

## Engines of Order

The book series RECURSIONS: THEORIES OF MEDIA, MATERIALITY, AND CULTURAL TECHNIQUES provides a platform for cutting-edge research in the field of media culture studies with a particular focus on the cultural impact of media technology and the materialities of communication. The series aims to be an internationally significant and exciting opening into emerging ideas in media theory ranging from media materialism and hardware-oriented studies to ecology, the post-human, the study of cultural techniques, and recent contributions to media archaeology. The series revolves around key themes:

- The material underpinning of media theory
- New advances in media archaeology and media philosophy
- Studies in cultural techniques

These themes resonate with some of the most interesting debates in international media studies, where non-representational thought, the technicity of knowledge formations and new materialities expressed through biological and technological developments are changing the vocabularies of cultural theory. The series is also interested in the mediatic conditions of such theoretical ideas and developing them as media theory.

#### **Editorial Board**

- Jussi Parikka (University of Southampton)
- Anna Tuschling (Ruhr-Universität Bochum)
- Geoffrey Winthrop-Young (University of British Columbia)

# Engines of Order

*A Mechanology of Algorithmic Techniques*

*Bernhard Rieder*

Amsterdam University Press

This publication is funded by the Dutch Research Council (NWO).

Chapter 1 contains passages from Rieder, B. (2016). Big Data and the Paradox of Diversity. *Digital Culture & Society* 2(2), 1-16 and Rieder, B. (2017). Beyond Surveillance: How Do Markets and Algorithms ‘Think’? *Le Foucauldien* 3(1), n.p.

Chapter 6 is a heavily reworked and extended version of Rieder, B. (2017). Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality. *Information, Communication & Society* 30(1), 100-117.

Chapter 7 is a reworked and extended version of Rieder, B. (2012). What Is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index. *Computational Culture* 2, n.p.

Cover illustration: The full text of this book, represented as a *feature vector*. © Bernhard Rieder

Cover design: Suzan Beijer

Lay-out: Crius Group, Hulshout

ISBN 978 94 6298 619 0

e-ISBN 978 90 4853 741 9

DOI 10.5117/9789462986190

NUR 670



Creative Commons License CC BY NC ND

(<http://creativecommons.org/licenses/by-nc-nd/3.0>)

© B. Rieder / Amsterdam University Press B.V., Amsterdam 2020

Some rights reserved. Without limiting the rights under copyright reserved above, any part of this book may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form or by any means (electronic, mechanical, photocopying, recording or otherwise).

Every effort has been made to obtain permission to use all copyrighted illustrations reproduced in this book. Nonetheless, whosoever believes to have rights to this material is advised to contact the publisher.

# Table of Contents

Acknowledgements	7
Introduction	9
<b>Part I</b>	
1. Engines of Order	25
2. Rethinking Software	51
3. Software-Making and Algorithmic Techniques	81
<b>Part II</b>	
4. From Universal Classification to a Postcoordinated Universe	145
5. From Frequencies to Vectors	199
6. Interested Learning	235
7. Calculating Networks: From Sociometry to PageRank	265
Conclusion: Toward Technical Culture	305
About the Author	347
Index	349



# Acknowledgements

This book has been long in the making and has benefited from many different inputs. I would first like to thank the Recursions series editors – Anna Tuschling, Geoffrey Winthrop-Young, and, in particular, Jussi Parikka – for their many valuable remarks and suggestions. Maryse Elliott from Amsterdam University Press has been an invaluable help in guiding me through the whole editorial process. Eduardo Navas’s constructive comments on the manuscript were much appreciated. I am also grateful to Carolin Gerlitz, Sonia de Jager, Janna Joceli Omena, Niels Kerssens, Emillie de Keulenaar, Thomas Poell, Gernot Rieder, Guillaume Sire, Michael Stevenson, and Fernando van der Vlist for reading drafts at various stages of completion and providing critical feedback.

I want to thank Thomas Brandstetter, Dominique Cardon, Mark Coté, Nick Couldry, José van Dijck, Nigel Dodd, Matthew Fuller, Paolo Gerbaudo, Paul Girard, Andrew Goffey, Olga Goriunova, Sanne Kraijenbosch, Camille Paloque-Berges, Jean-Christophe Plantin, Thomas Poell, Barbara Prainsack, Theo Röhle, Anton Tantner, Leon Wansleben, and Hartmut Winkler for conference and workshop invitations that allowed me to develop the ideas that run through this book. My thanks also go to my colleagues at the Mediastudies Department and the Digital Methods Initiative at the University of Amsterdam as well as my former colleagues at the Département Hypermedia and Laboratoire Paragraphe at Paris VIII University for the many stimulating conversations that shaped the following chapters.

Particular thanks are due to Richard Rogers and the Dutch Research Council (NWO) for making it possible to release this book through open access.

I dedicate this book to the memory of Frank Hartmann, whose passion for thinking technologies as media echoes through these pages.





# Introduction

## Abstract

The introduction chapter positions algorithmic information ordering as a central practice and technology in contemporary digital infrastructures, a set of techniques that serve as 'levers on reality' (Goody). While algorithms used in concrete systems may often be hard to scrutinize, they draw on widely available software modules and well-documented principles that make them amendable to humanistic analysis. The chapter introduces Gilbert Simondon's mechanology and provides an overview of the structure and argument of the book.

**Keywords:** algorithmic information ordering, information search and retrieval, mechanology, software-making

Over the last decades, and in particular since the widespread adoption of the Internet, encounters with algorithmic procedures for 'information retrieval' – the activity of getting some piece of information out of a collection or repository of some kind – have become everyday experiences for most people in large parts of the world. We search for all kinds of things on the open web, but also for products, prices, and customer reviews in the specialized databases of online retailers, for friends, family, and strangers in social networking services or dating sites, and for the next thing to read, watch, play, listen to, or experience in quickly growing repositories for media contents. There are at least three remarkable aspects to this spread of information seeking. First, computer-supported searching has sprawled beyond the libraries, archives, and specialized documentation systems it was largely confined to before the arrival of the web. Searching, that is, the act of putting a query into a form field, has become such a fundamental and ubiquitous gesture that a missing search box on a website becomes an almost disturbing experience. Second, what retrieval operates on – information – has come to stand for almost anything, from scraps of knowledge to things, people, ideas, or experiences. Digitization, datafication, and the capture of

always more activities in software are, in the words of Netscape founder and venture capitalist Marc Andreessen (2011), 'eating the world'. Search has become a dominant means to access and order the masses of digital and datafied bits and pieces that clutter the environments we inhabit. Third, the deliberate and motivated act of formulating a query to find something is only one of the many forms in which information retrieval nowadays manifests itself. Automated personalization, localization, recommendation, filtering, classification, evaluation, aggregation, synthetization, or ad hoc generation of information are similarly pervasive practices that do not require explicit user input to select, sequence, arrange, or modulate some set of digital items. And retrieval techniques are no longer limited to producing result lists: they generate scores, suggest items, discard or promote messages, set prices, arrange objects and people in relation to each other, assemble texts, forbid or grant access, fabricate interfaces and visualizations, and even steer objects in the physical world. In short, various activities or gestures this book addresses under the broad notion of 'information ordering' have become both pervasive and subtle in terms of how they operate in the thickening layers of digital mediation.

The proliferation of these algorithmic practices has been accompanied by considerable efforts in the humanities and social sciences to investigate techniques and applications in terms of power and social significance. Early analyses of search engines already highlighted their political dimension, claiming that 'there is no such thing as algorithms without their own weight' (Winkler, 1999, p. 36). This meant that one could examine 'the wide-ranging factors that dictate systematic prominence for some sites, dictating systematic invisibility for others' (Introna and Nissenbaum, 2000, p. 171) from a point of view concerned with social impact and public interest. Beyond search, authors have called attention to 'moments of algorithmic judgement' (Graham, 2005, p. 576) that abound when 'code-based technologized environments continuously and invisibly classify, standardize, and demarcate rights, privileges, inclusions, exclusions, and mobilities' (Graham, 2005, p. 563). Terms like 'automated management' (Kitchin and Dodge, 2011), 'algorithmic ideology' (Mager, 2012), 'algorithmic governmentality' (Berns and Rouvroy, 2013), and, more recently, 'algorithmic accountability' (Diakopoulos, 2015) all subscribe to 'the central premise that algorithms have the capacity to shape social and cultural formations and impact directly on individual lives' (Beer, 2009, p. 994). This broad recognition of the 'relevance of algorithms' is not, however, a symptom of a sudden curiosity for the fundamentals of computational theory. It stems from a more specific interest in the particular instances where algorithms serve as 'a means to know what there is to know

and how to know it, to participate in social and political discourse, and to familiarize ourselves with the publics in which we participate' (Gillespie, 2014, p. 167). Most of the techniques that sit at the center of these questions and concerns directly relate to the field of information ordering.

Search engines remain the most instructive illustration for the issues at hand since the tensions between their remarkable practical utility, their technical prowess, and their political relevance are so clearly visible. We intuitively understand that ranking web pages – and thus the services, contents, and viewpoints they stand for – is delicate business. But, as Grimmelmann (2009) argues, search engines face the 'dilemma' that they *must* rank in order to be useful. This imperative collides with the uncomfortable observation that there is arguably no technical procedure that can lay serious claim to producing assessments concerning ambiguous and contested cultural matters in ways that could be broadly accepted as 'objective'. In fact, whenever data are processed algorithmically, the transformation from input to output implies a perspective or evaluation that, through the coordination between data and what they stand for, is projected back into spheres of human life. Techniques for information retrieval become engines of order that actively intervene in the spaces they seek to represent (cf. Hacking, 1983).

The need to better understand the specificities of these processes becomes even clearer if we broaden the scope beyond everyday online experiences to activities where algorithms evaluate and inform decisions that can have dramatic effects, for example, in hiring, credit assessment, or criminal justice (cf. O'Neil, 2016; Christin, 2017; Eubanks, 2018). These emblematic and troubling applications point to a myriad of instances in business and government where procedures from the broad field of information ordering are used to inspire, choose, or impose a specific course of action.

The technical procedures involved are loaded, often implicitly, with specific ideas and attitudes concerning the domains they intervene in. Search engines evaluate the 'relevance' of information, news aggregators generate front pages according to various measures of 'newsworthiness', dating sites calculate 'compatibility coefficients' between members and order them accordingly, social networking sites filter friends' status updates based on quantified ideas of 'interest' or 'closeness', and microblogging services give prominence to 'trending' topics. In each of these cases, there is a framing of the application domain that implies various kinds of conceptual and normative commitments. This can involve a general allegiance to the broad epistemological 'style' (Hacking, 1985) of computation as a means of knowing; but it can also take more specific forms, for example, when

psychological research on partnership satisfaction flows into the design of a matching algorithm or when the optimization objectives for a machine learning system are being selected on the basis of business considerations.

At the same time, technical procedures are more than just a means to efficiently enact values and ideas that are themselves nontechnical. Jack Goody (1977) argued that list-making, from the start an essential part of writing, 'gives the mind a special kind of lever on "reality"' (p. 109) by supporting mnemonics and, more importantly, by facilitating different operations of ordering and reordering pieces of text and, by extension, the things these pieces refer to. As Goody knew all too well, the advent of list-making meant not just a quantitative extension in cognitive capacity. More fundamentally, it stimulated the production and recording of knowledge, spurred modes of classificatory and hierarchical thinking, and supported more complex forms of social organization. As Peters (2015) argues, '[i]n list writing, serial order loosens its hold' (p. 290), with wide-ranging consequences. The information ordering techniques that have become so pervasive today share the transversal character and broad applicability of list-making and may prove to have equally fundamental repercussions for how we construct and relate to the world around us.

Like list-making, algorithmic ordering comes with a genuine operational substance that rarely boils down to a simple transposition of a manual method into computational form. A web search engine, for example, orders documents through iterative processing of vast amounts of distributed signals and the specific way it produces an aggregate appreciation of these signals defines an epistemic substance and character that has little to do with the knowledge practices that have defined libraries, encyclopedias, or archives over the last millennia. As Edsger Dijkstra, one of the central figures in the history of software, remarked about computers over 40 years ago:

[T]he amount of information they can store and the amount of processing that they can perform, in a reasonably short time, are both large beyond imagination. And as a result, what the computer can do for us has outgrown its basic triviality by several orders of magnitude. (Dijkstra, 1974, p. 608)

Computers' capacity to run billions of data points through billions of iterations of small calculative steps means that they 'think' (Burrell, 2016) in ways that are not only opaque, but potentially *strange* and hard to fit into established categories. Techniques like machine learning, network algorithms, or relational database management systems are not just powerful

means to produce and apply knowledge, to enact value preferences, or to control practice; they participate in the very definition of what knowledge, value, and practice mean and *can* mean, both through the conceptual resources they propose to think with and the actual interpretations and orderings they generate when applied in practice. We should consider the possibility that they challenge cultural modes and social institutions in more fundamental ways than the necessary discussions of algorithmic opacity or bias can lead us to believe.

The methods and procedures involved in actual practices are often hidden from our sight by technical and legal means, latched not even in black boxes but somewhere in the 'black foam' (Rieder, 2005) of systems whose contours are hard to delineate. But, paradoxically, they have also become highly accessible, in the sense that concrete implementations draw heavily on open reservoirs of technicity and knowledge that find their expression in scholarly publications, software libraries, and communities of practice gathering on websites like Stack Overflow. These reservoirs are neither hidden nor closed off and we are free to examine a steadily growing archive of techniques that enable computers to accomplish tasks that seem increasingly 'cultural' or 'intelligent' in nature. This book is an expedition into this archive and more specifically into the areas that deal with information ordering.

The actual makeup of Google's search ranking may indeed be 'unknowable' for a number of practical, commercial, and legal reasons, but, as shown in Chapter 7, the content, history, and substance of its most famous algorithm, PageRank, stands wide open. We may never get access to the concrete specifications of the machine learning methods behind the personalized filtering Facebook applies to its users' News Feed, but we can ask, as in Chapter 6, where machine learning comes from, what concepts and ideas it builds on, and how it operates in general terms. The second part of this book is thus dedicated to a series of investigations into specific 'algorithmic techniques', that is, into the defined-yet-malleable units of technicity and knowledge developers draw on when designing the function and behavior of computers acting in and on the world. Offering many different ways to order and organize information, they serve as levers on the 'reality' of a world eaten by software.

While this book draws heavily on work situated in the 'cultural techniques' tradition, an approach coming out of German media scholarship, there is at least one important difference. Unlike Young's (2017) inspirational take on the list, which follows a particular cultural form through various societal settings, I examine a set of techniques as they traverse what is maybe not a single cultural domain but nonetheless a somewhat demarcated practice:

software-making. The broader theoretical perspective guiding these probes will be discussed at length in part one, but the particular focus on technical creation calls for some background and clarification.

## Toward Mechanology

This book is largely motivated by the remarkable spread of algorithmic information ordering but also translates a feeling of hesitation or uneasiness toward the way software is often presented and discussed in media studies and associated fields, or, more specifically, toward the emphasis on *code* as software's quintessential technical quality or substance. To be clear, understanding how written instructions produce machine behavior is fundamental to understanding software, but it is also a comparatively small step into the massive world of technicity software constitutes. Code is neither trivial nor transparent, but for any experienced developer it is a familiar means to access a domain of *function* that is vastly more complex than the term is able to address. Building a program or system is to craft a composite technical object, 'a being that functions' in the words of French philosopher Gilbert Simondon, who plays a central role in what follows. This may entail, today more than ever, the assemblage of many preexisting chunks of software. Code serves as the means to draw on an archive, to 'build-with', and to create in ways that are deeply relational and embedded. As I will argue over the following chapters, the world of software-making is structured around 'techniques', expressions of knowledge and technicity that enable developers to make computers *do* things that are more involved or complex than their 'basic triviality' suggests. This book does not presume any practical technical knowledge or experience, but it addresses algorithmic information ordering from the perspective of technical creation.

My own background plays an important role in this setup. While I have little formal training in any technical discipline, I have been developing software on a regular basis for a long time. I started to program when I was still in high school, worked as a web developer during my university studies, and taught programming to students ranging from beginners to computer scientists at master's level for about a decade. I continue not only to code but to make software, nowadays mostly in the domain of digital methods for Internet research (Rogers, 2013). The part of the software landscape under scrutiny in this book, algorithmic information ordering, is not only socially relevant but also closely connected to the technical practice I have been pursuing over the last 20 years. As a web developer, I worked

extensively with relational database management systems (Chapter 4) and I encountered advanced information retrieval techniques (Chapter 5) during my PhD in information and communication science at Paris 8 University when I was investigating the possibilities for ‘society-oriented design’ (Rieder, 2006). This work led to a system, *procspace* (Rieder, 2008), which used a variety of algorithmic methods to generate navigational pathways between documents to support a logic of connection, enrichment, and overview that breaks with the serial forms of order dominating search. The encounter with information retrieval, an established technical field that comes with a large body of well-documented methods, came as a shock: as an autodidact programmer I felt very comfortable when it came to writing code, but I was not fully aware how much I was missing. The techniques I discovered gave me a new sense of possibility and opened the door to forms of technical expression that have stimulated my imagination ever since. Although often more heavily mathematized than what I was used to, these techniques were relatively simple to implement and, like clay, could be modeled in countless ways. The entanglement between information ordering and the politically, culturally, and economically significant matters it is increasingly involved in became my principal research interest. This eventually led to work in digital methods, where I focused on studying online platforms that rely on algorithmic techniques in fundamental ways and, paradoxically, to a situation where I would apply similar techniques as analytical instruments to make sense of large sets of empirical data. The chapters about machine learning (Chapter 6) and network algorithms (Chapter 7) draw on this work.

The reason I mention these details is not to claim technical authority but to introduce and situate a perspective that has been fundamentally shaped by these experiences. This perspective is still uncommon in media studies and in the broader discussions of software or, to use the buzzwords of the day, of ‘algorithms’ or ‘artificial intelligence’. Following Johanna Drucker’s (2013) suggestion to give ‘[m]ore attention to acts of producing and less emphasis on product’ (n.p.), my conceptual vantage point is software-making, a series of practices that increasingly revolve around the use of packaged function as a means to extend programmers’ capabilities. It takes hardly more than an hour to install and set up PyTorch or TensorFlow, powerful open-source libraries for machine learning, and to have a first classifier trained. While some people will want to peek under the hood of these artifacts to make adaptations or simply out of intellectual curiosity, developers often draw on technicity and knowledge that they understand only in broad terms or not at all. What programming languages, software libraries, and similar

artifacts do is to enable software-makers to step further faster, not merely regarding resource efficiency but in terms of what can be considered possible in the first place. Such packages widen the spaces of expressivity, broaden the scope of ambitions, but also structure, align, and standardize. Spelled out, stabilized, and 'frozen', algorithmic techniques spread through technical imaginaries and artifacts, and further into application logics and business models. They are *means of production*, not simply outpourings of computational principles or scientific ideas.

Algorithmic techniques are ways of making computers do things, of creating function, and their history is characterized to a greater extent by accumulation and sedimentation than by paradigm shifts or radical breaks. Certainly, methods and approaches are regularly superseded or fall out of fashion, but it is clear that the archives that inform and constitute software-making have grown vastly over time. While this book entertains a somewhat complicated relationship with the field of media archeology, another prominent approach coming out of German media theory, it indeed follows a selection of techniques into their historical trajectories to excavate some of the fundamental ideas that resonate through our technical present. But throughout these historical probes, I strive to keep an eye on the possibilities for variation, combination, and divergence that invariably emerge when a technique becomes part of a concrete technical object. The developer, in contrast to the computer scientist, philosopher of science, or science historian, neither looks at the reservoir of techniques from below, as an emanation of foundational mathematical principles, nor from above, as outpourings of scientific progress. The developer is right in-between, surrounded by technicity coming in all shapes and forms, and thus '*among* the machines that operate with him' (Simondon, 2017, p. 18).

To interrogate technology both in terms of its fundamental nature and from the perspective of technical practice is the task Simondon laid out for 'mechanology', a discipline or mode of thinking that would serve as a 'psychology' or 'sociology' of machines (Simondon, 2017, p. 160), capturing their 'interior life' and 'sociability' in terms that do not reduce them to an exterior finality or effect. As a general science of technology, mechanology would approach technical function as human gesture, examine technical creation as mediation between human beings and nature, and interrogate the *values* implied in mechanical operation itself. This book, suffice to say, is an attempt to develop a mechanological perspective on software and to apply it to the engines of order that increasingly adjudicate (digital) life.



## Organization and Overview

The book is divided into two parts. The first part is dedicated to the theoretical and methodological foundations that inform and support the examination of four clusters of algorithmic techniques for information ordering in the second part.

The first chapter discusses central terms like ‘information’ and ‘order’, and it proposes the concept of ‘engine’ to point toward the infrastructural embeddings that have allowed techniques initially conceived for document retrieval to become pervasive mediators in online environments. While this book constitutes a humanistic exploration of technical substances rather than their practical application, the chapter pays tribute to the fact that the techniques under scrutiny have become prevalent in a specific situation, in *this* world and not another.

The second chapter then formulates a conceptual perspective on software, starting from an attempt to situate the project in relation to existing takes on the subject. But it is mainly dedicated to the presentation and appropriation of Simondon’s philosophy of technology, which reserves a central place to technical creation and evolution. Here, we find an understanding of technicity as a domain of life that constitutes its own substance and regularity, whilst remaining a fundamental form of human gesture. Simondon’s inductive view, which frames technology as multitude of technical objects rather than idealized *techne*, grounds the conceptual and analytical apparatus I then bring to the analysis of algorithmic techniques.

Chapter 3 builds on central ideas from Simondon’s work, such as the distinction between invention and concretization and the delineation of technical elements, individuals, and ensembles, to conceptualize algorithmic techniques as the central carriers of technicity and technical knowledge in the domain of software. In dialogue with the cultural techniques tradition, it addresses them as methods or heuristics for creating operation and behavior in computing and discusses how they are invented and stabilized. Algorithmic techniques, in this perspective, are at the same time material blocks of technicity, units of knowledge, vocabularies for expression in the medium of function, and constitutive elements of developers’ technical imaginaries.

The second part of the book then launches a series of probes into the history of algorithmic information ordering. These probes do not follow a single lineage or logic and cover different periods of time, but they come together in staking out an ‘excavation ground’ (Parikka, 2012, p. 7) that marks the 1960s and 1970s as the period where the fundamentals of contemporary

information ordering were laid out. While Simondon's understanding of technology as human gesture and my emphasis on adaptation and variation lead away from certain core tenets of media archeology, I seek 'to investigate not only histories of technological processes but also the current "archaeology" of what happens inside the machine' (Parikka, 2012, p. 86). The goal is to excavate select roots of an increasingly technological present. The four clusters of algorithmic techniques examined share the characteristic that they are highly relevant to contemporary information ordering while remaining fundamentally understudied, both in their historical and conceptual dimension. Looking at the inception and evolution of algorithmic techniques allows us to examine them in a state of relative 'liquidity', where they have not yet been fully stabilized or 'frozen' into the canon, remaining precarious propositions that have to be explained and justified in terms that are absent from contemporary publications in the computing disciplines.

Chapter 4 serves as a topic-focused introduction that situates contemporary information ordering in a historical lineage that is largely absent from dominant narrations. Although the story starts off from standard takes on knowledge organization and classification in libraries and encyclopedias, it zeros in on the field of information retrieval, which develops in fundamental opposition to even the most visionary of library techniques, not merely in terms of technology and method, but regarding the idea of order itself. Coordinate indexing, the first and defining technique in this lineage, is explicitly designed to eliminate the influence of librarians and other 'knowledge mediators' by shifting expressive power from the classification system to the query and, by extension, to the information seeker. Order is no longer understood as a stable map to the universe of knowledge but increasingly as the outcome of a dynamic and purpose-driven process of ordering. Although equally foundational for the statistical tradition in information retrieval, the chapter closes by discussing coordinate indexing as a precursor of the relational model for database management, which underpins large swaths of contemporary information handling, from enterprise software to web platforms.

Chapter 5 investigates the early attempts in information retrieval to tackle the full text of document collections. Underpinning a large number of contemporary applications, from search to sentiment analysis, the concepts and techniques pioneered by Hans Peter Luhn, Gerard Salton, Karen Spärck Jones, and others involve not only particular framings of language, meaning, and knowledge, they also introduce some of the fundamental mathematical formalisms and methods running through information ordering, preparing the extension to digital objects other than text documents. The chapter specifically seeks to capture the considerable technical expressivity that

comes out of the sprawling landscape of research and experimentation that characterizes the early decades of information retrieval. It also documents the emergence of a conceptual construct and ‘intermediate’ data structure that is fundamental to most algorithmic information ordering at work today: the feature vector.

Chapter 6 examines one of many areas where feature vectors play a central role. Machine learning is currently one of the most active domains in computer science and the wide availability of datasets and increasingly robust techniques have led to a proliferation of practical applications. The chapter uses the Bayes classifier as an entry point into the field, showing how a simple statistical technique introduced in the early 1960s is surprisingly instructive for understanding how machine learning operates more broadly. The goal is to shed light on the core principles at work and to explain how they are tweaked, adapted, and developed further into different directions. This chapter also develops the idea that contemporary information ordering represents an epistemological practice that can be described and analyzed as ‘interested reading of reality’, a particular kind of inductive empiricism.

Chapter 7 ventures into the field of network algorithms to discuss yet another way to think about information ordering. While Google’s PageRank algorithm has received considerable attention from critical commentators, the vast intellectual landscape it draws on and contributes to is less well known. Graph algorithms are used in many different settings, not least in the social sciences, yet the technical and epistemological commitments made by graph theoretical formulations of ‘real life’ phenomena are hardly a subject of discussion beyond specialist circles. The chapter shows how algorithmic ordering techniques exploit and integrate knowledge from areas other than information retrieval and demonstrates how the ‘politics’ of an algorithm can depend on small variations that lead to radically different outcomes. The context of web search means that the various techniques covered in the second part of the book can be brought together into a shared application space, allowing for a more concrete return to earlier discussions of variation and combination in software.

The conclusion, finally, synthesizes algorithmic information ordering into a denser typology of ordering gestures, paying particular attention to the modes of disassembly and reassembly that inform the underlying techniques. The attempt to distill an operational epistemology from the cacophony of techniques begs the question whether we are witnessing the emergence of a new *épistémè* (Foucault, 2005), a far-reaching set of regularities that characterize how we understand and operationalize the very notion of order at a given time and place. Independently from how we answer this

question, it is clearly impossible to avoid the more immediately pressing need to understand how the capacity to arrange individuals, populations, and everything in-between in highly dynamic and goal-oriented ways relates to contemporary forms of capitalism. To face this challenge, I come back to Simondon's mechanology and its broader cousin, technical culture, as a means to promote a 'widening' of technical imagination and appropriation. While certainly not enough to solve the many concrete issues surrounding advanced algorithmic techniques, an understanding of technicity as human gesture – albeit of a specific kind – can sharpen our view for the many instances where technology has become complicit in domination, for the reconfigurations of power relations that occur when new levers begin to operate in and on society, and for the increasing interdependence between technical critique and social critique.

## Bibliography

- Andreessen, M. (2011). Why Software Is Eating the World. *Wall Street Journal*, 20 August. Retrieved from <http://online.wsj.com>.
- Beer, D. (2009). Power through the Algorithm? Participatory Web Cultures and the Technological Unconscious. *New Media & Society* 11(6), 985-1002.
- Berns, T., and Rouvroy, A. (2013). Gouvernementalité algorithmique et perspectives d'émancipation. Le Disparate comme condition d'émancipation par la relation? *Réseaux* 177, 163-196.
- Burrell, J. (2016). How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3(1), 1-12.
- Christin, A. (2017). Algorithms in Practice: Comparing Web Journalism and Criminal Justice. *Big Data & Society* 4(2), 1-14.
- Diakopoulos, N. (2015). Algorithmic Accountability. *Digital Journalism* 3(3), 398-415.
- Dijkstra, E. W. (1974). Programming as a Discipline of Mathematical Nature. *American Mathematical Monthly* 81(6), 608-612.
- Drucker, J. (2013). Performative Materiality and Theoretical Approaches to Interface. *Digital Humanities Quarterly* 7(1), n.p.
- Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- Foucault, M. (2005). *The Order of Things: An Archaeology of the Human Sciences*. London: Routledge.
- Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, and K. A. Foot (eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-195). Cambridge, MA: MIT Press.

- Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.
- Graham, S. D. N. (2005). Software-Sorted Geographies. *Progress in Human Geography* 29(5), 562-580.
- Grimmelmann, J. (2009). The Google Dilemma. *New York Law School Law Review* 53, 939-950.
- Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.
- Hacking, I. (1985). Styles of Scientific Reasoning. In J. Rajchman and C. West (eds.), *Post-Analytic Philosophy* (pp. 145-163). New York: Columbia University Press.
- Introna, L. D., and Nissenbaum, H. (2000). Shaping the Web: Why the Politics of Search Engines Matters. *The Information Society* 16(3), 169-185.
- Kitchin, R., and Dodge, M. (2011). *Code/Space: Software and Everyday Life*. Cambridge, MA: MIT Press.
- Mager, A. (2012). Algorithmic Ideology. *Information, Communication & Society* 15(5), 769-787.
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
- Parikka, J. (2012). *What Is Media Archaeology?* Cambridge: Polity Press.
- Peters, J. D. (2015). *The Marvelous Clouds: Toward a Philosophy of Elemental Media*. Chicago: University of Chicago Press.
- Rieder, B. (2005). Networked Control: Search Engines and the Symmetry of Confidence. *International Review of Information Ethics* 3, 26-32.
- Rieder, B. (2006). *Métatechnologies et délégation: pour un design orienté-société dans l'ère du Web 2.0*. PhD dissertation, Université Paris 8. Retrieved from <https://tel.archives-ouvertes.fr/tel-00179980/>.
- Rieder, B. (2008). An Experimental Collaboration Tool for the Humanities: The Proospace System. In P. Hassanaly, T. Herrmann, G. Kunau, and M. Zacklad (eds.), *Supplement to the Proceedings of the 7<sup>th</sup> International Conference on the Design of Cooperative Systems, Carry-le-Rouet, France, May 9-12* (pp. 123-130). IOS Press: Amsterdam.
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA: MIT Press.
- Simondon, G. (2017). *On the Mode of Existence of Technical Objects* (C. Malaspina and J. Rogove, trans.). Minneapolis: Univocal Publishing.
- Winkler, H. (1999). Search Engines: Metamedia on the Internet? In J. Bosma (ed.), *Readme! Filtered by Nettime: ASCII Culture and the Revenge of Knowledge* (pp. 29-37). New York: Autonomedia.
- Young, L. C. (2017). *List Cultures: Knowledge and Poetics from Mesopotamia to BuzzFeed*. Amsterdam: Amsterdam University Press.



# Part I





# 1. Engines of Order

## Abstract

The chapter discusses central terms like ‘information’ and ‘order’, and it proposes the concept of ‘engine’ to point toward the infrastructural embeddings that have allowed techniques initially conceived for document retrieval to become pervasive mediators in online environments. While this book constitutes a humanistic exploration of technical substances rather than their practical application, the chapter pays tribute to the fact that the techniques under scrutiny have become prevalent in a specific situation, in *this* world and not another. To this end, the chapter discusses three critical trends: computerization, information overload, and social diversification.

**Keywords:** information ordering, computerization, information overload, social diversification, digital infrastructures

Although the various practices described as ‘information ordering’ have become ubiquitous parts of online experiences, the two notions making up the term are far from self-evident. Instead of providing strict definitions, however, I take ‘information’ and ‘order’ as starting points for an investigation into a domain of techniques that intervene in deeply cultural territory in ways that come with their specific framings and epistemological perspectives. Instead of asking what information and order *are*, I am interested in the operational answers enacted by algorithmic techniques. This means remaining at a certain distance from common uses of the vocabulary and concepts that characterize the fields associated with information ordering, itself already a somewhat uncommon term. Information scientists and readers familiar with volumes such as Svenonius’s authoritative *The Intellectual Foundation of Information Ordering* (2000) or Glushko’s recent *The Discipline of Organizing* (2013) will notice that my interpretative lens can differ substantially, despite the shared subject matter. This begins to manifest in seemingly small gestures, for example,

when glossing over paradigmatic distinctions between classification and categorization or between data, information, and knowledge. Instead of committing to particular definitions of these and other terms, I am interested in understanding how they inform and coagulate around specific ‘problematizations’ (Foucault, 1990, p. 10f.) of the domains they refer to and how they are strategically deployed in the construction and justification of techniques that produce epistemologically distinctive outputs. So far, I have used the term ‘information ordering’ very broadly, connecting it to tasks such as searching, filtering, classifying, or recommending items in online systems. The following section discusses information and order in sequence to address – rather than resolve – their vagueness.

## Information Ordering

The techniques and practices discussed in this book hinge to a great extent on the term ‘information’ and the key role it plays in and around computing. My concern, however, is not the ontological question of what information is, but rather its practical role in different discourses and ‘its apparent ability to unify questions about mind, language, culture, and technology’ (Peters, 1988, p. 21). In the already somewhat restrained domain I will be investigating, the term has become a central instrument in the endeavor to bridge the gap between human practice and the workings of computing machinery. Here, the fact that information has no shared definition,<sup>1</sup> both in and across different epistemological sites, that it remains ‘a polymorphic phenomenon and a polysemantic concept’ (Floridi, 2015, n.p.), should not be seen as a failure or deficit but, on the contrary, as a strategic benefit when it comes to smoothening conceptual differences and bringing entire domains into the fold of computing.

As AI-researcher-turned-social-theorist Philip Agre has shown in great detail in his critique of artificial intelligence, polysemy – or, rather, the strategic arrangement of precision and vagueness – plays a productive role in technical work because it helps in binding human affairs to the technical

1 ‘Information is not just one thing. It means different things to those who expound its characteristics, properties, elements, techniques, functions, dimensions, and connections. Evidently, there should be something that all the things called information have in common, but it surely is not easy to find out whether it is much more than the name’ (Machlup and Mansfield, 1983, p. 4f.).

world and the other way around. The following paragraph summarizes his pivotal argument:

It is frequently said that technical practice employs an especially precise and well-defined form of language, but this is misleading. In fact, terms like 'knowledge,' 'planning,' and 'reasoning' are simultaneously precise and vague. Considered as computational structures and processes, these terms are as precise as mathematics itself. Considered as descriptions of human life, however, they are profoundly *imprecise*. AI continually tries to assimilate the whole of human life to a small vocabulary. (Agre, 1997a, p. 48)

Agre's analysis details how artificial intelligence reduces the complex and ambiguous phenomenon of human 'action' to the much more contained notion of 'execution of plans', thereby opening up concrete pathways toward implementation in a working system, a fundamental requirement of the discipline (Agre, 1997a, p. 12). This involves conceptual work: plans are defined as mental structures that consist of subplans, going down a compositional hierarchy to a set of basic operations. The decomposition into small steps prepares a proclamation of equivalence between plans and computer programs (Agre, 1997a, p. 5f.). What is essential, here, is that this reductive, operational understanding of planning is used in such a way that it keeps the initial starting point, the rich world of human action, as a referent. If plans are programs and action the execution of plans, one can now – by definition – simulate human action. The gesture is supported by the idea that 'the proof is in the programming' (Agre, 1997b, p. 140), which leads to a form of tautological reasoning: a technical idea is true if one can build it, and if one cannot build it, it is not a technical idea and therefore has no merit in the field.

We can find comparable semantic operations in many areas of computer science, and the term 'information' often plays a pivotal role in connecting the worlds of humans and machines in similar ways. A well-known example can be found in Warren Weaver's introduction to Claude Shannon's *A Mathematical Theory of Communication*, published as a joint book in 1948 (Shannon and Weaver, 1964). Here, Weaver distinguishes 'three levels of communication problems', beginning with the *technical problem* (A), which is concerned with the fidelity of symbol transmission and thus the level where Shannon's mathematical definition and measure of information are situated. But Weaver then also postulates a *semantic problem* (B) that refers to the transmission of meaning and an *effectiveness problem* (C) that asks

how conduct is affected by meaning. While he is somewhat prudent in this regard, he clearly wishes to extend Shannon's model from level A to levels B and C, which should only require 'minor additions, and no real revision' (p. 26). The statistical framing of information on level A finds its equivalence in 'statistical *semantic* characteristics' on level B, and the 'engineering noise' that troubles Shannon's technical transmissions becomes 'semantic noise' (p. 26). The communication of meaning is framed in similar terms as an encoding/decoding type operation. The engineering communication theory 'has so penetratingly cleared the air that one is now, perhaps for the first time, ready for a real theory of meaning' (p. 27). If meaning 'behaves' like information, it is to be investigated and conceptualized in similar terms, which, very concretely, suggests and requires 'a study of the statistical structure of language' (p. 27). What we end up with resembles the transformation Agre describes: a definition of meaning that does not fully reduce it to Shannon's notion of information but postulates a somewhat vague equivalence that enables and authorizes the transposition of the conceptual and analytical apparatus from one to the other. And, as an additional benefit, since that apparatus is mathematical in nature, there is now a clear path toward building a running system, for example, for the practical task of machine translation. The field of information retrieval broadly follows this program from the 1950s onward.

However, an important nuance has to be introduced at this point. The movement of 'absorption' or 'incorporation' of various aspects of human life into the space of computation is often discussed as *formalization* and critiqued as a reduction of an overflowing richness into the cold language of mathematical logic. Golumbia (2009), for instance, takes Chomsky's attempts to model the fundamental rules of language as a finite set of algorithms as his main example to show how 'computationalism' installs formal logic as both an analytical tool and a model for the workings of the mind itself. While Chomsky's work does not seek to build working systems for machine translation but to understand the fundamental principles of cognition (Katz, 2012), such explicit instances of 'high rationalism' have indeed radiated throughout the field of computing. But in many domains, for instance in information retrieval, the conceptual apparatus driving formalization can be surprisingly unambitious, subscribing to the pragmatic mindset of statistics rather than the rationalistic purity of logic. In the paper that first laid out what is now known as a Bayes classifier (Chapter 6), M. E. Maron (1961) programmatically states 'that statistics on kind, frequency, location, order, etc., of selected words are adequate to make reasonably good predictions about the subject matter of documents containing those words'

(p. 405), and this is basically all he has to say about the nature of language in that text. Although a logician himself, he considers the modeling of human language in mathematical logic to be an impasse and instead promotes Weaver's probabilistic perspective.<sup>2</sup>

Information retrieval shares AI's practical goal 'to make computers do humanlike things' (Swanson, 1988, p. 97), but it takes a different route to achieving it. The key referent on the 'human side' in tasks like document search is clearly something having to do with meaning and knowledge, but there is an almost comical desire to *not* develop any serious theory of these concepts and to stick to commonsense uses instead. Lancaster's (1968) classic definition of information retrieval creates even more distance by arguing that an 'information retrieval system does not inform (i.e. change the knowledge of) the user on the subject of his inquiry [but merely] on the existence (or non-existence) and whereabouts of documents relating to his request' (p. 1). Rather than commit to a theory of knowledge, information retrieval sits comfortably in a space where the relationship between knowledge and information is implied, but remains vague.<sup>3</sup> In the end, information's designated role is to be 'the essential ingredient in decision making' (Becker and Hayes, 1963, p. v) and this results-oriented epistemic 'attitude'<sup>4</sup> runs through the field to this day. For example, the famous Text REtrieval Conference (TREC) series, which has been organizing competitions in retrieval performance since 1992, is based on comparing participants' systems to known 'right answers', that is, to classifications or rankings that were manually compiled by experts. The primary goal is to attain or exceed human performance in situ rather than furthering deeper understanding of cognitive processes. Chomsky indeed argues that 'Bayesian this and that' may have arrived at some degree of practical proficiency, but 'you learn nothing about the language' (Katz, 2012, n.p.). His deep disdain for the statistical approach to machine translation is an indicator that the field of computing is characterized by real epistemological variation and disagreement. As Cramer argues, '[c]omputation and its imaginary are rich with contradictions, and loaded with metaphysical and ontological speculation' (Cramer, 2005, p. 125).

2 'Thus the goal of processing ordinary language by translating it (first) into a logical language brings with it more problems than prospects, and raises more questions than it answers' (Maron, 1963, p. 139).

3 'To impose a fixed boundary line between the study of information and the study of knowledge is an unreasonable restriction on the progress of both' (Machlup and Mansfield, 1983, p. 11).

4 I take this term from Desrosières (2001).

When it comes to the concept of 'order', we could again pursue formal definitions, pitting it against notions like entropy, but keeping a loose understanding means remaining open to the practical propositions made in the field. The OED broadly suggests that order is 'the arrangement or disposition of people or things in relation to each other according to a particular sequence, pattern, or method'. Order, in this definition, does not have the connotations of Cartesian regularity, uniformity, or immutability. And, indeed, the types of 'ordering' the techniques discussed in this book perform can be fuzzy, fragmented, and dynamic. They generally subscribe to probabilistic frameworks but also draw on other mathematical fields to deal with complexity and variation. Indeed, computing has been instrumental in shifting the problem of 'arrangement and disposition' from static conceptions of *order* to dynamic processes of *ordering*.

One way to think about such changing conceptions leads through Michel Foucault's *The Order of Things* (2005) and Deleuze's reading of that text merits particular attention. Here, the central term to delineate historical formations, each carrying its own specific understanding of order, is that of *épistémè*. Deleuze (1988) reads the classic *épistémè*, situated roughly in the seventeenth and eighteenth centuries, through the notion of 'unfolding' and couples it with what he refers to as the 'forces that raise things to infinity' (p. 128). Epitomized by Linnaeus's *Systema Naturae* (published in twelve editions between 1735 and 1767), divided in the kingdoms of animals, plants, and minerals, this *épistémè* is organized around categorization into a timeless system. Following the logic of representation, there is an incessant production of two-dimensional tables that establish the bounds of the order of things; concrete entities do not define this space, they are merely positioned on it through the attribution of identity and difference with other entities, in infinite variation.

Around 1800, the modern *épistémè* first appears as a perturbation of the classic order. There are irreducible and contingent forces – life, work, language – that break through the preset representational grids ordering the entities these forces are entangled with. In Darwin's work, for example, there is no predefined *regnum animale* ('animal kingdom') that covers all animals and their infinite variations. On the contrary, the tree of life starts with a single organism and the way it evolves is contingent and dependent on interactions between individuals and their specific environments. There is no eternal plan or order: life sprawls and disperses in different directions through successions of abundant yet finite variations. According to Deleuze (1988, p. 126f.), the modern *épistémè* is marked by an empiricism organized around the continuous 'folding' of the forces of life, work, and

language. History is not simply variation on a constant theme, but a process of becoming. The order of things is the result of that process and no longer the unfolding of an eternal blueprint.

Rather than stopping at this point, Deleuze attempts to address a question Foucault famously evokes at the end of *The Order of Things*, asking what comes beyond the modern *épistémè*. It makes sense to quote the central passage of Deleuze's argument in full:

Biology had to take a leap into molecular biology, or dispersed life regroup in the genetic code. Dispersed work had to regroup in machines of the third kind that are cybernetic and informatic. What would be the forces in play, with which the forces within man would then enter into a relation? It would no longer involve raising to infinity or finitude but a fini-unlimited, thereby evoking every situation of force in which a finite number of components yields a practically unlimited diversity of combinations. (Deleuze, 1988, p. 131, translation amended)

This notion of the 'fini-unlimited'<sup>5</sup> provides a compelling way to address the question of order – 'the arrangement or disposition of people or things in relation to each other' (OED) – and how it connects to the algorithmic techniques under scrutiny here. Foucault's *épistémès* are not only connected to particular visual forms of arranging, such as the table or the tree, but they contain specific ideas about the nature of order itself. In the classic period, order is thought to be pre-given, a 'God-form' (Deleuze, 1988, p. 125) that runs through the things themselves, constantly unfolding according to eternal, unchanging principles. The scholar observes, designates, and takes inventory; and although words and things are considered to be distinct, a well-built analytical language or taxonomy keeps them from falling apart by producing a correct account of a world 'offered to representation without interruption' (Foucault, 2005, p. 224). In the modern period, however, order is an 'outcome', something that is produced by the processes of life, work, and language.

How does the notion of the fini-unlimited incubate a third understanding of order? The crucial element, here, is the idea that a limited number of elements can yield an (almost) unlimited number of combinations or arrangements. As shown throughout the second part of this book, permutative

5 While the common translation of 'fini-illimité' as 'unlimited finity' may be more elegant than 'fini-unlimited', this amounts to a rather drastic change in emphasis. For a discussion of the topic from a different angle, see Galloway (2012).

proclivity is indeed a central characteristic of algorithmic information ordering: for any sufficiently complex dataset, the idea that ‘the data speak for themselves’ is implausible; developers and analysts select from a wide variety of mathematical and visual methods to *make* the data speak, to filter, arrange, and summarize them from different angles, following questions that orient how they look at them. Rather than ideas of a natural order, there are guiding interests that drive how data are made meaningful.

This argument is indeed central to two popular books by David Weinberger, *Everything Is Miscellaneous* (2008) and *Too Big to Know* (2012), which are almost manifestos for a fini-unlimited *épistémè*. Even if Weinberger’s epistemic attitude and historical trajectory differ substantially from my own, we share the fundamental diagnosis that information ordering increasingly revolves around gestures of disassembly and reassembly that follow specific interests and desires: ‘How we choose to slice it up depends of why we’re slicing it up’ (Weinberger, 2008, p. 82).

Indeed, it has become widely accepted that computers, whether we think of them as computing machinery or as digital media, encourage ‘disaggregation and disassembly, but also reaggregation and reassembly’ (Chadwick, 2013, p. 41). The central idea informing the relational model for database management, for example, is to cut data into the smallest parts possible to allow for dynamic recombination at retrieval time with the help of a powerful query language that makes it possible to make selections, calculations, or ‘views’ on the data. Outputs are selected and ordered based on the ‘question’ asked. The machine learning techniques discussed in Chapter 6, to give another example, provide the means to create information sieves inductively. By ‘showing’ a spam filter which emails are considered undesirable, the classifier ‘learns’ to treat each word or feature as an indicator for ‘spamminess’. But no two users’ classifier profiles will be exactly the same, not only because they receive different emails but also because they will have different definitions of what constitutes an *unwanted* message. This book traces such instances of a fini-unlimited in a manner that remains attentive to commonality yet refrains from singularizing a space of variation into a totalizing assessment.

My purpose, however, is not to postulate a new *épistémè*, a new understanding of order that would have emerged sometime after WWII, and then to show how this new formation has ‘found its expression’ in a range of algorithmic techniques. In line with the cultural techniques tradition, and in particular with Bernhard Siegert’s (2013) radical formulation, I consider that order, as a concept, does not exist independently from ordering techniques and that any broad shift would have to be considered, first and foremost,



as a consolidation in the network of ontic operations established by the techniques themselves. From a methodological perspective, this means that the concrete gestures of ordering and the technical, functional, and epistemological substance they carry are the necessary starting points.

## Engines Ordering *This* World

While one can look at algorithmic information ordering techniques as a series of technical ideas, their role as ‘epistemological operators’ (Young, 2017, p. 45) acting on the world in significant ways cannot be understood without consideration for their embedding in ever-expanding infrastructures that play fundamental roles in mediating and constituting lived reality (Burrows, 2009, p. 451). As Peters argues, ‘[m]edia are not only devices of information; they are also agencies of order’ (Peters, 2015, p. 1) in the sense that they support and organize social, political, and economic systems in specific ways. The functional substance of ordering techniques cannot be separated from their application to the bits and pieces of the ‘real’ world. They have become part of ‘the connective tissues and the circulatory systems of modernity’ (Edwards, 2003, p. 185) and their integration into larger ‘operative chains’ (Siegert, 2013, p. 11) binds their broad technical potential into more specific roles. My emphasis on technicity is therefore not in opposition to the perspective Peters (2015) calls ‘infrastructuralism’ (p. 33) but approaches the large systems that define and support modern life from the perspective of their smaller components.

The term ‘engine’ indeed serves to link the work done in particular locations or instances to its broader infrastructural embeddings. Donald MacKenzie’s *An Engine, Not a Camera: How Financial Models Shape Markets* (2006) studies financial markets in these terms, connecting fine-grained attention for the substance or *content* of calculation with an appreciation of its role and performativity in larger systems. Financial theory, understood as a series of conceptual and mathematical models, is analyzed as ‘an active force transforming its environment, not a camera passively recording it’ (MacKenzie, 2006, p. 12). How investment markets are framed conceptually and methodologically has concrete consequences for individual (e.g., investment decisions) and collective (e.g., regulation, market design) choices and behavior. The performative dimension of a financial model, method, or theory is strengthened further when it becomes reified in software that defines operative modes directly (MacKenzie, 2010). Both the ‘cognitive’ and the ‘mechanical’ understanding of performativity can be fruitfully applied

to information ordering, but the latter calls increased attention to forms of operation and automation that are particularly relevant.

Following Adrian Mackenzie's (2017a) take on machine learning, one could emphasize information ordering as a field of academic inquiry and an epistemic practice that is organized around mostly well-delineated steps, where a deliberately selected technique is applied to a contained dataset at a specific moment in time to generate a classificatory output. While this is certainly a common setup, the infrastructural perspective emphasizes a scenario where large-scale platforms capture, support, and channel human practice continuously and information ordering becomes a pervasive arbiter of real-life possibilities. Indeed, the degree to which calculative processes have penetrated into the fabric of contemporary societies is striking, although historiographical work (Beniger, 1986; Yates, 1989; Gardey, 2008) has clearly shown that data collection and analysis techniques have a long history, becoming steadily more central to organization, coordination, and control in business and government over the course of several centuries. Even modern-sounding approaches such as graph algorithms or machine learning have been around since at least the 1960s but were only widely taken up over the last two decades. The question why this has not happened earlier and why this is happening now on such a large scale can serve as an entry point into a deeper appreciation of the context algorithmic information ordering operates in. In the remainder of this chapter, I will thus establish a broader picture, beginning with an assessment of what has been called 'computerization' and followed by a discussion of 'information overload', the problem most often put forward by early information retrieval specialists. Taking a more sociological angle, I will then single out social diversification as a contextual factor that cannot be ignored.

### Computerization

One of the reasons for the somewhat delayed adoption of algorithmic information ordering could be that computers were simply not powerful enough before the turn of the century, making the exponential growth in speed and capacity the principal driver. In his acceptance speech<sup>6</sup> delivered on receiving the Turing Award in 1972, Dijkstra (1972) noted that 'as the power of available machines grew by a factor of more than a thousand, society's ambition to apply these machines grew in proportion' (p. 862)

6 In computer science, award speeches are one of the few publication formats where broad 'discoursing' is not only allowed but encouraged.

and his argument cannot be easily dismissed: processing brawn is indeed a prerequisite for making certain applications of information ordering a feasible option. Another technical explanation could call attention to the growing availability of algorithmic techniques beyond university labs and specialized documentation centers. But instead of singling out individual 'causes', it makes sense to think about these elements as parts of a larger, self-reinforcing process of 'computerization'.

While the term has fallen out of fashion after its heyday in the 1970s and 1980s, speaking of computerization reminds us that digital media are not just sleek graphical interfaces for making and accessing various kinds of 'content' or 'data' or, but also machines that vary in shape and ability, offering a variable computational basis for the implementation of all kinds of forms, functions, and autonomous operation. The capacity to connect ever-expanding capabilities for storage, transmission, and processing to rich and sophisticated input and output interfaces connected to the world in myriad ways has allowed the computer to infiltrate and to constitute a large number of practices. This can be understood as a process of progressive mediatization, a 'deepening of *technology-based interdependence*' (Couldry and Hepp, 2016, p. 53) that is not limited to consumer devices and includes countless activities in business or government. While the term 'infrastructure' is not reserved for technical systems, it is clear that fewer and fewer practices are not channeled through computing in one way or another.

The web still constitutes the prime example for a pervasive, general-purpose infrastructure that affords access to media content and social interaction as well as myriad services that rely on its technical malleability to organize activities through end-user interfaces and backend coordination. The rapidly expanding entanglement of practices related to communication, coordination, consumption, and socialization with computing is realized through the design and adoption of 'activity systems that are thoroughly integrated with distributed computational processes' (Agre, 1994, p. 105). Facebook, for example, can be understood as a highly complex amalgamation of various layers and instances of hardware and software that, together, form a global infrastructure for 'socializing online' (Bucher, 2013). Agre (1994) argues that an activity is 'captured' in the technical and conceptual vocabularies computing provides when it is enabled and structured by software-defined and computer-supported 'grammars of action'. Since the way this happens is clearly not a mere transposition of previous forms of 'socializing' into a new environment, computerization must be seen as an 'intervention in and reorganization of [human] activities' (Agre, 1994, p. 107). Facebook is not a neutral or transparent means to make, maintain,

and enact social relationships, but, 'by organizing heterogeneous relations in a specific way, constitutes a productive force' (Bucher, 2013, p. 481) that operates and mediates through an arrangement of deliberately designed forms and functions. Information ordering techniques become engines of *social* order when they operate and intervene in such environments, where '[t]hey may change social relations, but [...] also stabilize, naturalize, depoliticize, and translate these into other media' (Akrich, 1992, p. 222).

To consider the evolution of computing hardware from the mainframe to personal computers and further to mobile, networked, and integrated devices would be one way to analyze the deep incursions into the frameworks of human life computers have made. Notions like computerization and grammatization, however, seek to address the many different ways broad technical possibilities have been connected to a large variety of practices. If we follow Turing (1948) and Manovich (2013c) in framing computers both as universal machines capable of simulating all other machines and as 'metamedia' uniting various media forms in a single screen, software stands out as the principal means to create the fine-grained structures capable of capturing the components of highly complex activities such as online gaming or project management.

More recently, scholars have used the term 'datafication' to call attention to the process of 'taking information about all things under the sun – including ones we never used to think of as information at all, such as a person's location, the vibrations of an engine, or the stress on a bridge – and transforming it into a data format to make it quantified' (Mayer-Schönberger and Cukier, 2013, p. 15). This is clearly an important aspect to consider. The result of datafication has been the rapidly increasing production and availability of very large datasets that often comprise transactions (logged events or behavior) or other forms of nontraditional data such as traces of movement in navigational or physical spaces, social interactions, indications of cultural tastes, or sensor readings. This, in turn, stimulates demand for analytical capabilities. The accumulation of complicated yet highly expressive unstructured data in the form of textual communication, for example, has fueled interest in techniques like topic modeling or sentiment analysis that seek to make them intelligible and 'actionable', that is, applicable to decision-making.

However, speaking of computerization rather than datafication emphasizes that data accumulation enables forms of 'immediate' management that operate through interface modulation. The direct *application* of algorithmic ordering is made possible by the emergence of digital infrastructures and environments that allow for both data collection and output generation,