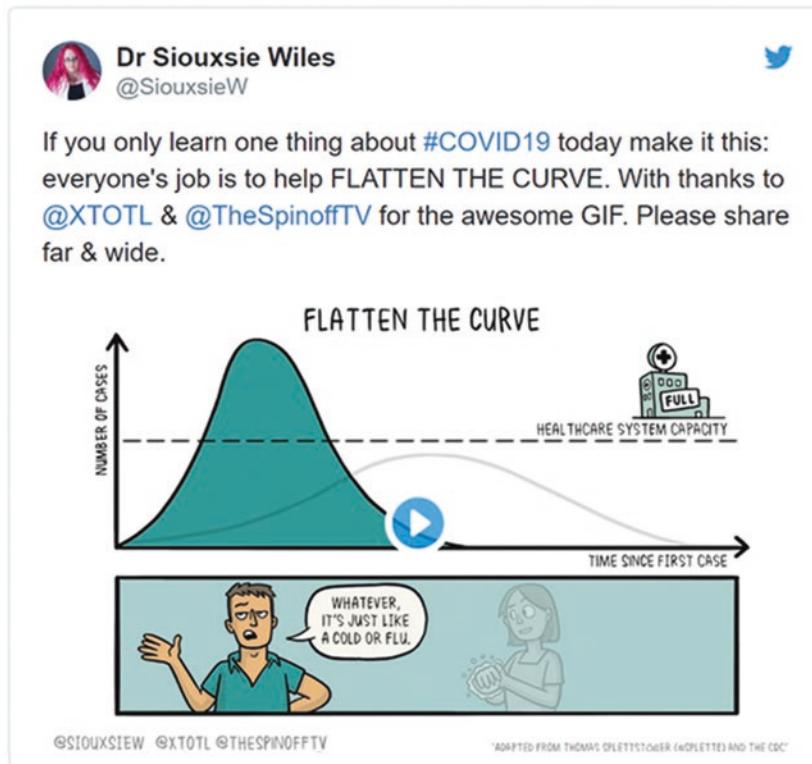


Fig. 4.8 The dashed line marks the capacity limit of the healthcare system. With this crucial twist, microbiologist Siouxsie Wiles illustrates the reason for keeping the infection curve flat. (Source: Twitter)

more points, until a whole bundle of bundles is visible on the right side. In the second pass, the graphic shows how this chain can be interrupted by concrete actions: “Worked from home”, “Didn’t go to that BBQ”, “Didn’t fly”, “Stayed home”. These interventions now colour the originally pink dots and lines grey as they progress. The message: everyone can stop the spread of the virus with their actions.

These graphics in times of cholera and Corona could save lives, albeit in different ways. Wiles’ Corona visuals on Twitter stand on their own.

#StopTheSpread - Siouxsie Wiles

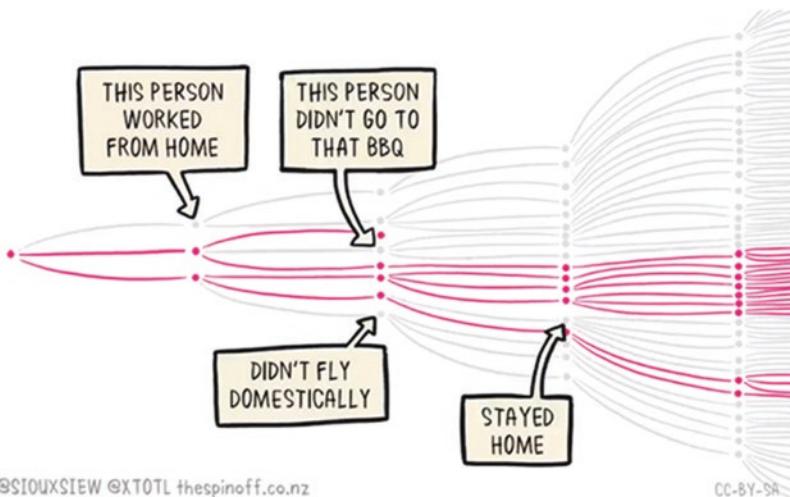


Fig. 4.9 Concrete examples show how anyone can prevent the spread of the virus and thus curb its exponential growth. (Source: Twitter)

They are the story and provide the specific instructions for action. Wiles is using her animated graphics to help change patterns of public behavior to slow the spread of the pandemic. The graphic execution gets to the heart of the idea behind it. The visual argument in this case is stronger and faster than any text.

In the case of the cholera epidemic, John Snow's map of London with the dead provided a clue to the accuracy of his thesis of germs as the disease spreaders of cholera. In my model, the Snow graph is on the exploratory side – it brought John Snow to an understanding of the routes of infection and provided him with the crucial arguments to convince local government of his theory of infection. It is only in the context of the medical debate about miasms and germs that it comes into its own. John Snow needed this map as a protagonist to establish the context and convince his counterparts. Only with it was he able to persuade those responsible in London to build a sewage system (see Fig. 4.10 and the explanations in Sect. 1.4).

Storytelling with Data - Explore or Explain?

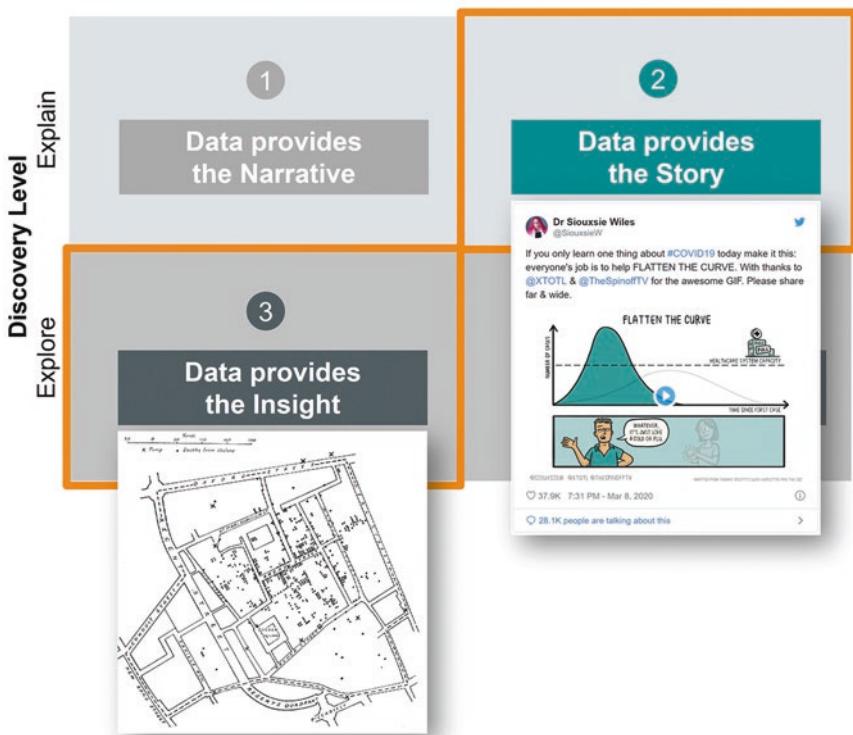


Fig. 4.10 The difference between exploring and explaining – from the example of John Snow and Siouxsie Wiles’ graphics. (Source: Own representation)

4.5 Data Storytelling Is Teamwork

Journalistic practices have long since found their way into companies. Long before the triumph of data journalism, companies began to act as media houses. They were fascinated by the idea of getting in touch with target groups directly and no longer having to rely on journalists for mediation. This approach was called corporate publishing. The term emerged early in the twentieth century, but it gained momentum in Germany in the 1990s. Publishing houses had the opportunity to market their journalistic expertise to companies. Customer magazines, employee

magazines and editorially prepared annual reports were the flagship products of corporate publishing (Fux 2019).

With the advent of digitalization, content marketing took over from corporate publishing. It started with the claim to offer journalistically prepared content on all channels accessible to companies. In addition to the already mentioned regularly recurring publications, other, primarily digital channels were added. And the term also made it clear that it was now marketing, rather than the press, that was claiming leadership, and that in addition to magazines and journals, corporate blogs, newsletters and social media sites were now also being supplied with content.

The Wikileaks example has made it clear: if you want to turn data into stories, you need a team with very different skills. Wherever data are the sources of wisdom, a Pythia is needed to open up access to these sources and a priest to interpret the Pythia's words. In today's corporate world, multiple functions vie for this interpretive authority. The CEO or managing director does not always give the communications and marketing managers the most ear. Depending on the industry, market focus and corporate culture, the power of interpretation tends to lie with IT, finance, product development, service or sales. In B2B companies, communications and marketing leaders rely on good interplay with sales and product development, while their counterparts in B2C companies with strong brand orientation, broad target groups, and large budgets can more easily take the lead. However, digitization gives all communications and marketing leaders an important asset – access to the customer. The availability of knowledge about customers, on the one hand, and customers' changing expectations of companies, on the other, are leading to a growing conviction in many companies that innovations must be driven primarily from the customer's perspective.

In order to be successful, the communication or marketing area is dependent on cooperation with other specialist areas, i.e. above all IT, product development, service and sales. In order to speak at eye level, today it needs a deep understanding of the data sources and their potential. This requires different skills in the team than ten years ago. Of course, there is now someone there who can confidently play the keyboard of social media channels. But where does the content come from, who controls the campaigns, analyzes them, and networks them with other activities?

At branded companies and retailers, a performance marketing manager will develop the sales-oriented campaigns. But as the example of

Adidas has shown, there needs to be a counterweight on the side of brand management that can explain which communication activities pay off for the brand and in what way, and which can damage the brand.

The technology lead will probably be based in the IT area of the company. A good connection there is of utmost relevance. Ideally, the technology lead will be in charge of the business intelligence application, into which all data from sales, marketing, purchasing, service, etc. flows. Valuable insights for the planning of future campaign and sales activities can be derived from this.

However, many companies have not yet established central data storage. Then it is even more important to bring together the information from the various systems such as the online shop, CRM, ERP and the channels managed in the communication units such as the website, social media channels and the email system. A data strategist helps identify the right sources and discover correlations. When tapping into source material, it helps if someone knows scraping, which is the automated extraction of data from websites, applications or documents, and can turn it into a set of structured raw data. For example, information can be extracted from websites of potential customers (see Gervalla 2020). The analyst, in turn, examines the data and identifies trends, develops predictions, and thus transforms data into information. To do this, he or she must master databases and how to query them, be proficient in business intelligence tools, and also know something about visualizing data.

The data strategist or data scientist (the terms are not always used uniformly and also change) in turn translates the questions of his team into a strategy for the development and preparation of the data sources, he knows or has mastered the methods of machine learning, including, for example, software such as Matlab or programming languages like Python and R.

Exciting new roles are emerging at the boundaries of existing disciplines, for example between business and technology or between data analysis and its visualization. On the visualization side, there is the so-called UX designer, who takes care of the most positive user experience (UX) possible. For the UX designer, it's all about usability and the interaction design of a product or service. An infographic designer is familiar with the techniques it takes to make connections visible to the viewer, highlighting the important things and leaving out the unimportant ones. He or she knows the importance of scales, colors, chart types, and is also

aware of the possible manipulations that can result from truncated scales or certain color schemes. He or she will have learned more than the UX designer about data preparation in his or her training and will have analytical thinking, but he or she also lacks the data analysis side.

A border crosser between the two worlds of data and visualization would be a data UX designer, for example. He or she must understand something about data analysis and data design workflow and be able to master the common visualization tools with graphic types, color palettes, etc. as well as data storytelling with a view to the expectations and wishes of the customer (cf. Münster 2019).

Given the large number of possible issues, it makes sense to first approach the topic as pragmatically as possible. A generalist approach helps here. Overview is initially more important than focus. It is important to know the relevant players in the company and to develop a role allocation with them. This means that a data strategist or an analyst does not necessarily have to be on board in the communications team if this role already exists in the company. And if it doesn't already exist, you don't have to hire someone with that title to begin with. All you need to do initially is develop someone in this direction and continually build the skills within the team. Depending on the size of the department and the distribution of tasks in the company, these functions can be distributed among several people or bundled in one person.

Many of these developments and the resulting new requirements are technologically driven. The skills required for this are not necessarily among the core competencies of communications and marketing managers. They would therefore be well advised to expand their network within the company and create a common understanding of how the handling of data affects collaboration with IT, sales, research and development, and service, for example. In order to develop a data culture, a good start would be to overcome silos and divisional thinking and work on a common understanding of the challenges. After all, the other areas are also changing as a result of digitalization. So why not learn from each other instead of hiding behind outdated tasks and roles?

If they use this correctly, communicators and marketers can make an important contribution to the company's success and thus improve their own relevance within the interpretive structure of a company. Depending on the company's orientation and the role it assigns to the brand, this role

will lie with the head of communications and/or brand management. This person develops the story of the company or brand and translates it into the language of the target groups. As a storyteller, he looks at the interaction with the eye of a dramaturge: Are the themes set correctly, are the characters right, is the story clearly drawn and does it develop? In order to be able to do this, however, he ideally has all communication topics in view and does not sort them into pigeonholes such as press and marketing, but recognizes the interaction of the disciplines as an opportunity.

References

- Anscombe FJ (1973) Graphs in statistical analysis. *Am. Statistician* 27(1):17–21
- Barr C, Chalabi M, Evershed M (2019) A decade of the Datablog: ‘There’s a human story behind every data point’. *The Guardian*, 23. März. <https://www.theguardian.com/membership/datablog/2019/mar/23/a-decade-of-the-datablog-theres-a-human-story-behind-every-data-point>. Accessed: 11. Dez. 2019
- Beitzer H, Ebitsch S, Endt C, Öchsner T, Schories M, Zajonz M (2018) Deutschlands Mietmarkt ist kaputt. *Süddeutsche Zeitung*, 26. Juli. <https://projekte.sueddeutsche.de/artikel/wirtschaft/miete-wohnen-in-der-krise-e687627/>. Accessed: 1. Juni 2020
- Biermann K (2011) Was Vorratsdaten über uns verraten. *Die Zeit*, 24. Februar. <https://www.zeit.de/digital/datenschutz/2011-02/vorratsdaten-malte-spitz>. Accessed: 11. Mai 2020
- Bildesheim O (2019) So finden und nutzen Sie offene Daten. *Computerwoche*, 6. Dezember. <https://www.computerwoche.de/a/so-finden-und-nutzen-sie-offene-daten,3548143>. Accessed: 6. Mai 2020
- Casselman B (2019) In data journalism, tech matters less than the people. *New York times*, 13. November. <https://www.nytimes.com/2019/11/13/technology/personaltech/data-journalism-economics.html>. Accessed: 12. Dez. 2019
- Casselman B, Dougherty C (2019) Want a house like this? Prepare for a bidding war with investors. *New York times*, 20. Juni. <https://www.nytimes.com/interactive/2019/06/20/business/economy/starter-homes-investors.html>. Accessed: 12. Dez. 2019
- Cleveland WS, McGill R (1984) Graphical perception: theory, experimentation, and application to the development of graphical methods. *J Am Stat Assoc* 79(387):531–554

- Dykes B (2016) Data storytelling: the essential data science skill everyone needs. Forbes, 31. März. <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs>. Accessed: 12. Dez. 2019
- Evershed N, Ball A (2020) How coronavirus spreads through a population and how we can beat it. The Guardian, 22. April. <https://www.theguardian.com/world/datablog/ng-interactive/2020/apr/22/see-how-coronavirus-can-spread-through-a-population-and-how-countries-flatten-the-curve>. Accessed: 11. Mai 2020
- Funke J, Frensch P (2006) Handbuch der Allgemeinen Psychologie – Kognition. Hogrefe, Göttingen
- Fux K (o. J.) Corporate publishing – eine definition. Mediapunk.org. <https://www.mediapunk.org/corporate-publishing-definition>. Access: 29. Nov. 2019
- Gervalla A (2020) Erweitern Sie Ihr Kundenwissen mit Hilfe von B2B Web Scoring. B2B Smart Data, 3. Dezember. <https://www.b2bsmartdata.de/blog/erweitern-sie-ihr-kundenwissen-mit-hilfe-von-b2b-web-scoring>. Accessed: 10. Dez. 2020
- Github (o. J.) Awesome public datasets. <https://github.com/awesomedata/awesome-public-datasets>. Accessed: 6. Mai 2020
- Hamann G (2017) Macron ist schwul, NOT! Die Zeit, 24. Februar. <https://www.zeit.de/politik/2017-02/fake-news-emmanuel-macron-russland-rekonstruktion>. Accessed: 11. Mai 2020
- Heath C, Heath D (2008) Made to stick. Why some ideas survive and others die, Random House, New York
- Höffe O (2009) Der wahre Aristoteles. FAZ, 27. Januar. <https://www.faz.net/aktuell/feuilleton/buecher/rezensionen/poetik-der-wahre-aristoteles-1759205.html>. Accessed: 5. Juni 2020
- Kane L (2018) Designing effective infographics. Nielsen Norman Group, 12. August. <https://www.nngroup.com/articles/designing-effective-infographics/>. Accessed: 17. Jan. 2020
- Material Design (o. J.) Data visualization beta. Material Design. <https://material.io/design/communication/data-visualization.html#principles>. Accessed: 11. Mai 2020
- Menden A (2020) Das Ende des großen Gestanks. Süddeutsche Zeitung, 3. Mai. <https://www.sueddeutsche.de/kultur/stadtplanung-das-ende-des-grossen-gestanks-1.4895361>. Accessed: 4. Mai 2020
- Münster E (2019) 9 design pitfalls on the way to a successful data product. Towards data science, 27. Nov. <https://towardsdatascience.com/9-design-pitfalls-on-the-way-to-a-successful-data-product-6ea5a3e6842>. Accessed: 16. Febr. 2020

- National Research Council (1995) On the full and open exchange of scientific data. The National Academies Press, Washington, DC. <https://doi.org/10.17226/18769>. Accessed: 6. Mai 2020
- Neumann A (2018) 25 Jahre: Wie R zur wichtigsten Programmiersprache für Statistiker wurde. heise.de, 3. Aug. <https://www.heise.de/developer/meldung/25-Jahre-Wie-R-zur-wichtigsten-Programmiersprache-fuer-Statistiker-wurde-4127034.html>. Accessed: 9. Mai 2020
- Nussbaumer Knafflic C (2020) Storytelling with data – let's practice. Wiley, Hoboken
- NYC Open Data. <https://opendata.cityofnewyork.us/overview/>. Accessed: 6. Mai 2020
- Rivière (2017) The state of d3 Voronoi, 3. Januar. <https://visionscarto.net/the-state-of-d3-voronoi>. Accessed: 23. Febr. 2020
- Rogers S (2010a) Wikileaks' Afghanistan war logs. The Guardian, 27. Juli. <https://www.theguardian.com/news/datablog/2010/jul/27/wikileaks-afghanistan-data-datajournalism>. Accessed: 27. Nov. 2019
- Rogers S (2010b) Afghanistan IED attacks – 2006 to 2009. SCRIBD <https://de.scribd.com/document/34850058/Afghanistan-IED-attacks-2006-to-2009>. Accessed: 27. Nov. 2019
- Rogers S (2013) John Snow's data journalism: the cholera map that changed the world. The Guardian, 15. März. <https://www.theguardian.com/news/datablog/2013/mar/15/john-snow-cholera-map> accessed: 19. Jan. 2020
- Schories M (2020) So haben wir den Bundestag ausgerechnet. Süddeutsche.de (ohne Datumsangabe). <https://projekte.sueddeutsche.de/artikel/politik/so-haben-wir-den-bundestag-ausgerechnet-e893391/>. Accessed: 11. Mai 2020
- Seel NM (2012) Gestalt psychology of learning. In: Seel NM (ed) Encyclopedia of the sciences of learning. Springer, Boston
- Tufte E (1983) The visual display of quantitative information. Graphics Press, Cheshire CT, Edward R. Tufte
- Tufte E (2003) PowerPoint is evil. wired, 9. Januar. <https://www.wired.com/2003/09/ppt2/>. Accessed: 7. Mai 2020
- Valentino-DeVries J, Singer N, Keller MH, Krolik A (2018) Your apps know where you were last night, and they're not keeping it secret New York Times, 10. Dezember. <https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html>. Accessed: 11. Mai 2020
- Vinten-Johansen P (o. J.) The John snow archive and research companion. <https://johnsnow.matrix.msu.edu/index.php>. Accessed: 1. Juni 2020
- Visualising Advocacy (o. J.) Visualisation tools. Visualising Advocacy. <https://visualisingadvocacy.org/resources/visualisationtools.html>. Accessed: 11. Mai 2020



5

Fair Play: What Counts in Data Stories

Abstract It is through selection and interpretation that data gain their meaning. This chapter is about the false certainties and deliberate manipulations that influence us in the interpretation of data and how we can protect ourselves from them. The more power is attributed to data, the more ethical questions gain relevance: the protection of privacy, the role of algorithms in making decisions and the bias of machines come into view.

Data is the new smoke: The thesis formulated at the beginning of this book (see Sect. 1.2) aims at the fact that today we ascribe a large and growing role to data in explaining the world. It should have become clear that it is not the data themselves that contain this information, but we attribute it to them. Only through selection and interpretation does the data acquire meaning. To what do we direct our attention? This question is more relevant than ever in view of the overabundance of possible information. We do, after all, have a limited attention capacity. Every story that is told and shared claims our attention and crowds out countless other stories that are not told and shared. What narratives does the story

being told dock onto? The more deeply a story is embedded in already known and accepted narratives, the more likely it is to generate resonance itself. Context is an important factor in its success. But beyond that, so is its novelty value, that is, its deviation from familiar patterns. The task of discovering meanings in the data and making sense of it is for the pythias and the priests of the data oracles. And only when we believe the prophecies and forecasts and act on them do they become effective.

Precisely because this process of meaning-making has so much influence on our actions, it is worth taking a look at misinterpretations and manipulations. This chapter is about outlining which false certainties and deliberate manipulations influence us when interpreting data and how we can protect ourselves from them. Finally, we look at voluntary initiatives and legal frameworks for storytelling with data.

5.1 False Certainties and Deliberate Manipulation

One of the biggest sources of error in storytelling with data is that we believe there is something like objective, incontrovertible truth in the data. The more we believe this, the more powerful the influence of the stories we draw from the data. At the same time, we should keep in mind that our perception works quite differently. The human brain is not a rational calculator. It loves drama and the emotions it evokes: The more of it, the better it can remember the content. Stories with a clear downfall spread well. This is how our ancestors were able to pass on vital experiences.

This was helpful in making quick decisions on the steppe: Fight or flight. Sometimes fractions of a second were decisive in the hunt. At the sight of a sabre-toothed tiger, there was no need for a sophisticated set of subjunctives and possibilities. It was drama in its purest form. This pattern is deeply inscribed in our brains, much like our preference for sugar and fat, which provide our bodies with vital energy. This has ensured our survival for the past tens of thousands of years. For living in our world today, these patterns are not always helpful and are sometimes counterproductive. Our cravings for sugar and fat fuel entire industries that

provide us with food that is more dangerous to our lives than all the wars in the world (see Harari 2017, p. 14). And our instinct to divide things into good and evil, black and white, rich and poor, sometimes blocks the view of differentiations that give us valuable insights into our world and form the basis for new stories and options for action.

Over the millennia of its development, our brain has been programmed for survival. Rationality and reason are not among its vital functions, but fast intuitive action is all the more important. Thus, in the course of evolution, a kind of division of labour between two systems has developed in our brain. The psychologist Daniel Kahneman has called these systems 1 and 2, or “fast thinking” and “slow thinking”. “Fast thinking” allows us to act intuitively and emotionally. It works like a kind of autopilot. Without pause, it makes judgments about distances, dangers or moods, for example, and gives us the confidence in everyday life that we are always in control of the situation. Experienced experts can make precise intuitive decisions. For example, a fireman can sense the danger of an impending explosion, a chess master knows the next appropriate move without thinking. System 1 is receptive to simple judgments, clear polarizations, great drama. However, it is of no use in understanding more complex relationships.

This is where “slow thinking” comes into play. This type of thinking, also called system 2 by Kahneman, switches on when something complex or unexpected arises in the stream of fast thinking. Slow thinking consumes more energy, it is exhausting and has only limited capacities. Above all, however, it is characterized by its laziness: it usually lets system 1 take the lead in interpreting sensory impressions, but always believes that it is in control.

According to Kahneman, the interaction of both systems determines our thinking and our perception with all its false certainties and distortions. For it is precisely slow thinking that lulls us into a sense of security and makes us believe that we are in control of the situation, while fast thinking has long since made the decision. In short, we make our decision emotionally, only to justify it rationally (Kahneman 2011, pp. 19–30). Applied to our topic, this means that we provide the rational justification with the data. In many cases, however, we decided long beforehand to believe the story and that is why we chose it in the first place.

Storytelling can address both systems, the fast thinking with dramatic effects, clear polarization and great fall, the slow thinking with irritations, disruptions and breaks. It is in our hands to choose which story we tell. In doing so, it is helpful to be aware of the traps we can fall into. These traps consist of the distortions of our perception that are due to the construction of our wetware, i.e. our brain.

The Swedish physician Hans Rosling has devoted an entire book to these distortions of our perception. In *Factfulness*, Rosling illustrates these patterns of our thinking shaped by instincts, that is, what Kahneman called quick thinking (Rosling 2018). He makes clear how to overcome these narratives and retell stories. The technique is also called reframing; familiar concepts are put into a new context and result in a different story.

A frequently recurring pattern is thinking in polarities. In many cases, the division of our world into two parts prevents us from understanding complex interrelationships. Rosling shows this in the distinction of the world into rich and poor, into developing and industrialized countries, into North and South. All these are variants of the same narrative that divides the world into two and determines our view of politics, economics and society. Rosling recounts an experience in the 1990s when he was discussing infant mortality in the world with his students and uncovered this pattern: the “us” of the rich industrialized countries versus the “them” of the developing countries. To the fight against this “mega misconception” he had dedicated himself. Above all, Rosling’s approach makes one thing clear: you don’t fight such a powerful narrative with individual facts. To be truly convincing, you need a completely new narrative that gives us a new perspective on things.

Rosling develops this narrative on the basis of the distribution of wealth. For this purpose, he chooses per capita income and life expectancy as indicators. His famous chart (Fig. 5.1) shows income (gross national product per person and country) on the X axis and life expectancy on the Y axis. The countries of the world are shown in circles, the size of which represents the size of the population.

As might be expected, when the countries are arranged with the current data, a diagonal line emerges that points to a correlation between wealth and life expectancy. Or to put it more simply: the wealthier I am, the longer I live.

Lifespan and Income

Hans Roslings World Health Chart - Data from 2019

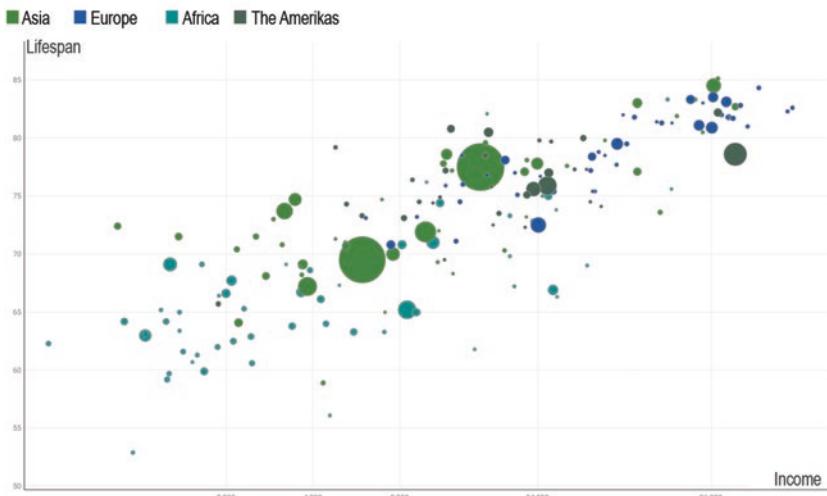


Fig. 5.1 With this illustration, Hans Rosling shows that wealth in the world cannot be divided between rich and poor, but is rather concentrated in the middle. Interactive graphic at www.data-storyteller.de. (Source: Rosling, own representation)

But this chart makes something else clear: there is no discernible polarity of a distribution into rich and poor. For if there were, the picture would look different, showing most or the largest circles (i.e. the countries with the largest populations) on the X-axis on the left (i.e. the “poor”) side, while there would be a few rich countries on the right. But there should be a gap in between if the picture of poor vs rich is correct. But the picture shows a different distribution, namely a relatively strong middle and a decrease towards the edges.

This leads Rosling to replace the dichotomy of “poor” vs. “rich” with four development levels, which are also a simplification, but better reflect the distribution of wealth (Fig. 5.2). Here, Level 1 (greatest poverty) and Level 4 (greatest wealth) each have roughly equal numbers of people, at one billion, while the bulk of humanity (5 billion) is distributed across

World Population: Four Income Groups

Per capita income in US-Dollars per day

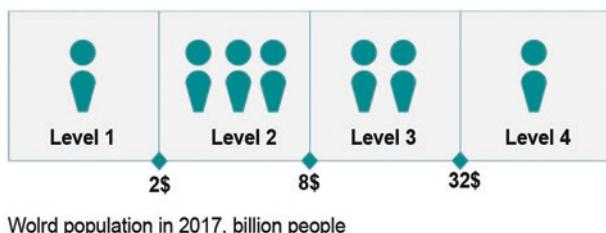


Fig. 5.2 According to Hans Rosling, four income levels correspond better to the distribution of wealth than the division into rich and poor. (Source: Rosling 2018)

the two middle levels.¹ Rosling thus replaces our polar pattern of thinking with a picture of four development levels in the form of a normal distribution: the centre of gravity in the middle and significantly less at the two poles “poor” and “rich”.

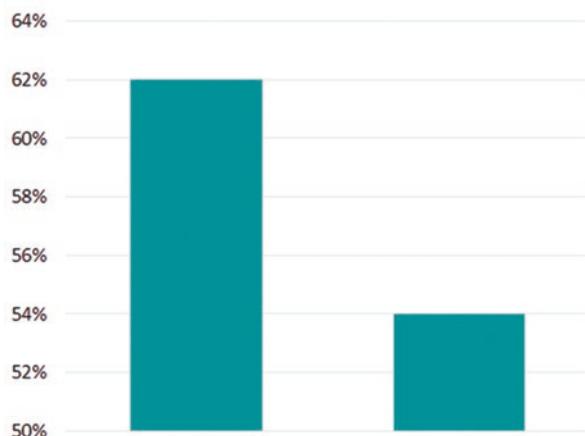
The rich vs. poor thought pattern is based on a tendency to polarize, which appeals to quick thinking. Polarization makes life easier by drawing clear boundaries. Such a tendency of our brains to see clear patterns can, of course, be used specifically to emphasize differences that are not present in the data. Such overemphasis can be implemented by simple means, such as cutting off a scale. The boundaries from focusing to manipulation are blurred here. The cropping forms an enlargement of the actual image. To illustrate this, here are two representations of values in a bar chart – one with a heavily cropped Y-axis and different colour markings and one with the complete scale from 1 to 100 and the same colours (Fig. 5.3).

The truncation of the Y-axis highlights differences much more than they actually are in the percentages, namely 62% versus 54%. To avoid such manipulations, Datawrapper, for example, always lets the scale of the Y-axis start at 0. Only at the top is there the possibility to limit it.

¹ Rosling (2018, p. 33). The graphic on the endpaper of the book is available as an interactive representation at www.gapminder.org. An implementation of the graph in the R programming language has been developed by Keith McNulty: McNulty (2019).

Trimming Scales

Scale with trimmed y-axis



Scale with full y-axis

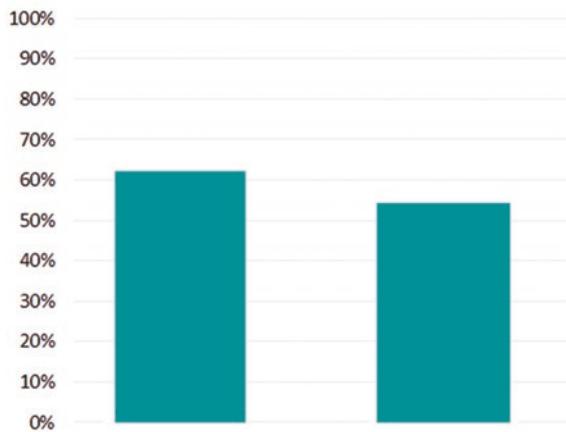


Fig. 5.3 Trimming the Y-axis amplifies the differences. (Source: Own representation)

Excel, on the other hand, presents a cropped view with the same data and can only be expanded to a full scale by intervening in the axis options.

However, there is a lot of potential for misinterpretation or deliberate manipulation not only in the cutting of the scales, but already in the preparation of the data. Hans Rosling shows this with the example of the use of averages.

Averages can emphasize differences that are not present in this form in the raw data. Rosling explains this with the example of math skills of women and men. Multi-year comparisons reveal differences in math skills between women and men based on SAT (Scholastic Assessment Tests) tests in the US. Men perform slightly better than women, although the gap has been steadily narrowing since the 1980s. However, when looking at the within-year (in this case 2016) distribution rather than the yearly averages for math scores, a picture of two hump-shaped distributions emerges that are slightly skewed. The majority of women are just as proficient in math as men. Polarities built up by means of averages, his message goes, are therefore not always a source of insight. They are especially not so when there is a normal distribution between the two poles, as in this case (Rosling 2018, pp. 40–41).

Deliberate manipulation or an invitation to misinterpretation can also occur through the use of colors. The meaning of colors is deeply anchored in our culture: we see red, make blue or are green with envy. The use of these colors in a graphic cannot be separated from the context of color meanings. It is debatable whether colors have a universal meaning, or whether their roots lie in our respective cultures. Probably both aspects overlap (see the empirical study by Jonauskaite et al. 2020). For example, the meanings of colors differ by culture. In China, red is the colour of life, which is also said to bring good luck. In our Western European cultural circle, on the other hand, red stands for danger or aggression, among other things. It should therefore always be clear in which cultural context a graphic is used.

This also applies to color combinations and contrasts: For example, the color pair red/green stands for stop/go or negative/positive. In graphics, this is used in analogy to the traffic light system, for example, for rising and falling directions in price trends or dos and don'ts in recommendations for action. The use of these colours as accents will therefore always

have such a connotation. The designer should consider whether this is desired or possibly misleading.

The contrast red/blue has been given a further context by the climate debate: in this case red stands for warm, blue for cold. This has given rise to a new form of graphic representation: Series of vertical stripes representing temperature corridors. The climate scientist Ed Hawkins from the National Centre for Atmospheric Science (NCAS) at the University of Reading has made this form of graphic known with his Warming Stripes (Fig. 5.4): a colour bar stands for a specific average temperature segment. In this way, long-term changes can be made visible to the eye. As simple as the representation is, as complex are the calculations behind it, because each colour nuance stands for a certain temperature corridor in a previously defined period of time. Yet the result is so captivatingly intuitive and has been so successful that the graphic format, which was first introduced in 2018, is now available for a large number of countries and time periods and is still being further developed (cf. Müller-Jung 2019).

In June 2019, the endeavor resulted in the “Show your stripes” initiative, which now features graphics for nearly every country from 1901 to 2018 at <https://showyourstripes.info/>.

Ed Hawkins - Warming Stripes

Global temperatures 1850 - 2018

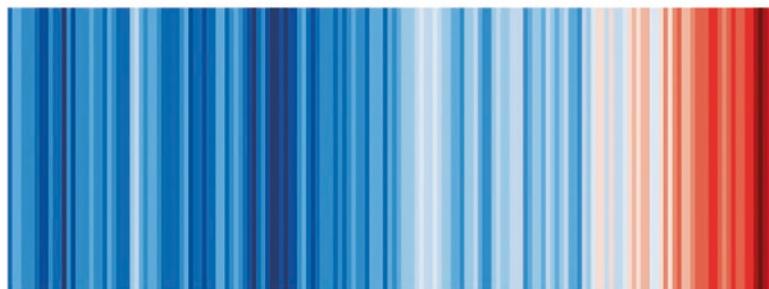


Fig. 5.4 Making long-term changes visible: Climate scientist Ed Hawkins' Warming Stripes document climate change. (Source: Wikimedia Commons [https://commons.wikimedia.org/wiki/File:20181204_Warming_stripes_\(global,_WMO,_1850-2018\)_-_Climate_Lab_Book_\(Ed_Hawkins\).png](https://commons.wikimedia.org/wiki/File:20181204_Warming_stripes_(global,_WMO,_1850-2018)_-_Climate_Lab_Book_(Ed_Hawkins).png))

The language of this representation is intuitive: blue stands for cooler temperatures, red for higher. As temperatures rise, it follows that the bar on the right-hand side of the image contains more and more red. This is a deliberately chosen color strategy that immediately makes the underlying narrative clear to any viewer: it's getting warmer.

In the meantime, climate change deniers have also taken up this colouring and are trying to counter this reading with a counter-strategy. Birgit Schneider, a professor of media ecology in Potsdam, is investigating the images of climate change and is working on a study for which she collects images of scientists and climate change deniers and analyses the respective colour strategy. She can prove that climate change deniers rely primarily on the color blue. When they adopt red, it is to show how the opponents, i.e. the alleged ‘simulators’, have been deceived (Schneider 2020).

The more we know about these processing techniques and our brain's secret weaknesses, the better: when we know them, we not only read images and stories differently, but we also find new ways to find and tell stories. This also applies to the following heuristics:

- Everything used to be better: many narratives are based on this pattern. The dying of the forest, the distribution of wealth, the rise of crime. It's related to idealizing our memories and overweighting present woes in comparison. “In human memory, the good is often stronger than the bad,” psychologists Constantine Sedikides and John Skowronski write about this. One reason, they suggest, is that good memories have stronger affective content: Feelings associated with rosy nostalgia have more emotional power than feelings with negative thought content, according to the paper (Sedikides and Skowronski 2020). In Factfulness, Rosling goes against this reading and narrates the development of humanity as a process of steady improvement – less poverty, falling crime rates, more education, rising prosperity and less inequality.
- Extrapolation – the extension of the straight line: An obvious but often incorrect assumption is that an observed development continues to progress uniformly: This is not true for the outbreak of an epidemic, nor for the price trend of a stock. We find it difficult to predict expo-

nental growth or even the reversal of directions. For many years, the development of globalization also seemed to know only one direction: towards increasingly interconnected international value and sales chains. The financial crisis, the strengthening of China and the rise of nationalist movements in Europe and the USA have put the brakes on this development. The outbreak of the Corona pandemic at the latest showed how fragile this order is.

But can such developments actually be modelled using mathematical methods and thus forecasts developed for the further course? Curve fitting describes the procedure for describing a curve from data using mathematical functions – for example by interpolation or regression analysis. Most visualization programs, such as R, Matlab, and tools based on them, provide a range of curve fits for different scenarios. They lull us into the certainty that mathematical functions can predict developments.

- Simple numbers: A number without context means nothing. Are 805 million tonnes of greenhouse gases in Germany in 2019 a lot or a little? Only in relation does the number gain meaning: for example, in comparison to the previous year or over a longer period of time. Or in comparison to other economies, to the size of the population, etc.? The same applies to sales, profits of companies, to the evaluation of website statistics and engagement rates of social media posts: The comparison raises the anchor and determines the rest of the narrative.
- Generalization: Generalizing helps us to orient ourselves in everyday life and is an essential characteristic of quick thinking. We use it to transfer familiar patterns to new situations. However, it obscures the view of other perspectives. Therefore, it is sometimes useful to question the categories into which we divide our world: be it consumer habits, educational level or financial needs. It is worth looking at differences within large groups, as well as similarities and differences between groups. A critical look also applies to orders of magnitude: For example, is a majority made up of 51% or 99%? This is a big difference when it comes to assessing how relevant this group is compared to other groups.

- Confirmation bias: People tend to trust information that supports their attitudes. Information is selected, ascertained and interpreted in a way that fulfils or confirms one's own expectations. This is particularly the case in complex and emotional debates, such as climate change or elections. But confirmation bias also plays an important role when deciding on a product, because this usually consists of a bundle of benefits and cannot simply be compared with another product from a competitor. When analyzing and interpreting data, it is important to be clear about one's hypotheses as early as the search stage.² In communication, however, confirmation bias is used very deliberately. Brand strategies come into play here: they offer the reduction in complexity that System 1 "wants". Brands build trust by confirming and supporting us in our attitudes. They prevent system 2 from intervening and critically questioning how much sugar is in the spread or what is in the fine print of the insurance conditions.
- Standstill: We have no sense of slow change. So some changes extend over such a long period of time that they seem like stagnation to our perception. Our instinct then tells us that these things are unchangeable. For example, ideas about nations, cultures, and economies. Much data is available for longer periods of time, but media and communicators need many occasions to produce news. That's where shorter time periods help. Best-seller lists work like this: Weekly updates produce news, whereas if viewed over centuries, the Bible and Koran would always be ahead. How boring. But when assessing the relevance of current topics such as globalisation, industrialisation or digital transformation, it is worth looking at longer periods of time to see what momentum the terms have and how relevant they will be in the future. As in the example of the term field of artificial intelligence (see Sect. 3.3.3): Are mentions and search entries rising to new heights, is the curve flattening out or is it already sinking? Are there new topics from the field whose attractiveness is increasing or is the whole technology already heading for the next AI winter? And of course, one should also

²Then a positive test strategy can help to limit the set of all possible test parameters to a plausible and practically verifiable selection, cf. for example Pohl (2004).

be aware that the viral course with a hump-shaped development is again only an assumption from the pattern of an infection course.

- Simple is attractive: Simple solutions appeal to our System 1. Finally getting clarity about a connection triggers feelings of happiness. The world becomes explainable and causes can be traced to an effect. The idea of the free market must declare all responses of government intervention to be false. It simply eliminates all arguments from the opposing side. If we have only one tool, we have a solution for everything.
- The Availability Trap: Anyone who has ever read statistics about burglaries in their neighborhood tends to overestimate the frequency of such incidents. We estimate the likelihood of something happening based on the examples that come to mind. It is a shortcut to estimating the situation. Anyone can fall into an availability trap by looking at numbers and data and using existing material. For example, does the engagement rate of my social media posts allow me to make a solid statement about the success of my content strategy, or do other factors need to be included? Such as: the target groups reached, their purchasing power, the conversion rate, etc.?
- Survivorship Bias: Successes are systematically overestimated because they are more visible than defeats. The term goes back to Allied engineers in the Second World War. They examined the bullet holes of fighter planes. These were mainly in the area of the wings, tail units and in the middle of the fuselage. Therefore, they initially reinforced these areas, but this did not improve the rate of return. From this the insight was formed that just the places must be strengthened, which were not affected with the returning airplanes: above all pilot pulpit and engines. Because these weak points ensured that the other aircraft did not return.

The examples show that even if a part of our thinking prefers simple and quick truths, we do not have to accept them immediately. A close look at the data can protect us from misinterpretation and help us discover new narratives. It helps to check whether the data really support our interpretation or whether we have fallen for false certainties and the distortions of our own perception.

5.2 Machine Bias

The more machines take over the processing of numbers, texts and images, the more we also hand over part of the power of interpretation to them. However, a machine is only as good as the data it works with. Among computer scientists there is the saying: “Garbage in – Garbage out”. But the garbage we produce with data can also be toxic. Jim Balsillie, founder of Blackberry’s manufacturer Research in Motion (RIM), sums up this danger with a drastic comparison: “Data is not the new oil – it’s the new plutonium. Amazingly powerful, dangerous when it spreads, difficult to clean up and with serious consequences when improperly used.”³ For Balsillie, this danger is particularly great when the data ends up in the hands of large technology corporations: Corporations like Facebook and Google would even replace the press as the fourth power in the state and are a danger to liberal democracy.⁴ Even if one might consider this view to be exaggerated, it is worth taking a look at the toxic side effects of our data-centric economy.

When processing data, there are two main sources of error: the selection and the processing of the data. The biases that arise from these errors are collectively known as machine bias or algorithmic bias. One of the originators of the critical view of computer science was Joseph Weizenbaum, who analyzed the methodological basis of bias in his 1976 book *Computer Power and Human Reason* (Weizenbaum 1976).

Prejudices can already be contained in the selection of data. Many methods of data processing are also based on interpretations, sometimes and possibly very often also on premises that are not highlighted clearly enough or are even erroneous. As soon as machines automate the processing of large amounts of data, this effect is potentiated. This is why a critical look at the sources and their processing is so important. In the case of data that a company collects from its own sources, it is still easier to trace the origin and therefore also its meaningfulness. With data sourced from

³ Balsillie before the International Grand Committee on Big Data, Privacy and Democracy in Ottawa on May 18, 2018.

⁴ “Technology is disrupting governance and if left unchecked could render liberal democracy obsolete. By displacing the print and broadcast media in influencing public opinion, technology is becoming the new Fourth Estate.” ibid.

third parties, this is much more complex. The more Big Data consists of recycled and highly aggregated data, the greater the risk of such errors.⁵

Data selection: A common argument for data mining with algorithms is that human bias and prejudice play no role in decision making. Yet data mining, if used without reflection, can reproduce the very patterns of discrimination, biases of decision makers or even of parts of society. Decisions, for one, can be made on the basis of inaccurate, incomplete, or even non-representative data. Even high-quality data can contain statistical biases if they do not accurately represent the proportions of certain groups. For example, to detect road damage, the city of Boston has developed an app that uses a smartphone's accelerometer. If someone drives through a pothole, the app records it. However, as efficient as this approach is, there is a risk that the recording will disadvantage poorer parts of the city because fewer people there have smartphones with corresponding sensors.⁶

Second, the data itself may already contain biases. Many intelligent systems today learn from data that comes from the internet and social media channels. Biases contained therein are then reproduced in the processing by the algorithms or AI. A well-known case is the scandal Google produced with an earlier version of its facial recognition software, Google Photos. In 2015, the software classified images of dark-skinned people as "gorillas." Google was shocked and improved the system. But Google is not alone in the bias. Other systems also recognize white faces better than anyone else. For example, facial recognition software from Microsoft, IBM, and the Chinese company Face++ is also particularly good at recognizing white males (Kaltheuner and Obermüller 2018).

Meanwhile, language has overtaken image recognition as the branch of AI with the greatest appetite for data and computing power. Again, there are numerous biases reproduced from the data used and incorporated

⁵ "Nowadays, we don't have direct data. We are recycling data and using proxy data – things like how you click on websites, what you purchase, what you say on Twitter, who your friends are on Facebook – to infer things that we are interested in. The promise of big data is that we will be able to use all this proxy information to determine with increasing accuracy the things that we care about." Interview with Cathy O'Neill – Burack (2017).

⁶ See, for example, the detailed account of errors in data selection and preparation in Barocas and Selbst (2016), here p. 685.

into further processing. Open AI, a non-profit organization that addresses the “existential threat of AI,” also develops its own open-source software. One of their most powerful tools is Generative Pre-Trained Transformer 3 (GPT-3), a software that writes text. The system is capable of producing impressive literature of a wide variety of genres – short stories, sketches and even poems – that are sometimes barely distinguishable from texts written by humans.

The system is based on a model that calculates and creates words, sentences and paragraphs based on statistical predictions. And therein lies the problem: statistical methods are no substitute for a coherent understanding of the world. GPT-3, like all other AI systems, has no internal model of a world, its values, or its narratives. Therefore, it cannot ultimately provide reasoning that requires such a model.

Rather, this system also reproduces the prejudices with which it has been fed. And since it obtains the necessary quantities of texts from the Internet and social media channels, its “view” of the world also comes from these sources. Because of its statistical approach, texts containing terms such as “black,” “Jew,” “woman,” and “gay” are often associated with prejudices that include racism, anti-Semitism, misogyny, and homophobia. Given the amount of training data required for such a language engine, any attempt to establish adequate quality control is doomed to failure.

The processing of the data: Biases do not only arise from the data base. Further processing via algorithms or AI also make it necessary to ask questions about the ethics, transparency, and accountability of the systems used. Pro Publica’s editorial board investigated software widely used in the US criminal justice system to determine whether a convict should receive parole. Because the software company was unwilling to disclose the details of how the software worked, Pro Publica used convict data to do a kind of “reverse engineering” of the software system. It found that this software reproduced racism. The software rated the likelihood of people of color to recidivate significantly higher than for whites (Angwin et al. 2016).

Cathy O’Neill has written a book about the biases that US government agencies and corporations use to make decisions based on data, thereby cementing prejudices and disadvantaging social groups. In

Weapons of Math Destruction, she shows how these procedures are used in the labor market, education, banking and insurance, policing and justice, and, of course, advertising (O’Neil 2017). All procedures are based on large amounts of data and the belief of authorities and companies that by modelling this data they can not only automate decisions but also eliminate human influences. But the data collected also contains biases on the basis of which decisions are made about loans, access to education, or even arrests. Thus, the exact opposite happens: namely, the software draws its conclusions based on the data put there by humans and their human-programmed processing, which not only contains errors, but also reproduces them and thus further reinforces them.

Reminiscent of the movie Minority Report is the controversial PredPol system that the LA Police Department implemented in 2011, which was only suspended after massive criticism under the pretext of the Corona crisis. PredPol is crime-fighting software that uses historical data to calculate the likelihood of future crimes. The system remains in use in many police departments around the world. According to the manufacturer, only three data points are needed for prediction: Type of crime, location and time. No demographic or ethnic data would be used in the analysis (Lobe 2019, p. 207). But they are not needed, because they are already included in the three data points through correlations.

The reason why US-American examples are listed here is that automated decision-making processes in politics, administration and business are particularly widespread there and correspondingly many examples of such distortions are also documented. The questions that arise from this are also relevant for communication and marketing managers in German companies:

- Anyone using a large collection of sources from the internet and social media for their analysis should be aware that these sources contain bias.
- Many tools for automated speech, image, and text recognition and prediction using AI may contain processing errors that arose from the test data used.
- When analyzing their own data (such as CRM, website, shop, sales, service), communication and marketing managers should always ask themselves how representative these results are and where there may be

blind spots, for example because certain user groups are underrepresented. This can lead to certain target groups not even coming into view because they have hardly come into contact with the company so far and have therefore not left any relevant data.

That's why it's important that today's pythias and priests in companies understand how the data they work with came about and what the premises are. They need to be comfortable with technologies that allow them to tap into and interpret their own data sources. They need knowledge of sources, market knowledge of common tools, methodological skills in analyzing and preparing data. And they should be familiar with common heuristics.

5.3 Ethical Issues of the Data Oracle

Dealing with the distortions of our perception quickly leads into the realm of ethics. For it is generally about questions that guide our actions: What goals are we pursuing? What reasons do we have for doing so and what means do we use to achieve them? Precisely because data has such a great influence on our world through digital technologies, it is necessary to consider developments also in terms of those problems and conflicts that are triggered by them. The discussion about ethical issues in dealing with information technology is as old as the discipline itself. Norbert Wiener, one of the founding fathers of artificial intelligence, already addressed the ethical dimensions of his discipline in his standard work *Cybernetics* (1948). Even during the war, Wiener foresaw enormous social implications of technology. In this context, he spoke of a second industrial revolution with enormous potential for good and evil. It was clear to Wiener that this would bring with it a host of ethical challenges and opportunities. In his subsequent works, he explored a range of ethical issues that computers and information technology would raise. He identified a variety of issues that continue to inform data ethics today: These included implications for security, for the labor market, for the responsibilities of computer specialists, for information networks and globalization, for virtual communities, telecommuting, transhumanism,

and robot ethics (Wiener 1948, pp. 169–180; Bynum 2018). However, the topic did not really go viral until the beginning of the new millennium, when the impact of technologies was clearly felt with the advent of cloud technology and the enormous availability of data (Big Data) (Fig. 5.5).

The more data is traded as a commodity of the twenty-first century and processed by machines, the more important it becomes for data subjects to know what happens to their data. Questions about informational self-determination are more relevant today than ever before. Business models worth billions are based on the analysis of user data, the Internet of Things is increasingly finding its way into our everyday lives in the form of wearables or smart homes, gigantic amounts of data are managed in data warehouses and we can all pay for supposedly free services by giving up our data.

Data Ethics in the Context of Big Data

Percentage of search terms (n-gram) (*1E9) in Google Ngram Corpus English 2019

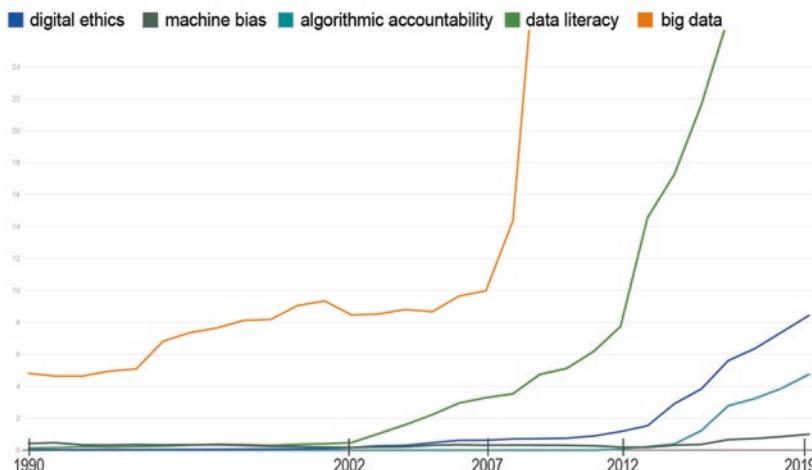


Fig. 5.5 For a long time, the ethical handling of data was a topic for specialists. The topic only gained greater relevance at the beginning of the new millennium, when with the advent of cloud technology and the enormous availability of data, the effects of the technologies became clearly noticeable. Interactive graphic at www.data-storyteller.de. (Source: Google Ngram Corpus English 2019)

With its growing importance, data ethics has also left the field of specialists and gained general relevance. At the latest with the revelation of the activities of the US intelligence services by Edward Snowden in 2013, governments have recognised the importance of the global cloud infrastructures of the data economy. Even if the focus here was not primarily on the question of personal rights, but rather on the defensibility of states, this has initiated a rethink. The fact that the preoccupation with the ethical consequences of technological developments has arrived in the industry is also documented by the market research company Gartner, which has selected “digital ethics” as the top technology trend for 2019.

And things are also moving at the political level: With the General Data Protection Regulation (GDPR), which came into force in May 2018, Europe has taken a stand for the protection of personal rights and thus set global standards. The German government has launched an AI Enquiry Commission as well as a Data Ethics Commission. In October 2019, the Data Ethics Commission presented its report, which contains 75 recommendations on the handling of data and algorithmic systems, including artificial intelligence (Report of the Data Ethics Commission 2019).

5.3.1 The Protection of Privacy

Since May 2018 at the latest, the next page on the Internet is no longer one but two clicks away. In the meantime, we have long since become accustomed to first confirming the privacy notices of each website when surfing the internet before moving further on the page. This is the most visible, but by no means the only consequence of the introduction of the General Data Protection Regulation (GDPR).

The regulation, which is valid throughout the EU, is primarily concerned with the personal and fundamental right of informational self-determination. It is derived, for example, from the general right of personality described in the German Basic Law (Article 2 (1) GG). This refers to the right of every individual to be able to decide for himself or herself on the disclosure and use of his or her personal data. This is in tension with economic interests, which are a major driver of value creation with the collection and further processing of data. Thus, the GDPR

tries to balance the protection of personal data within the European Union on the one hand and the guarantee of a free movement of data within the European Single Market on the other hand.

The GDPR is polarizing: While some are of the opinion that the rules overshoot the mark, others consider data protection to be a good thing. However, criticism of the regulation is not so much ignited by the protection of personal data, but by the poor implementation of the directive. Marco Zingler from the German Digital Economy Association (BVDW) criticised: “When it came into force in May 2018, the GDPR put a damper not only on the digital industry, but on the entire economy in Germany and Europe. Not because of overly strict data protection regulations. The most serious problem is the legal uncertainty caused by contradictory and unclear formulations of the regulation” (Eichsteller and Seitz 2019, p. 13). This uncertainty is also reflected in the results of the Digital Dialog Insights 2019 survey.

While less than a third of the study participants have doubts about the ROI of data-driven communication, concerns about the permissibility of data collection and uncertainty caused by legal developments are the main hurdles to the expansion of data-driven marketing communication (Fig. 5.6).

Challenges for Data-Driven Marketing Communication

Question: In your opinion, which topics pose the major challenges to the expansion of data-driven marketing communication?

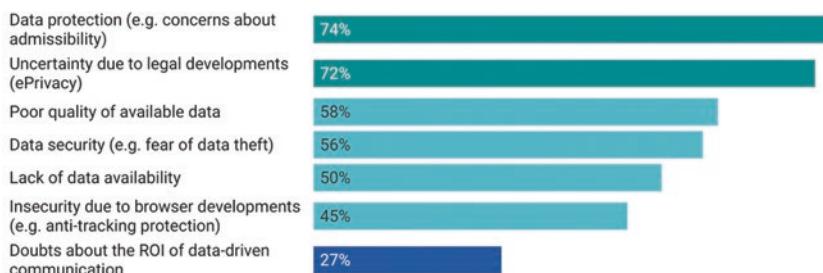


Fig. 5.6 Doubts about the legitimacy of data collection and uncertainty caused by legal developments are the main hurdles to the expansion of data-driven marketing communication. In contrast, only a few question the ROI of data-driven communication. (Source: Eichsteller and Seitz 2019)

Criticism has also been levelled at the high costs and the not always practical requirements that implementation of the requirements has entailed, especially for smaller companies. However, it is an important milestone for securing the rights to one's own data and a step in the right direction – namely positioning the EU as a pioneer in the protection of personal rights in the digital age. The GDPR is based on “privacy by default”: primarily, strict protection of personal data applies. Only with the consent of the data subject may these be weakened. The General Data Protection Regulation is generally aimed at the handling of data – online and offline.

The GDPR is shaped by the idea of a single European market for data. Thus, data may only leave the EU if companies have adequate security measures in place or if the destination country has an “adequate level of protection”. Such a level of protection is no longer provided by the EU-US data protection agreement, as ruled by Europe's highest court, the ECJ. In their ruling of 16 July 2020, the Luxembourg judges found that the so-called “Privacy Shield” is invalid. When transferring data of European consumers to a third country, a level of protection at the level of the GDPR had to be maintained. The US legislation does not provide a basis for this. They demand that the transfer of personal data must comply with the level of protection required by the EU.

But a regulation does not create a market. The EU is therefore developing the next steps for a single market for data, along the lines of the single market for goods. The EU website says: “The European Data Strategy aims to put the EU at the forefront of a data-driven society. By creating a single market for data, it will be possible to share it within the EU and across sectors for the benefit of businesses, researchers and public administrations.” (European Commission 2020). A European cloud solution is also intended to contribute to this: with GAIA-X, representatives from politics, business and science are developing a proposal for the design of a data infrastructure for Europe. The aim is to create a secure and networked data infrastructure that makes Europe independent of non-European cloud providers and promotes innovation.⁷ The question

⁷ See the website of the Federal Ministry for Economic Affairs and Energy <https://www.bmwi.de/Redaktion/DE/Dossier/gaia-x.html>

remains whether the EU will succeed in standing up to the de facto dominance of US and Chinese cloud providers.

After all, when it comes to protecting personal data and limiting the flow of data, the GDPR has found many imitators. India keeps payment information within its own country and may soon require that certain types of personal data not leave the country. Russia requires that data be processed and stored on servers within its territory. China blocks most international data flows. Most importantly, California has enacted the California Consumer Privacy Act (CCPA), which went into effect in January 2020, a regulation that rivals the European model in importance and scope ([Economist 2019, 2020](#)).

5.3.2 Algorithmic Accountability

In addition to collecting data, companies are also increasingly in the public eye when it comes to dealing with the machine processing of data. The question of responsibility for machine results and their use in public authorities, politics and companies is gaining social relevance. The communications researcher and computer scientist Nicholas Diakopoulos has coined the term “algorithmic accountability” for this. His report on the study of algorithms as black boxes appeared at the beginning of 2014, and since then the term has had a firm place in the public debate (Fig. 5.5).

In it, Diakopoulos describes a new task for journalists to understand software systems as objects of reporting. In addition to the programming code, the data sets with which these systems are trained must also be examined. Otherwise, their modes of operation cannot be understood ([Matzat 2017](#)).

The FAIR Data Initiative focuses on the processing of data by machines. It emerged from an association of data scientists who developed the FAIR principles in 2016. FAIR stands for

- Findable
- Accessible
- Interoperable
- Reusable

Germany, France and the Netherlands have taken up the initiative and created GO FAIR, a body to support it (Wilkinson et al. 2016).

The industry association for the digital economy, Bitkom, has made recommendations for the responsible use of AI and automated decision-making and advocates that companies develop internal guidelines on the use of algorithms. “This can take the form of corporate ethics, for example. Impact assessments that are developed in relation to the use of algorithms should be incorporated into the development of algorithms. An efficient and at the same time agile process must be anchored that regularly adapts these guidelines to new technologies and the issues that arise with them.” (Bitkom 2018).

The European Union has taken up the issue and is in the process of defining the framework. In the 2019 study “A governance framework for algorithmic accountability and transparency”, it outlines four policy areas for action:

1. Awareness raising: Education, watchdogs and whistleblowers,
2. Accountability in the use of algorithmic decision making in the public sector,
3. Regulatory oversight and legal liability in the private sector, and
4. The global dimension of algorithmic governance (European Parliament 2019).

5.3.3 Voluntary Commitments

The most important step in achieving algorithmic accountability is the willingness of companies to take legal and ethical responsibility for it. First and foremost, this includes an understanding that there are sources of error and bias in this area. Awareness of this, however, has room to grow, as a study by consultant Deloitte shows. In its 2020 AI study, respondents see algorithm bias as much less relevant than, say, security concerns and transparency. It is little consolation that the sensitivity of the respondents from Germany is two percentage points higher than worldwide – especially since the respondents were experts in companies that already use AI technologies (Fig. 5.7).

Major Ethical Risks in the AI Context

Survey among AI experts in companies with more than 500 employees

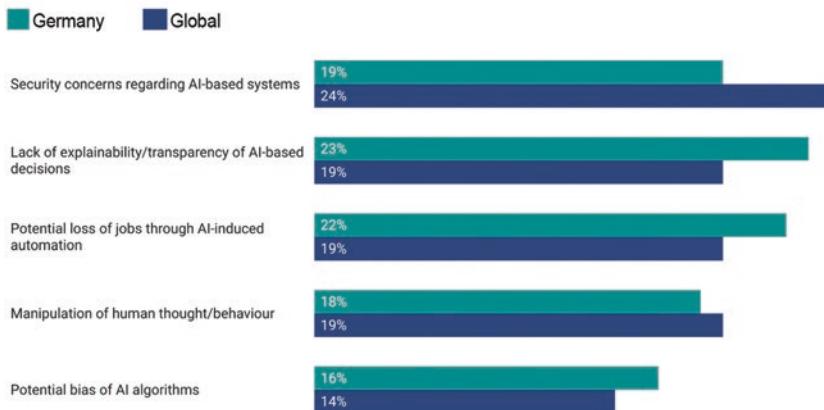


Fig. 5.7 Security concerns and lack of transparency dominate the evaluation of AI, while a possible bias of the algorithms is assessed as less relevant, Germany n = 201, Global n = 2737. (Source: Deloitte 2020)

One of the first corporate initiatives in this area was the Partnership on AI initiative (<https://www.partnershiponai.org/>) by the major US tech companies. Amazon, Alphabet, Facebook, IBM and Microsoft announced it in September 2016. It has since been joined by more than 100 partners. “In support of our mission to benefit people and society, the Artificial Intelligence Partnership conducts research, organizes discussions, shares insights, provides thought leadership, consults with relevant third parties, answers questions from the public and media, and creates educational materials that advance understanding of AI technologies, including machine perception, learning, and automated thinking.” Right after safety-critical AI (especially in health and transportation), Machine Bias is addressed here as one of six topic areas: fair, transparent, and accountable AI – This is about uncovering hidden assumptions and biases in data, especially in the disciplines of biomedicine, healthcare, security, criminal justice, education, and sustainability.

Salesforce, a US cloud software provider and member of the initiative, created a dedicated position within the company for the ethical use of

data in August 2018. Kathy Baxter is the architect for the ethical application of artificial intelligence (AI). To do this, Baxter works with Salesforce's research team, which develops the models for the company's AI application, Einstein. Here are the questions:

- Is there enough training data available?
- What are the possible reorganizations that we are shaping with the model?
- And how do we mitigate that?

Subsequently, the product teams consider how to deliver a responsibly usable product to the customer. Particularly sensitive information such as a person's age, ethnicity, and gender cannot be used to make decisions in regulated industries (health insurance and financial services) in the United States. But even if these fields are disabled, it may be possible to draw relatively accurate conclusions about these details through correlations with, say, place of residence. A module in the Salesforce software therefore checks precisely these correlations between the fields and identifies potential problems (Kutsche 2019).

In Germany, the debate on voluntary commitments by companies is taking place under the term Corporate Digital Responsibility (CDR). The term ties in with corporate governance and corporate responsibility structures that are already established in many German companies. Since 2015, various companies have been developing their understanding of a responsible approach to digitalisation and data. The paths and approaches are as diverse as the companies themselves.

Companies from the retail, financial services, IT, media and telecommunications sectors have emerged as the main drivers of this topic. These either have a particularly large amount of data or particularly sensitive data, or play a role in the further processing of data or the development of digital products. Depending on the market and the company's business model, the focus is either on the handling of personal data in general or on the algorithmic processing of data in particular.

In 2018, the Federal Ministry of Justice, under the leadership of Katarina Barley, launched a CDR initiative in which representatives of Telekom, Miele, Otto Group, Telefónica Deutschland as well as DIE

ZEIT participated and to map out the framework conditions of corporate digital responsibility (cf. the press release of the Federal Ministry of Justice “Responsibly shaping digitalisation” of 9 October 2018).

These considerations are based on the common understanding that sustainable development begins where legal requirements end. As a voluntary commitment by companies with a (partly) digital business model, CDR addresses the side effects and risks of digitalisation that the community and individuals currently have to deal with, including the following

- the “unethical” use of customer data and corruption of digital self-determination,
- a non-transparent digital world whose rules cannot be controlled by the community, e.g. the functioning of artificial intelligence

In other words, it is equally about exploiting opportunities and averting risks. The conviction behind this: Digital ethics, along with technological competence, is part of the foundation of innovation and determines the economic success of a company. The German Digital Economy Association (BDWV) has taken up the topic and developed a framework to support companies in the strategy and implementation of CDR (see <https://www.bvdw.org/der-bvdw/gremien/corporate-digital-responsibility/cdr-building-bloxx/>).

5.3.4 Data Literacy

But how can such a large topic be implemented in practice? Ethical standards can only be established if there is a broad discussion and corresponding competence in wide circles of society. What is required, therefore, is a responsible approach to data. Just as reading, arithmetic and writing are among our basic skills, in the information age we need skills for dealing with data. The term data literacy, best translated as data competence, is used to describe the abilities to deal with data appropriately, to interpret it and to present the results. “Data Literacy is much more than a broad and deep detailed knowledge of constantly changing

methods and technologies. Rather, the dimension of data ethics, motivation and value attitude plays a central role in being able to deal with data successfully and confidently in the future.” This is what Katharina Schüller, Paulina Busch and Carina Hindinger write in their paper Future Skills: a Framework for Data Literacy.

This requires skills from several disciplines – mathematics, statistics and programming. In their interplay, they help us achieve the necessary competence in dealing with data. Thus, data literacy is considered a key competence of the twenty-first century, essential for our society and working world. Not at all a specialized discipline of computer scientists and statisticians, but a cross-disciplinary skill for students of all disciplines, the authors urge. “The process of knowledge creation involves several steps:

- A. Establish data culture
- B. Provide data
- C. Evaluate data
- D. Interpret results
- E. Interpret data
- F. Deduce action

In order to systematically create knowledge or value from data, the ability to deal with data in a planned manner and to be able to consciously use and question it in the respective context will therefore be of decisive importance in all sectors and disciplines in the future.” (Schüller et al. 2019, p. 10).

The communications disciplines have not yet played an active role in establishing ethical data standards and building data literacy. The topic has also not yet found its way into the standards of the industry associations. Neither the Press Council, the Association of Professional Journalists, the German Advertising Council nor the German Council for Public Relations deal with data.⁸ In the sub-discipline of data journalism, on the

⁸ See Press Code of the Press Council <https://www.presserat.de/pressekodex.html>, Code of Ethics of the German Association of Professional Journalists DFJV <https://www.dfv.de/ueber-uns/ethik-kodex>, Guidelines of the German Advertising Council <https://www.werberat.de/content/leitfaden-zum-werbekodex-des-deutschen-werberats> and German Communication Code of the DRPR <https://drpr-online.de/deutscher-kommunikationskodex/>

other hand, there is no code, but there is a great sensitivity towards the collection and processing of data. Here, people are very well aware of the dangers. Behind this is the conviction that visualizations are only trusted if the process of their creation is as comprehensible as possible. In addition to citing sources, describing the research, selecting and preparing the data, access to the raw data also plays an important role in transparency (Greussing 2019). It is therefore most likely that important impulses for the communication disciplines and their handling of data can be expected from this side.

5.4 Less Is More: An Opportunity for Storytelling

The collection of personal data is to be further restricted by the European ePrivacy Regulation (ePVO). Even though it will not come into force until 2022 at the earliest, the course has long been set for a world without cookies from third-party providers (so-called third party cookies). The interpretation of the GDPR by courts and the default settings of browser providers have now largely anticipated the effects of the regulation.

The regulation aims to ensure that the analysis of user data through cookies is only permitted if users have given their express consent beforehand. Why does this have such major consequences? In a nutshell: This will deprive the advertising industry and advertising companies of what is currently their most important lever in the collection and personalised targeting of advertising.

If you do not click on “accept all” when entering a website, you will get an overview that looks something like Fig. 5.8. The user can decide which categories of cookies he accepts. Information from necessary, performance-oriented and functional cookies remain with the website operator, the cookies marked with advertising are the so-called third-party cookies, which pass on information to advertising partners and their partners.

Since third-party cookies can be used to track users’ surfing behaviour across many websites and the collected data can also be resold, they are in the focus of criticism. On this point, the ePrivacy Regulation is intended to further specify the general “Privacy by Default” requirement of the

Data Protection: First Party and Third Party Cookies

First Party Cookies

These cookies are set on the website where a visitor is currently surfing. They are not shared with third parties.

- Necessary:** These cookies are necessary for the functionality of the website.
- Performance:** These cookies collect statistical information of website usage and are used to measure and improve performance.
- Functional:** These cookies are used to improve functionality and enable personalisation. These include language and location preferences, the collection of browser settings, also information for the integration of shop systems, live chats and videos

Third Party Cookies

- Advertising:** These cookies are used to track a visitor's surfing behaviour over a longer period of time. They are used for targeted advertising and are inserted on the website by third-party providers.

Fig. 5.8 The user can decide to whom his data may be passed on. (Source: Own representation)

GDPR and apply it to digital channels.⁹ This stance is also reflected in two rulings: in October 2019, the European Court of Justice (ECJ) ruled that users must actively consent to the setting of these cookies. In May 2020, the German Federal Supreme Court (BGH) reaffirmed this ruling and ruled that non-essential cookies may not be activated on websites without the user's consent.

The major browser providers Apple with Safari and Mozilla with Firefox have already rejected third-party cookies. Google wants to follow suit with Chrome 2022. This poses dramatic challenges, especially for performance-based business models: Agencies, data traders, media operators and marketers will lose an important basis for their business.

But also all companies that want to address their customers more specifically by means of data are affected by this. Following a user through his specific customer journey is no longer possible. Digital campaigns can then only be implemented on the remaining data-based models and in

⁹ It states in recital 30: "Natural persons may be associated with online identifiers such as IP addresses and cookie identifiers provided by his or her device or software applications and tools or logs, or other identifiers such as radio frequency identifiers. This may leave traces that can be used, especially in combination with unique identifiers and other information received by the server, to profile and identify individuals." (<https://dsgvo-gesetz.de/erwaegungsgruende/nr-30/>)

the environment planning with the classic advertising media. The companies that will benefit most from this are those that have already collected a large amount of their own data, especially Amazon, Facebook, and Google.

This weakens the position of online retailers, e-commerce platforms and all other companies that use digital channels to contact their target groups. Personalized communication will rely more on predictions and will have to develop appropriate models. The industry is in the process of developing alternative personalization models that do not require user identification. Semantic techniques are increasingly being used in the contextual analysis of advertising fields. This used to be called contextual targeting, but has become more in-depth and meaningful with the semantic analysis techniques shown in Sect. 3.3.4. Whereas in the past it was certain keywords that appeared in a text, today text content can also be evaluated statistically and linguistically. With these methods, content can be analysed, meaning correlations can be established and linked to advertising messages. To do this, significantly larger volumes of data must be analysed in order to be able to form patterns for addressing target groups. Machine learning plays an important role here.

The exclusion of third parties forces a change of strategy in communication: away from the performance-focused approach with an eye on the last click towards a holistic view of the customer relationship. It breaks the fixation on the end of the sales funnel with one-to-one personalisation and creates space again for the development of customer groups and a process of permanent optimisation of the approach. The focus is then less on the channels and more on the interests and needs of the customers. All channels in the customer approach can thus be coordinated and mapped in a dramaturgy. In the future, this will lead away from rule-based marketing automation (with classic if-then rules) to an orchestration of all contact points and communication channels.

However, this is only possible if the company's own data is consistently collected and linked with one another. Communication managers are only on the safe side if they build up data records from their own sources (i.e. first-party data). This requires a planned approach, as this data cannot be bought, but has to be developed by the company itself. Knowledge about customers is available in many places in the company (especially sales, service) and stored in some systems (especially ERP and CRM). It

can also be collected and further enriched in the many contact points of the communication units (especially with data from the websites, social media channels and email marketing). Above all, it is important to have a common understanding of the objectives to be achieved by collecting and using the data. Restricting the use of third-party data is a good opportunity to take another general look at the reasons for building and expanding one's own data base and to start by looking for the questions that are to be answered by using the data. To use the image used at the beginning: There are enough sources of one's own that are worth developing and processing. For the statements of the oracle, however, it is not the quantity of smoke that decides, but the quality of the question and the abilities of the team of pythia and priests that delivers the answers. The latter should be aware that not only their own thought patterns are susceptible to distortion, but also the tools they use to process the data.

References

- Angwin J, Larson J, Mattu S, Kirchner L (2016) Machine bias. ProPublica (23 Mai) <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed: 23. Mai 2020
- Balsillie J (2018) Data is not the new oil – it's the new plutonium. Financial Post, 18 Mai. <https://business.financialpost.com/technology/jim-balsillie-data-is-not-the-new-oil-its-the-new-plutonium>. Accessed: 22. Jan. 2020
- Barocas S, Selbst A (2016) Big data's disparate impact. 104 California Law Review 671. <https://ssrn.com/abstract=2477899>. Accessed: 10. Aug. 2020
- Bitkom (2018) Empfehlungen für den verantwortlichen Einsatz von KI und automatisierten Entscheidungen, Berlin
- Burack C (2017) Algorithms are ‘existential threat’ to shared reality, says Cathy O’Neil, in: DW 21.08.2018 <https://www.dw.com/en/algorithms-are-existential-threat-to-shared-reality-says-cathy-oneil/a-40167802>. Accessed: 2. Mai 2020
- Bynum T (2018) Computer and information ethics. The Stanford encyclopedia of philosophy, HG. Edward N. Zalta. <https://plato.stanford.edu/archives/sum2018/entries/ethics-computer>. Accessed: 20. Aug. 2020
- Deloitte (2020) State of AI in the enterprise – 3rd Edition. Deloitte, Juni 2020

- Economist (2019) Companies should take California's new data-privacy law seriously. Economist, 18. Dezember. <https://www.economist.com/business/2019/12/18/companies-should-take-californias-new-data-privacy-law-seriously> Accessed: 23. Dez. 2019
- Economist (2020) Governments are erecting borders for data. Economist, 20. Februar. <https://www.economist.com/special-report/2020/02/20/governments-are-erecting-borders-for-data>. Accessed: 23. Febr. 2020
- Eichsteller H, Seitz J (2019) Digital dialog insights 2019. Bundesanzeiger, Köln
- Europäische Kommission (2020) Europäische Datenstrategie. Website der Europäischen Kommission, 19. Februar. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_de. Accessed: 21. Aug. 2020
- European Parliament (2019) A governance framework for algorithmic accountability and transparency, EPRA Brüssel. [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRA_STU\(2019\)624262_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRA_STU(2019)624262_EN.pdf)
- Greussing E (2019) Datenvizualisierung vom Publikum her denken. European Journalism Observatory. <https://de.ejo-online.eu/digitales/datenvizualisierung-vom-publikum-her-denken>. Accessed: 21. Aug. 2020
- Harari Y (2017) Homo deus: a brief history of tomorrow. Harper, New York
- Jonauskaitė D et al (2020) Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. Psychol Sci. <https://doi.org/10.1177/0956797620948810>. Accessed: 29. Sept. 2020
- Kahneman D (2011) Thinking, fast and slow. Penguin, London
- Kaltheuner F, Obermüller N (2018) Diskriminierende Gesichtserkennung: Ich sehe was, was du nicht bist. Netzpolitik.org. <https://netzpolitik.org/2018/diskriminierende-gesichtserkennung-ich-sehe-was-was-du-nicht-bist/>. Accessed: 9. Aug. 2020
- Kutsche K (2019) Wenn Daten in die Irre führen. Süddeutsche Zeitung, 17. Dezember. <https://www.sueddeutsche.de/wirtschaft/kuenstliche-intelligenz-wenn-daten-in-die-irre-fuehren-1.4726877>. Accessed: Sept. 2020
- Lobe A (2019) Speichern und Strafen. Die Gesellschaft im Datengefängnis, Beck, München
- Matzat L (2017) Algorithmic accountability. Der nächste Schritt für den Datenjournalismus Datenjournalist, 18. Dezember. <https://www.datenjournalist.de/algorithmic-accountability-der-naechste-schritt-fuer-den-datenjournalismus/>. Accessed: 23. Mai 2020
- O'Neil C (2017) Weapons of math destruction. Penguin Books, London
- Rosling H (2018) Factfulness. Sceptre, London

- Schneider B (2020) „Wir können nicht die ganze Zeit hinschauen“, Süddeutsche Zeitung Magazin. <https://zeitung.sueddeutsche.de/webapp/issue/SZM/2020-10/8/index.html>. Accessed: 10. März 2020
- Schüller K, Busch P, Hindinger C (2019) Future Skills: Ein Framework für Data Literacy, Hochschulforum Digitalisierung Arbeitspapier Nr. 47. Berlin
- Sedikides C, Skowronski J (2020) In human memory. Good can be stronger than bad. In: current directions in psychological science 29/1, S 86–91
- Weizenbaum J (1976) Computer power and human reason: from judgment to calculation. Freeman, San Francisco
- Wiener N (1948) Cybernetics. MIT Press, Cambridge
- Wilkinson MD et al (2016) The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3:160018. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>. Accessed: 21. Aug. 2020