Human capabilities in a datafied society

Empirical approaches to studying the interplay between digital communication and internet infrastructures

Signe Sophus Lai
PhD Dissertation 2021
Human Capabilities in a Datafied Society:
Empirical approaches to studying the interplay between
digital communication and internet infrastructures

Supervisor: Associate Professor Rasmus Helles
Co-supervisor: Professor Klaus Bruhn Jensen

Submitted on October 1st 2020

Number of characters (incl. bibliographies): 82,139

All images in the dissertation are non-copyrighted

The project was funded by The Carlsberg Foundation
Human capabilities in a datafied society.
Acknowledgements

To describe the process of acquiring a PhD degree, people often use the metaphors of different extreme sport triumphs like running a marathon, climbing a mountain, or even (in my home) winning the Tour de France. Like much of academia, these metaphors usually focus on solitary, individual achievements: There can be only one to set the marathon record, one to reach the summit first, and one to finish in the yellow jersey. Yet these achievements are – like most scholarly ones – shared accomplishments: marathon runners pace and support each other and bystanders cheer them on, mountain climbers have
fellow climbers and carriers, and the Tour de France is raced in teams. These next lines are far from enough for paying the kind of tribute I owe to the people that have raced next to me, encouraged me in the last sprint, pushed me uphill, and believed that I’d cross the finishing line.

First, I want to thank the PIN (Peoples’ Internet) team. Klaus Bruhn Jensen, for being inspiring, kind, and always generous with your time and advice. My two fellow PhDs, Fiona Huijie Zeng Skovhøj and Jesper Pagh, for excellent companionship and teamwork. And Jacob Ørmen, Jun Liu, Chris Su, Verena Brändle, and Miklós Sükös for thought-provoking discussions and comments on the dissertation as it progressed.

A special thanks goes to Rasmus Helles. For years of motivating collaborations, hours of coding the worst television shows or disturbing tracking-companies, for challenging and trusting in me, having my back, being open to far-out ideas and to not sticking with the plan. And for your friendship.

I also want to thank Stine Lomborg. For your generosity, ambition on my behalf, and encouragements over the past years. Some would call you a role model, and I wouldn’t disagree with them.

A special salute goes to Sofie Flensburg. For laughing, crying, sharing much-appreciated motherhood wisdom, and for being the coolest human scraper I’ve ever met. It would have been an uphill ride, to say the least, if it wasn’t for you.

I also have to thank Sofie Thorsen, for making this dissertation an absolute piece of eye-candy and for caring co-mentorship. And Cecilie Astrupgaard, for offering elaborate, somewhat research-related, and truly life-saving procrastinations along the way.

Lastly, I’m endlessly grateful to my family. My mom, for the community spirit. Mads, for room and time to work hard but also to be anything but a serious researcher, and for believing in me when I had my own doubts. And Luca, for being exactly who you are – a marathon and a mountain peak wrapped in a yellow jersey – and for reminding me on a daily basis why we race in the first place.

Signe Sophus Lai, October 1st 2020
## TABLE OF CONTENTS

### PART I

| 01 | Introduction: there is no problem technology cannot fix | 7 |
| 02 | Article previews: home-away-home | 10 |
|    | Article 1: Tracing communicative patterns | 12 |
|    | Article 2: Mapping digital communication systems | 13 |
|    | Article 3: Networks of power | 14 |
|    | Article 4: Infrastructures of tracking | 15 |
|    | Article 5: A proxy for privacy | 16 |
|    | Article 6: Appscapes in everyday life | 17 |
|    | Article 7: ‘She’s the communication expert’ | 18 |
| 03 | Theoretical framework: a capabilities perspective | 19 |
|    | The Capability Approach | 20 |
|    | Sen’s basic framework | 23 |
|    | Capabilities in communications | 26 |
|    | Communication as capability | 39 |
Methodology: following individuals, following infrastructures

Made

Found

Contributions: bits of data, bits of power

Theoretical

Methodological

Empirical

PART II

Article 1: Tracing Communicative Patterns

Introduction

Studying the internet in everyday life

Sampling maximum-variation networks

Diaries: structured and un-structured approaches

Interviews: introductory and elicitation

Following people

Concluding remarks: comparing communications

Acknowledgements

Article 2: Mapping digital communication systems

Introduction

Communication system analysis & the infrastructural turn
Conceptualizing digital communication systems .............................................. 107
The Danish case .............................................................................................. 110
Conclusion ...................................................................................................... 125

08 Article 3: Networks of Power ....................................................................... 131
Introduction ..................................................................................................... 133
The evolution of large technological systems ............................................... 135
Method, data sources, and processing ............................................................ 138
The evolution of the internet in Denmark ....................................................... 142
From socio-technical transfer to digital momentum ....................................... 152
Conclusions ..................................................................................................... 159

09 Article 4: Infrastructures of Tracking .......................................................... 167
Introduction ..................................................................................................... 169
State of the art and rationale for the study ..................................................... 171
Methods and dataset ...................................................................................... 176
Findings ........................................................................................................... 177
Discussion ....................................................................................................... 187
Conclusion ....................................................................................................... 189

10 Article 5: A Proxy for Privacy ...................................................................... 197
Introduction ..................................................................................................... 199
State of the art and theoretical foundations ................................................... 201
Methods .......................................................................................................... 205
Analysis .......................................................................................................... 211
Discussion ....................................................................................................... 229
Introduction: there is no problem technology cannot fix

This dissertation is, at its core, about the relationship between individuals and technological systems. While it is certainly not about the corona virus, the pandemic testifies to the timeliness of the dissertation’s critical discussions of and empirical approaches to big data, digital infrastructures, power, and the role of the internet in everyday life.

In the first months of the spread of the virus, the by now redundant conceptualisation of data as the new oil (e.g. Yonego, 2014) was replaced by the idea of data as one of the most prominent weapons against COVID-19. The examples of state- and data-driven mechanisms for fighting the virus are manifold: a Polish app requires that people infected with the corona virus take frequent selfies in order to prove that they are staying indoors. Taiwanese authorities employ a so-called digital fence in order to receive alerts if diseased people circumvent quarantine. Using geolocation data, South Koreans are able to localise infected citizens on an online map made publicly available by the authorities. In Israel, cell phone data, usually employed in fighting terror attacks, is used for tracking people that have tested positive for COVID-19. Also in Denmark, cell phone data relay information on the movements of the Danish population. A Chinese health-rating app assigns colour codes to citizens, thereby determining whether or not they can leave their homes, ride on public transportation, or go shopping. And apps using Bluetooth technology allow for contact tracing of individuals living in Norway, Singapore, and a string of other countries. Various private corporations that were just months ago acting according to the oil metaphor, constantly trying to increase data harvesting around all aspects of life and thereby their individual data banks, are now invested in these national initiatives towards countering the spread of the virus. This includes for instance the unusual collaboration between Apple and Google to develop contact tracing technologies, or Amazon’s assistance to nation states that are scrambling to come up with solutions for hosting the large quantities of data that follow from the initiatives. As such, in the midst of the pandemic, crisis measures and quick fix solutions draw
attention away from the commercial aspects of data harvesting towards the health benefits of it.

The examples testify to the forcefulness of the belief that “there is no problem technology cannot fix” (Klein, 2020), and the failure to recognise how digital technologies come with their own suite of complex and unanticipated problems. They show the various potentials in fighting the pandemic through digital technologies and not least individuals’ data, while simultaneously demonstrating ways general human and citizen rights are being overturned or renegotiated as part of fighting the spread of the virus. This includes for instance anti-discrimination rights and the EU General Data Protection Regulation (GDPR). Although much of what is being instated is framed as emergency legislation and often with sunset clauses, one cannot help but wonder what the long-term consequences of these initiatives will be (see Taylor et al., 2020); how these data might be used for alternate and creepy purposes in the future; or whether we are right now living out a situation equivalent to the historical impact points that Shoshana Zuboff (2019) refers to when questioning how we arrived at surveillance capitalism. Zuboff calls attention to, for instance, the year 2001, where legislation on privacy that would outlaw some of the emergent, and now routine, practices of online surveillance was being discussed in the US congress. After 9/11, concerns about privacy were, however, traded in for a fixation on gathering as much information as possible as part of the renowned war on terror, and the legislative initiatives were side-lined for good.

The specific data processes and infrastructures that are being invoked in the different initiatives outlined above are key to understanding their implications. A good part of this dissertation is devoted to empirical studies of digital infrastructures as well as of economic and other incentives for increasing the datafication of everyday life. This includes reverse-engineering the opaque and less regulated tracking mechanisms of the web and apps as key components in contemporary infrastructures for datafication. The ways technologies and data are employed in this rather unprecedented time will have consequences for big data practices in the future, also in the realms of user profiling, targeted advertisement, and content filtering and curation. In other words, it matters what states do when it comes to commercial spaces of operation (Jørgensen, 2019, p. xix) – that is why 9/11 was fatal for individuals’ right to privacy, and why the handling of the
pandemic crisis through complex private and public arrangements will influence future commercial initiatives aimed at trading in ever-increasing amounts of data.

Since the infrastructural foundations for these systems are firmly engrained in people’s everyday lives as well as across societal sectors, opting out of the digital initiatives aimed at fighting the spread of the coronavirus (or of Facebook for that matter (Portwood-Stacer, 2013)), is challenging at best, and impossible at worst. Often, when it comes to the variety of digital services that play prominent roles in people’s everyday communication, there are no actual alternatives (and if there ever were, they were bought up, competed out of business, and so forth).

It is the aim of this dissertation to contribute to emerging understandings of this “world with no off switch” (DeNardis, 2020) and its implications for people’s abilities to live the lives they value and to make their own futures – or perhaps more aptly, for the good – digital and datafied – life. As such, motivated by a strong feeling of a call to action (or, perhaps, arms), the dissertation deals with a three-fold problem of understanding: what individuals are capable of in and through digital communication; how this communication is enabled, constrained, and negotiated in a datafied society; and, in turn, the implications for individuals’ everyday lives, the choices they (do not) have, and the inequalities they live with. Digital communication is, in this respect, understood as communication types that are supported by and distributed through internet-based technologies and infrastructures. Although datafication was always a part of communication technologies (Beer, 2016), the use of the term datafied in this context refers to big data in particular and the ways they render aspects of our everyday lives that were never quantified before into vast measures of commodifiable and monetizable data and metadata (Cukier & Mayer-Schoenberger, 2013).

To approach this problem, I introduce, in this first part of the dissertation, a perspective on communication as capability. The capabilities perspective (Sen, 1980) offers a conceptual and normative lens through which to approach and understand what individuals can and cannot do, why, and with what implications. Standing on the shoulders of a limited body of work in communications and elsewhere, I develop a framework around capabilities and communication that sets the agenda for the articles in the dissertation and for future research to engage with aspects of digital communication that are rarely studied in combination, and thereby push the boundaries for existing fields within media and
communication research. Following this theoretical framework, I outline the ensemble of methods for finding and making data that were employed – and indeed necessary – in order for the dissertation to empirically approach capabilities at the intersection between digital communication and datafication and at macro, meso, and micro levels of inquiry. Lastly, the remainder of Part I is devoted to the findings of the dissertation as they traverse the individual articles and to how they contribute to contemporary and future research agendas through studies that are empirically grounded, often holistic and macro-oriented, critical and oriented towards impact.

Before embarking on the theoretical framework, it is, however, necessary to introduce the very body of the dissertation, namely the articles that represent my theoretical and methodological discussions as well as empirical explorations, so that I am able to refer to them throughout this first part of the dissertation.

**Article previews: home-away-home**

An article-based dissertation is a curious thing. For the articles to be published in international peer-reviewed journals\(^1\), they must contain individual and comprehensive arguments that can stand on their own. As such, the articles that form the body of this dissertation were not written (or researched) in order to make up a perfect theoretical, methodological, or empirical whole. Rather, each of them puts forth its own perspective on different aspects of the overall problem of understanding what individuals are capable of in and through digital communication, why, and with what implications. Yet, the article-based PhD format requires that the articles are stitched together into a coherent and meaningful entity that can be read from A via B to C and so forth. It is, therefore, necessary for this first part of the dissertation to provide a reading instruction that links the different articles through connections that might otherwise be overlooked, and shows how common trajectories stand out when reading across them.

\(^1\) Five of the articles have at the time of writing this been published as individual research outputs, one has been accepted for publication, and the last one is still under review (see the publication status for each title in the article previews).
Although the articles are strongly connected (they, for instance, focus consistently on infrastructural tenets, employ communication perspectives, and present research that is empirically and systemically grounded), they could have been more so. If I had, for the purpose of fighting the fragmentation that article-based dissertations are so often criticised for, stayed with one theoretical perspective or one empirical aspect, the stitching together might have been easier, yet the particular richness in the multiple perspectives put forth in the articles would have been lost. I argue that because this dissertation deals with issues that are, by all comparisons, under-researched, emergent, fast-moving, and often viewed in isolation, the diversity of the studies comprised in the dissertation is necessary. I do not claim that the articles combined provide the whole picture, but I believe that their different perspectives and focus areas are valuable both for covering as much ground as possible, when it comes to approaching capabilities at the intersection between digital communication and datafication; and for traversing different theoretical perspectives and levels of inquiry, thereby pushing the boundaries for established fields and methods rather than reproducing them.

The seven articles could have been selected and ordered according to different logics. One would be going from macro-oriented quantitative analyses to micro level case studies of digital communication in everyday life (or the other way around). Another would be to bulk the more theoretical inputs, the methodological, and the empirical. Yet the following outline of the articles follows a trajectory that moves from the micro level of engaging with individuals’ capabilities – what they are and what they do – in and through digital communication; to the macro level of large technological systems that these beings and doings are embedded in; and back to the implications of these systems for individuals’ everyday life. Taken together, they thereby chart how macro structures manifest themselves in big data harvested from micro and mundane everyday communications, and how the everyday life is embedded in and conditioned by the same structures. There is, beneath the common trajectories across the articles, also a commonality in that the outline follows the chronology of my work. That is, the first article was written first, we wrote the second article second, and so forth – although there is work in-between that I will refer to in this dissertation, which is not featured anywhere as part of it. The following makes a short outline of each of the seven articles, their research aims as well as key findings. The outline,
in turn, provides a basis for discussing the ways the dissertation addresses the problem introduced above.

**Article 1: Tracing communicative patterns**


https://doi.org/10.2478/nor-2019-0019

This article represents the first part of my PhD research as a fellow in the People’s Internet project (PIN), where my initial mission was to carry out an ethnographic field study that would complement the survey and big data components of the PIN project. The article offers a research design for qualitative comparative studies of digital communication in everyday life – *a comparative ethnography of communication*. It argues that if we are to approach and understand the intermediacy (Jensen, 2008) of people’s everyday communication, we need methods for following them across their everyday contexts, the media they own, and the platforms they use. The suggested research design combines maximum variation and network sampling in order to produce small sample networks of individuals that are connected through their relationships but also disconnected by the inherently disparate starting nodes for each of the sampled networks. The design then combines elicitation interviews with respondent-driven communication diaries in order to generate knowledge on all types of communication happening throughout a normal day, the interaction and intersection of internet-based communication with other types of communication, as well as the respondents’ reflections and understandings of “what the devil they think they are up to” (Geertz, 1973). This combination of methods allows for the collection of thick data traces that stay close to communication tools, purposes, and situations. It also allows for the capture of individuals’ understandings of the communication possibilities available to them and their choices as they relate to contexts, media, and platforms. The article, as such, offers a way of approaching human capabilities on the basis of what empirical individuals do, and what they are, in and through communication.
Article 2: Mapping digital communication systems

https://doi.org/10.1177/0163443719876533

After six months of ethnographic fieldwork and hours of conversation with people about their everyday communications, the second article represents a shift in perspective – from the micro context of everyday communications to the technological, economic, and political conditions that frame the same communications. The article takes individual internet users in a given societal context as its theoretical point of departure and asks how their capabilities for communicating digitally are enabled and constrained by the infrastructural, market, and policy structures of the context in question. The article thereby contributes to existing studies of infrastructural power by developing a framework for mapping digital communication systems and making comparative analyses of how and why structural conditions differ across national and historical contexts. This framework contrasts other systemic approaches that typically focus on (legacy media) sectors, institutions, or policy programs (Bar & Sandvig, 2008), and instead identifies the system components through their ability to support particular activities (like digital communication). Specifically, the framework is developed from the simple question: “What possibilities do I have when I pick up my smartphone (or any other digital device) and what are the structural consequences of my actions?” (Flensburg & Lai, 2019, p. 5 [article 2]). Digital communication is, as such, understood as dependent on a chain of events with infrastructural, economic, and political implications: when communicating digitally, an individual activates an internet-enabled device and an access network, backbone network mechanisms, specific applications or platforms, and lastly digital communication content. In exemplifying the application of the framework and the methodological steps the analysis entails, the article maps the components of the Danish communication system that frames my research respondents’ everyday lives. The broad communication system concept takes into account monumental changes in the system, for instance, the development and eventual dominance of tracking technologies leading to the data economy as we know it, and includes them as significant components of the system with important repercussions for what individuals’
are capable of. The framework provides a much-needed basis for future comparative studies of digital communication systems in different national contexts, while also offering a critical foundation for understanding and contextualising, for instance, comparative ethnographies, survey studies, or quantitative country indexes.

**Article 3: Networks of power**

https://doi.org/10.1080/24701475.2020.1759010

The third article digs deeper into the Danish context and the digital communication system that frames the everyday communications of my respondents as well as any other individual living there. The article demonstrates the comparative potentials of the digital communication system (DCS) framework developed in the previous article by analysing the evolution of the internet infrastructure in Denmark through comparisons across three impact points in Danish history, 2002, 2012, and 2019. It thereby contributes to historical studies of infrastructural developments and of how power accumulates over time, through a number of empirical analyses that start out from the material infrastructures in order to analyse how they are controlled and regulated. The article revisits and develops Thomas P. Hughes’ momentum theory (1994) and argues that, like other large technological systems, the internet has evolved in different phases reflecting a shift from being mainly influenced by socio-economic conditions to having a determining influence on the development of societal structures. In other words, in Denmark, the implementation and early development of the internet infrastructures reflected historical market structures and institutional characteristics, yet the later development challenges and transforms the same structures and characteristics, eventually giving international, digitally-born, and private corporations quite some leverage on the future of digital communication in Denmark.
Article 4: Infrastructures of tracking


The fourth article focuses on a particular layer – or subsystem – of the digital communication system analysed and discussed in the previous two articles; namely the content layer, and more specifically, minute content bits known as third-party cookies used to track individuals across the web. This article thereby represents another shift in perspective from the Danish context to the European, but also from fibre-optic submarine cables and incumbent telcos to an advanced system built to support the surveillance capitalism business model. The shift was, in particular, motivated by the lack of theoretical frameworks for understanding – as well as empirical knowledge on – online tracking mechanisms as important aspects of the forces that enable and constrain individuals’ digital communication, what they can and cannot do, not least when it comes to retaining control over their data and making their own futures. The article focuses on the top-150 most used websites from each of the 28 EU countries and their outsourcing of ad tracking, security functions, content hosting, and so on to third-party companies. The services performed by these companies, in turn, ground the material functionality of the web as well as the commodification and monetisation of internet users. Empirically, the article shows how usages of third-party services vary across the European region as well as across different types of sites – variations which can be traced back to differences in languages, regulatory regimes, and digital business models. These findings add to ongoing debates about surveillance capitalism and big data by connecting broader social and cultural structures to the specific and varied forms of data practices and user commodification at play across the identified third-party infrastructures. The article, along with the one that follows, is first and foremost an empirical contribution to existing studies on the political economy of surveillance infrastructures. Both articles also represent a call for new empirical agendas in communication research, namely, studies that make visible otherwise invisible, yet tremendously influential infrastructures for datafication of everyday life.
Article 5: A proxy for privacy


Similar to the subject area of the previous article, article five looks at surveillance infrastructures, but focusing on apps rather than websites. My fieldwork, and work before that, has shown the immense integration of mobile apps into people’s everyday communication, yet, we know relatively little about mobile data collection practices in general and their implications. Also, compared to the web, we know even less about the embedding of third-party services in apps and the work that they do. Addressing this research gap, the article contributes to an emergent field of app studies (e.g. Gerlitz et al., 2019) by developing a methodological and empirical approach for gauging the ways users’ data is collected and shared through mobile apps. The approach has a dual focus on 1) the access and permissions that apps request and 2) the third-party corporations they work with. In other words, the infrastructural components of apps that condition the disclosure of (big) user data and thereby the commodification of and trade in future user behaviours. The article explores the surveillance ecology of smartphone apps and the privacy implications of everyday app usage through three analytical perspectives: first it looks at ‘appscapes’ of individual smartphone users, that is, their apps, their permissions, and third-parties in order to get at the consequences of particular app repertoires. These appscapes belong to ten of my fieldwork respondents and are deducted from screenshots of their smartphone home screens. As such, after three articles that operate at a meso (national) or macro (regional) level in order to approach the structures that condition everyday communication, this article returns to the empirical individuals and to how their mundane and mobile communications are embedded in the larger ecology of apps and the data economy. The second perspective looks at different types of apps within the sample (N=173) and compares them in order to explore relationships between apps and third-party service providers. From the third perspective, the article focuses on a particular app category and discusses the functional and commercial motivations for permissions and third-part collaborations. It thereby offers an empirical and critical perspective on mobile
communication, app ecologies, and data economies, as well as an approach for gauging the intrusiveness of mobile apps.

Article 6: Appscapes in everyday life


The use of ‘free’ mobile apps (or any other free digital services) comes with a price, paid through data rather than cool cash, but a price all the same. Article six looks into this ‘trade-off’ by returning to the fieldwork participants and digging deeper into the fieldwork data in order to approach the implications of large technological systems like the app ecology mapped out in the previous article for the everyday lives of my respondents as regular smartphone users. The article combines user and infrastructure studies – qualitative interviews and mappings of infrastructures for datafication – to explore both how the respondents understand the complex mechanisms of online tracking as well as how their individual appscapes frame and impact mundane and mobile communications that in large parts hinge on app-based services. Hence, re-introducing and operationalising the concept of appscapes coined in article five, it links the understandings of my fieldwork respondents with the extents to which their everyday life is subject to datafication, thereby exploring and nuancing knowledge on the so-called privacy paradox (Barth & Jong, 2017), digital resignation (Draper & Turow, 2019), and corporate obfuscation (Ibid.). Finding a disconnect between what users believe happen to their data, and the actual data harvesting and distribution mechanisms of their apps, the article argues that users’ resigned attitudes should be understood in light of the material conditions of the app ecology. By engaging empirically with the implications of commercial mobile infrastructures and the commodification of app use(rs), it offers a tool whereby these implications can be made accessible to individual users, and thereby enhance user empowerment, corporate transparency, and not least critical debate.
Article 7: ‘She’s the communication expert’

Signe Sophus Lai. “‘She’s the communication expert’. Digital labour and the implications of datafied relational communication”. Feminist Media Studies, in review.

The seventh and last article contributes to ongoing discussions about gender inequality, social justice, online tracking, and algorithmic bias. Like the previous article, it returns to the micro level of the everyday life, and looks into the implications of infrastructural systems mapped out in the previous articles for individuals’ mundane communication. It does so through discussions of digital labour as well as the role played by communication in social reproduction. Although there are many distinct conceptions of labour at play in the digital media and communication literature, hardly any address the gendered dimension of digital labour (for exceptions see e.g. Arcy, 2016; Jarrett, 2015). The article critiques this absence and suggests communication as a missing link in understanding the role of everyday interactions in the data economy. In doing so, it specifies relational communication as a particularly important but also under-researched aspect of social reproductive labour, and argues for a reconceptualization of the sociological concept of “the second shift” (Hochschild & Machung, 2012), by introducing “the digital shift”. Lastly, the article charts the implications of digital labour by debating if and how the commodification and monetisation of relational communication can be understood as exploitation that reproduces and reinforces existing gender inequalities. The article draws on experiences and understandings from the ethnographic fieldwork, while also tying together and taking inspiration from the empirical investigations featured in article 4, 5, and 6 on the online tracking mechanism of particularly websites and apps – both of which play a prominent role in facilitating the relational communication activities of my respondents and beyond. As a key contribution, it thereby pushes feminist critiques from the outskirts to the centre of critical data studies.

The selection and ordering of the articles for this dissertation follows both the trajectory of my research experiences as well as a logic of first engaging with what individuals are capable of – what they are and what they do – in and through communication, then with the macro level of the large technological systems that these beings and doings are embedded in, and last with the
implications of these systems for individuals’ everyday lives, the choices they (do not) have, and the inequalities they live with. The remainder of Part I shows how the dissertation, if not as a ‘whole’, then as a many-sided story, approaches the problem outlined in the introduction by offering a theoretical framework around communication as capability, a variety of methods that overall complement each other as they make and find relevant data, and a number of empirical analyses that testify to the ways digital communication is enabled, constrained, and negotiated.

Theoretical framework: a capabilities perspective

In the introduction to this first part of the dissertation, I briefly mentioned that I will approach communication from a capabilities perspective. A perspective which, in turn, originates in the Capability Approach — a normative and people-centred framework about social justice, freedom, and wellbeing (e.g. Nussbaum, 2011; Sen, 1980). In the following, I outline how a perspective on communication as capability — and a particular understanding of enabling and constraining forced embedded in that perspective — frames the articles in my dissertation; and how it motivates studies at the macro level of social and technological structures, at the meso level of particular contexts like that of Denmark, as well as the micro level of individual actors and the ordinary and extraordinary things they accomplish. The following includes: an introduction to the tenets of the Capability Approach, an outline of Sen’s basic framework for studies of capabilities, then a review of applications of the approach in (digital) media and communication research, and lastly a development of a systematic framework around communication as capability and research agenda for this dissertation as well as future studies.
The Capability Approach

Capabilities – what individuals are possibly capable of achieving in life and the structures that frame their achievements – have been central for my engagements with infrastructures, markets, and policies that condition digital communication, as well as with the ways individuals navigate and negotiate the same structures in and through their mundane communications. The Capability Approach was pioneered by Amartya Sen (1982; 1985, 1995, 1999) and Martha Nussbaum (2000) before the internet was part of most people’s everyday lives. Coming out of welfare- and development economics, the approach is particularly concerned with what individuals are capable of and why, as well as how capabilities vary within and across contexts. The approach thereby broadens concerns about wellbeing beyond economic or utilitarian measures of happiness, arguing that such measures are insensitive to the unequal distribution of wealth in societies or even households, and that people may adapt their expectations of life to fit discriminating structures (Robeyns, 2016). Employing a capabilities perspective in this dissertation therefore means focusing attention to what individuals achieve (referred to as their doings and beings or functionings in the capability literature), what they could possibly achieve (their capabilities), as well as the environment that frames their (possible) achievements (the conversion factors). In other words, the approach “directs us to examine real lives in their material and social settings” (Nussbaum 2000, 71). As capability, communication is both a constituent of being a certain individual and a condition of doing particular things (Jensen, in press). As a theoretical framework, capabilities, in turn, offer a systematic approach to the problem outlined in the introduction.

The Capability Approach makes two key normative claims: first, that the freedom to achieve wellbeing (rather than just the achievement in itself) is central, and second, that this freedom is to be assessed through people’s capabilities (Robeyns, 2016, p. 1). Originating in the Aristotelian concept of dunamin, meaning “capability of existing or acting” (Garnham, 1997, p. 27), capabilities correspond to “the actual ability of a person to achieve things that she has reasons to value” (Sen, 2002, p. 10). Sen (2009), in his work, emphasises the “reason to value” (p. 180), urging researchers to investigate the underlying motivations for valuing specific types of lives along with the situatedness and diversity of possible ways
of making choices. To give an example from my work, this entailed investigating, for instance, both if, how, and why individuals did (not) take measures towards protecting their personal data, as well as the infrastructural, economic, and political structures that enable and constrain the possibilities for taking such measures in the first place (Lai & Flensburg, 2020b [article 6]).

Two central characteristics of the Capability Approach motivate, in particular, the capabilities perspective in this dissertation. Namely that the approach is normatively (or ethically) individualistic, but ontologically non-individualistic (Robeyns, 2003, p. 65). That is, the units of normative judgement are individuals and, as properties of individuals, their capabilities and functionings. This outset in individuals, be it theoretical, empirical, or both, contrasts other approaches that focus on households, communities, sectors, institutions, policies, and so forth, and it forces us to look at communication as capability from a cross-sectoral and cross-institutional perspective, thereby pushing existing boundaries in both research and legislation on digital media and communication. The framework then fosters comprehensive analyses that follow the Kantian move (Kleine, 2013, p. 30) of placing oneself in the place of the other, and seeing structures as means and individuals as ends (rather than the other way around). In other words, the Capability Approach puts the individual centre stage, focuses on what she is capable of and why, and uses this as a basis for discussing how societal structures should be organised and regulated so as to enhance her freedom to achieve well-being and, thereby, to live a good life. The approach, as such, works from a positive conception of freedom (Berlin, 1969), not as in freedom from, for instance interference from state governance, but freedom to accomplish things in life that are valuable and to make one’s own future:

[The Capability Approach, ed.] is focused on choice or freedom, holding that the crucial good societies should be promoting for their people is a set of opportunities, or substantial freedoms, which people then may or may not exercise in action: the choice is theirs. It thus commits itself to respect for people’s powers of self-definition (Nussbaum 2011, 18).

To give another example from my work, the dissertation approaches the lived lives of empirical individuals through in-depth fieldwork and conversations with them about their everyday lives, their doings and beings in and through digital
communication, and their understanding of what they are capable of, the choices they have, and the ones they make (Lai et al., 2019 [article 1]).

At the same time, however, the approach is not ontologically individualistic, but quite the opposite: individuals and their capabilities are not seen as independent from the actions of others or their concerns for them, nor from the structures that characterise the contexts that people live out their everyday lives in (Robeyns, 2003, p. 65): “[T]hey [capabilities, ed.] are not just abilities residing inside a person but also the freedoms or opportunities created by a combination of personal abilities and the political, social, and economic environment” (Nussbaum 2011, 20). In other words, following this ontology, capability studies are particularly sensitive to the material and social structures that frame people’s everyday lives. They are also, however, concerned with going well beyond aspects of access to means (for instance digital technologies) and skills (to use them), and instead focus attention to the outcomes – materialised in capabilities and functionings – of both:

\[I]\textit{t is not sufficient to know the resources a person owns or can use in order to be able to assess the well-being that he or she has achieved or could achieve; rather, we need to know much more about the person and the circumstances in which he or she is living. Sen uses “capability” not to refer exclusively to a person’s abilities or other internal powers but to refer to an opportunity made feasible, and constrained by, both internal (personal) and external (social and environmental) conversion factors (Robeyns, 2016, p. 7).}

From a capabilities perspective, the dissertation thereby approaches infrastructural, economic, and institutional conditions as structures that frame what individuals can and cannot do, in and through communication. The last quote captures this underlying principle, which also guides the development of the digital communication system framework in article 2 (Flensburg & Lai, 2019) and article 3 (Flensburg & Lai, 2020a). Here, focus is on what Ingrid Robeyns refers to in the quote above as social and environmental factors that frame what individuals in Denmark are capable of and why.

Lastly, unlike other approaches, the capability framework does not assume that people are equally influenced by the same personal, social, or environmental
circumstances, or the same resources. In Sen’s words: “Human diversity is not a secondary complication (to be ignored or to be introduced ‘later on’); it is a fundamental aspect of our interest in equality” (1995, p. xi). Put differently, we might say the approach acknowledges individual diversity by default: it dismisses the notion that access to means and resources alone – like the promises of connecting the unconnected (Mansell, 2017a) – will result in increased quality of life for all, and instead argues that researchers need to focus on what comes after – in my case, connectivity – and on how (possible) outcomes differ between different individuals that do not make the same from the same resources: “It seems reasonable to move away from a focus on goods as such to what goods do to human beings” (Sen 1982, 30).

**Sen’s basic framework**

The Capability Approach is a complex framework that only increases in density once one dives into the various views on what the approach entails, how it might be conceptualised, and for what purposes. In order to show the ways the capabilities perspective – and my operationalisation of it – frames the articles of the dissertation and its contributions, I will introduce Sens’ basic capabilities framework. Figure 1 visualises the components of the Capability Approach as well as their inherent relationships.
In short, the figure shows how means to achieve – capability inputs – like goods and other resources are converted through three conversion factors – personal, social, and environmental – to freedom to achieve things in life that are valuable – the capability set. It also shows how choices – informed by personal history and psychology as well as preference formation mechanisms and social influences on decision making stemming from the conversion factors – eventually lead to achievements in the form of functionings.

The small arrow in the figure from freedom to achieve to achievement is essential for understanding the implications of the approach. If capabilities, as described above, correspond to an individual’s real freedoms, then functionings are the outcome of what that individual chooses to be and to do with her freedom: “while travelling is a functioning, the real opportunity to travel is the corresponding capability. The distinction between functionings and capabilities is between the realized and the effectively possible” (Robeyns, 2016, p. 4). This distinction emphasises how capabilities are a precondition for any functioning, but also why it is important to scrutinise the choices people make, since an apparently similar outcome in the space of functionings might come from quite distinct capability sets and choices, while dissimilar realised functionings might be the results of similar capabilities:
In contrasting capabilities with functionings, we should bear in mind that capability means opportunity to select. The notion of freedom to choose is thus built into the notion of capability. To use an example of Sen’s, a person who is starving and a person who is fasting have the same type of functioning where nutrition is concerned, but they do not have the same capability, because the person who fasts is able not to fast, and the starving person has no choice (Nussbaum, 2011, p. 25).

All three components – capabilities, choices, and functionings – follow, in the figure, from the two remaining key constituents, namely means like goods and services as well as conversion factors. Conversion factors enable and constrain individuals’ abilities to translate resources – that are in themselves enabling and constraining by their absence or presence – into freedom and achievements: “the relation between a good and the achievement of certain beings and doings is captured with the term ‘conversion factor’: the degree to which a person can transform a resource into a functioning” (Robeyns, 2016, p. 7). Sen distinguishes between three types of conversion factors, namely personal, social, and environmental (Robeyns, 2016, p. 7). Personal conversion factors include “metabolism, physical condition, sex, reading skills, intelligence” (Ibid.), and other internal aspects. Social conversion factors include “public policies, social norms, practices that unfairly discriminate, societal hierarchies, or power relations related to class, gender, race, or caste” (Ibid.), and other societal aspects. And lastly, environmental conversion factors materialize from the physical and built environment and include “climate, pollution, the proneness to earthquakes, and the presence or absence of seas and oceans” along with “the stability of buildings, roads, and bridges, and the means of transportation and communication” (Ibid.).

By way of example, I can explain Sen’s understanding of the relationship between resources and conversion factors and its implications for capabilities with reference to a smartphone (see also Sen (1985) and Robeyns (2005, 2016) for the example of a bicycle)). Any resource has specific and typical characteristics that makes us pay attention to it. Hence, individuals are not interested in smartphones because they are technical objects that consist of certain materials like processors and batteries, but because they make them capable of doing an array of things that are valuable to them, like, for instance, reaching out to others.
across time and space through text messages. However, the extent to which the smartphone contributes to this functioning, depends on the person’s ability to read and write and to use a smartphone (personal), the existing cultural traditions and policies that decide if an e-inclusion plan is in place to secure (also so-called digital) literacy (social), and the availability of a physical infrastructure in the form of a telephone network and, in the case of messages send over the internet, a mobile broadband connection (environmental).

Taken together, the means and conversion factor components of the framework underlines the importance of scrutinizing the context that frames individuals’ capabilities and the choices they make from them: “the freedom of agency that we individually have is inescapably qualified and constrained by the social, political and economic opportunities that are available to us” (Sen, 1999, pp. xi–xii). Sen thereby offers a typology and terminology for how to approach people’s scopes for action and conditions of life. However, because the approach does not focus on communication or digital technologies, it raises a subsequent challenge of conceptually developing and identifying the components – or subject areas – that ought to be studied when approaching communication as capability.

Capabilities in communications

Sen hardly ever talks about the internet (for an exception, see 2004, p. 79) and only makes sporadic or implicit references to communication (most notably when addressing freedom of speech (Sen, 2009, p. 63; see also Nussbaum, 2007, p. 80)). This is symptomatic of what has been a general disconnect between the capabilities community on the one hand and the media and communication field on the other (Birdsall, 2011). The reasons for this disconnect are perhaps quite straightforward: different social conditions suggest different priorities, and the framework was not initially developed with affluent and highly digitised contexts like Denmark in mind (see Sen, 2004, p. 79). The framework is, however, dynamic and adaptable, dependent on what we are investigating and why:

*The capability approach is a framework of thought, a normative tool, but it is not a fully specified theory [...] It is not a mathematical algorithm that prescribes how to measure inequality or poverty, nor is it a complete theory*
of justice. The capability approach, strictly speaking, only advocates that the evaluative space should be that of capabilities (Robeyns, 2003, p. 64).

In other words, like affordances of media (Hutchby, 2001), human capabilities should be assessed in relational terms (Jensen, in press).

In recent years, a small but increasing number of media and communication scholars have applied the capability approach to the understanding and regulation of (digital) media and communication (Birdsall, 2011; Comim et al., 2018; Couldry, 2019; Garnham, 1997; Hesmondhalgh, 2017; Jacobson & Chang, 2019; Jacobson, 2016; Jensen, in press; Litschka, 2019; Mansell, 2002, 2017a; Moss, 2018; Shomron, 2020; Sourbati, 2012). While these studies might disagree on the specific role of communication in the capability approach – as a means or a resource, a capability, a functioning or an end – they principally agree that communication is fundamental to what humans are capable of (e.g. Birdsall, 2011; Couldry, 2019). Jensen (in press) argues that communication can be considered a particularly architectonic capability on par with, as well as conditional for and constituent of, the central capabilities in Martha Nussbaum’s work (2011) of ‘affiliation’ to other people and ‘practical reasoning’. In this view, communication is both a capability in itself and a basis for other capabilities, a way of being and a way of doing, a means towards particular ends and an end in itself (Jensen, in press). The capabilities perspective thereby broadens concerns about wellbeing and (in)equality from common questions in the field concerning access to digital technologies or consumption of digital content to the ways the same technologies frame human existence and everyday communications:

(...) what the capability approach argues is the need to develop and agree more fine grained measures and indices that reflect what people in practice can or cannot do with these services, the benefits they do or do not derive from them, rather than measuring mere access or expenditure” (Garnham, 1997, p. 34).

As Robin Mansell (2017a) contends with regards to existing investigations of so-called digital divides, the approach focuses attention to aspects that go well beyond the first (access) or second (skills and knowledge) level of the divide, to a third level that is more concerned with the outcomes and implications of (not)
obtaining both access and skills in the first place (see also Mihelj et al., 2019): “[n]ot only do so-called digital divides trace spatial and social inequalities (Norris, 2001), they may also increase social polarization further” (Kleine, 2013, p. 5).

Common to the recent applications of the Capability Approach is a moral or philosophical aim of arriving at what constitutes capabilities and functionings that are specifically valuable in relation to media and communication, and therefore relevant to media and communication research and policy. This includes paying attention to and discussing new and emergent capabilities as entitlements that change when the structural conditions that frame people’s lives change:

> While I personally hate the mobile phone, we have to recognise that mobility widens the capability set and is a functioning that is increasingly widely available. Are those without it now sufficiently disadvantaged to make it an entitlement? Similar questions are already being raised in relation to Internet connections, usage of the World Wide Web, etc. (Garnham, 1997, p. 34).

The development of the internet entails, in this respect, a fundamental structural change, in need of new and coherent perspectives on what constitutes a good – digital and datafied – life:

> We have never needed more a sharp focus on what counts as good, on what complex, layered states of affairs cannot be regarded as good: states of affairs that threaten the fragile and deeply interconnected mass of possibilities that we would want to include among the conditions of a good life (Couldry, 2019, p. 53).

In a recent discussion of how Sen’s approach and complex account of ‘the good’ can benefit media and communication research on the implications of datafication and increasingly automated decision-making, Nick Couldry (2019) develops a list of functionings that ought to be valued in relation to media and communications, namely: “(i) not being harmed/humiliated, misrepresented through media, (ii) basic access to media resources, (iii) opportunity to be in some broad sense represented as a type, plus (iv) if one wants the chance to speak, the opportunity of voice” (Couldry, 2019, p. 48; see also Garnham, 1997), as well as (v)
recognition, and (vi) being listened to (Couldry, 2019, pp. 49–50). Albeit a valuable starting point, the list leaves a wide and critical gap for capability research concerned with both the opportunity to be left alone and what is commonly referred to as the right to privacy in communication (Birdsall, 2011, p. 99). In one of Nussbaum’s (2000) central capabilities, the ‘control over one’s environment’, she touches upon this issue by elaborating on the capability as “having the freedom from unwarranted search and seizure” (p. 80), yet this particular capability does not play a central part in existing applications of the capability approach to media and communications (with the exception of Mansell’s work, e.g. 2017a). Although invisibility – being misrepresented or not represented as a type (Couldry, 2019, p. 48) – can certainly be a form of injustice, so can visibility, particular once we start inquiring who is visible to whom. A capabilities perspective on the entitlements to be left alone and privacy, thereby fosters entirely different questions about the relationship between digital communication and capabilities, choice and power, and visibility and invisibility:

The introduction of a particular technology involves the extension of the capabilities of some, empowering them while making others disempowered or even redundant. Thus a central task will be to question whose capabilities [...] are being extended, and what the implications of this might be (Lawson, 2010, p. 220; in Oosterlaken, 2011, p. 430).

What individuals are capable of in and through digital communication is, thus, enabled and constrained by the capabilities of others – individuals, commercial corporations, state bodies, and so forth – to track the increasingly datafied lives of consumers and citizens. Capabilities are, equally, influenced by individuals’ knowledge (or lack thereof) of the choices they have, and so the choices they make, which can result in quite different functionings and outcomes than the ones promoted by corporate logics:

There needs to be an awareness that the technologies we are creating carry embedded ideologies within them which need to be subject to scrutiny so that users can at the very least recognise what choices have already been made for them if they choose to use a particular technology (Kleine, 2011, p. 126).
The quote above emphasises the need for studies that scrutinise the underlying values that govern choices about choices, or choices that have been made for users whenever they communicate digitally. In other words, the technologies and applications that individuals use and the people who own and profit from them. It also challenges quantitative politics stating that “[i]f information wants to be free, then more connectivity equals more freedom” (Thacker, 2004, p. xv) by representing a contradiction that is central to the understanding of capabilities at the intersection between digital communication and datafication put forth in this dissertation, namely that:

Platform operators and infrastructure providers will continue to seek public legitimacy for their practices in order to ensure the financial sustainability of their business models (Andersson-Schwarz 2016), but these will amount to small shifts in strategy [...] Such tinkering is unlikely to address the principle contradiction [...] that the more digitally mediated benefits we have, the fewer opportunities there are for humans to exercise their control and authority (Mansell, 2017a, p. 156, my emphasis).

The capabilities perspective is thereby a way of moving beyond established and exhaustively discussed contradictions like the ‘privacy paradox’ (Barth & Jong, 2017; Norberg et al., 2007; Young & Quan-Haase, 2013) and towards questions concerning people’s fears of what they will not be capable of as a result of datafication. That is, beyond a loss of privacy (since this might not be what people are afraid of losing per se), and towards a loss of the opportunity to not be put in a certain box, to be able to recognise their data as it returned to them, and so forth.

There are, besides the few that apply the Capability Approach to communications, studies that do pursue research agendas adjacent to questions about what people can and cannot do in and through their digital communications and why, although they never mention the concept of capabilities let alone the approach. This includes research on datafication related to so-called critical data studies; studies of digital infrastructures; and political-economy research on digital business models and market structures. I will briefly go over existing research in these domains in order to outline how they motivate my operationalisation and development of the Capability Approach into a research agenda for the
dissertation as well as future research. I also review the three domains in order to, later in Part I, show how research in each of them can benefit from the capabilities perspective, and thereby how the dissertation contributes by addressing gaps in existing research when it comes empirical and coherent approaches to digital communication that traverse institutional boundaries as well as macro, meso, and micro levels of inquiry.

**Critical data studies**

In recent years, an increasing amount of scholarly work, which can be collected under the nascent field of critical data studies (Dalton & Thatcher, 2014), has looked beyond the hype and seduction surrounding big data in both research and society (boyd & Crawford, 2012), and discussed the implications of increasing datafication (Mayer-Schönberger & Cukier, 2013) of all aspects of life. Broadly defined, critical data studies entail:

> research and thinking that apply critical social theory to data to explore the ways in which they are never simply neutral, objective, independent, raw representations of the world but are situated, contingent, relational, contextuality, and do active work in the world (Kitchin & Lauriault, 2018, p. 7).

Though interdisciplinary and brought together by rather loose ties, critical data studies thus share profound concerns about the societal consequences of datafication as well as the lack of adequate responses to data harvesting on the part of individual users. Inspired by the Aristotelian notion of “education for the common good”, Iliadis and Russo (2016, p. 5) summarise three basic principles of critical data studies: namely 1) “the identification of social data problems”, 2) “the design of critical frameworks for addressing social data problems”, and 3) “the application of social solutions to increase data literacy”. Thus, sharing Sen’s focus on enhancing freedom and choice, they go on to stress that critical data studies ought to emphasise applied, participatory, and interactive approaches towards empowering people to act on datafication:

> CDS [critical data studies, ed.] should provide individuals with the necessary tools for becoming more informed and the ability to organize efforts
around data justice issues. By maintaining these orientations and principles, CDS should encourage us to think about Big Data science in terms of the common good and social contexts (2016, p. 5).

In a definition of the field, Dalton and Thatcher (2014) offer seven provocations for comprehensive critiques that 1) situate data in time and space rather than detached from their precursors and ancestry; 2) enforce the understanding that data are never neutral but always serve certain interests; 3) uncover the complex relationship between data and society and confront hard technological determinism; 4) show the ways data are always cooked according to some(one’s) recipe; 5) challenge the delusions that big data will replace small data or that ‘raw’ data speak for themselves; 6) investigate how data and counter-data may be used in socially progressive ways; and lastly 7) emphasise the potential in combining big and small data in critical analyses. Summarising their critique, Dalton and Thatcher (2014) further pose five central research questions, which might be asked of many societal phenomena beyond big data: namely, in my own rough translation: How did we get here? Who controls it? Who produces it? How is it applied? And what can we do about it?

Taken together, although studies in this vein often acknowledge the benefits of big data to society (think of the ways data are used to e.g. confront humanitarian and environmental crises), critical data studies usually focus on the “darker side to much data work” (Kitchin & Lauriault, 2018, p. 14). Coalesced under the heading of uncovering this darker side of big data, three research areas in particular approach datafication as an enabling and constraining force. One focuses on the commercial aspects of datafication and looks at different commodification processes, in which personal data are monetised and sold as inputs for advertisers and other stakeholders that are in the business of profiling, filtering, curating, categorising, and targeting users (e.g. Beer, 2016; Dijck, 2014). From this perspective, datafication is seen as fuelling a market that thrives on both the increasing availability of data as well as the ways digital surveillance of everyday life is justified and naturalised through limited transparency and knowledge surrounding data practices, eventually leaving people with the understanding that there is no other way of organising society (Dencik & Cable, 2017, p. 777; see also Couldry & Yu, 2018, p. 4474; Draper & Turow, 2019, p. 1833). That is, as
a result of the work of powerful actors invested in enhancing commercial surveillance, data harvesting and reuse come to be understood as a natural and inevitable part of life. Outlining how the same practices can and should be, almost was, but effectively are not outlawed, Zuboff (2019) further argues that, in effect, the large surveillance capitalists are left with immense power over human futures. Another research area is concerned with the ontology and epistemology of data and data ethics (van Dijck, 2014; Lomborg, 2013; Zwitter, 2014). This includes critical perspectives on what Helen Kennedy aptly describes as “a desire for numbers” (2016, p. 11): in contrast to policies aimed at minimising data collection (e.g. the GDPR), big data science relies heavily on generating an excess of data — data maximization — with or without a defined purpose (Kitchin & Lauriault, 2018, p. 15). Relatedly, discussions surrounding ‘data justice’ (Dencik et al., 2019) and data activism (Kazansky et al., 2019; Milan, 2017) have extended the debate on the societal implications of datafication beyond issues of security and privacy, and towards new and existing inequalities, democratic processes, discrimination and exclusion, diminishing working conditions, automated decision-making around sensitive topics, and algorithmic bias (Eubanks, 2017; Noble, 2018; O’Neil, 2016; Sandvig et al., 2016). Lastly, research on internet governance (DeNardis, 2012; Flyverbom et al., 2019; Kitchin, 2014; Milan & ten Oever, 2017) has identified a governance gap (Jørgensen, 2019; MacKinnon, 2012) in which big data corporations operate, a failure of ‘old’ legislation to govern ‘new’ actors in the digital realm (Flensburg & Lai, 2020a [article 3]; Just, 2018), and a disconnect between the invisible and the governable (Flyverbom, 2016).

Critical data studies, and the research agendas they pursue, have thus contributed with important theoretical insights and rich conceptual work regarding the characteristics of big data (volume, velocity, resolution, scalability, and so forth (see. e.g. Kitchin, 2013)), central research questions (how did we get here, how is it applied, etc. (see e.g. Dalton & Thatcher, 2014)), and the need for collective responses to political and economic processes that hinge on big data (powered by informing users, organised efforts around social justice, etc. (see e.g. Iliadis & Russo, 2016)). However, although some empirical work has been carried out in particular user groups, contexts, or platforms, there are consistent calls for more “empirical research to underpin and flesh out critical data studies” (Kitchin &
Lauriault, 2018, p. 18). In other words, more coherent empirical work that can support and validate critiques as well as political and regulatory responses to datafication and emergent power structures. To this I might add the need for more studies that answer, rather than just pose questions, and apply theory to empirical explorations. By empirically interrogating the infrastructural tenets of big data, a bulk of research that can be collected under the field of infrastructure studies goes some of the way towards answering that call, and, as such, towards understanding how digital infrastructures enable and constrain what individuals are capable of in and through digital communication.

**Infrastructure studies**

The scarcity of coherent empirical work to substantiate the critical data studies field can in part be traced back to a notion of digital infrastructures as *invisible* (Bowker et al., 2009; Star & Ruhleder, 1996) and so inherently difficult to study and regulate. Algorithms are for instance articulated as hidden, proprietary, and black-boxed, thereby leaving people with nothing but their own imagination of how they might work (Bucher, 2017). Recently, however, following the ‘turn to infrastructure’ (Musiani et al., 2016) in internet governance studies, media and communication researchers have shown an increasing interest in studying “the Internet not as ‘what people say with it’ but as ‘how it works’” (Sandvig, 2013, p. 90). This entails foregrounding and making infrastructures like the wires of ‘wireless’ internet technologies (Starosielski, 2015) or the sizeable data centres for hosting ‘cloud’ services (Holt & Vonderau, 2015) visible. Studies are thereby focusing on the structural rather than the symbolic aspects of digital communication, and critical research questions concerning ‘who controls it’ and by what means (DeNardis, 2012) as well as political-economy analyses of digital infrastructures (e.g. Dijck et al., 2018; Plantin & Punathambekar, 2019) take centre stage.

Infrastructure studies can be devised along two lines of inquiry following, on the one hand, a ‘relationist’ (Sandvig, 2013) perspective, and, on the other, a materialist perspective on infrastructure as a key concept. When defined as a “fundamentally relational concept” (Bowker et al., 2009, p. 99), infrastructure is tied to the specific social practices revolving around, and constructing, it. “[T]hey
[infrastructures] are not things themselves” (Weltevrede & Jansen, 2019), but are dependent on particular practices and perspectives: “the daily work of one person is the infrastructure of another” (Bowker et al., 2009, p. 98). Relationist approaches are, as such, useful when analysing specific engineering and innovation processes from qualitative perspectives (Harris, 2015; Parks & Starosielski, 2015). Studies in this vein have contributed with a multitude of insights that are particularly valuable for addressing how infrastructures as enabling and constraining mechanisms can be and are negotiated. These count for instance Lisa Park’s (2015, 2016; Parks & Mukherjee, 2017) work on the practices around digital infrastructure scarcity in Zambia, or Ramesh Srinivasan’s (2019) conversations about a democratic internet with designers in rural, low-income, and indigenous communities as well as public figures in activist and labour organisations. Yet, the relationist approach is difficult to apply to more systematic and large-scale studies aimed at analysing the general features or structures of, for instance, digital communication networks, regardless of how they are used in particular practices. In contrast, a materialist perspective on infrastructure returns to a somewhat more conventional use of the term, which emphasises infrastructures as physical resources that can be built, owned, used, and regulated. Inspired by medium theory (Innis, 1949; Meyrowitz, 1985; Peters, 1999) and existing analyses of the societal consequences of infrastructural changes as well as of the interplay between institutional structures and material conditions, studies in this vein usually focus on material aspects of digital infrastructures and on uncovering the ownership and power structures regulating access to and development of them as critical resources (e.g. Helles & Flyverbom, 2019):

The Internet has a complex technical architecture beneath the layer of applications and content and generally out of public view. This architecture includes a considerable ecosystem of Internet governance technologies, meaning the digital systems and processes inherently designed to keep the Internet operational (DeNardis, 2012, p. 721).

DeNardis’ way of understanding digital infrastructures as control mechanisms is mirrored in for instance Dwayne Winseck’s (2017, 2019) research on core elements of the global internet backbone – fibre-optic submarine cables, autonomous system numbers (ASNs), and Internet Exchange Points (IXPs) – or
Alexander Galloway’s (2004) analyses of digital networks and internet protocols (see also Pötzsch, 2017 for a comprehensive review of material approaches to digital media). These studies are, however, in a minority position insofar as digital infrastructural systems are still rarely studied from material and coherent macro-perspectives (Lievrouw, 2014).

There is, within infrastructure studies, only a limited body of work that deals explicitly and empirically with what is perhaps best described as infrastructures of tracking (see Helles, Lomborg & Lai, 2020 [article 4]) or surveillance infrastructures (Lai & Flensburg, 2020a [article 5]). These infrastructures are both a foundation for and a result of the immense datafication of all aspects of life that follow from digitisation of infrastructures and commodification of data. The fact that they are heavily understudied might be because they too seem quite invisible, and deliberately so: research in the domain indicates how corporate obfuscation serves to keep tracking infrastructures under the radar – for users, researchers, and policy-makers alike (Draper & Turow, 2019). This calls for studies, both of how surveillance infrastructures operate, technologically and economically, as well as of their ramifications in terms of conditioning individuals’ contemporary and future capabilities: their communications, access to information, and privacy. Existing empirical studies on surveillance infrastructures chart the mechanisms invoked in the surveillance of individuals as they move around (Perng et al., 2016), browse the web (Falahrastegar et al., 2014; Kalavri et al., 2016), use dating apps (Weltevrede & Jansen, 2019), shop in retail stores (Turow, 2017), and so forth. Some focus on particular case-studies like in Wood’s (2019) analyses of policing apps or Helmond’s (2017) mapping of third-party cookies on The New York Times site over time. Others, largely computer-science based approaches, focus more broadly on large-scale mappings of tracker ecologies (e.g. Atkinson et al., 2015; Binns et al., 2018; Libert, 2015). Krishnamurthy and Wills (2009), for instance, employ a historical perspective and map the consolidation process of the market for web tracking, and Falahrastegar (2014), Roesner (2012), and colleagues develop typologies and categorisation schemes to differentiate between trackers and the work that they do. As a common denominator, many of the latter studies have a normative aim of reverse engineering different tracking mechanisms and thereby developing tools for circumventing surveillance. The studies are, however, also largely descriptive and call for more
theoretical engagement towards understanding and explaining digital infrastructures and data flows as parts of wider social, economic, and political structures. One approach for how to engage with infrastructural power – as the capacity for controlling the material foundations of a digital communication system – can be found in the political economy of communication tradition.

**Political economy**

Political-economy studies can be defined as research on the distribution of goods in society, as well as of corporate and political power structures around this distribution. This entails studying how critical communication resources – like “the internet [infrastructure, ed.] as one of today’s most important forms of world property” (Mazepa & Mosco, 2016, p. 163) – are organised and controlled, how the distribution reflects and sustains existing and emergent power structures in society (Mansell, 2017b), and its implications for what people are capable of in their lives. Mosco (2009, 2014) outlines four central characteristics of critical political economy that all resonate with the research aims and contributions of the dissertation: 1) continuing the tradition of classical political economists, political economy consistently aims to uncover social change and historical transformation as part of the dynamics of capitalism; 2) political economy also strives to study the totality of social relations, that is the economic, political, social, and cultural spheres of life and their interrelations; 3) in its commitment to moral philosophy and interest in both the values that partly create social behaviour and the ones that ought to guide efforts to change it, political economy is both descriptive and normative; and finally 4) following a social praxis perspective, political economy is committed to transcending the divide between thought and action, thereby calling for interventions that foster actual social change and practice through for instance policy initiatives grounded in research (Mosco, 2014, pp. 37–38; Wasko, 2004, p. 310).

In the two sections above, I referred to studies which might as well have featured here (e.g. DeNardis, 2012, 2020; José van Dijck et al., 2018; Flyverbom, 2016; Flyverbom et al., 2019; Plantin & Punathambekar, 2019). They focus on the business models and strategies, ownership structures, and control mechanisms that characterise digital markets and data economies, and thereby structure the
ways communication resources are distributed in society. Although they can be collected under different variations of political-economist perspectives within the field – political economy of media (Hardy, 2014), of communication (Mosco, 2009; Wasko, 2004), of information (Garnham, 2011), of new media (Mansell, 2004), or of the internet (Mazepa & Mosco, 2016) – these studies centre around unusual, perhaps subordinate or even overarching phenomena: namely the interplay between digital infrastructures on the one hand and political and economic power structures on the other. DeNardis (2012), for instance, challenges a consistent focus in internet governance research on the role of institutions in establishing public policies, and she refocuses attention towards core governance functions “carried out via arrangements of technical architecture and through policy decisions of private industry” (p. 721). Mansell (2017b) shows that a policy focus on net neutrality (the end-to-end principle) tends to cloud over discussions about how control over data has changed and the ways “citizen interests have been subordinated to corporate and state interests historically” (p. 3). Furthermore, she critiques how focus remains on the responsibility of either individual users or (un)compliant corporations rather than legislators and policy-makers (see also Cammaerts & Mansell, 2020). Finally, Andrejevic (2011) argues that privacy – as a philosophical and user-centric concept – overshadows discussions about who controls communication infrastructures and for what ends (p. 279).

Taken together, these studies can be said to map a number of distraction or cover-up stories that detours research away from digital infrastructures as critical resources and mechanisms of power. There are, however, only a few empirical studies that uncover these cover-ups by mapping out the contours of infrastructural power, and even fewer that do so outside the established and enduring sectoral silos of media and communication research and policy.

As mentioned, the research agendas pursued in the fields above all contribute with important knowledge on the implications of big data, the relations and materiality of infrastructures, and distribution and power structures around digital infrastructures and data. My adaption and development of the Capability Approach hinges on this knowledge insofar as the different domains focus attention to aspects that are particularly important when approaching capabilities at the intersection between digital communication and datafication. Aspects that do not, as of this writing, play prominent roles in the existing applications of the
Capability Approach to media and communications. And aspects that are also usually studied within the confines of the particular fields and through established methods within them, thereby offering important theoretical discussions yet without following them through in empirical analyses. A discussion of the “big data divide” by Andrejevic (2014) highlights this issue. The idea of a big data divide rests on boyd and Crawfords (2011) notion of the ‘big data rich’, who can “generate or purchase and store large quantities of data” (Andrejevic, 2014, p. 1675), and the ‘big data poor’ who are “excluded from access to data, expertise, and processing power” (Ibid.). This is also a divide between the ‘sorters’ “who are able to extract and use un-anticipatable and inexplicable [...] findings” (p. 1683) and the ‘sortees’ “who find their lives affected by the resulting decisions” (Ibid.). Citing David Weinberger (2011), Andrejevic appropriately argues that, if “in the era of big data ‘the smartest person in the room is the room’, then much depends on who owns and operates the room” (2014, p. 1676). Yet, he then turns to survey and focus group data in order to discuss “what people talk about when they talk about privacy” (p. 1682). These user perspectives are important and also documented in, as Andrejevic himself puts it, “[s]urvey after survey, including my own” (p. 1677), yet, they are also far from enough when it comes to decrypting ‘the room’ and its infrastructural tenets, let alone who owns and operates it.

The final section of my theoretical framework outlines a systematic approach to communication as capability that is both inspired by Sen and Nussbaum’s work as well as by research in the domains reviewed above. As a research agenda, the systematic approach focuses attention to research areas that are seldomly studied together and research perspectives that are seldomly combined.

**Communication as capability**

The Capability Approach represents a particular way of understanding wellbeing and agency. Although it is not explicitly applied in most of my articles, it serves as a common denominator for how the research questions of each article are approached – namely from the normatively individualistic but the ontologically non-individualistic perspective of capabilities. This entails taking a step back
from established fields of inquiry, methods, and empirical sources, and instead (re)considering, in comprehensive and systematic ways, the emergent and fast-moving structures that enable and constrain what individuals are capable of in and through digital communication. As Dorothea Kleine (2013) argues, “if individuals themselves conceptualize the lives they value in complex ways” (p. 15), then “[t]he Internet as a multifunctional and transversal set of technologies is best understood in a holistic and systematic analysis” (Ibid.).

Although media and communication research has contributed with important perspectives on the relationship between the Capability Approach and digital communication, the majority of studies, as mentioned, remain at a theoretical level. This also means that they do not offer ways forward for operationalising the approach or empirically exploring capabilities as constituents of a good life in datafied societies. This might be because the advantages of the Capability Approach – its nuancing of what entitlements to a good life are, its conceptual richness, and its commitment to engaging with people’s capabilities in context – are also its disadvantages, making it particularly difficult to operationalise (Kleine, 2013, p. 4). Coming out of the ICT4D community, which is, as implied in the name, concerned with researching the ways internet and communication technologies (ICTs) and development interact, there are, however, a number of studies that do translate the approach into empirical frameworks (Alampay, 2006a, 2006b; Hatakka et al., 2019; Hatakka & Lagsten, 2012; Kleine, 2013). Taking inspiration from previous operationalisations, figure 2 below shows my amended version of Sen’s framework and a systematic approach to communication as capability by mapping out the components – or research areas – that are crucial to cover in any empirical engagement with capabilities at the intersection between digital communication and datafication. I will explain each of the components as they are motivated by previous adaptations of the approach to media and communications as well as the existing research on datafication, digital infrastructures, and digital business models and market structures. The concept of capabilities, and the individual perspective it entails, is thereby a way of identifying a research agenda and creating attention around structures that frame how digital communication is enabled, constrained, and negotiated in a datafied society. The empirical studies that target these structures and mechanisms are, in turn, featured in the articles of the dissertation.
As a representation of a research agenda, figure 2 shows the interaction and co-dependence of research areas within media and communication studies that are usually not connected. Although the framework follows the basic logic of the Capability Approach (see figure 1), it also exhibits a number of important alterations that are consistent with an adaption of the framework to studies of digital communication.

The first component of the framework covers digital technologies as means to achieve. As meta-media (Jensen, 2013), the internet, its hardware and software, incorporate and modify other media and other types of communication, while also enabling new ways of communicating (Ibid.). When approaching digital communication from a capabilities perspective, access to digital technologies – connection through digital networks and devices – is thereby an outset for discussing potential achievements and outcomes following from it:

*Technology plays a very important role in expanding human capabilities [...] and the CA [Capability Approach, ed.] should ascribe causal efficacy not only to individuals and social structures, but also to technical artefacts. All three form a constitutive element of human capabilities (Oosterlaken, 2011, p. 431).*
In contrast to Sen’s framework, the amended framework sets apart internal and personal conversion factors from the external, social and environmental conversion factors, thereby demarcating different research areas that are addressed in my articles. Taking inspiration from Kleine’s (2013, p. 44) adaption of Sen, personal conversion factors are conceptualised as resources that an individual can have more or less of, and which give them more or less clout over their freedom. I divide these resources into five rough types that include psychological-educational (like self-confidence, resilience, imagination, positivity, and formal and informal education and skills), material-financial (material objects and income, savings, etc.), temporal (the degree of control over one’s own time), cultural (status and power), and social (social networks and group membership through friendship, kinship, etc.) resources (see Kleine, 2013, pp. 46–48). As aspects of agency, an individual’s resources are enabling and constraining when it comes to her scope for action, but also, importantly, the individual uses her resource-based agency to navigate and negotiate the social structures that frame her everyday life. The dissertation focuses on these different aspects of agency in three particular instances: by approaching intermediality – the systemic relationship between all available communication types (Helles, 2013, p. 17) – on the basis of empirical individuals’ beings and doings, and the choices they make as part of navigating and negotiating the multitude of media that are available to them in their everyday communication (Lai et al., 2019 [article 1]); by engaging with the infrastructural and economic implications of mobile app usage from the perspective of individual users and their educational and social capacities for understanding the choices they have and the ones that have been made for them within that environment (Lai & Flensburg, 2020b [article 6]); and by discussing the ways individuals’ psychological, cultural, and temporal resources interact with infrastructural mechanisms, digital business models, and gendered divisions of labour (Lai, in review [article 7]). Individuals’ resources are, in this sense, in a complex, enabling, or constraining relationship with structural factors.

The external conversion factors are, in turn, conceptualised as infrastructural, market, and policy structures. This conceptualisation mirrors my understanding of digital communication systems but is equally motivated by the materialist approaches to digital infrastructures and political economy analyses of infrastructural power that were outlined in the sections above. The mappings comprised
in articles 2 (Flensburg & Lai, 2019) and article 3 (Flensburg & Lai, 2020a), thus, provide crucial information on the context that frames my respondents’ everyday lives and the structural conditions that are more or less negotiable. The understanding of digital infrastructures as a set of conversion factors contrasts previous operationalisations of the Capability Approach insofar as ICTs goes, where infrastructure is considered solely as a means or a capability input (Haenssgen & Ariana, 2018, pp. 100–101; for an exception see Kleine, 2013). As the underlying assets that communication markets and policies revolve around, digital infrastructures thereby come to play an equally important role on par with economic and political structures in analyses of what people are capable of in and through digital communication, and why. Digital infrastructures enable and constrain the conversion of means – like the smartphone – to capabilities and functionings – like the possibility of micro-coordinating everyday events on the go, as well as the corresponding achievement of making ends meet through mobile communication. As infrastructures for datafication – and commodification and monetisation of data – infrastructural mechanisms like the cookie scripts and software development kits (SDKs) analysed in articles 4 (Helles, Lomborg & Lai., 2020), article 5 (Lai & Flensburg, 2020a), and article 6 (Lai & Flensburg, 2020b) enable users to, for instance, save their previous online beings and doing (like items in an online shopping cart2) or use logins. They, however, also constrain individuals’ abilities to control their personal data flows: the ways their data are processed, distributed, and eventually turned back upon them. Big data, thus, enhance the capabilities of some (usually for trading in and profiting from human futures (Zuboff, 2019)), while diminishing those of others (to choose their own futures independent of the ways they are profiled and targeted through filtering, curation, and discriminatory practices). Empirically grounded studies of how big data is made and remade, as well as of the economic incentives for the increasing datafication of everyday life, thereby become critical to any comprehensive study of how digital communication is enabled and constrained.

---

2 The HTTP cookie was originally implemented by Netscape Communications in 1994 for this purpose (Kristol, 2001). It thereby offered a solution to a crisis of control (Beniger, 1986) when it came to the World Wide Web and its lack of infrastructure for recording and remembering past uses.
Taken together, individuals’ resources and the structural conditions interchangeably impact on the freedom to achieve. Some capabilities are turned into functionings and others are not, depending on the choices individuals make from their combined capability sets. Choices are, in turn, both enabled and constrained by resources and structures. Inspired by Kleine’s (2013, p. 44) four ‘degrees of empowerment’, the *choice* component of the amended framework in figure 2 is elaborated on by way of distinguishing between *existence*, *sense*, *use*, and *achievement* of choice (Ibid.). The *sense of choice* is particularly important to this dissertation, as it points to issues surrounding knowledge gaps when it comes to datafication, and thereby a lack of sense of choice and an abundance of digital resignation (Draper & Turow, 2019) when it comes to online surveillance (see elaboration in Lai & Flensburg, 2020b [article 6]). While both Mansell (2017a) and Couldry (2019) emphasise the need for studies that scrutinise *choices about choices*, also research in the critical data studies domain dive into the implications of various datafication mechanisms like automated-decision making systems for the *existence of choice* in the first place. While article 1 (Lai et al., 2019), article 6 (Lai & Flensburg, 2020b) and article 7 (Lai, in review) look into the choices people (do not) make, why, and with what outcomes, article 5 (Lai & Flensburg, 2020a) charts the choices that have been made for people and the ones that are up for negotiation by analysing mechanisms of mobile infrastructures for datafication.

Finally, the arrows in the framework emphasise the causal power of, for instance, inequalities in means and access to generate inequalities in capabilities and functionings. It also shows the ways structural conditions as well as individuals’ personal resources interchangeably affect the conversion of means into capabilities and functionings. Further, it shows how the choices people make, and the ones that have been made for them, result in realised functionings that in turn reflect back upon both structures and resources, thereby negotiating their conditioning powers.

Motivated both by Sen’s basic framework (figure 1) as well as by the emphases of recent critical data studies, infrastructure studies, and political economy analyses, figure 2 outlines a systematic approach to understanding digital communication from a capabilities perspective. As a research agenda, the systematic approach focuses attention to components – or research areas – that are important
to analyses of communication as capability. The articles of the dissertation, in turn, comprise a total of seven empirical studies that traverse the different research areas and make comprehensive accounts of them. Having established a theoretical framework around capabilities, the same framework thereby raises questions concerning how the different components and their interdependencies, as they are visualised in figure 2, may be approached through established and emergent methods for making, finding, and analysing data. Following some of the general epistemological tenets of the Capability Approach, the studies that I will describe below are empirically grounded, often holistic, macro-oriented, and critical, aimed at approaching the good – digital and datafied – life. In other words, they are aligned with Sen’s argument that we have to approach capabilities in their empirical contexts and return to individuals and their diversity as fundamental aspects of our engagements. Also, rather than seeing “the perspective of recipients and users, as a counterpart to the perspective of [e.g. ed.] media systems, institutions and professionals” (Sandvik et al., 2016, p. 9), the studies scrutinise the structures that frame individuals’ capabilities, thereby reconciling the normatively individualistic with the ontologically non-individualistic perspectives of the approach in holistic and/or macro-oriented investigations. And lastly, they follow the impact agenda of the Capability Approach by applying critical perspectives on issues of power and control in order to pinpoint unrecognised divisions of labour that exacerbate existing inequalities and to augment individual freedom.

3 The most direct and significant results of Sen’s work can be found in the United Nations Development Programme’s (UNDP) Human Development Index (HDI). The HDI challenges the adequacy of measures like GNI and life satisfaction in assessments of development and freedom, and instead broadens the analytical scope to also include other aspects of what makes a good life – including health and education.
Methodology: following individuals, following infrastructures

The sections above outline a particular understanding of capabilities the intersection between digital communication and datafication. They also outline a research agenda that calls for discussions of how we then approach and collect data about capabilities and the ways they are conditioned and negotiated. The following sections therefore outline the methodology of the dissertation and explain the particular methods that ground the studies comprised in it; the ways in which the research objects determined the choice of methods; as well as how they relate to covering different aspects of the research agenda. I will focus more on how the methods were developed as part of an ongoing research effort – how they relate to each other and to the systematic approach sketched out above – and less on the particularities of each method, which can be read in the individual articles’ method section.

The articles comprised in the dissertation all depend upon different research designs, methods, and empirical data(sets). All the approaches do not capture the same thing, yet, they complement each other by engaging with different aspects of the problem of understanding what individuals are capable of in and through digital communication; how this communication is enable and constrained; and its everyday implications for individuals. The data collection methods span qualitative and quantitative approaches for collecting both found and made types of data (Jensen, 2013, 2012): while ethnographic fieldwork data were made, data on the internet backbone or cookie-scripts embedded in websites were found. The data analysis methods, in turn, amount to open coding of fieldwork data, thematic interview analysis, quantitative content analysis, statistics, and big data visualisations. As such, the dissertation blends both big and thick data sources (Lai, 2017).

In all of this, Denmark, as the context for the majority of my empirical explorations, is both a curious and an extreme case (Flyvbjerg, 2006; Yin, 2009). Curious because the capabilities perspective is most often applied to developing contexts, even though it can be, as demonstrated in Robeyns’ (2003) studies of gender inequality in Western societies, applied to contexts where the most immediate
entitlements to a good life are met. *Extreme* because communication infrastructures and the use of digital communication services are exceptionally widespread and highly developed in Denmark compared to most other countries outside the Nordic region; and because the emergence of digital markets, global governance, and so forth challenges established welfare regimes and the positive conceptions of freedom that they are built upon. The purpose of extreme case studies are often “[t]o obtain information on unusual cases, which can be especially problematic or especially good in a more closely defined sense” (Flyvbjerg, 2006, p. 72). That is, Denmark as a case of extreme connectivity is useful for scrutinising the implications of datafication as well as the trade-off (Draper & Turow, 2019) inherent in the many ‘free’ digital services that are ubiquitously weaved into Danish individuals’ everyday lives.

**Made**

Some of the analyses comprised in this dissertation involved the making of (a lot of) data. Most prominently, the ethnographic fieldwork that is thoroughly described in article 1 (Lai et al., 2019). The fieldwork is designed as a multi-sited ethnography (Marcus, 1995, 2011), which can be constructed according to different strategies. George Marcus (1995) outlines how ethnographers *follow*, for instance, “things”, “metaphors”, “conflicts”, or “people” in order to develop a multi-sited fieldwork (p. 106-110). Similarly, I find that researchers studying the role of the internet in everyday life delimit their research (and wisely so) by adhering to one of four general demarcation strategies, including following 1) particular *contexts* like the home (Silverstone, 2006) or the workplace; 2) specific *media devices* like mobile phones (Thorhauge & Lomborg, 2016) or laptop computers (Anderson et al., 2009); 3) particular communication applications like World of Warcraft (Poels et al., 2015), Twitter (Marwick & boyd, 2014), or Facebook (Miller et al., 2016); or 4) individual users (Beneito-Montagut, 2011; Hasebrink & Domeyer, 2012; Helles, 2012) (see Lai et al., 2019 [article 1]). Consistent with the latter strategy – as well as with the turn to individuals inherent in the Capability Approach and my aim of approaching digital communication holistically – the fieldwork focus on a total of 20 individuals as they traversed
the multitude of sites that provide the context for their everyday lives and individual capabilities.

In order to follow these individuals across contexts, media, and platforms that are usually not studied holistically but in isolation, the ethnography integrates recurrent qualitative interviews with communication diaries. This combination of methods allowed me to examine intermediality on the basis of the empirical user and in close relation to communication situations and networks. I was able to approach the respondents’ digital communication in the workplace, social gatherings, the home, during transportation or even bathroom breaks, and anywhere in between. I was also able to study how their social uses of the internet interacted with other ways of reaching out, how they combined and experienced networked communication as a coherent system of possibilities; and how they chose between different media and applications for different communicative purposes; and in accordance with both the affordances of those media and applications, but also with their knowledge, skills, and sense of choice. In the interviews and the diary entries, the respondents would talk about who they were, what they were doing, and why they did it (and did not do something else), what they thought themselves to be capable of, their understandings of the choices they had, and the ones they made, as well the outcomes of their particular beings and doings. The conversations showed me how their digital communications were enabled, constrained, and negotiated through their individual resources, be they material and financial, educational and psychological, temporal, social, or cultural.

Despite all the merits of doing ethnographic fieldwork and the findings that came from making and analysing ethnographic data, other aspects of the overall problem also call for a set of different methods and data sources that are, instead, dependent on how to approach the structural conditions that provided the context for my respondents’ everyday lives and their digital communications. In other words, methods that enable me to approach the ways the respondents’ capabilities were framed by the market and policy structures governing the Danish context as well as by the global and local internet infrastructures – networks, applications, and services – through which they live out so much of what counts as social life.
Some of the analyses comprised in this dissertation involved the finding of larger and big data. These analyses address the research problem from a different position in the systemic framework of figure 2, namely that of macro structures approached through quantitative methods. In and of their operation, digital networks generate a constant stream of meta-data – data about data – and thereby document aspects of communication that were often not captured by the registration mechanisms of previous communication media (Jensen, 2012, p. 436). These data can, in turn, be found and used for research. Chances are, however, that data found are also made for someone else to do something that is most likely different from the research aims of the researcher that finds it. It other words, although, for instance, meta-data generated through digital networks can be found in big measures, the data do not offer “a view from nowhere” (Jurgenson, 2014), and they also require a lot of remaking before they can be employed in empirical studies. In other words, in order to engage with the aspects of the systematic approach, visualised in figure 2, that were not concerned with individuals’ resources or their functionings at the micro level, but rather with the structural conditions of the Danish context, I left the field and returned to the desk in my office in search of data that might be found.

Approaching the infrastructure, market, and policy structures that create a common ground for digital communication in Denmark (and elsewhere) and frame individuals’ capabilities involved the development of the Digital Communication System (DCS) framework that is foundational for article 2 (Flensburg & Lai, 2019) and article 3 (Flensburg & Lai, 2020a). When developing this type of framework, a crucial step is to determine what indicators and measures to use and how to amass and process the relevant data (Flensburg & Lai, 2020b). The development of the framework follows both the turn to infrastructure in internet governance studies as well as the normatively individualistic perspective of the Capability Approach by mapping common structural and material conditions for digital communication from the perspective of individuals living in a specific societal context. This entails mapping infrastructures, markets, and policies as they relate to the specific infrastructural chain of events following whenever an individual communicates digitally: that is, she activates a digital device and an
access network, backbone network components, applications, and content. As such, the framework mirrors the openness inherent in Sen’s insistence on making capabilities (rather than GNP, personal income, or life satisfaction) the space for evaluation, by taking individuals’ (rather than specific institutions or policy programs) as its outset. It also, importantly, makes the argument that because any type of digital communication hinges on the existence and availability of digital infrastructures, then analyses of what people are capable of in and through digital communication have to start with digital infrastructures as the underlying resources that communication markets and policies revolve around. By focusing on the infrastructural conditions for and implication of any digital communications, the framework thereby enables analyses that cut across established institutional boundaries between sectors, policy fields, and research areas, and emphasise that the internet is not confined to or governed by any one of them alone. The mappings complement the ethnographic fieldwork insofar as they provide important knowledge on the contextual and structural conditions that are enabling, constraining, and sometimes negotiable (think of non-negotiable terms of service or hardwired functions in online tracking) when it comes to individuals’ digital communication.

The final analyses of enabling and constraining conditions of digital communication focused on digital tracking mechanisms. Mechanisms that are supported by access and backbone networks, digital applications, and small and rather unnoticeable content types like web cookies (Helles, Lomborg & Lai, 2020 [article 4]) or software development kits (SDKs) in apps (Lai & Flensburg, 2020a [article 5], Lai & Flensburg, 2020b [article 6]). Digital technologies as means for achieving things in life that are valuable, are more than often also means for other ends than (the ones imagined by) the user. While apps for findings dear and long-lost friends, killing time by crushing digital pieces of candy, or searching for information online contribute with valuable beings and doings to the respondents (and any other individual), they are, following the ontologically non-individualistic Capability Approach, also contingent on the extension of an entirely different set of capabilities. Namely that of commercial service providers and distributors that offer the services free of monetary charge, yet in exchange for the ability to track their users through the trails of data that they leave behind. These data are, in turn, sold to the highest bidder on the online ad exchange as means
towards the commercials ends of affecting people’s future choices and function-ings through the tweak of a curated feed or a precisely targeted ad. Following these constituents of data flows required the use and development of a constellation of methods, which can be collected under the digital methods framework (Rogers, 2013). That is, methods concerned with re-appropriating and re-engineering digital infrastructures, predominantly at the layer of applications and content, for studying the social through a growing availability of large-scale meta-data that can be found. Digital methods leverage the continuous advances in computational power and digital tools to move freely between micro and macro levels in data (Ruppert et al., 2013), and thereby redirects focus away from traditional methodological separations of the qualitative and quantitative.

Taken together, the methods – quantitative and qualitative – complement each other, and data – found and made – are combined as the studies comprised in the dissertation interchangeably address the research problem. They do so by focusing on different aspects of the systemic approach to communication as capability – be it means, resources, structuring conditions, freedom, choices, or achievements. The final section of this first part outlines how the dissertation contributes to and develop research agendas within media and communications research on datafication, digital infrastructures, and power.

**Contributions: bits of data, bits of power**

The dissertation makes three key contributions through the development of theories, methods, and empirical analyses that push the boundaries for existing fields within media and communication research, traverse macro, meso, and micro levels of inquiry, and engage with aspects of digital communication that are rarely studied in combination.

**Theoretical**

The theoretical framework developed in Part I of the dissertation on the one hand offers a capabilities perspective on digital communication, and thereby also
a way of understanding connections between communication, freedom, and the
good life. On the other, it also offers a development of the Capability Approach
insofar as digital communication goes that is captured in the generation of the
systemic approach in figure 2. Figure 2, in turn, also outlines the contours of a
research agenda that is concerned with combining infrastructure and user studies
in order to comprehensively account for the implications of the ways digital
communication is enabled, constrained, and negotiated in a datafied society.
Both at the micro level of what individuals are capable of – what they are and
what they do – in and through communication, and at the macro level of large
technological systems that these beings and doings are embedded in.
Furthermore, a number of theoretical frameworks that can be interrogated in
the individual articles of the dissertation combine theoretical perspectives in un-
usual ways that pave the way for new insights on how macro structures manifest
themselves in big data, harvested from micro and mundane everyday communica-
tions. The dissertation, for instance, combines the Capability Approach with
media ecology, political economy, and momentum theory in order to develop
the concept and the approach of digital communication systems as an alternative
to existing media systems analyses. It combines perspectives from infrastructure
studies, political economy, privacy research, and the emergent field of app stud-
ies in order to engage with the infrastructural mechanisms of the app economy.
And it combines feminist theory, sociology of emotions, and communication
theory in order to discuss (gendered) digital labour, inequality, and commercial
exploitation from a different vantage point than that of existing studies.

**Methodological**

The dissertation applies a range of methods that complement each other by un-
covering different aspects of digital communication and serving different pur-
poses when it comes to approaching the empirical objects of analyses. A com-
mon denominator, for most of the articles, is the fact that many of the empirical
objects are quite new and emergent in media and communication research –
think of, for instance, mobile permissions or Internet Exchange Points (IXPs) –
and as such, we have no established methods and frameworks for how to study
them. In effect, the dissertation comes up with new tools, methodological constellations and frameworks for exploring nascent and fast-moving phenomena in comprehensive ways. It thereby contributes by developing and demonstrating the efficacy of different research designs, comprising various methods and theoretical frameworks that combine otherwise disparate traditions and fields to approach empirical phenomena that are usually studied in isolation.

Seconding the notion that “[f]uture communication research needs all the data it can find and make, qualitative as well as quantitative” (Jensen, 2012, p. 445), the dissertation represents a call for more empirical and systematic approaches to studying internet regulation, more holistic approaches to studying the role of the internet in everyday life, more digital methods aimed at reverse engineering the practices of powerful digital market actors and dissolving mysteries surrounding datafication and big data through, in turn, more visual methods that make otherwise invisible infrastructures visible, thereby grounding further analyses of how they are controlled, by whom, and with what implications for individual users. It contributes by showing potential ways forward for answering these calls and identifying new and crucial research areas.

Empirical

Lastly, the dissertation makes a number of empirical contributions that spill from the findings of the different articles and document how capabilities can (sometimes not) be turned into valuable functionings and good – digital and datafied – lives. Rather than going into each article, as I did in the article previews earlier, I bulk these contributions into three main categories that are consistent with the theoretical and methodological contributions outlined above, as well as with the underlying principles of the capabilities perspective. Namely, empirical research that substantiate important theoretical and conceptual work on big data and its implications; holistic and macro-level studies of digital infrastructures – including surveillance infrastructures – and their control mechanisms; and, impact-oriented analyses of existing and emergent power structures around digital infrastructures and data, their origins, and assumed inevitability. Taken together, these contributions are fundamental for critically interrogating the ways
individuals’ freedom is conditioned by digital infrastructures built around the logics of surveillance capitalism.

First, the dissertation answers the recent calls for more “empirical research to underpin and flesh out critical data studies” (Kitchin & Lauriault, 2018, p. 18) through empirical and systematic studies of infrastructural tenets for datafication. As bits of power (Mansell, 2017b), bits of data give those who have it in accumulated and big measures quite some leverage on what the future of surveillance, commerce, social justice, and internet governance will look like. Although they are mass producers of small bits of data, users rarely hold the distributive power over these assets: “technological innovations that occur through time in the media and communications sphere are associated with power relations that privilege the interests of the owners of production, especially through the control they exert over information (Innis 1950; Babe 2015)” (Mansell, 2017b, p. 5). The empirical studies in the dissertation approaches this control over information by following the trails of bits of data, from the users to the stakeholders that ground so much of their business in the continued availability of constant streams of data. The empirical findings contribute to critical studies of big data by enhancing the transparency of data processes and power structures that otherwise thrive on their intransparency (Libert, 2015) and apparent inevitability (Couldry & Yu, 2018; Dencik & Cable, 2017; Draper & Turow, 2019). By producing empirical evidence on how digital infrastructures work, who controls them, and for what ends, as well as how they might be regulated so as to protect individuals, the articles uncover central mechanisms of big data and demystifies them. This particular approach to critically interrogating the tenets of big data is, in turn, aligned with and contributes to a different, yet related, research field, namely that of infrastructure studies.

Second, the dissertation joins and contributes to the existing, but few, empirical studies that approach digital infrastructural systems from material and coherent macro-perspectives. It also contributes to largely descriptive studies of surveillance infrastructures by offering theoretical frameworks for understanding and explaining digital infrastructures and data flows as parts of wider social, economic, and political structures. The dissertation’s insistence on the materiality of digital infrastructures – and thereby the ability to monitor and regulate them, even if they (deliberately so) seem intangible – is key to understanding these
contributions. Whereas the relationist perspective on infrastructure as a key concept allows for important analyses of people’s particular practices in and with infrastructures, the materialist approach is foundational for the findings comprised in the dissertation regarding how companies obtain structural power through infrastructures, and for what ends. In other words, infrastructures can and do mean different things to different people – they can be imagined in various ways, which also became evident to me during the ethnographic fieldwork – yet, that does not make them less real or consequential. Although they often work in the background, are taken for granted, and go unnoticed, infrastructures – as “the guts of the internet” (Winseck, 2019, p. 94) – are in the dissertation made visible, well beyond moments of breakdown and particular micro contexts (Star, 1999), and subjected to empirical analysis as they enable and constrain individuals’ digital communication and capabilities.

Third, the dissertation contributes with empirical studies of infrastructural power – as the capacity for controlling the material foundations of a digital communication system. It also approaches what we might best refer to as the political economy of free (stuff, services, lives...) (Helles, Lomborg & Lai, forthcoming) through critical engagement with one of the most fundamental business models in the digital market: services that are free of monetary charge, yet exchanged through ever-increasing amounts of data, thereby pushing the datafication of all aspects of life to new limits. That is, in the political economy of free, “[t]he assets of interest are the bits of power that mediate company-customer relationships and support the advertising industry in “guiding one of history’s most massive stealth efforts in social profiling” (Turow, 2011, p. 1 in Mansell, 2017b, p. 6). The findings comprised in the dissertation testify to how control over infrastructures for datafication along with the institutionalisation and naturalisation of the trade-off (Turow et al., 2015) – between convenience and price (free) in exchange for control over data – are evidently critical aspects in obtaining and retaining power in the digital ecosystem. As such, the dissertation produces knowledge on infrastructures as critical resources, but also, significantly, as communication systems, ecologies, and division of labour. And finally, following the impact agenda of the Capability Approach and political economy tradition, the findings are, as of this writing, being put to active use in governmental bodies that are scrambling to get on top of how to hold corporations accountable for their data
commodities as well as how to handle the use of commercial, digital services in educational contexts.

The following part II comprise the seven articles that were previewed in the beginning of part I, and further introduced throughout its discussions. Part III, in turn, features the conclusion to the dissertation. Now, happy reading!

References


https://doi.org/10.1017/CBO9780511841286


Poels, K., Ijsselsteijn, W. A., & de Kort, Y. (2015). World of Warcraft, the aftermath: How game elements transfer into perceptions, associations and (day)dreams in the everyday life of massively multiplayer online role-playing
game players. New Media & Society, 17(7), 1137–1153. https://doi.org/10.1177/1461444814521596


Tracing Communicative Patterns 06
Mapping Digital Communication Systems 07
Networks of Power 08
Infrastructures of Tracking 09
A Proxy for Privacy 10
Appscapes in Everyday Life 11
She’s the Communication Expert’ 12
Publication

Nordicom Review, 2019
DOI: https://doi.org/10.2478/nor-2019-0019

Authors

Signe Sophus Lai, Jesper Pagh & Fiona Huijie Zeng
Tracing Communicative Patterns

A comparative ethnography across platforms, media and contexts
Abstract

This article outlines a research design for a qualitative comparative study of communication across platforms, media and contexts – in China, the US, and Denmark. After addressing limitations in previous research on digital media in everyday life, we argue in favour of a comparative ethnography of communication that emphasizes the study of intermediality by taking a people-centered approach. The methodological design combines network sampling and maximum variation sampling with communication diaries and elicitation interviews. This design makes it possible to collect small and deep communicative trace data, to capture individuals’ unique linking of all the communication tools and channels available to them and, in turn, to identify the role of the internet as it interacts and intersects with other forms of communication.

Keywords

Communicative patterns, comparative ethnography, diary, elicitation interviews, sampling strategies.

Introduction

In July 2017, we set out to conduct qualitative fieldwork in China, the US, and Denmark, following a shared approach that allowed for comparative studies of communication practices across culturally distinct contexts. Here we present that particular methodological setup, which facilitates the mapping of complicated, ubiquitously embedded, and mundane communication as it occurs in various everyday contexts. In other words, the article focuses on the small, even deep data that enable the analysis of a person’s communication traces, and the motives and incentives that trigger them. We do this with particular attention to how communicative acts involving the internet, interact and intersect with the total range of communicative practices of any individual. Applying an understanding of internet uses as communicative instances, allows us to study connections and interdependencies between the myriad of ways in which people can communicate and the ways in which people understand, differentiate, distinguish and choose, more or less consciously, between those possibilities. The method
provides researchers with a way to place digital trace data in the context of a person’s total communicative repertoire by making it possible to discern how and why people move between digital and analogue communication. As such, the article offers a unique methodological design for comprehensively studying peoples’ everyday Internet use by capturing all communicative acts moving across different platforms, media, and contexts.

Researchers focusing on digital communication often define their research area by the level of online or offline integration (Garcia et al. 2009), or delimit their scope by focusing on specific platforms for (online) communication (such as social media: Miller et al. 2016), specific media (such as smartphones: Thorhauge and Lomborg 2016), or specific contexts (such as the home: Bakardjieva 2005). Although these limitation strategies define important areas for in-depth research, they also rule out the comprehensive study of connections and interdependencies between different platforms, media and contexts. Researchers studying social media cannot integrate this with knowledge about how and why people will move a conversation from a social media platform to a phone call. Researchers studying smartphones cannot integrate this with knowledge about how and why people use the same calendar software on their laptops as on their smartphones, but in an entirely different manner. Researchers studying specific contexts like the home cannot integrate this with knowledge about how and why people use their tablets differently when they commute as opposed to when they are at home.

Taken together, the different strategies for limiting research scopes leave a gap in communication studies when it comes to analyzing digital trace data in a cross-platform (software), cross-media (digital/analogue hardware) and cross-context (physical/social situation) communication. In order to address this gap, we propose a qualitative methodological framework that, through the capture of small and deep data, enables mapping and analyses of the communicative practices that make up a person’s daily communication. Rather than looking at a single platform, digital media in particular, or online and offline contexts in isolation, we “follow the people” (Marcus 1995) and focus on the proliferation of ways in which people can communicate. Essentially, by focusing on people, we are able to follow in their footsteps. This allows us to not only look at their digital communicative traces, but also to understand why they were left there in the first
place, and to analyze them as they relate to other forms of communication that a person engages in. As such, digital trace data are not studied in and of themselves, but are contextualized and understood as part of a communicative environment where people seamlessly switch between online and offline communication practices. In effect, the research design addresses the interwoven nature of people’s online and offline presence (Simon and Ess 2015), and targets intermediality (Jensen 2008) on the basis of the individual respondent and in close relation to their actual communication purposes, networks, and situations. It does so by departing from the question: how do we, empirically, study the situated enactment of communicative practices under a comparative framework?

Participants for the study were recruited via a combination of network sampling (Heckathorn and Cameron 2017) and maximum variation sampling (Bryman 2016). The approach combines recurrent interviews with 24-hour communication diaries in an iterative interview-diary-interview-design. Interviews and diaries intersect by anchoring findings from one in the other. The diaries are time and context bound rather than reliant on memory, while the subsequent interviews contextualize the diary data points and allow for the respondent to infer meaning from, make sense of, and co-interpret their communicative practices as they figure in their individual diaries. Unlike many other cross-national comparative studies that focus on exploring differences, our research departs from the notion that despite large cultural, political and economic differences between the three nations (and indeed within them), communicative actions are driven by the same fundamental social mechanisms – such as a desire to maintain strong social ties (Granovetter 1977). The shared methodological framework provides small and deep data that is comparable across contexts precisely because it focuses on those fundamental incentives in a way that produces a similar type of data. The comparisons, then, open up for an analysis of the interactions between different cultural, political and economic contexts and how those fundamental social mechanisms take shape.

The article is structured into three parts. First, a review of previous studies on the internet in everyday life. Second, a discussion of different sampling strategies and qualitative data collection methods and an argument for the particular combination of network and maximum variation sampling with communication diaries and elicitation interviews. Third, examples from our fieldworks that explain
how the methodological design produces data that can comprehensively account for people’s individual understandings of their communication practices across platforms, media and contexts.

**Studying the internet in everyday life**

The “ethnographic turn” in media studies heralded a shift in focus from the specific content of media to the everyday contexts in which media was used. The Social Uses of Television (Lull 1980) and Reading the Romance (Radway 1984) were among the studies that introduced this focus on contexts of use, which later re-emerged in studies of internet use in everyday contexts (Bakardjieva 2005). In the first generation of internet studies, however, a specific understanding of the internet as a placeless cyberspace (Slater 2002) resulted in studies targeting online contexts such as MMOs (Turkle 1995), MUDs (Markham 1998) and discussion forums (Baym 2000) rather than the empirical contexts of use.

By the end of the millennia, scholars began challenging the notion of the internet as a cyberspace that could be examined and understood as a context in and of itself. Through extensive fieldwork in Trinidad, Miller & Slater (2000) demonstrated that the internet rather consisted of numerous technologies, used by different people in quite different real-world contexts, which were both shaped by and shaping the internet. Trinidadians were using the internet in one of three contexts: by the stationary computer in the home, at work or in the internet café. With the development in mobile media, uses of the internet have diffused into a myriad of other daily life contexts. With uses of the internet entering into nearly all aspects of everyday life, communication researchers are, as we outlined earlier, resorting to various strategies in order to delimit the scope of their research objects. These strategies can be divided into four categories as summarised below.
First, scholars delimit their studies of the internet by focusing on specific contexts. In one strand of research, the domestication tradition, scholars are increasingly attentive to the home as a particular empirical context wherein the social norms and physical surroundings of specific households govern the appropriation of different media including the internet (Silverstone, Hirsch, and Morley 1994; Silverstone 2006). In line with this tradition, Maria Bakardjieva (2005) challenges the dichotomy between virtual and real lives. Looking at people’s interactions in online communities, she develops a typology of five modalities of virtual togetherness, all of which are closely intertwined with people’s offline realities. Second, the overall research scope can be delimited by focusing on a specific medium like mobile and smartphone use in everyday life (Ling and Lai 2016; Thorhauge and Lomborg 2016) or laptop computers (Anderson et al. 2009). Third, other researchers focus on specific platforms of communication that originated online like MySpace (Boyd 2006), SecondLife, World of Warcraft (Poels, Ijsselsteijn, and de Kort 2015), and Twitter or Facebook (Marwick and boyd 2014; Miller et al. 2016). Fourth, some scholars take the individual as the unit of analysis and follow his or her communication across platforms, media and contexts. Striving for a holistic account of individuals’ communication, uses of the
internet are understood only as they relate to everyday life and to the entire assemblage of communicative practices that amounts to the individuals’ communicative pattern. Applying this limitation strategy, Hasebrink and Domeyer (2012) map individuals’ media repertoires, Beneito-Montagut’s user-centred ethnography focuses on interpersonal communication (2011) and Helles (2011) connects lifestyle types with different configurations of media and internet use. It is within this fourth strand of research strategy that we place our study. In effect, we examine communicative patterns, with or without the internet, through mundane actions across any platform, medium and everyday context. This requires an intense and comprehensive relation to the individual participants, which in turn constraints the number of research subjects. Hence, we get to know a lot (deeper data) about a few (smaller samples).

To sum up, this study is placed in the tradition of everyday life Internet studies and strive to get at people’s communicative practices in a comprehensive manner, while accounting for an increased complexity, mobility and fluidity in those practices. In the following part, we discuss sampling strategies and qualitative data collection methods as they relate to the research design proposed in this article.

**Sampling maximum-variation networks**

Although the function of any sampling strategy is to lend validity to the study at hand (Morse and Niehaus 2016, 66), researchers often neglect to argue thoroughly in publications for their choice of sampling strategy, if at all. Rather, sampling is shallowly treated as a technical research state, without explaining the ways in which the qualities of the sampling strategy coincide with the aims of the study.

With the purpose of gaining in-depth knowledge on communicative practices and their relationships with culture, regulation, market, and technological infrastructure across comparative research contexts, we rest our argument on Marcus’ study (1995), which note the pitfalls in attempting to explain macro structures through a single case micro-perspective. To avoid this, Marcus calls for an ethnography that is multi-sited, so that each site differentiates enough from the
other to be able to give a unique perspective on the same macro structures (be they for instance cultural, political or economic). In Marcus’ framework multisitedness is the result of following things, metaphors, stories, biographies, conflicts or, as is the case here, people across diverse sites. We propose doing this by sampling respondents through a combination of network sampling with maximum variation sampling. This section outlines the values in the combining these specific strategies.

**Network sampling**

Network sampling is a purposive, non-probability sampling strategy. As such, network sampling targets specific groups or types of people, and the sample will not be reflective of a wider population or enable statistical generalizations. Additionally, recent ethnographic research (Cook, Laidlaw, and Mair 2009) challenges traditional notions of the ethnographic field site as a geographically bounded location and suggests network sampling as way for researchers to construct the field site by tracing the nodes of a social network (Burrell 2009).

In order to bridge different social networks, researchers can network vertically through strong ties only or horizontally through both weak and strong ties (Granovetter 1977). The latter essentially entails “casting the sampling and recruitment net wide rather than deep” (Geddes, Parker, and Scott 2017, 2), and as such, the former is often put forth as the ideal that correlates with an effort to make in-depth descriptions of a given phenomenon. Networking horizontally, however, can prevent sampling stasis and also enable heterogeneity not only between networks but also, to some extent, within networks. In our study we used both networking strategies dependent on the specific network and sampling situation.

Following Noy’s concept of sampling knowledge (2008, 3), the method of sampling networks can generate important social knowledge of an interactional character in and of itself. The sampling then provides additional knowledge to the one produced through interviews and diaries – for instance in the form of different perspectives on the same communicative relationship between two individuals in a network.
**Maximum variation sampling**

One way of constructing the spatial components of a networked field site is by applying a maximum variation principle to the sampling of the network(s). This method helps to locate cases that are heterogeneous on selected characteristics which are thought to be important influences to the object of study (Bryman 2016; Morse and Niehaus 2016). As such, the term maximum variation refers to the variation between cases, rather than to the variation of the full population of the field site(s). While being aware of the limitations that a small n-study puts on the ability to achieve maximum variation on a high number of variables, we sought to have the maximum achievable variance on seven variables when sampling our participants, namely 1) age, 2) gender, 3) place of residence and household, 4) ethnicity, 5) educational level, 6) religious beliefs, and 7) relationship and parental status in order to obtain the broadest possible representation on those factors. Applying maximum variation as a sampling strategy in a qualitative study can, firstly, provide deep descriptions of the individual cases and show the diversity between them, and secondly, highlight patterns that traverse variations and prove to be important precisely because they are unaffected by diversity (Patton 2015). Figure 2 illustrates the different networks in our studies.
FIGURE 2. Sample networks: The structures of the various social networks across the three research fields.
From going places to tracing relations

Although Marcus, in his argument for a multi-sited ethnography, emphasized place as a physical location, our adaptation of his argument to the analysis of field sites as networks (Burrell 2009) resituates place as less dependent on physical properties and more on relational properties. Maximum variation can therefore be seen as a tool to combine the deep description of usage patterns with the, often shallow, superimposition of macro-influences at the analytical level. That is not to say that maximum variation in itself meets the requirements of the kind of experimental ethnography that Marcus called for. It is, rather, to propose a sampling strategy that addresses those same issues: that a disorganised system (Lash and Urry 1987) cannot be adequately represented through homogenous cases.

The combination of maximum variation and network sampling in our study generates what can be described as constructive interferences, where differences between the approaches can simultaneously provide a stronger understanding of communicative practices both in isolation and in relation to structural influences. While limiting the sampling to people connected through a social network potentially restricts maximum variation due to homogeneity within networks, we obtained heterogeneity across networks by finding inherently different starting nodes for the networks. In effect, each network acts as a distinct place, wherein people’s relationships, and the ways in which their communication influences these relationships (and vice versa), can be analyzed. In the next part, we discuss the combination of communication diaries and elicitation interviews as an approach to follow people in their communicative networks in order to collect the digital trace data left behind from their communicative moves across platforms, media and contexts.

Diaries: structured and un-structured approaches

Diary studies have been an established approach in social science research since the 1920s, with time-use diaries in the 1960s and onwards as the most widely recognized form (Couldry, Livingstone, and Markham 2007, 45). Critics have identified many limitations in time-use diaries, particularly with regards to low
response rates and participants’ abilities to reliably estimate time spent on mundane activities. However, advantages of the method are evident in that diary data is collected close to the time of the registered event (Ibid.) and that self-produced diaries generate data on the everyday context of mundane practices that would be difficult to obtain otherwise (Zimmerman and Wieder 1977). Incorporating diaries in our design supports the research aim by providing a broad representation of participants’ communicative practices while minimizing our interference with their daily lives.

Research diaries can be devised on a continuum ranging from fully structured to entirely unstructured, which is also a distinction between ‘objective’ and ‘subjective’ approaches. Whereas time-use diaries belong to the more structured and ‘objective’ approach, ‘subjective’ unstructured diary studies emphasize not confining the reflection of the participants and making the researcher as unobtrusive as possible in the data collection process. For instance, in a highly structured diary study, Hargittai and Carr prompted their participants recurrently at random times asking them to: “Please reply with your location, current activity, people you are with (number, your relationship to them, gender) and any media you are using now” (2009, 20). In contrast, Markham and Couldry, using unstructured weekly diary entries, explain how: “The priority was to access people’s processes of reflecting on media use rather than cataloguing consumption habits” (2007, 688).

Couldry and colleagues (2007) describe how research tends to favour either structured or unstructured approaches, with unstructured diaries being the common component in ethnographic studies. The diary design we argue for here, however, is innovative as it combines the two approaches in one: an objective approach with subjective measures. That is, while our diary method is quite structured in what to report and when, it is fully unstructured when it comes to how to report. By this coupling, the type of data collected is manifold although it reports on media use instances each time, and the diary is more organically incorporated into the existing communication practices of the participants. By the end of the introductory interview, participants were given oral and written instructions on what to report in the communication diary, allowing for questions to be answered in person. We asked them to register and report answers to the following questions whenever they used media (analogue or digital),
throughout the day and as temporally close as possible to the actual uses, for one or two full days.

1. When was it and for how long (if this is not implied)?
2. Where were you (if this is not implied) and who were you with (if any)?
3. What were you using the media for (for instance texting, social network sites, e-mail, applications, TV or radio programs, phone calls)?
4. Who were you communicating with (if any)?

These rather structured instructions enabled the data collection to stay close to the actual instance of media use as well as the potential trace data left in its wake, giving greater accuracy to the data reported. Therefore, we were able to infer communicative purposes of the media uses as they were happening, and to ask clarifying questions then and there. It also allowed for the small, mundane acts of communication – such as the many times a person notices messages on the lock-screen of a smartphone or the ritual text messages containing nothing but phatic conversations – to enter into the foreground. To further structure what to report, a definition of what we meant by media, which had been open for interpretation during the introductory interview, was given to the participants, so that all participants shared a similar idea of what media use entailed when reporting back to us.

For each report instance, the participants were then free to decide on the report platform and medium (e.g., social media, email, phone calls, tablets and paper records) and report type (e.g., texts, visuals, as well as audio and video messages). Additionally, they were introduced to automated data collection apps and software for different digital devices, which they could opt in and out of. Hereby, the diary reporting was anchored in their existing media habits and contingent on their regular media uses, which minimized the effect of the reporting on the routines of the participants (Carter and Mankoff 2005). Hence, the specific amalgamation of structured registration with unstructured reporting allows for the researcher to stay close but also out of the way during the data collection. Moreover, the ways in which participants reported to us in terms of report types, platforms and media provided valuable insights into their general communicative choices and patterns.
We recognize, however, that the diary method is not able to catch every instance of communication, which was also not the purpose. The data collection depends on the participants’ ability to remember to register and report whenever they use a medium. Even though the diary was structured to allow for the reporting of trivial and habitual media use, it was inevitable that some uses, to some of the participants, were so engrained in habit that they did not register the acts themselves. Or in other cases, participants became too self-conscious and in effect changed their media habits. We argue, however, that the design of the diary registration and reporting process minimized our intrusion into participants’ routines while maximizing the participant feedback. Also, the combination of diaries and interviews allowed us to ask about media uses gone unregistered as well as about face to face communications throughout the diary day. Hence, we employ the concept of communication diaries since the specific fusion of diaries and interviews capture both mediated and non-mediated communication. The diaries, then, represent a different type of trace data than the one found in mining uncontextualized and non-consented Facebook comments, retweets or YouTube likes. The situated enactment of the participants’ communicative practices amounts to different traces that they themselves decide to pass on to us (or not to). As such, rather than being left out of the research process, the participants own, reflect upon, and co-interpret the traces of their communicative actions, thereby adding information on contexts of use, motives and purposes. In the following, we discuss advantages and challenges in interchangeably augmenting diary data with interviews.

**Interviews: introductory and elicitation**

In the interview-diary-interview structure of our study, the purpose of the introductory interview is to get to know the participants, their everyday life and media use in general, whereas the second interview is an elicitation interview taking prompts from both the diary and the introductory interview.

The methodological combination of diaries and interviews in social science can be dated back to the 1970s, when Zimmerman and Wieder (1977) propose a diary-interview method in order to document participants’ activities in a more
feasible manner than through participation observation. Similarly, in recent media repertoire research, Hasebrink and Domeyer (2012) explore the diversity and differences between participants’ media repertoires by documenting their media use through diaries and in follow-up interviews generating detailed and descriptive data of participants’ media use as well as facilitating further interpretations about those uses. In other words, the diary-interview approach enables researchers to employ the diary data in two ways. First, as data in and of itself, and second, as prompts for the elicitation interview questions. The diaries in our study—the ways in which participants choose to report, the platforms and media they use, their frequency of reporting, the people they communicate with and the ways in which they reach out to us—provide important data on the communicative practices of those respondents. The follow-up interviews allow for participant reflections and a line of questioning that can be rooted in the actual communications of each participant.

In the introductory interview, we get to know the participant (asking them about their daily routines, their close network, things that are important to them and the media they use). We ask for ‘tours’ of their smartphones, tablets and computers, posing questions concerning notifications and sound settings, particular apps as well as the general organizing principles of their devices (app locations, folders, pinned apps).

Elicitation interviews are commonly described as interviews where another data source is used to elicit or spark a conversation within a mixed-methods framework (Gundelach, Nielsen, and Frederiksen 2014). Examples of different constellations of interviews and elicitation data include observational studies (Beneito-Montagut 2011), screenshots and GPS data points (Thorhauge and Lomborg 2016), open-ended reflexive diaries (Markham and Couldry 2007) and log data (Anderson et al. 2009).

Although we had a preliminary interview guide for the elicitation interview, it was comprehensively rewritten to combine discussion topics with concrete prompts from the introductory interview as well as the diary. The tailored line of questioning made it easier for participants to relate to each topic and encouraged them to fill out the gaps in the diaries and to co-interpret their data. As such, the diary data allow us to ask more abstract questions, as the questions are tied to
the participants’ own concrete acts of communication. It also brought attention to otherwise ubiquitous and indiscernible communication instances, allowing participants to elaborate on the circumstances and contexts of those specific instances. Finally, by asking questions concerning the blank spaces and periods of non-media use in the diaries, we targeted the communication that was not present in the diary entries, namely face to face interaction. As part of a comprehensive study of the role of the Internet in everyday communication, face-to-face interactions play a significant role in the multitude of ways people can communicate.

Having outlined the sampling strategies and empirical methods, the last part of the article elaborates on our design by providing data examples that show how the combination of communication diaries and interviews enables us to follow people and analyze their communicative patterns.

**Following people**

We have argued that, unlike research that delimits its scope by specific platforms, media, or contexts, our research design enables us to follow communication flows wherever they go: when participants suddenly switch between the affordances (Hutchby 2001) of different platforms and media, when they communicate in several contexts at once or when they find themselves in private contexts where researchers cannot usually follow them. We now present data examples from three participants, who are maximally different to each other as they move across platforms, media and contexts. The examples show how we empirically approach the situated enactment of the participants’ communication patterns insofar as they can be traced in the diaries as well as in the subsequent conversations with them. Ena in Denmark, a 46-year-old first generation immigrant from eastern europe, who works as a part time nursery teacher and lives with her husband and children; John from the US, a 56-year-old painter of catholic icons who lives alone in Chicago and works part time as a church janitor; and Chun from China, a 19-year-old college student that lives in Beijing and majors in Japanese studies.
**Cross-platform communication**

In any individual communication diary in our sample, several platforms are represented throughout the day, and we asked participants to reflect on instances of digital cross-platform communication. In Denmark, Ena, abruptly shifted from writing with her friend in Eastern Europe on Facebook Messenger, to calling her over FaceTime. She described how she had called her friend because the conversation turned to a serious topic that she did not feel like talking about in text. Being able to hear her friend’s voice, she managed to console her and felt more comfortable that she could say the right things to her. In the US, John used different platforms to sustain his business of selling religious paintings. For instance, he would switch interchangeably throughout the day between replying to inquiries and checking incoming orders on his email and on his personal Facebook profile. He used to only sell his paintings through exhibitions and word of mouth, but friends encouraged him to employ the digital platforms as a way to enhance business by reaching a wider range of possible buyers. In China, Chun, explained how she had two online personalities on two different social media platforms. On the first one, WeChat, everyone knew her real identity and she only shared positive content within this network. On the second, Fanfou, she used an anonymized identity and to this network she would pour out her negative thoughts without these postings being related to her real identity.

In the diaries, we could observe participants switching between different platforms, and through the reflections and co-interpretations of the participants in the subsequent elicitation interview, we were able to infer meaning from the stand-alone diary entries. It was only through these entries that it became possible to capture the mundane, minute, and at times complex moves between platforms – as well as moves between *media.*

**Cross-media communication**

The communication diaries contained the different media that the participants own and switch between or use in parallel. In China, Chun reported on a situation where she was watching and discussing a Japanese TV series on her laptop while having lunch with her roommate. In a case of parallel media use, Chun
took a picture of the laptop screen with her phone and posted this to her WeChat network, commenting that she was missing out on the series’ details because she was distracted by the lunch situation. In the US, one of John’s tasks for the church involved assisting a 73 year-old deacon with his email exchanges. The deacon had a stationary computer that he did not know how to operate, so John would log into the deacon’s system, check his emails and then print them for the deacon to read. Subsequently, the deacon would write hand-written replies that John could then type into the email program, and send out on behalf of the deacon. As such, John functioned as an intermediary between the deacon and his network as well as between the deacon and the computer. In Denmark, Ena, doing the diary on the day of new year’s eve, spent most of her time in the kitchen. She used her phone to look up recipes for the evening’s dinner only to switch to her tablet once she found a suitable recipe. The tablet, she explained, had a bigger screen and was more easily incorporated into the cooking situation in the kitchen.

Like for cross-platform communication, the concrete acts of switching between different analogue or digital media served as a point of entry into the individual’s unique communication pattern. Being able to ask concrete rather than abstract questions to those media uses made it possible for participants to express elaborate systems of thought and behavior grounding these exchanges. Using different media in sequence or in parallel were often not based on conscious choices. However, when asked, the participants were able to provide reasons that rested on the affordances of the media they were using and of the contexts they were in.

**Cross-context communication**

The in-situ self-reporting format of our communication diaries allowed for communication instances taking place in intimate contexts to be included in the study. In the US, John used his text messages from his phone to coordinate a meeting with his brother regarding an intimate family issue. Their mother had recently passed away and John and his brother were meeting to discuss how to take care of their 90-year-old father. Later that day, after hours of no diary updates, John explained how the two of them had a rule never to use media when they were together. In Denmark, Ena who suffered from post traumatic stress...
disorder after fleeing her home country during war time many years ago, explained how she took her tablet to bed each night and put on white noise videos from YouTube in order to fall asleep. In China, Chun went on a romantic date with another participant from her network, Yu, and they both used their smartphones for taking pictures of the food and store memories from the date.

What these contexts have in common is that they are highly intimate and the actual presence of another person, let alone a researcher, would have impinged on the situation and changed it altogether. The diaries, being our conceptualization of “being there”, thus enable contexts like these to be part of the elicitation interview and the study as a whole.

These three participants exemplify how the small and deep data collected encompass the cross-platform, cross-media and cross-context communications of an individual. Taken together, the communication instances map the full spectrum of communicative possibilities relating to the individual participant, and the interviews contextualize these practices by adding aspects of everyday life, personal values, attitudes and aspirations as well as participant interpretations. The data examples illustrate how the coherent methodological design was implemented in the three very distinctive national fields, and laid the ground for future comparative analysis. Upcoming work will include a thorough analysis of cross-cultural comparisons. We do, however, also acknowledge that approaching the field with the aim of capturing a person’s entire communication repertoire carry with it certain limitations.

**Limitations of following people**

We identify at least four limitations of our research design. First, the methodological set-up is demanding in time and effort on both participants and researchers because of the recurrent and extensive contact throughout the research period. Second, because of the intensive research participation, we end up with a comprehensive snapshot of the participants’ communicative pattern that is susceptible to their current life circumstances and that does not enable us to make any longitudinal inferences. Third, we lose depth when it comes to any one individual platform, media or context while trying to follow connections across them. Fourth, as opposed to observational studies where the researcher is
entirely in charge of the data collection, our data quality varies in that it depends heavily on the ability of our participants to report their communicative instances.

**Concluding remarks: comparing communications**

This study was designed to enable comparisons of communicative patterns in each field and across national contexts. *Within each field*, the coupling of network and maximum variation sampling produced a number of inherently different networks of inherently different people whose communicative patterns can be comparatively analyzed relative to their everyday lives. By applying the same combination of sampling strategies *across the three fields*, the diverse individual communicative patterns can also be comparatively analyzed in relation to the cultural traditions, regulatory regimes, market structures and technological infrastructure pertaining to each field.

The article presented an answer to the methodological question of how to, empirically, study the situated enactment of communicative practices under a comparative framework. In this context, also the question of how to not delimit the research area to specific platforms, media or contexts, was addressed in the introduction of the novel approach of following participants’ cross-platform, cross-media and cross-context practices through their self-reports as well as their subsequent elaborations on these data traces. Any methodological selection is simultaneously a deselection, which the above-mentioned limitations are also a testament to. We do not argue that this approach is a way to capture *everything*, rather it is an attempt to study communication comprehensively through the collection of small and deep data. The research design makes it possible to capture intermediality of the basis of an individual’s unique linking of all the communication tools and channels available to him or her and, in turn, a way to identify the role of digital communication tools as they interact and intersect with other forms of onlife communication. As such, the methods we present address how to study the internet in its increasingly ubiquitous presence in everyday life by mapping communication – online, offline or both at once – as it happens, wherever it happens. The approach is a novel iteration on how digital media use and digital trace data can be mapped and analyzed as part of a holistic
view of individual communicative practices. It has been suggested time and again that online and offline experiences cannot be explained separately. We take this argument one step further and suggest that future research into digital communication needs to take the interwovenness seriously, not just between online and offline contexts, but between communicative practices in general, as they are performed across platforms, media and contexts.

Acknowledgements

This study is part of the research project *The Peoples’ Internet*, which is funded by The Carlsberg Foundation.

References


Markham, Annette N. 1998. Life Online: Researching Real Experience in


Publication

Media, Culture & Society, 2019
DOI: https://doi.org/10.1177/0163443719876533

Authors

Sofie Flensburg & Signe Sophus Lai
Mapping Digital Communication Systems

Infrastructures, markets and policies as regulatory forces
Abstract

This article presents a framework for mapping digital communication systems and thereby analyzing how and why structural conditions differ across national contexts. Following the ‘turn to infrastructure’ in Internet studies, we conceptualize communication systems as made of infrastructural, market, and policy structures that enable and constrain mediated communication in a given society. As opposed to media system analyses that typically focus on legacy media institutions, we take individual Internet users as our theoretical point of departure and ask how their communicative capabilities are regulated. In order to exemplify the application of the framework, we describe the methodological steps in a mapping of the components of the Danish communication system. In conclusion, we discuss the overall findings that the method uncovers and its implications for future comparative research.

Keywords

communication system, infrastructure, media ecology, media systems, political economy, regulation

Introduction

On New Year’s eve 2016, a massive TV signal blackout sent shock waves through the Danish population as they were waiting for the broadcast of the Queen’s annual New Year’s speech. Unsurprisingly, the fact that nearly 25% of the Danish population suddenly found themselves gathered around darkened screens caused a national uproar against the biggest Danish infrastructure provider, TDC, and its network operator YouSee. Those with the capabilities to do so left their TVs and turned instead to their broadband connections in order to access the speech via the national public service broadcaster DR’s streaming service. The DR app, in turn, experienced the highest amount of traffic ever measured and suffered an overload.

The New Year’s blackout illustrates how more or less mundane communicative activities are enabled and constrained by devices, access networks and backbone
networks that influence the use of different types of platforms, and ultimately
the access to communication content. It also shows how the Internet as a com-
mon communication infrastructure increasingly substitutes previous distribution
systems by serving the same communication purposes — be it through
Chromecasts and Smart TV’s instead of satellites and cable TVs or by Snapchat
and iMessage instead of the SMS.

The mundane choices between streaming and flow TV or between internet-distrib-
uted messages and traditional text messages seem inconsequential and minor
to the individual, yet the structural implications of these actions are significant.
While communication activities involving legacy media depend upon mainly na-
tional infrastructures, market actors, and policies, digital communication in-
volves global infrastructures and markets that call upon international govern-
ance. While the former is neatly distinguishable into legacy market sectors and
legislative categories, the latter traverse existing sectoral and institutional catego-
ries and policy fields. Taken together, this poses a challenge to established busi-
ness models, media and communication policies and scholarly media systems
analyses.

Consequently, we argue in favor of a reconceptualization of what we mean by
‘system’ — from media systems to digital communication systems. This conceptualiza-
tion rests on a functionalistic and materialist approach that identifies the com-
ponents of a system on the basis of how they support particular activities (such
as digital communication). The broader communication system concept consid-
ers monumental systematic changes, for instance the prominence of tracking
technologies leading to the rise of the data economy, and encompasses them as
important components insofar as they have repercussions for individuals’ com-
munication.

On the basis of this conceptualization, the article presents a methodological
framework for mapping digital communication systems that enables further
analyses of the structural forces that regulate internet-based communication.
The framework maps the components of a digital communication system by
identifying four communication levels (devices and access networks, backbone networks,
platforms, and types of content) that are mapped across three structuring dimen-
sions (infrastructure, market and policy). Applications of the framework ground
descriptions and discussions of digital communication systems and qualify comparative analysis of regulation. In other words, what do we mean when we talk about communication systems and how might we compare the system components? In turn, it provides a much-needed base for understanding communicative patterns in different societal settings and contextualizing for instance ethnographic comparisons across contexts (Miller et al., 2016) as well as quantitative country indexes (The Economist, 2016).

Communication system analysis & the infrastructural turn

The article builds on and extends research in three different domains, a) media systems analysis usually comparing institutional models and interplays between media and politics across national contexts, b) media ecology with a specific emphasis on how material communication infrastructures condition social environments, and c) political economy studying the distribution of goods, and commercial as well as political power structures in society.

Analyses of media systems rest on the premise that the institutional organization and use of communication media vary across national settings due to different structural and historical conditions. Systemic analyses of media in different contexts then aim at identifying the structural forces that frame specific media organizations (on a micro level), sectors (on a meso level), and entire media systems (on a macro level) (McQuail, 1992). The most common type of media system analysis relies on the framework developed by Hallin and Mancini (2004), and studies differences between news media systems and the conditions for practicing journalism in different parts of the world Terzis, 2007; BNrüggemann et al., 2014; Psychogiopoulou, 2014). However, as argued by Norris 2009 – and acknowledged by Hallin and Mancini (2012, 2017) – the framework does not capture the complexity of digital systems and the all-encompassing influence of the internet on practices inside and outside the sphere of legacy media institutions:

*if the larger theoretical point concerns national variations in access to different types of mass communications, then the 800-pound gorilla in the room, concerns the role of new information and telecommunication technologies,*
which are not featured anywhere as part of the classification’ (Norris, 2009: 332).

Also, as the framework was developed with the press in mind, it does not suffice in a digital reality wherein everyone is potentially a journalist and each their own small news outlet (Hardy, 2012). It fails to account for the tremendous influence that market actors such as Alphabet (Google) and Facebook have (also on news media) and falls short of defining what is and what is not covered by the classification today. This underlines the need for a more comprehensive framework that includes communication types outside the sphere of news as the boundaries set up by the initial one is, to say the least, blurry.

Hence, we suggest a macro perspective, which defines digital communication systems as ecologies that enable and constrain digital communication. Applying a media ecology perspective implies uncovering the ‘often implicit and informal’ infrastructural conditions that ‘specify what we can do and what we cannot’ (Postman 1970 in Scolari, 2012: 205). Rather than limiting the research scope to certain types of content (e.g. ‘journalism’), communication forms (e.g. ‘political communication’), or institutional categorizations (e.g. ‘news media’), we map the various types of technologies that enable digital communication. In line with governance research emphasizing ‘the currently neglected role of technology in media governance constellation beyond technological determinism’ (Katzenbach, 2012: 119), we see infrastructures as important regulatory forces that both condition and are conditioned by economic and political structures. This theoretical reframing of systems analysis follows the current ‘turn to infrastructure’ (Musiani et al., 2016; Parks and Starosielski, 2015; Plantin and Punathambekar, 2019) in internet studies by ‘turning away from the symbolic and investigating the structural—this is the Internet not as “what people say with it” but as “how it works”’ (Sandvig, 2013).

In this perspective, infrastructures are the material basis for communication systems as they enable and constrain mediated communication and make up the underlying resources that communication markets and policies revolve around. In other words, if a population does not have access to the internet, there is no market for web products or any need for internet policies. This is not to say, however, that infrastructures have a determining role for the emergence of social structures, but rather that the material foundation for digital communication is
an important starting point for identifying and understanding the political economy of digital communication systems. The digital market actors should thus be identified and understood in the light of the basic communication resources that they control the access to – but the development, organization, and distribution of these resources should also be seen in the light of existing and emerging power structures. We thus follow Mansell’s (2004) call for a revitalization of the political economy perspective in digital media studies. This approach to market analysis is broader than case studies of ownership, market dominance, or monopolization, and refer to how power structures are institutionalized and naturalized (Chang 2001). Put differently, when we map the actors that own and administer basic digital infrastructures, our ultimate goal is to be able to discuss how power and control of internet-based communication is institutionalized.

Taken together, our theoretical approach to communication system analysis follows Lessig’s (2006) call for ‘(…) a more general understanding of how regulation works—one that focuses on more than the single influence of any one force (…)’ (p. 121). When we talk about ‘regulation’, we refer to the structuring forces that frame common communication conditions in a specific societal context and contribute to the shaping of individuals’ communicative capabilities.

**Conceptualizing digital communication systems – the individual perspective**

The communication system concept calls for a change of perspective that acknowledges the limitations of studying specific institutions, market sectors, and policy programs (Bar and Sandvig, 2008; Braman, 2004). We argue that digital communication systems should therefore be approached from the analytical perspective of individuals, as the application of the individual perspective entails cutting across historically defined sectors and policy schemes and instead identifying the infrastructures, economic conditions, and policies that frame digital communication.

This does not mean (necessarily) studying actual individuals but rather mapping the common structural conditions for internet-based communication from the perspective of individuals in a specific national context. This, in turn, can serve
as a backdrop for understanding how and why communication patterns differ between specific groups or individuals, as they are influenced by various personal attributes (media literacy, resources, creativity, desires, etc.) and cultural norms and traditions. In line with Sen’s capability approach (1983), we acknowledge human diversity and the fact that people act differently within the same context, but at the same time underline that external and environmental conditions create a common ground.

**Four levels of digital communication**

Mapping the conditions that frame digital communication capabilities in a society, we simply start out by asking: what possibilities do I have when I pick up my smartphone (or any other digital communication device) and what are the structural consequences of my actions? We understand digital communication as dependent on a chain of events with infrastructural, economic and political implications. That is, when communicating digitally, an individual activates:

1. an internet-enabled device and an access network that give the end-user access to the global internet and thereby to:
2. the backbone network that connects local access networks and thereby enable the use of:
3. platforms for digital communication purposes that carry and aggregate:
4. digital communication content of any kind from individuals’ interpersonal communication to private or public broadcasting and beyond.

Similar to other layering models – e.g. Bratton’s *stack* model (2016) – the four levels of digital communication are listed hierarchically as the existence of devices as well as access networks is a fundamental condition enabling (or constraining) the following dimensions; e.g. if my optical fiber connection is lost, my subscription to broadband is useless and so is my access to platforms such as Netflix and thereby to whatever content is there. That is, unless I can connect to an alternative infrastructure such as mobile broadband, accessed through my phone subscription, and so on.

The four levels are used for mapping the infrastructural, market and policy structures that make up a digital communication system. As such, the analytical
strategy is to start from the device and access network level and from the infrastructural dimension, which, taken together, translates into the upper left corner in Table 1.

<table>
<thead>
<tr>
<th>INFRASTRUCTURE</th>
<th>MARKET</th>
<th>POLICY</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEVICE &amp; ACCESS NETWORK</td>
<td>What are the existing devices and access networks for digital communication?</td>
<td>Who owns and controls devices and access networks?</td>
</tr>
<tr>
<td>BACKBONE NETWORK</td>
<td>What are the existing backbone networks for digital communication?</td>
<td>Who owns and controls backbone networks?</td>
</tr>
<tr>
<td>PLATFORMS</td>
<td>What are the existing platforms for digital communication?</td>
<td>Who owns and controls digital communication platforms?</td>
</tr>
<tr>
<td>CONTENT</td>
<td>What are existing types of digital communication content</td>
<td>Who owns and controls digital communication content?</td>
</tr>
</tbody>
</table>

Table 1: Framework for communication system analysis

Table 1 serves as a guiding framework for identifying the interdependencies between devices and access networks, backbone networks, platforms and content on the one hand and infrastructures, market and policy structures on the other. We argue that answering each question in the matrix based on the analysis of a particular research context leads to a mapping of the communication system.

**Empirical material and analytical strategies**

In order to exemplify the application of framework, we map the Danish communication system. Denmark then serves as a case for exploring the structural components of a concrete digital communication system and for identifying relevant variables that could – and should – be studied across different contexts. We focus on Denmark for a number of reasons: first, Denmark can be described as an ‘extreme case’ (Yin, 2009), since all types of communication infrastructures
are highly developed and digital communication services are extensively used, making Denmark a clear example of a digitalized communication system. Secondly, Denmark (as well as other Northern European countries) is interesting as the emergence of digital markets, global governance, etc. challenge the traditional welfare state policies. Thirdly, being native Danes, we are familiar with the research context, which – especially in small language states – is crucial to the mapping process.

Applying the framework requires a broad collection of information from different fields. In the case study below, we consulted text books and journal articles, investigated databases, and conducted source critical analysis of official reports and evaluations, legislative acts and court rulings, and news articles.

The following sections describe how we map first the infrastructural implications of digital communication (table 2), secondly the market actors that control (and profit from) these infrastructures (table 3), and finally, the various policies that regulate digital communication (table 4).

The Danish case

Mapping infrastructures

Table 2 illustrates the key elements in a digital communication system as exemplified in the mapping of the Danish communication infrastructures. The mapping is based on scholarly research on internet technology – especially network components and layers – and market reports, statistics, white papers etc. describing the prominence of different material infrastructures.
Mapping the Danish case, all types of existing internet-enabled devices (from stationary computers to smartwatches and beyond) are present – most prominently PCs (91% of the population) and smartphones (88%) (Tassy, 2018). We identify four types of access networks that enable the digital communication of individuals: DSL broadband, cable TV broadband, fiber broadband and mobile broadband. Broadband connections are dependent on the material infrastructures insofar as one cannot for instance access a DSL broadband connection without having access to cobber wire or one can only access mobile broadband such as 4G if electromagnetic radio frequencies have been allocated for this. The presence (now and historically) of various types of highly developed and widespread access networks, means that 97% of households connect to the internet and eight out of ten use their mobile phone for internet purposes (Tassy, 2018). In turn, people are cancelling their landline phones and traditional TV subscriptions, which puts increasing pressure on the demand for broadband. The high degree of access to different infrastructures in Denmark should be seen in the light of the country’s geographic characteristics: By all comparisons, Denmark is a relatively small, flat and densely populated country suitable for infrastructure projects. Also, historical conditions (e.g. the extensive electricity system and early telecommunications network) have eased the implementation of new communication technologies. This is further underlined by global datacenters that are being built in Denmark, exploiting the country’s geographic position, stable energy supply, and well-functioning IT infrastructure.
As the existing access networks have been gradually refunctioned for internet purposes, the underlying **backbone networks** are increasingly important to the Danish communication system. Not only do the central fiber cables connect the national access networks and thereby Danish individuals, they also create the link to the global internet and thereby to an indefinite number of IP addresses and domains. The backbone network components in Denmark include first the fiber cable bundles and routers that connect the local, regional, and global networks, the internet exchange points (IXPs) where network providers interconnect, servers and data centers where content is physically stored, and content delivery networks (CDNs) that create local caches to store content temporarily and close to the end user.

The components of the **platform** dimension – operating systems, apps, and websites – are characterized by enabling any type of mediated communication over the internet. In Denmark, the decline in TV subscriptions goes hand in hand with a decline in the viewing of flow TV (more than 25% since 2011) and the use of VOD services is massively increasing (48% use a streaming service each week) (DR Media Research, 2018). Similarly, the extensive use of smartphones and mobile apps is moving telephone calls and messages to internet-based platforms (e.g. facetime and iMessage), and multi-purpose platforms in general supplement or replace former single-purpose technologies (The Danish Business Authority, 2017). As communication moves to mobile platforms (smartphones, tablets etc.), the use of apps has challenged the former position of websites and thereby the browsers used to locate the website-urls in the communication systems.

Lastly, **content** types, provided by different types of platforms, accessed through the backbone and access networks and activated through the use of a digital device, are divided into text, image, sound, video and lastly data. Data should be understood here as the specific configuration of web applications that enable an unprecedented, constant and instant flow of (meta)data via analytics, cookies, fingerprinting and other tracking technologies. Due to the high internet penetration, the generally high bandwidth, and the common use of digital services, almost all Danes have access to all types of content through the internet and, in turn, corporations have access to unprecedented amounts of data on people’s mediated communication.
In sum, from an infrastructural perspective, Denmark serves as an example of a digital communication system in which the internet can (and gradually does) substitute all former distribution systems. The mapping outlines a country characterized by highly developed and wide-ranging infrastructures and by general accessibility to different networks, which also reflects socio-economic conditions as well as political and economic strategies. Digitalization has thus increased the access and options to choose between various communication media. At the same time, these infrastructural changes have also had a significant impact on the market actors controlling and profiting from communications.

**Mapping economic structures**

Table 3 illustrates the key elements in a digital communication system as exemplified in the mapping of the Danish market actors who own and control access to different (increasingly intersecting) infrastructures. Similar to the previous infrastructural mapping, the table as a whole exemplifies how market actors can be mapped out at particular levels (e.g. the market for devices and access networks at the first level or the market for content at the fourth), as well as across levels (e.g. the way the market for infrastructure influences the platform market structures). This enables analyses of the most important power structures and competing interests at play in the Danish communication system.

We conducted the mapping through various stages of research: first, we identified the variety of actors that operate in the Danish market on the basis of mainly market reports from the Danish business and competition authorities. Following from this, we uncovered the often unclear and complex (co-)ownership structures. This led us to a further analysis of the different actors’ market assets and their subsequent market positions. In contrast to the access network level, the backbone network, platform and content actors make up a much broader and manifold sample consisting of national as well as international actors, where only the most prominent and exemplary are included in the mapping.
The key market actors are divided into two groups dependent on their ownership and control of respectively device and, the principal focus here, access networks. While the device manufacturers consist entirely of global actors, the access network providers are regional ‘giants’ characterized by having expanded their original business models through acquisitions in areas outside their original domain. For instance, TDC, the legacy telecommunications company, is placed at the center of the Danish marketplace due to its former position as the public monopoly telephone company, and thereby its ownership of the nationwide copper wire network. TDC has used their original market position to expand into mobile telephony, TV distribution and broadband services. Another example from the Danish context is the presence of various energy suppliers that have merged into the market for optical fiber on the basis of their existing networks. The seemingly wide range of different companies that offer for instance mobile telephony and broadband is to a large degree owned by the four licensed electromagnetic radio frequency holders: the mobile operators 3, Telenor, Telia, and TDC. The Danish market is, then, characterized by intense competition between a few large market actors, which keeps subscription rates at a minimum and leads to a continuing pressure on the supply of faster and stronger internet connections. However, TDC (and other national market actors) is both losing former revenues from TV distribution and telephone subscriptions and increasingly

<table>
<thead>
<tr>
<th>MARKET</th>
<th>DEVICE &amp; ACCESS NETWORK</th>
<th>Backbone Network</th>
<th>Platforms</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Apple, Huawei, Samsung, Lenovo, Asus, Dell &amp; TDC, Telia, Telenor, Waoo, Hiper, SE</td>
<td>Globalconnect, i2, TDC, Telia, Telenor, Alphabet (Google), Facebook, Akamai, AT&amp;T, Verizon, Amazon, Centurylink, etc.</td>
<td>Apple, Google, Microsoft, Facebook, Godaddy, DR, Netflix, JP/Politikens Hus, public authorities, etc.</td>
<td>Miscellaneous (users), Google, Facebook, Nordisk Film &amp; TV, Blu, Netflix, JP/Politikens Hus, Ritzau, public authorities, etc.</td>
</tr>
</tbody>
</table>

Table 3: The Danish communication system 2019 from a market perspective – most prominent actors
competes with, often global, infrastructure actors supplying international fiber connections, data centers, etc. (e.g. Akamai, Amazon, etc.).

The **backbone network** market is perhaps the most difficult to map out since infrastructures at this level are complex and opaque and often co-owned by multiple international actors with different levels of control over assets and inherent dependencies (this makes it an obvious, although neglected, field for further research). The most powerful actors at this level are global tier 1 networks such as CenturyLink, Verizon, and AT&T that mainly handle data traffic through extensive global fiber networks. Lower level backbone networks (tier 2 and down) rely on and pay transit fees to the tier 1 companies. These include the national and regional actors TDC, GlobalConnect, i2, Telia (Sweden) and Telenor (Norway) that control national fiber cables, routers and switches, and for GlobalConnect and i2, also own the existing Danish IXPs. Akamai, and Amazon also figure in the mapping of the backbone market, as they provide data storage in the form of cloud hosting and CDNs, and lastly Alphabet and Facebook own data centers and servers, but also co-own central sub-sea cables such as the Havfrue [Mermaid] cable as well as hosting services.

At the **platform** level, we identify a diverse sample of different prominent market actors that can be divided into two groups dependent on either their legacy positions in the communication system or their status as international, and mainly interned-based, businesses. The most powerful actors include Apple, Alphabet, and Microsoft that operate proprietary operative systems, browsers and apps for telephone calls, text-messaging, navigation, etc., that are preinstalled in Android phones (Alphabet), Iphones (Apple), and PCs (Windows). Although national websites and apps continue to rank high in web statistics (e.g. e-boks – the public digital mailbox, Mobilepay – the key mobile payment service, and DR – the national public service broadcaster), international players dominate the top lists (e.g. Facebook (also Messenger), Instagram (Facebook), Google (Alphabet), YouTube (Alphabet), and Snapchat). In general, the internet has fundamentally altered the national market conditions as e.g. Netflix and YouTube both reach larger proportions of the population each week than any of the national VOD players (DR Media Research, 2018). Social media platforms challenge the business models of traditional news media institutions (funded by subscription fees and advertising), but are simultaneously crucial distribution platforms –
Facebook e.g. reaches 65% of Danish adults on a daily basis (DR Media Research, 2018). Dominated by a few large actors (Alphabet and Facebook), personal data increasingly replace former sources of income, placing the data traders at the center of the Danish 2019 communication system (Helles, Lomborg and Lai, 2019). Lastly, internet domain registrars such as DK Hostmaster and GoDaddy act as necessary gatekeepers for the establishment of interned-based services.

In accordance with our theoretical point of departure, the definition of content producers extends from individuals engaged in private, semi-public or public communication to professional content producers. For search engines, social media applications and general online tracking enterprises, the list of producers is a varied and in principle endless assemblage of actors – we label these ‘miscellaneous’. In order to find examples of the key producers in Denmark, we looked at each of the actors in the platform level of the market mapping and asked: who produces content for this particular platform (e.g. DR), and who profits from it (e.g. Nordisk Film)? In other words, the strategy is to look for ‘the money arrow’ (DeFleur, 1971) in the content production system. Examples of professional content producers in the mapping include public and commercial national actors as well as international actors that to a large extent reflect the power structures at the platform level. In general, producers that used to have exclusive access to national audiences (e.g. publishers and TV producers) by now compete with and depend on global market actors such as Alphabet and Facebook, who take over large amounts of the former revenues through data-based business-models (DACP, 2017).

In sum, by mapping the dominant marked actors onto the infrastructure mapping, we gain knowledge on the existing actors’ assets, business models, competition relations, and interdependencies. The mapping shows on the one hand how the fast and wide-ranging digitalization of all parts of the Danish communication system has been led by national legacy actors: Former monopoly players such as TDC and DR built their market positions on historically determined distributions of assets (e.g. frequency allocation, ownership of telecommunications infrastructures etc.), and have used these as stepping stones for positioning themselves in the digital landscape. However, digitalization has, on the other
hand, led to a massive globalization of the Danish communication market and an erosion of traditional business models and value chains.

**Mapping policy structures**

Table 4 illustrates the key elements in a digital communication system as exemplified in the mapping of the Danish communication policies. We identify the overall structures of policy regulation and the categorizations that continue to determine how digital communication media are regulated.

Based on the previous mappings of infrastructural and economic structures – as well as a review of the existing research on media and communications policy – we collected the bulk of policy documents (e.g. acts, regulatory contracts, political agreements) that regulate the digital communication system in Denmark. After a qualitative content analysis, the entire collection was divided into four piles representing each of the four analytical levels – device and access networks, etc. – and later into ministerial departments and policy areas – respectively frequency and wired infrastructure policies. We then identified ideological values and regulatory strategies and analyzed how these influence and are influenced by the infrastructure and market structures.
Table 4: The Danish communication system 2019 from a policy perspective

Mapping the policies for **devices** and **access networks**, it is important to notice how Danish political regulation hinges onto EU legislation (e.g. the EU telecommunications directive, The Digital Single Market, etc.) as well as international internet governance (ITU, IETF, ICANN, BEREC, etc.). The mapping explicates differences between the regulation of the four types of material infrastructure identified in the access network level of the infrastructure mapping: ‘Wired infrastructure policy’ – regulating copper wire, coaxial cable, and optical fiber – promotes a market-based development of communication infrastructures since wired infrastructures are (in principle) an abundant resource. However, the Danish government grants funding to co-operative projects aimed at improving access to high speed internet through the so-called **Broadband Pool** as part of a general digitalization strategy. ‘Frequency policy’ – regulating air-borne infrastructures – implies a higher degree of political involvement as electromagnetic frequencies are a limited resource and the market demand for frequencies is larger than the available spectrum. The Danish state and international committees (e.g. ICANN) therefore play a more direct part in regulating the allocation and use of radio frequencies. Cultural policies have historically been dominant in this policy field, reserving a large part of the spectrum for public service broadcasting. The growing use of spectrum for mobile telephony and broadband has led to an
increased market orientation and since the 1990s various commercial mobile operators have been assigned licenses for radio frequencies on increasingly commercial terms. ‘Telecommunications policy’ revolves around ensuring universal access and is influenced by the liberalization of the telecommunications sector in the 1990s; and oblige the incumbent operator (TDC) to give competing operators access to its infrastructures and thereby create conditions for free competition. As a result, cross-country access is secured and subscription fees are kept at a minimum compared to for instance in the US. The telecommunications act also determines the rules for net neutrality (as required by EU telecommunications regulation), regulated through an independent forum of professional stakeholders and consumer representatives. Both ‘telecommunications policy’ and its historical opposite, ‘broadcasting policy’, balance on the one hand socio-cultural values based on universalistic welfare principles and on the other hand commercial interests encouraging competition and economic growth. Radio and TV, however, continue to be primarily regulated in the realm of cultural policy, while telecommunications is regulated as a matter of communication carriage.

Similar to the access network level, backbone networks are subject to telecommunications policy (net neutrality), competition policy, etc. However, although for instance Danish competition laws aim at ensuring that no actor becomes large enough to dominate any market, the enforcement of competition policy is limited since the actual backbone market has not yet been adequately identified. Moreover, the US Federal Communications Commission do not monitor US backbone competition. In result, the global (and often American) backbone companies (e.g. Akamai, Facebook, etc.) operating on Danish ground are not subject to competition regulation.

The mapping of platform policies shows that the Danish state, policy-makers and authorities have been actively involved in the digitalization of the communication system through the development of a range of public websites and apps. Early digitalization strategies were built on a demand-driven logic assuming that the population would acquire internet access (and skills) if the digital products were available and if they were encouraged to use them. However, Danish media and communication policies are challenged by the multifarious nature of web-based platforms that can combine formerly separated communication forms and thus blur the basic categorizations that these legislative frameworks are built on.
For example, legislators categorize some market players as ‘TV-like’ and hence subject to broadcasting policy, while others (such as YouTube) are excluded from cultural policy legislation. National legislation is in general far more successful at regulating national legacy players and newcomers than it is at targeting international, internet-based platforms (such as Google and Facebook) that act from outside Danish jurisdiction, and are built on entirely different data-based business models. Following the EU ePrivacy Directive, however, the Danish Business Authority oversees that platforms follow the ‘cookie rules’ (information and consent). Lastly, whereas DK Hostmaster administering .dk domains is subject to Danish domain policy, GoDaddy (and other international actors) handling .com domains are not. Consequently, national policies do not regulate the globalized communication system that Danes actually act within, as it is still oriented towards the subsystem made up by national platform actors.

The level of content policy resonates with the platform level, as the state has paved the way for the extensive digitalization of Danish communication content, but at the same time struggles to amend the legislative frameworks to effectively address digital communication. Content policy traditionally distinguishes between telecommunications developed for private one-to-one communication, subject to limited and more general regulation, and broadcasting media distributing public one-to-many communication, characterized by higher degrees of state involvement (e.g. through extensive public service schemes). Digital content policies are generally challenged by the use of social media and many-to-many communication that blur the distinctions between when a speech act is made in public or private, which leads to emergent regulation that hinges on court rulings, re-interpreations of existing legislation, etc. An example of recent attempts at increasing the political regulation of digital content is the extensive GDPR program which – at least in principle – limits the ways in which data can be collected, stored and traded across Europe. In general, digital information flows have challenged the existing interpretation and enforcement of key policies related to freedom of expression, copyright, etc.

In sum, applying the framework to the Danish case sheds light on the different and sometimes contradictory initiatives, strategies and values that influence the political regulation of the digital communication system. Crucial to the Danish communication system from a policy perspective is the dichotomy between
socio-cultural policies, based on universalistic welfare state values that legitimate relatively strong state regulation, and competition policies encouraging market and self-regulation. On the one hand this duality has contributed to the development of a thriving telecommunications sector characterized by high access and low prices where market and consumer interests go hand-in-hand. On the other, the global nature of the internet has challenged national regulation and evoked conflicts between capitalist governance structures and universalistic welfare state principles, which complicates efforts to reform existing policy schemes. Digital communication services and content forms are still in the process of being institutionalized and continue to fall between historically defined policy fields. As a result, digital communication content is subject to less political regulation than its analog predecessors and the digitalization of the Danish communication system has all in all led to a general decline in the impact of political regulation.

**Communication scenarios illustrative of the Danish digital communication system**

The three example mappings of the Danish context have so far been described separately, however, the next and crucial step is to synthetize them. The four levels – devices and access networks, backbone networks, platforms and content – recur in each of the three versions of the mapping so that the technological features of the existing access networks resonate with market structures and policies regulating access to these networks etc. This also makes it possible to assess how the infrastructural, economic and political regulatory forces interact and condition each other. Hence, on the basis of the description of the components of the Danish 2019 communication system, we now exemplify how the mapping method enables further analyses of the power structures and control mechanisms that frame communication capabilities within a specific societal context. Figure 1 establishes three communication scenarios where each chain of events has different regulatory implications.
Figure 1 (a, b, c). Illustrates three different communication scenarios. Each of the three subfigures show infrastructural, market and policy implications of specific communication activities across the four analytical levels (device and access network, backbone network, platforms and content).
In figure 1a, an individual logs into her Acer laptop to get a news update from the Danish newspaper Politiken’s website. Her PC connects to the internet through a DSL broadband connection running on the Danish cobber wire infrastructure. This is owned and controlled by the network operator TDC and both the infrastructure and the market for it are politically regulated by for instance Danish cable laying access and expropriation policy and telecommunications policy. In order to access the website and its content, the DSL broadband connects to a national backbone network and the servers hosting Politiken’s content. The network is still owned by TDC but the servers belong to Amazon (Cloudfront). Less legislation applies to this level altogether, and the national actor TDC is subject to more political regulation than its international colleague. This enables the connection to a browser, owned by Microsoft (Internet explorer), and a website owned by JP/Politikens Hus – the latter being politically regulated under national domain policy, the European E-commerce rules, consumer policy, etc. Lastly the individual receives news in the form of both textual and audiovisual content and simultaneously cookies on the site harvest data on the individual. As politiken.dk is a national website and politically categorized as ‘news media’, it is subject to various forms of content regulation (e.g. the guiding rules on press ethics grounded on the media liability act, media subsidy requirements, marketing policy, copy right etc.). The actors who profit from data tracking on the site (Alphabet, Facebook and numerous others) are in principle subject to the European GDPR directive, but, unlike in other European contexts, no ruling has so far enforced the rules of the direction on Danish ground.

In a different scenario, figure 1b, the individual unlocks her Huawei Smartphone to check the recent US NFL scores. Her smartphone connects to the 4G network enabled by the electromagnetic radio frequencies and in this case owned by the mobile operator Telenor. Telenor’s mobile broadband is politically regulated under for instance spectrum policy and telecommunications policy. The 4G network connects to a fiber cable, still owned by Telenor, and from there to GlobalConnect’s IXP, which connects to the undersea Havfrue [Mermaid] cable, owned by several market actors including Alphabet and Facebook. While the local actors, Telenor and GlobalConnect, are subject to Danish telecommunications policy, competition policy, etc., the latter part of the infrastructural chain of events – from the undersea cable stakeholders to the NFL – is not.
Once the NFL website metadata is retrieved (via the Google Chrome browser), then consumer policy and the European E-commerce rules apply. The individual now accesses all five content types on the NFL site, NFL and again various trackers profit from her visit and are, again in principle, subject to EU data protection policy.

In the third scenario, figure 1c, the individual picks up her iPad to watch a TV series on Netflix. Her tablet connects to her fiber broadband running on the optical fiber network. This is owned by the assembly of former energy supply companies Waoo, and politically regulated by the same array of policies that applied to the access network level in the first scenario. The fiber broadband connects to Waoo’s national backbone network, regulated by telecommunications policy, and to a local CDN owned by the Netflix subsidiary OpenConnect, which in principle is regulated under Danish competition policy, although, as mentioned earlier, the rules are not enforced. This, in turn, allows her to open the Netflix application enabled by Apple IOS, both of which, again, only in principle are subject to e.g. E-commerce rules, consumer policy, and competition regulation. Lastly, she accesses the desired audiovisual content and data is retrieved on her visit and preferences. Netflix profits from her visit and its interface is refined to match the newly harvested data and push personalized content. Netflix is also in principle subject to data privacy policy and the producers, internal and external, are subject to audiovisual policies.

The scenarios exhibit that digital communication rests on not just national access networks but also international backbone networks, IXPs, CDNs, cloud servers, etc. that are increasingly vital to any type of mundane communication – be it reading the news, engaging in sports, or watching TV. Across the communicative chain of events, digital communication systems such as the Danish are critically reliant on both global infrastructures controlled by international actors and the adaption to business models that rely on the massive collection of digital data – be it the Danish operator TDC at the access network level or the news site Politiken at the platform level. Increasingly, dominant international market actors, such as Apple, Alphabet, and Netflix, take over functions and revenues from national legacy players and establish new markets that have not yet been adequately identified and thereby not subjected to political regulation. Policy reforms have focused on regulating national legacy market sectors and institutions
transitioning online – such as broadcasters, news media and telcos – rather than emergent internet-based markets, international corporations and digital content.

In other words, in the scenarios we glimpse the analytical potentials of the mapping method as it supports systemic empirical analyses of the ways in which infrastructures, economic structures and policy structures regulate digital communication. These analytical perspectives include: how mediated communication is rematerialized as it increasingly rests on digital infrastructures, and how data traffic is influenced by the geo-political circumstances surrounding material infrastructures; how new markets and market actors emerge and become dominant, and how business models and market interdependencies develop; and lastly how legislation is in the process of being reformed and adapted (more or less successfully) to address emerging, urgent political issues surrounding digital infrastructures and markets.

**Conclusion**

This article has identified a research gap in terms of theoretical and methodological frameworks for analyzing and understanding the interplay between infrastructures, markets and policies in digital communication systems. By applying an infrastructural and materialist perspective to system analysis, we have developed a methodological framework for analyzing the structural forces that regulate common communication capabilities in different societal settings. The case study described above serves as a starting point for and a guide to mappings and analyses of contexts beyond the Danish. Going back to the matrix between regulatory structures and communication chains of events that make up table 1, the three mappings (table 2, 3, 4) provide answers to each of the 12 matrix questions. Apart from generating original knowledge on the Danish digital communication system, the lessons learned from this analysis also translate into three main guidelines for future comparative studies that all underline benefits of an ‘infrastructural turn’ in systemic media and communication analysis:

First, the framework draws attention to the material conditions that influence the structural characteristics of communication systems. Physical characteristics such as country size, topography, etc., as well as historic infrastructures have
significant impact on internet development as both market structures and political initiatives are fundamentally based on the resources at hand. Characteristics associated with democratic corporatist media systems (Hallin and Mancini, 2004) or media welfare states (Syvertsen et al. 2014) should thus be seen not only as reflections of institutional features and ideological traditions, but also as dependent on material conditions that frame and have framed (de-)institutionalization processes. Hence, future comparative studies should seek to analyze similar as well as discrete contexts that differ in terms of geography, demography and infrastructural development – e.g. where digital infrastructures are built from scratch rather than established on the basis of existing communication networks.

Second, and following from the first guideline, we stress the importance of studying the specific infrastructures that digital communication systems are based on. In order to explain quantitative differences between countries in terms of internet quality and use, affordability and competition structures, platform use and content supply, etc. it is crucial to explore how internet access is supplied. For example, even though wired and air-born infrastructures can be used for the same communication purposes, they have different technological and regulatory implications and thus give rise to different commercial power struggles and political questions. The types of communications infrastructure that provide a country with broadband are thus key to understanding the competition structures and policies and ultimately the structural conditions that frame digital communication capabilities.

Finally, rather than focusing on the activities and market strategies of particular institutions (such as Facebook or legacy players), we draw attention to the implications that digitalization has for individuals and thereby for societies as a whole. Taking individuals’ communication capabilities as a point of departure and asking how they are enabled and constrained by material conditions, in turn makes it possible to identify the basic communication resources that key market players control and thus to explain power structures and competition conditions. This approach also enables further discussions and explorations of the general features that make internet-based players fundamentally different from legacy media, and the ways in which they challenge the existing policy structures originating in the structural conditions of the past.
In conclusion, the framework and mapping method create a common ground for future comparative analyses of communication systems with a greater attention to digitalization than that of previous studies. In future studies, the framework will also serve as a common ground for developing indicators that can assess how infrastructures, markets and policies regulate digital communication across the globe.

References


Publication

Internet Histories, 2020
DOI: https://doi.org/10.1080/24701475.2020.1759010

Authors

Sofie Flensburg & Signe Sophus Lai
Networks of Power

Analysing the evolution of the Danish internet infrastructure
Abstract

This article studies the evolution of the internet infrastructure and assesses emerging digital power structures and regulatory dynamics. We revisit and develop Thomas P. Hughes’ momentum theory (1994) and contend that the internet, as other large technological systems, has evolved in different phases reflecting a shift from being mainly influenced by socio-economic conditions to having a determining influence on the development of societal structures. We argue that contemporary internet infrastructure studies can benefit from Hughes’ theoretical approach, but also need to strengthen their methodological and empirical strategy. The article contributes to this by approaching the changes in digital infrastructures, markets, and state policies in Denmark from 1992 to 2019. Building on database material, we analyse the development of digital devices and internet connections, submarine fibre optic cables and internet exchange points, websites, and digital content. We conclude that the identified changes confirm Hughes’ momentum hypothesis: in the beginning the implementation of the internet reflected historical market structures and institutional characteristics, while the later development challenges and transforms the established regime.

Keywords

infrastructure, internet history, momentum, technological systems, political economy

Introduction

Of the great construction projects of the last century, none has been more impressive in its technical, economic, and scientific aspects, none has been more influential in its social effects, and none has engaged more thoroughly our constructive instincts and capabilities (…) from a study of these years one can perceive the ordering, integrating, coordinating, and systematizing nature of modern human societies (Hughes, 1983, p. 1).

This is how the technology historian Thomas P. Hughes begins his extensive exploration of the electrification of Western societies. The invention and gradual
introduction of electricity started a monumental transformation of people’s everyday lives and fundamentally altered the physical and socio-cultural organization of modern societies. Electricity is, on the one hand, an infrastructural precondition for the introduction and diffusion of electronic communication media. On the other hand, the electrification processes bear a striking resemblance to the development of the internet and provides a valuable inspiration for analysing the digital infrastructure. In this article we apply and extend Hughes’ (2012) theory on “the evolution of large technological systems” as a framework for mapping the complex ways societal structures and digital infrastructures mutually influence and transform one-another.

We focus on the evolution of the internet and digital communication services in Denmark, asking how the structural conditions and control mechanisms framing internet-based communication have developed since the first commercial internet connections were established in the early 1990s. Our aim is to analyse and discuss how the basic power structures have been reconfigured as the internet has evolved into a dominant communication infrastructure. We therefore stress the changing business models and market structures as they relate to both the infrastructural development and changing conditions for regulation and governance. In other words, we focus less on the social and cultural aspects of digitisation in order to emphasise the arenas where infrastructural, economic, and political power is negotiated and established.

We choose Denmark for a number of reasons: First, being a highly digitised society, we can think of Denmark as an ‘extreme case’ (Yin, 2009) that is suitable for exploring how the internet has evolved into a general purpose technology (Naughton, 2016); Secondly, Denmark is a particularly interesting case for studying the (changing) role of state institutions as these have historically played an active role in establishing and operating communication networks (see Flensburg, 2020); And finally, the article is a contribution to the relatively limited scholarly work on the historical development of the Danish internet.

The article is divided into three main sections: first, we outline Hughes’ theory on technological systems and discuss how it can be applied to historical analyses of the internet; the second section comprises an empirical case study analysing
the structural development of the internet in Denmark; and lastly, we discuss how the analysis relates to the phases identified by Hughes.

**The evolution of large technological systems**

Different theoretical traditions have explored the influence of new technologies on societies and vice versa under the headlines of for instance mediatization (Hjarvard, 2013) or medium theory (Briggs & Burke, 2009; Meyrowitz, 1994). In recent years, the growing field of internet infrastructure studies have explored how digital governance forms are developing in the intersection between technology and regulation (see Musiani et al., 2016). Across these research fields, debates continue to revolve around the basic question: whether or not humans shape communication technologies or if it is the other way around.

Despite the growing interest in the mutual influence of material infrastructures and institutional arrangements, research continues to struggle with this question of causality. Institutional analyses, on the one hand, tend to forget “the simple, but crucial fact that each communication technology is a material resource whose distinctive features help to explain the media institutions and communicative practices that have emerged” (Jensen, 2013, p. 216). Technologically-oriented analyses, on the other hand, tend to focus on specific innovation or implementation processes and thereby underestimate the impact of institutional structures. As argued by internet infrastructure researchers (Parks & Starosielski, 2015; Sandvig, 2013; Schroeder, 2019), Hughes’ study of the electrification of Western societies was pioneering as it uncovered “the political and social choices in arenas that had previously seemed only technical and mundane” (Sandvig, 2013, p. 94), while at the same time stressing the “soft determinism” (Hughes, 1994) of technologies. The momentum theory thereby provides a useful “middle ground” between social constructivism (young systems) and technological determinism (mature systems), emphasising the processual nature of technological systems (Jensen, 2013, p. 216).

Hughes (2012) argues that technological systems pass through four general development phases that are characterized by particular configurations of determinism: in the first phase, the invention and early development of a technology
is dependent on social conditions, political decisions, and allocation of money (p. 52-57). This leads to the transfer phase, where a technology is transformed into a complex system of “manufacturing, sales and service facilities” (Hughes, 2012, p. 57) and extended beyond its original context. In the case of both the electricity system and the internet, the transfer phase involves a geographical transfer (e.g. from the US to Europe) as well as an adaption of the system to fit into the new context. As technologies are adapted to various societal contexts, they acquire particular ‘styles’ that make them “differ in characteristics from time to time, from region to region, and even from nation to nation” (p. 63).

A successful transfer will often result in a technological as well as commercial acceleration moving the system into a phase of significant growth characterized by increasing popularity, extensions of the usability of the technology to support a growing number of activities, and economic surplus that often leads to increasing competition as well as consolidation of market powers (Hughes, 2012, p. 64). As systems are used for a growing range of purposes and subjected to an increasing load, inherent weaknesses and limitations, referred to as ‘reverse salients’, appear (p. 67). These will typically slow down the growth of a system for a period, but eventually lead to a breakthrough as the problem is solved and can, in turn, an acceleration in the public diffusion of the technology and to the final phase of ‘technological momentum’.

In the momentum phase, a technological system reaches a ‘point of no return’ where it serves as a foundation for societal functions and institutions. When a technology has reached momentum, a breakdown in the system is catastrophic since a wide range of external activities and systems critically depend on it. Hughes notes that momentum entails a high degree of integration between societal structures and the particular system. While economic and political structures are external forces influencing technologies at the early stages of development, these structures are increasingly influenced by the material infrastructures of a given technology reaching momentum.

This understanding of technological development provides internet researchers with a theoretical lens through which the complex processes of digitisation can be understood. It surely cannot capture or explain all aspects of the development as it focuses on the broader, environmental, and structural development, rather
than on the more particular forms of use, the content, and the “social spaces” that the internet creates (Abbate, 2017). It points our attention to the dialectic ways that technology and institutional structures mutually shape each other and is thereby an important reference point for studying the infrastructure and political economy of the internet. However, despite the frequent references to Hughes in digital infrastructure studies, none of the above-mentioned examples – including Hughes’ (1998) own work on the development of ARPANET – specify how the theoretical framework can be applied to empirical analyses of the internet.

**Operationalising the momentum theory**

When operationalizing Hughes’ theory, the first challenge is to delimit the scope of analysis and thereby establish what is meant by a ‘technological system’ – a definition that Hughes is deliberately and admittedly vague about (Hughes, 1983, p. 6). For the purposes of the following analyses, we argue for a narrow understanding, similar to when Hughes at one point (1983) refers to technological systems as “constituted of related parts or components [that] are connected by a network, or structure” (p. 5). Large technological systems – such as the internet – are, in this perspective, made up by material components that are organized according to different economic and political logics. While these structures have a significant influence on how the infrastructure develops (and vice versa), they are external forces and not part of the system itself.

This materialist approach means that we study the introduction and diffusion of the internet in Denmark as first and foremost a technological development emphasising the physical changes it entails and assessing what activities and practises the digital infrastructure enables at different points in time. When analysing the political economy of the internet, we identify the market actors and regulatory institutions that control the material resources of this infrastructure. In other words, we focus on the development of internet infrastructures, digital markets, and state policies as three analytical dimensions for studying digitisation processes. The following analysis is devised along the lines of these dimensions as well as across four analytical layers that translate into key components of the digital infrastructure, namely: devices and access networks; backbone networks;
applications; and content (see Flensburg & Lai, 2019). The four layers are hierarchical in so far as one cannot access a particular application and send or receive data without having access to a digital device connected to a local access network as well as to the global backbone infrastructure. By studying each layer separately, we gain insight into the horizontal power structures (e.g. the actors controlling and profiting from the access networks); by comparing the different layers, however, we shed light on the vertical power structures (how various businesses control and profit from resources across the value chain).

In the case study, we analyse key empirical indicators from each level: the gradual diffusion of PCs and smartphones and the use of different types of internet and broadband connections; the build-out of submarine optical fibre connections and the establishment of Internet Exchange Points (IXPs); the development in the most used websites; and the prevalence of high-capacity content as well as third party services. These examples do not reflect the entire infrastructure and thereby do not comprise all relevant market actors (e.g. the web is only one type of internet-based application), yet they do constitute critical components of the internet. While other analytical examples (e.g. mobile apps, datacentres, etc.) would be interesting for an analysis of the contemporary internet, the existing sources of information for uncovering these aspects are less reliable – and particularly difficult to apply in historical analyses.

**Method, data sources, and processing**

The empirical data for the analysis stems from a number of sources containing historical information on the above-mentioned indicators. More specifically, we reviewed textbooks, journal articles, and archival records, made calculations on the basis of existing statistics and databases, and conducted source critical analysis of official reports and evaluations, and legislative acts and court rulings.

In order to study the device development, we assess the share of internet users in the population and the prominence of the most typical digital devices (computers and smartphones) using the national statistics published yearly by Statistics Denmark (n.d.-b, n.d.-a). When assessing the development in access networks (the number of fixed and mobile broadband subscriptions), we use the
yearly Tele Statistics that are compiled by the national tele authorities (Danish agency for energy, n.d.). These numbers also make it possible to identify the relevant market actors and to calculate the market shares of respectively “national or international”, “state-owned or commercial”, and “brownfield or greenfield” actors. The latter concepts originate in management theory on infrastructure development (Hjarvard & Helles, 2015, p. 5) and is for the purpose of this article used to distinguish between market actors that base their digital endeavours on top of an already established legacy business on the one hand, and newcomers on the other. Lastly, we identify the relevant, existing policies that regulate PCs/smartphones and broadband connections as well as the market for them.

To assess the backbone development, we calculate the amount of submarine cables and the number of existing IXPs using Telegeography’s databases (n.d.-b, n.d.-a). Different from for instance Statistics Denmark, Telegeography is a commercial enterprise (financed by Huawei and Equinix). Nonetheless, the databases are to date the most substantial on backbone infrastructure, and are therefore also the default source used by established researchers (see e.g. Parks & Starosielski, 2015; Winseck, 2019). We analyse the market structures by coding the existing actors in the submarine cable markets and calculating market shares according to the same variables as listed in the section above (national vs. international, etc.). And finally, we approach the policy question by accessing the relevant policies that regulate the submarine cable and IXP infrastructures and the equivalent markets.

While the two first analytical layers, and the empirical examples derived from them, focus on actual ‘internet layers’ supporting the transport of data between different networks (Tanenbaum and Wetherall, 2013, p. 48), the following two layers and examples are concerned with the software and hosting of content. We focus on the web, as one of many applications running on the internet (email is another), since this has historically been the main arena for commercialisation and thereby for observing economic and political power struggles. More specifically, we analyse the development in the 50 most used websites in Denmark as measured by visitors and page views. We could also have looked at for instance the sites where people spend the most accumulated time or more broadly at for instance the most central sites in the link structure of the Danish web (Brügger
et al., 2020). However, the most visited websites where people also view the most pages are, by all comparisons, highly useful for comparing the most dominant actors in the web market. In other words, from a political economy perspective, the most used websites are crucial, even if we then miss out of a multitude of other sites down the long tail that would nuance the limited view of the top 50 websites. This analytical category is among the most difficult to assess empirically as we, in order to map the websites historically, rely on three different measurement systems: Kantar Gallup measurements in 2002, the association of digital media organisations in Denmark (FDIM) and the Gemius system in 2012, and Amazon Alexa in 2019. These data sources are not straightforwardly comparable as they employ different configurations of site and panel measurements, and at times have been skewed towards Danish sites (the member sites of FDIM) (see Lai, 2015 for methodological discussion). Nonetheless, each of them constitute the most established systems for monitoring web traffic at each of the three impact points. Common for the databases is that they only relay information on the top 50 most used sites (for non-members or otherwise protected behind a paywall), which delimits our scope of analysis to the same. As a result of these methodological challenges, the historical analysis of the changes in the Danish web (market) is indicative rather than exact. As opposed to analyses focusing on for instance the .dk domain (see e.g. Brügger & Laursen, 2019), we aim at targeting the highly globalized market structures in order to discuss the ways that international businesses control Danes’ web-activities (Flyverbom et al., 2017). We see the web development as crucial for assessing the changing governance forms and conditions for state regulation and therefore accept the methodological reservations that inevitably follow from this type of historical analysis.

The analysis of the content development is built on the top websites and is therefore subject to the same methodological considerations. We first access the historical versions of each website on the top-50 lists in the Internet Archive (n.d.), and code their front pages according to five content types – text, image, audio, video, and data. Although front pages cannot be considered representative of the entire website in question, they function as a key indicators of the purposes that the internet is used for (Flensburg & Lai, 2019) and the development of the infrastructure (streaming of audio-visual content is one of the most prevalent
high capacity tasks). In terms of data being harvested from the sites, we use two tools to extract the residing cookies on the front pages of each site: the Internet Archive, which allows us to access historical versions of website source codes, and the Tracker Tracker tool (n.d.), which in turn allows for the scraping of historical cookies that were present on the sites at the time of use. There are a number of reservations towards this method, most importantly regarding the issue of cookie databases and their historical coverage (see Helmond, 2017; Nielsen, 2019 for discussions of limitations of the Tracker tracker). Approaching the web tracking market, we look at the actors controlling the unique cookies on the sites. Lastly, we assess the existing policies that regulate web content and the online trackers.

With these indicators, the following sections outline how the various infrastructure, market, and policy structures have developed over time. We look specifically at the development from 1992, when the first private internet connections were introduced in Denmark, until 2019. We divide this timespan into three periods that reflect the infrastructural dimension and testifies to the hierarchical relationship between the four analytical layers (where devices, access networks, and backbone infrastructures serve as an initial foundation for the development of applications and content, although, as we discuss later, the use of various applications and content types can also increase the demand for built-out of the underlying infrastructure). As such, the first period from 1992 to 2002 represents an early phase where internet connections become publicly available, are used by a growing proportion of the population, and the backbone infrastructure is established; the second period from 2003 to 2012 is characterized by a massive digitisation process as various internet-based services and applications become an integrated part of most people’s everyday life; And in the final period from 2013 to the present, the internet becomes a dominating infrastructure that can replace a wide range of network services and distribution systems. In the discussion and conclusion, we assess Hughes’ hypothesis of technological momentum as it relates to the case study and discuss alternative approaches, interpretations, and periodisations.

We focus on the years 2002, 2012, and 2019 in particular and use these impact points to compare the infrastructural, market and policy structures across the
four analytical layers. This analytical approach is reflected in the organisation of the following sections.

**The evolution of the internet in Denmark**

In the period from 1992 to 2019, the internet went from being a niche endeavour reserved for the most technologically inclined Danes to playing a central role in most people’s everyday lives. Denmark is today one of the most digitised countries in the world, and an ever-growing amount of activities and services are by now – directly or indirectly – dependent on digital infrastructures. Below, we analyse this historical transformation process in order to discuss how the internet as a technological system has been shaped by existing societal structures, and how the material conditions following from the increased reliance on digital infrastructures changes power structures and regulatory dynamics of the past.

**Devices and access networks**

Figure 1 overviews the penetration of internet connections and internet-enabled devices over time, and thereby paints a compelling picture of the massive digitalisation of Denmark. In 1996, just 5% have access to the internet, while 23% own a PC. By 2002 more than half of the population is online and the diffusion of PCs is up to 72%. In the second period, from 2002 to 2012, internet connections and digital devices become an integral part of most peoples’ lives with 92% owning a PC and an internet connection. And from 2013 and onwards, the internet reaches its maximum penetration, and the diffusion of smartphones extends the general use of digital technologies.
As internet becomes common property, the supply and quality of broadband connections increases and, as shown in Figure 2, mobile broadband subscriptions come to dominate. In 2002, connections were relatively slow and poor quality (71% were dial-up and just 17% were fixed broadband). By 2012, the speed and quality of connections is improved and mobile broadband connections exceed fixed broadband. 2019 speaks to the continued growth in broadband, especially mobile. Mobile subscriptions surpass the number of inhabitants with 130 to 100 and 75% of all connections are mobile.
From a market perspective, the development in both devices and access networks reflects the growing importance of internet service provision. While the market for internet connections throughout the 1990s and in the early 2000s was subordinated to the legacy business models of landline telephony and TV-distribution, the supply of broadband becomes increasingly profitable and the competition increases as new broadband types are introduced. In spite of the explosive development in internet connections across the three periods, the former monopoly telecommunications operator TDC (privatized in 1994) remains dominant with largely unaltered market shares around 45% across fixed and mobile subscriptions. TDC’s key position as an internet service provider (ISP) is the result of the company’s ownership of the national copper wire network originally established for telephony and repurposed for dial-up and DSL connections. As new types of broadband are introduced, most prominently mobile, new market actors enter the stage independently of TDC’s leased lines. The Swedish and Norwegian incumbent actors, Telia and Telenor, as well the private, greenfield mobile operator 3 make up significant proportions of the Danish market.
Towards 2019 the market for broadband continues to grow (as seen in figure 2), but the business models in the telecommunications sector are rapidly changing: While internet service provision used to make up a profitable add-on business to existing businesses, by 2019, streaming, voice-calls, and messaging services significantly challenge the formerly lucrative telephony and TV trade (The Danish Agency for Business, 2017). As a result, the revenues of the telephone companies (telcos) are decreasing and the financial leeway for infrastructural investments is reduced.

From a political perspective, the establishment and spread of internet connections in the early 1990s was framed by the existing policy regime, where the state took active part in developing and managing basic infrastructures. For instance, the first private internet connections were established by public institutions (e.g. the later TDC) and various digitisation strategies were passed in the 1990s (see e.g. Forskningsministeriet, 1994). In the mid-1990s, the telecommunications sector was liberalised with the explicit aim of challenging TDC’s incumbent position while at the same time ensuring that all Danes had equal access to basic services. The role of the state was to monitor the competition structures and pricing and ensure that competing market players had access to TDC’s infrastructure. As internet service provision became increasingly profitable, this approach proved highly successful by keeping rates at a minimum and allowing new players to enter (although independent companies were frequently bought by the large market actors). This competition-oriented telecommunications policy, however, was dependent on a highly competitive and lucrative market and has been challenged by the emergence of over-the-top (OTT) services⁴, raising fundamental, ideological questions of how to ensure continued improvement of the digital infrastructure. As a result, in 2017 a public broadband pool was established, allocating state funding to broadband build-out. To sum up, the changing market structures following from the infrastructural development are putting a significant pressure on the established telecommunications policy developed for the purpose of regulating historical sectors and business models.

---

⁴ Over-the-top here refers to internet-based distribution services that can substitute specialized and dedicated network services (e.g. internet-based telephony, text messaging, VOD and streaming etc.)
Backbone infrastructures

The growing use of internet connections described above was highly dependent on the build-out of the underlying backbone infrastructure that gave Danes access to the global internet. Figure 3 shows how a substantial amount, 33,000 km, of the existing submarine cables landing in Denmark were laid during the first period, in which the first Danish IXP is also put in place. After a period of little activity in this domain, late in the third period, from 2018 and onwards, another 9000 km of cables are (being) laid. While the early build-out of the backbone was foundational for the general digitisation, the later build-out is a consequence of the heavy use of high capacity content (epitomized by streaming) and from expectations for future ventures (with 5G, Internet of things, artificial intelligence, and virtual reality).

![Internet exchange points and submarine cables landing in Denmark](image)

Figure 3: total length of submarine cables and accumulated number of internet exchange points (IXPs).

The early years of extensive submarine cable laying were spurred by high expectations of return on investment and financed by existing telcos positioning
themselves strategically at the core of the growing market for internet connections, and utilising their existing infrastructure and expertise. Figure 4 outlines the shares between different market actors involved in the build-out of submarine cables landing in Denmark in the three periods. The substantial submarine cable laying processes of the 90s and early 2000s involve primarily brownfield infrastructure actors (95%) such as TDC, and is divided evenly between state-owned and private players. In 2012, national, greenfield (60%) companies such as GlobalConnect overtake the majority of cable laying processes. However, by 2019, there are no national players involved, and the engaged actors are primarily private, greenfield companies (85%) like for instance Google (Alphabet) and Bulk Infrastructure (that own and control the significant mermaid cable established in 2019, which connects Denmark directly to the US servers). The market for IXPs follows the same tendency: the only Danish exchange point for network operators was until 2011 operated by a state institution and located at the technical university of Denmark. In the second and last phase, various competing IXPs have been established by commercial actors.

![Market actors involved in submarine cable laying over time](image)

**Figure 4:** Market actors involved in cable laying processes over time, cables landing in Denmark. For each binary market code, the figure equals 100%, as illustrated by the three colour shades.
From a political perspective, the state’s influence on the development and operation of the backbone infrastructure is generally weak as the absence of political recognition of the market entails limited monitoring and lack of regulatory standards compared to the access network market (Nuechterlein & Weiser, 2013, p. 6). Once the backbone network market changes hands from the established national telecommunications operators, regulated by existing national policies, to international actors, this regulatory challenge only increases. By 2002, backbone network components are included in existing legislation on telecommunications, but from 2002 to 2012 there are no further inclusion of backbone networks into existing policy nor any new initiatives aimed at politically regulating the internet backbone. Policies on internet security and net neutrality put in place by 2019 in principle include the backbone network components, still the backbone network market is not subjected to systematic monitoring and thus remains highly unregulated and black-boxed.

**Applications**

On the one hand, the development of both local access networks and global backbone networks is an underlying premise for the increasing use of digital applications over the three periods. On the other, the demand for high capacity networks is caused by an explosive growth in the world wide web (from 38 million in 2002 to 238 million in 2019) (Internet Live Stats, n.d.)). Figure 5 illustrates our coding of the 50 most used websites over time.
Looking at the top 50 websites in 2002, the majority of web-actors are national, private actors, and more than half are brownfield companies adding an online activity to their existing ones. For instance, the largest Danish newspapers and broadcasters all establish websites in the 1990s and TDC launches Opasia. Universities and other public institutions also held prominent positions in the top 50. By contrast, in 2012, more than half of the sites in top 50 are international and greenfield companies like Facebook and Google. By 2019 international, greenfield actors still dominate, yet, the prominence of public sites testifies to the substantial e-Government initiatives of the Danish state.

Across the three periods, political website regulation happens via either existing telecommunications policies that are extended to cover the internet in some form or another, or sector specific policies that target the online advances of brownfield actors in e.g. news or broadcasting. In 2002, a public digitalization strategy, e-commerce policy, and support for digital literacy have been passed, and media liability regulation has been extended to include web-based news outlets (by self-registration). By 2012, with international actors like Netflix and the large public and private Danish broadcasters having established streaming
applications, video-on-demand has been included in general broadcasting policy (on the basis of the EU Audiovisual Media Services Directive) and a national domain policy is in place. However, the impact of specialized policies such as media subsidies, public service policy, as well as the national domain policy decreases following the increased use of global sites, which do not fit into or abide to the existing regulatory frameworks. As such, parallel to internet-based communication increasingly happening on international applications, the state loses regulatory power, although, in recent years, attention towards regulating international digital applications has grown.

**Content**

The most occurrent types of content on the website front pages follows the development in the most used websites described above. In the early top 50 list, we encounter primarily text-based content in the form news (TV2, DR), dating (Scor.dk), discussion and chat forums (Opasia), and multipurpose portals. Later, we see a significant increase in the existence of audio and video content on the top listed sites’ front pages: from just 14% of the sites displaying audio-visual content in 2002 to 75% in 2019, with actors in streaming, social networking, and user-generated content such as Netflix, Facebook, and YouTube topping the charts.

By 2012, online advertising revenues levels out the revenues made from non-digital advertising, and by 2019 they have by far surpassed them (Slots- og Kulturstyrelsen, 2019). In terms of making money from online advertising, third-party cookies are essential. Figure 6 shows how, in 2002, the Tracker tracker finds just 13 cookies in total, 9 of which were owned by the US legacy media measurement company Nielsen on the Danish top-10 websites. In 2012, depicting an online advertisement gold rush, the number of unique cookies multiplies to 71 – 10% of them owned by an, at that point, emergent data-reseller giant: Google (Alphabet) – and the rest owned by individual, smaller actors. In the period from 2013 to the present, we see a manifest market consolidation in the form of buy-ups of smaller enterprises. As such, the 2019 top-10 has just 31 unique cookies, 40% of which are owned by Google. Had we looked at the entire web, the number of cookies would of course only have increased alongside the
explosion of websites, this analysis, however, testifies to the consolidation history of an emergent and extremely influential market (Zuboff, 2019).

Generally, political regulation of digital content is struggling to make an impact: Due to the scale of the web and the countless pages of content, it is, to say the least, difficult to enforce any actual policies, and as consequence, market
regulation prevails when it comes to what is deemed offensive or dangerous content or how freedom of speech is interpreted. In first period from 1992 to 2002, the existing policies regulating public service content, media liability, copyright, and freedom of expression are relatively straightforwardly extended to include online utterances, since digital content in this period reproduces existing communication forms and is controlled by brownfield institutions. From the mid 2000s, the established content regulation is challenged as the most popular sites do not fit into the existing businesses and sectoral definitions in Denmark, and contain content that also does not fall into the established policy categories (e.g. news, TV). A key example of the growing insufficiency of existing policies is the efforts to regulate the application and data markets that cut across and disrupts a wide range of sectors (Larsson & Vetter, 2018). Despite various EU initiatives like the eCommerce directive and the General Data Protection Regulation (GDPR), the increasingly mature and consolidated data economy continues to present massive challenges to both national legislation and global governance.

From socio-technical transfer to digital momentum

Having analysed the technological, market, and policy development of important infrastructural resources across the three periods, we now move to a synthesis of the analysis as it relates to the phases\(^5\) in the evolution of large technological systems (Hughes 2012): the transfer phase from 1992-2002, where the internet infrastructures are developed and transferred across the country, market actors approach the economic opportunities that the internet affords, and policy makers identify an emerging field; the growth phase from 2003-2012, where critical problems pertaining to internet capacity and speed, market strategies, and policy making are addressed, and the internet grows as a result of popular demand; and the momentum phase from 2013-2019, where the internet is irreversibly integrated into various social institutions and practices and becomes a regulating force in

\(^5\) We leave out Hughes’ first phase, invention, here. Not because it is not significant to a historical analysis of the internet, but because it relates to a different time and different place: namely the development of the ARPANET in late 1960s’ US.
its own right. The following sections synthesize the analysis into a number of conclusions that characterize the internet’s history in Denmark. These conclusions are summarized in key terms, marked with italics, for each phase and finally collected in a table.

**Transfer phase: The welfare state goes online**

Legislation and market are critical factors in transfer and adaptation, but there are other factors involved, including geographical and social ones (Hughes, 1983, p. 60).

When the first private internet connections were introduced in the early 1990s, different medium specific technologies and infrastructures supported different sectors, business models, and services, that were regulated by different policy frameworks (Bar & Sandvig, 2008). Hence, the existing communications infrastructures were controlled by national institutions subject to different types of welfare regulation that guaranteed all Danes access to electricity, a phone line, and broadcasting services. While transfer and adaptation processes can be challenging, these preconditions made the integration of the internet technology, which was originally developed in and for an American context, fairly easy: Internet connections could be established on the basis of existing infrastructures (e.g. the extensive copper landline telephone network, the cable TV system, etc.); strong national institutions had the financial muscles and economic inclination to position themselves online; and existing policies encouraged equal access to basic supply services.

However, the reliance on relatively slow dial-up connections meant that the internet in this period was a supplement to the existing communication technologies rather than an actual alternative. In the early years of digitisation, dedicated distribution forms still dominated as landline networks, and increasingly also mobile networks, mainly supported the use of telephones, while various types of TV-distribution networks (terrestrial, coaxial, satellite etc.) supported the use of TV-sets. The internet was still not universal, and the connections were unsuited for TV and other high-capacity services. Lastly, digital content and applications
reproduced the existing communication genres (e.g. news and other types of text- and image-based communication like contact ads, phonebooks, etc.).

Digital markets for internet connections and websites were emerging in the period from 1992 to 2002, but they generally reflected historical power structures and institutional frameworks: National, brownfield actors ventured into the digital market as reflected in the strong positions of for instance TDC and legacy publishers. The internet was generally seen as an opportunity to expand existing businesses digitally without jeopardizing established market positions. Since supplying digital services was quite costly, entry into the market was guarded by high barriers and existing market structures were stable. However, the internet’s potential for breaking up old monopolies and disrupting existing power structures was beginning to show.

The fact that established institutions controlled large parts of both digital and analogue communication flows, meant that the existing, historically rooted policy regimes were relatively intact. In the period, existing welfare state policies were expanded to also account for digital technologies, the internet was generally seen as a common good and the role of the state was that of a generally developmental and positive state (Henten & Falch, 2014): public institutions, including the universities (Leiner et al., 2009), built and managed the earliest access networks, backbone infrastructures, websites, as well as encouraged and provided funding for the production of digital content. As digital content across the top 50 sites in this period reflected the legacy genres, inscribing it into existing legislation was relatively uncomplicated. The strong legislative power, however, was challenged by the difficulty in targeting emerging actors such as search engines and the budding realm of online chat.

**Growth phase: Convergence and expansion**

*As technological systems expand, reverse salients develop. Reverse salients are components in the system that have fallen behind or are out of phase with the others (Hughes, 1983, p. 67).*

By the second phase, the internet had been transferred to all regions and sectors and it was commonly used by Danes for multiple purposes, which in turn created
a set of new problems. Using Hughes’ terminology, various types of reverse salients emerged that have turned out to be crucial to the continued digitisation, as the solutions came to determine the progression and use of the internet. A key challenge to the growth of the internet in the 2000s was the limited capacity of early internet connections. While web services giving access to games, video, etc. developed quickly, the dial-up and early ADSL-connections constrained the capacities of regular internet-users to fully take advantage of these possibilities. The introduction of new types of broadband such as cable TV, fibre, and mobile (all initiated in the second period from 2002 to 2012) addressed this issue, and especially the development of LTE/4G represented a key breakthrough enabling the increasingly omnipresent use of OTT services. As the potentials of the internet began to show, another challenge appeared as the original anonymity that characterized the early web architecture did not support the functionality of for instance logins, measurements, online shopping carts, etc. Once the commercial potentials of the web began to dawn on businesses, the development of cookies became a crucial innovation that has since then significantly shaped the digital market and political regulation thereof. A final example relates to the initial challenges in the development of online TV. Unlike the functionality of traditional broadcasting networks, content distributed through the internet has to be sent individually to all receivers. This means that the network is in high risk of congestion – especially in the case of high capacity audio-visual content. Apart from the development of increasingly strong access networks, the success of OTT TV also rested on content delivery networks (CDNs), that were designed to take the pressure off the internet backbone by establishing local or regional data caches where content could be stored close to the end-users (Helles & Flyverbom, 2019; Sandvig, 2013). So, while traditional TV broadcasting peaks in terms of viewership in 2011, the infrastructural foundation for the rich streaming environment as we know it today is established by the end of the growth phase. Taken together, the three innovations all contributed to the development of an increasingly digitised and convergent communication ecology where the internet became an alternative to all existing distribution technologies. Communication genres were increasingly being restructured as for instance social media and streaming enabled the users to cut across former boundaries of interpersonal and mass as well as synchronous and asynchronous communication (Bar & Sandvig, 2008).
These infrastructural challenges – and solutions – characteristic of the growth phase are also manifest in the market developments: Dominant actors like Google and Facebook based their positions on data harvesting and reselling, contributing to the establishment of data as a new currency as well as massive and escalating tracking and advertising; parallel with the mobile broadband development, mobile device manufactures such as Apple and Huawei as well as operators such as 3 entered the Danish market; and streaming players like YouTube and Netflix were (and are) highly dependent on – and contributed to the development of – CDNs and other technologies. For the strong, brownfield actors, digital undertakings were still seen as a profitable add-on to their existing business, but the low barriers to entry into the expanding digital market meant that a multitude of digital market actors surfaced, and so we begin to see a destabilization in the legacy power balance. Consequently, the digital market was in a phase of rapid expansion, where international market actors increasingly challenged national institutions originating in historical sectors.

The changing market conditions as well as the infrastructural restructuring also influenced the scope and efficiency of existing legislation. The governance logic was generally one of liberalisation as the expanding markets were seen as creating a prosperous competition environment, reducing the need for state intervention. The state therefore played a regulatory – as opposed to developmental – role where state authorities intervened only when the market structures were considered problematic. At the same time, the restructuring of communication genres and the introduction of new communication channels (e.g. social media) challenged the legislative power as new market actors and services do not fit into existing categories.

**Momentum phase: New order**

*Old systems like old people tend to become less adaptable, but systems do not simply grow frail and fade away. Large systems with high momentum tend to exert a soft determinism on other systems, groups, and individuals in society (Hughes 1983, p. 48).*
By the end of the third phase, the internet is fully integrated with Danish society and it reaches momentum as the dominant infrastructure in the communication system. In other words, in much the same way as the electrification described by Hughes, the material characteristics of the internet – the end-to-end architecture, the global nature of digital distribution, and the code that breaks down content into 0s and 1s – have transformed how people communicate. Over-the-top solutions increasingly substitute all former distribution systems – landline telephone lines are cut and traditional broadcast TV viewing decreases steadily, while all IP services are no longer experimental but increasingly standardized. That is, the conventions and practices tied to internet-based communication gradually influence more and more social situations.

Since the gold rush expansion of the digital markets, survival-of-the-fittest and buy-up processes have led to a consolidated digital market with a number of global giants dominating the new order: global web actors that were earlier dependent on regional access and backbone network actors, are now increasingly backbone providers themselves; device and software manufactures become powerful gatekeepers controlling the access to for instance the important app-market; and the pervasive data economy entails that large data resellers become indispensable to all industries that operate commercially online. The low barriers to entry have thus been replaced by walled gardens wherein the largest actors own components across the value chain.

This development means that the failing efforts of establishing a holistic internet policy regime that can include global, greenfield infrastructure actors become increasingly obvious to policy-makers. The fact that digital communication systems are highly commercial and regulated by market forces rather than by welfare ideology creates an institutional crises that call for basic policy discussions and more open approaches to regulatory reforms. However, the maturity of the digital system complicates these regulatory efforts as the business models and power structures are built into the material infrastructure. As a result, the role of the state seems to be rather reactive as efforts to regulate dominating market actors come at a time when their business strategies are established and competition structures are difficult to break up. While especially the EU continuously struggle to amend existing legislation and regain power from American tech-companies, the regulatory regime of the Nordic welfare states (Syvertsen et al.,
2014) has lost its previous power over the structural development in the media and communication sectors. As Danes are increasingly using services and websites supplied by global companies, the effect of public service, telecommunications regulation, and so on is weakened. Furthermore, Danish authorities (as opposed to e.g. their German colleagues, see Bundeskartellamt, 2019) have not yet made any efforts at extending competition regulation or market monitoring to digital market actors that operate from outside Denmark, but dominate Danish markets. While especially EU initiatives are likely to influence the structural conditions also for Danes, internet-based activities are at the moment controlled and monitored significantly less than the legacy media they increasingly replace. As such, the legislative power and effective regulatory influence is increasingly undermined.
Conclusions

Table 1 synthesizes the conclusions in the previous sections as key terms for each of the three phases described above.

<table>
<thead>
<tr>
<th>INFRASTRUCTURE</th>
<th>EVOLUTIONARY PHASES</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Materiality</td>
<td>Medium specific</td>
<td>Convergent</td>
<td>All IP</td>
</tr>
<tr>
<td>Network distribution</td>
<td>Dedicated</td>
<td>Digitalised</td>
<td>OTT</td>
</tr>
<tr>
<td>Internet status</td>
<td>Supplement</td>
<td>Alternative</td>
<td>Dominant</td>
</tr>
<tr>
<td>Communication genres</td>
<td>Reproduced</td>
<td>Restructured</td>
<td>Standardised</td>
</tr>
<tr>
<td>Digital market</td>
<td>Emergent</td>
<td>Expanding</td>
<td>Consolidated</td>
</tr>
<tr>
<td>Ownership</td>
<td>National</td>
<td>International</td>
<td>Global</td>
</tr>
<tr>
<td>Business entry</td>
<td>High barriers</td>
<td>Low barriers</td>
<td>Walled gardens</td>
</tr>
<tr>
<td>Power balance</td>
<td>Stable</td>
<td>Destabilised</td>
<td>New order</td>
</tr>
<tr>
<td>Policy regime</td>
<td>Welfare state</td>
<td>Liberalisation</td>
<td>Commercial</td>
</tr>
<tr>
<td>Ideology</td>
<td>Common good</td>
<td>Competition</td>
<td>Crisis</td>
</tr>
<tr>
<td>Role of the state</td>
<td>Developmental</td>
<td>Regulatory</td>
<td>Reactive</td>
</tr>
<tr>
<td>Legislative power</td>
<td>Strong</td>
<td>Challenged</td>
<td>Undermined</td>
</tr>
</tbody>
</table>

Table 1: Synthesis of phases in Danish internet history.

As the table and the analyses illustrate, Hughes’ phase model provides a useful theoretical framework for explaining the evolution of the internet in Denmark along with the general Danish digitization processes - especially when combined with an analytical strategy that focuses on infrastructure, markets, and policies. It confirms Hughes’ middle ground understanding of the socially constructed yet also determinant evolution of large technological systems and shows how the early introduction of digital services reflects socio-political and economic conditions existing prior to digitisation, while the more mature internet infrastructure
challenges institutional structures and regulatory logics of the past. The study has two main research contributions: First, it is an example of how the momentum theory can be operationalized and used for analysing and understanding the changing power structures and regulatory challenges that follow as the internet grows into a mature and dominant infrastructure. Second, it fills a gap in media and communication research by analysing a critical period in Danish internet history.

However, there are also a number of inherent limitations of the theoretical and analytical strategy: The focus on Denmark makes it difficult to draw more general conclusions on the evolution of the internet in other national settings or on a global scale. The highly developed communications infrastructures and strong market actors existing in Denmark before the internet as well as the general socio-economic conditions have surely had an impact especially in the transfer phase. In a US context, for instance, the conclusions would most likely be different due to the regulatory traditions and the fact that the internet technology as well as dominant market actors to a wide extent originate in the US. Nonetheless, we believe that the Danish case will have similarities with especially other Scandinavian or European countries and that it is a starting point for other single-country studies or comparative analyses testing the generalisability of the conclusions made here and identifying variations in the evolution of the internet in different societal settings (see e.g. O’Hara & Hall, 2018).

Our focus on the macro-structures and environmental conditions means that we pay less attention to how individual media users actually behave, how they make use of the digital opportunities (or do not), and how the gradual shift to internet-based communication is related to socio-cultural traditions. Our periodisation highlights the infrastructural ‘points of no return’ and argue that the economic and political disruptions should be understood in light of these. An analysis focusing on for instance the changes in everyday communication would most likely make alternative periodisations (e.g. emphasising the introduction of Facebook, YouTube, etc. around 2006). While acknowledging that other research approaches could identify relevant phases and turning points, we also believe that these studies can benefit greatly from an infrastructural perspective highlighting the broader digital ecosystem and material conditions that paved the way for, amongst others, large social networks sites. Just as the dominant market
positions and business models of such actors have wide-ranging consequences that can only be fully understood when applying both a vertical and horizontal perspective that cuts across traditional sectors.

**References**


Bundeskartellamt. (2019). Bundeskartellamt prohibits Facebook from combining user data from different sourcesBackground information on the Bundeskartellamt’s Facebook proceeding. https://www.bundeskartellamt.de/SharedDocs/Publikation/EN/Pressemitteilungen/2019/07_02_2019_Facebook_FAQs.pdf?__blob=publicationFile&v=6


Danmarks Statistik. (n.d.-a). FABRIT01: Familiernes adgang til pc og internet i hjemmet efter type, region, indkomst, antal børn (pct af familier). År: 2008-


Publication

New Media & Society, 2020
DOI: http://doi.org/10.1177/1461444820932868

Authors

Rasmus Helles, Stine Lomborg & Signe Sophus Lai
Infrastructures of Tracking

Mapping the ecology of third-party services across top sites in the EU
Abstract

Today, websites operate in a modular fashion, outsourcing the surveillance and datafication of users to outside companies, along with security functions, video hosting etc. These third-party services (TPSs) function as key enablers of the web, with respect to functionality and the monetization of user activity. Departing from critical data studies and media systems analysis, the article contributes to understanding TPS infrastructures by placing these in a wider context of markets, cultural differences and regulation. Through a study of top-150 websites from the 28 EU countries, the article demonstrates how the use of TPS services varies between different parts of the region and different types of sites, and traces this variation to issues of language, regulatory traditions and differences in online businesses. These insights may inform current debates about surveillance capitalism and big data, by linking different forms of commodification of users’ behavioural data to broader social and cultural structures.

Keywords

Big data, third party services, media systems theory, tracking infrastructures, online tracking, web spheres, ad tech, political economy, critical data studies, surveillance capitalism.

Introduction

Contemporary scholarship of digital media reflects a deep sense of worry about the so-called data industries which thrive on many-to-one communications (Jensen and Helles, 2017) – the collection of the communications of many users in digital systems, often through third-party services (TPSs) for aggregate level ‘big data’ analysis, user profiling, predictive marketing etc. (Falahrastegar et al, 2014a,b, 2016, Libert, 2015). A key critique of data-based business models revolves around the possible suppression of minority voices, surveillance practices and privacy infringements they entail (e.g. Harper, 2017; Baruh and Popescu, 2017), and this critique has led researchers to call for a stronger ethical commitment from global companies, e.g. by requesting transparency and measures of accountability in companies’ uses of data, increased user control with stored
information, and more detailed privacy settings on social media sites. While the exploitative practices of online tracking and profiling are a general source of worry, both media users and policy makers have struggled to find proper responses to keep checks-and-balances with the data industries. The General Data Protection Regulation (GDPR) implemented in EU in May 2018, so far represents the most comprehensive of such efforts to protect the privacy of EU citizens, of which we have yet to see the effects. One plausible reason why it has been difficult to mobilize collective responses to datafication is that the TPS infrastructures through which datafication operates are largely understudied and poorly understood in both scholarship and public debate.

This article aims to chart key tenets of infrastructures for datafication to sensitize research to the myriad theoretical and empirical implications of the widespread use of TPSs. Whereas earlier scholarship typically focuses on the global players (such as Facebook or Google) in the TPS market, or the importance of the advent of TPS based advertising for the historically and democratically important news industry, our starting point is the web as a whole. From the perspective of the market that the TPS infrastructure supports, it matters little which kind of site or business sector is involved: what matters is web traffic. If a site has traffic, its audience can be commodified, and different sites may leverage data for user commodification in different ways through the embedding of TPSs (Zuboff, 2019). It is a crucial empirical challenge to examine these market dynamics and their nuances at the baseline level of trackers. Hence, in this article, we present an empirical analysis of how the TPS infrastructure is mobilized in the top 150 websites across the 28 EU member countries. Even if Google and Facebook dominate the playing field of datafication, we show that the TPS infrastructure has a highly complex internal structure. Through a network analysis of the relationship between websites and their use of TPSs, we first demonstrate that websites in the EU are characterized by how they mobilize the TPS infrastructure. Second, our analysis testifies to a strong regionalization of the patterns of TPSs across websites in the EU. We suggest that this regionalization reflects more fundamental societal characteristics, specifically the national political and economic dynamics that have contributed to shaping the development of national web spheres during the past 25 years (Foot, 2006; Foot and Schneider, 2006). Finally, our shift of empirical focus from the top internet companies or news
industries to the web infrastructure as a whole—and the analytical insights that may be gleaned from this shift—indicate the need for rethinking media systems theory. If we are to explain the infrastructures for datafication in the digital media ecology, we may be better served by thinking of these not as media-specific infrastructures, but as general communication systems that reflect the societies they help to constitute.

**State of the art and rationale for the study**

**Third-party services (TPS)**

In recent years a substantial body of research has examined the voluntary, self-initiated tracking through digital media (Kristensen and Ruckenstein, 2018; Lomborg et al., 2018; Lupton, 2016; Nafus, 2016; Sharon and Zandbergen, 2017). In contrast, empirical research in the field of media and communication on web tracking carried out by TPSs has been sparse, even if it constitutes a key tenet in contemporary infrastructures for datafication, and has profound implications for the political economy of digital media. This is presumably, in part, due to the fact that infrastructure of TPSs is linked to complex nature of computer networks.

‘Third-party’ services are not operated by the website visited by the user. The source code of a website visited by a user (e.g. a news site) may embed calls to other servers, which provide the so-called third party services. They frequently do this in a way that is undetectable for the user. So, for example, the TPS may place a cookie on the user’s computer, or it may provide content such as images or video. The latter function is often used by news sites that offer streaming video content, mainly because the server infrastructure required to provide these services to many concurrent users is expensive and complex to operate (Libert and Nielsen, 2018).

The array of functions that can be provided by TPSs is vast. The hosting of content and constituent parts of a website such as images, graphics, sound, fonts, style sheets etc. are frequent uses of TPSs. So are functions for monitoring traffic on a site, for example to detect anomalies in user activities and trace them to
security violations or flaws in the design of the web interface. Even more common are the outright tracking services aimed at identifying and monitoring users, which make up the bulk of TPSs installed on the sites in our sample. These are part of a complex ecosystem of online advertising relying on big data.

In the following we take all potential functions of TPSs to be of interest, and shift the analytical focus towards the way websites integrate TPSs in their site architectures. We look for characteristic patterns in the way sites mobilize different parts of the TPS infrastructure, with the aim of identifying key determinants which shape these infrastructural choices.

In order to situate our study in a wider field of inquiry, we base it on research from three specific domains: a) research on datafication associated with so-called critical data studies and political-economy analysis of data flows and governance; b) computer-science based charting of online tracking, often with a normative goal of preventing or circumventing tracking; and c) media systems analysis, addressing policy and regulation issues.

**Critical data studies**

The emergent field of ‘critical data studies’ has developed in response to the hypses and seductive discourses of big data in research and society alike (Boyd and Crawford, 2012; Dalton, Taylor, and Thatcher, 2016). While loosely connected and interdisciplinary by nature, critical data studies are commonly undergirded by a deep worry about datafication and its possible outcomes as well as people’s apparent inability to produce appropriate responses to being tracked. Some problematize the commercial aspects of datafication: the commodification of users by way of using their data input to profile, categorize, and serve them to advertisers and other stakeholders (Beer, 2016; Couldry and Yu, 2018; van Dijck, 2014; Elmer, 2004). Others direct attention to exposing surveillance mechanisms and structures that not only pose threats to individual and group privacy, but also involve issues of potential discrimination through social sorting (Kitchin and Lauriault, 2014; Reigeluth, 2014). Moreover, the umbrella of critical data studies encompasses research into data ethics, including philosophical perspectives on the ontologies and epistemologies of data (Floridi and Taddeo, 2016; Iliadis and Russo, 2016), and research and interventions regarding
accountability in IT systems and as part of company CSR strategies (Ananny and Crawford, 2018; Metcalf and Crawford, 2016), as well as mobilizing civil society actors and alternative pathways to data justice (Dencik, Hintz, and Carey, 2018; Milan, 2018). Critical research into datafication has also been pursued in infrastructure and data governance studies (Flyverbom, 2016; Just, 2018; Kitchin, 2014) concerned with issues of power, regulation and communication policy. This latter field has affinities with political-economy analyses of digital media (Hardy, 2014; Fuchs, 2017; Mansell 2004; Mosco, 2009, 2014).

The critical research agenda pursued by these strands of research has contributed important theoretical insights into the politics and epistemic biases of technology, advancing new concepts and ways of positioning datafication as a political-economic process that calls for collective responses. While some empirical work has been pursued, mostly on specific platforms, user groups or contexts, and specific regulatory issues, there is a need for more coherent empirical work to substantiate critique and qualify political and regulatory responses to questions of datafication. The lack of an empirical baseline is perhaps linked to an often alluded-to assumption that infrastructures are invisible. For instance, algorithms are seen as ‘black boxes’ (Pasquale, 2016), the workings of which are not transparent to most people, who can only imagine their ways of operating (Bucher, 2018; Turow, 2013), thus hampering the production of empirical evidence on how they work, and how to regulate them. Bowker’s (Bowker, Baker, Millerdand, and Ribes, 2010) influential work in this area in fact proposes invisibility as the defining parameter for distinguishing infrastructures from other technological assemblages. We contend that although they might go unnoticed, be intransparent and taken for granted, it is in fact possible to make key aspects of the infrastructure the object of empirical analysis, including ownership structures, TPS types, co-occurrences of TPSs, etc. These infrastructure aspects may be seen as valid indicators of the wider infrastructural arrangement of web spheres, their political-economic logics, and its influence on the organization of society.
Tracking trackers

An important body of work that seeks to map the tracker portion of the TPS infrastructure largely stems from other fields of research. Computer science-based inquiries have approached online tracking empirically through either large-scale mappings that are methodologically similar to ours (Libert, 2015; Roesner et al., 2012; Falahrastegar et al., 2014; Krishnamurthy and Wills, 2009) or through studies of smaller user samples installing client-side software (Falahrastegar et al., 2016; Puglisi et al., 2015). These studies often share the normative aim of enabling the individual user to circumvent tracking through the development of tracker blocking tools (Kalavri et al., 2016). Many also work towards developing TPS typologies in order to distinguish between “good guys” and “bad guys” in online tracking (Roesner et al., 2012; Falahrastegar et al., 2014). While studies in this vein typically focus on national or regional top lists of sites and their associated trackers, some take as their point of departure individual, existing TPSs in order to evaluate their prominence. A longitudinal study (2005-2008) of the top third-party domains across 1200 sites and their acquisitions by parent companies (Krishnamurthy and Wills, 2009) found an increasing number of TPSs aggregating user-related data, which are owned by a steadily decreasing number of companies. Lastly, case studies of specific sites or groups of sites map the TPS ecology of particular sectors, such as all Dutch governmental sites (van der Velden, 2014) or sites such as The New York Times site in a historical perspective (Helmond, 2017). Perhaps owing to the normative commitment of these studies to reverse engineer tracking as a basis for developing circumventing measures, the research findings produced in this strand are largely descriptive, and lack a coherent theoretical commitment to understanding and explaining the TPS patterns as part of larger socio-cultural systems.

Albeit focusing on online news media, some recent studies share with us the aim of understanding the role of TPSs in the web and mobile ecosystems in a wider sense (e.g. Lindskow, 2016; Libert et al., 2018). In particular, Timothy Libert’s work (Libert, 2015; Libert and Nielsen, 2018) on TPSs, is similar to ours in scale and methodology, and we use the WebXray software (https://webxray.org) developed by Libert to collect data on TPSs. However, our aim of linking
differences in TPS formations to wider social and cultural structures differs from the aims of these studies.

**TPS ecologies and the web as a communication system**

We draw initial assumptions and ideas about the ecology of TPSs from media systems analysis (Hallin and Mancini, 2004; Brüggemann et al., 2014). However, this field is strongly tied to mass media and specifically the traditional news industry (print, radio, television) – the regulatory and business models of which have consistently failed to demonstrate transferability to the internet, prompting a blurring of the regulatory domains of media and communication (e.g. Just, 2018). Broadening the scope of media systems analysis, we contend that there is an urgent need to pursue research that considers the entire digital media ecology. Switching the conceptual lens from media to communication, we are able to see how municipalities, banks, NGO’s, retail sites, etc. have become media on a par with news websites, social media, streaming services and so on. They are media, because they channel and orchestrate online communication with users and between systems and users (Jensen and Helles, 2017, Lomborg and Frandsen, 2015) – and thus generate traffic and massive amounts of data that is useful for a wider market.

Following the conceptual lead of communication, we focus on web spheres across the EU as spheres of communication that reflect broader societal arrangements: we take the web spheres as indexes of the degree and character of the digitalization of a given society. For instance, the strong presence of public institutions in the web sphere of Denmark reflects not only the almost full diffusion of the internet in Denmark, but also a consistent political commitment to digitalization of society. If the web sphere is an index of society, our empirical study contributes to the understanding and theorizing of the intricate relationship between digital infrastructures and society. Focusing on the most used sites across all the European member nations and the totality of TPSs existing on those sites, we chart a system consisting of both big and small players in the TPS business framed by a theoretical interest in what can be learned about contemporary digital infrastructures and societies from the empirical study of online tracking.
Methods and dataset

The dataset comprises the TPSs found at the top 150 sites for each of the 28 European member states. The list of trackers was found by traversing the top lists for all countries and harvesting metadata about all TPSs that were loaded by each site using a multistep scan with the Webxray tool during June 2018.

The aggregated top-150 lists were extracted from the Alexa database and nominally comprise a total of \( N = 4,200 \) sites. Yet due to site duplication between some (ikea.com is on 18 lists) or all the lists (e.g. google.com), the number of unique sites in the sample is \( N = 2,150 \).

The harvested information about sites and their TPSs was used to form a network structure, where nodes represent the sampled websites, and where the presence of a link between two sites shows that they have at least one TPS in common. The strength of the links in the network represents the total number of TPSs shared between two sites.

The network was obtained by transforming the data about sites and TPSs from its original, two-mode network form to the ordinary, one-mode network analysed below (Luke, 2015). This required the omission of a number of the most common TPSs from the dataset before the transformation: Since any two sites are linked if they share a TPS, the presence of highly common TPSs (such as e.g. Google Analytics) in the dataset means that nearly all sites in the final network are linked if the prevalent TPSs are left in. This renders common network analytical tools such as community detection ineffective (Alupoaie and Cunningham, 2013), since the highly interconnected nature of the network obscures more subtle structures in the grouping of sites. To avoid this, the 50 most common TPSs were omitted from the dataset prior to transformation. The number was arrived at by conducting a stepwise comparison of the clustering coefficient of the resulting network against the number of omitted TPSs, aiming to achieve a stable clustering and keeping the number of omitted TPSs as low as possible.

The resulting network was subsequently subdivided into clusters using a modularity class analysis (Blondel et al. 2008), which found 35 clusters, ranging in size from 244 to two sites, with the median size \( \text{Mdn} = 14 \).
In the following sections we report the findings of the analyses of the network clusters. However, we begin the analysis by looking at the characteristics of the most popular TPSs, which were excluded from the network analysis.

Findings

Ownership of popular TPSs

The frequency of TPSs in the initial dataset is very unevenly distributed and reflect a classic longtail pattern with the global advertising conglomerates Google and Facebook playing a very dominant role. Table one shows the companies that own the most common TPSs. Even if some of the services provided by these companies (such as e.g. Google Analytics, which offer webmasters an analysis of traffic patterns on their sites) are not explicitly aimed at tracking audiences, they are part of corporate ecosystems which allow an exceptionally fine-grained insight into peoples’ online behaviour.

<table>
<thead>
<tr>
<th>Company</th>
<th>Sites with tps (n = 2150)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>86</td>
</tr>
<tr>
<td>Facebook</td>
<td>48</td>
</tr>
<tr>
<td>Adform</td>
<td>24</td>
</tr>
<tr>
<td>Appnexus</td>
<td>23</td>
</tr>
<tr>
<td>Gemius</td>
<td>19</td>
</tr>
<tr>
<td>Amazon</td>
<td>18</td>
</tr>
<tr>
<td>Mediamath</td>
<td>17</td>
</tr>
<tr>
<td>Rubicon project</td>
<td>17</td>
</tr>
<tr>
<td>Ipon web</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 1. Ten most prevalent TPS owners

Together, the top TPSs are extremely common across the sample. We also see regional services at this level: A prominent example is the Polish company Gemius, whose various TPSs are found on many sites in the sample, in particular
in the Eastern European countries. For some companies, such as Polish Gemius and Danish Adform, the uptake of the companies’ arrays of TPSs means that they rank among the top 10 TPS providers in Europe. Despite the consistent effort by Google in buying up competing companies, there is still some degree of differentiation even among the top sites. Beyond the 50 most popular services, the remaining 9,027 are present on between one and 206 sites in the sample, so the majority of TPSs are confined to substantially fewer sites. In other words, there is a whole wilderness of TPSs each playing some part in the big picture of the ecology. This suggests that even if the commercial tracking business is highly concentrated, many minor players maintain a market presence.

**Cluster analysis**

The clusters found in the analysis of the network are characterized by different constellations of TPSs. These, in turn, reflect different strategies in the way webmasters mobilize the available infrastructural resource that TPS providers deliver. Table 2 (below) shows the main characteristics of each cluster, including an assessment of which TPS are most distinctive of each cluster. The judgement of the distinctiveness of the various TPSs to each cluster was determined by calculating the tf-idf (Term Frequency Inverse Document Frequency) weight for each TPS per cluster (Alupoaie and Cunningham, 2013). For a given cluster, the weight is calculated by multiplying the frequency of the TPS in the cluster (the number of sites it occurs on) with the log of the ratio between the total number of clusters it occurs in and the total number of clusters in the data. This means that TPSs get a higher weight if they occur on many sites in a cluster, but are penalized if they also occur frequently in other clusters. We adopt this weighting scheme from computational linguistics (where it is frequently used to gauge the specificity of language used in large corpora of texts) as a way of identifying the centrality of TPSs to different clusters as a function of an active choice. Many TPSs can perform similar functions, so examining the choices of TPSs most specific to the sites that form each cluster reveals information about what has shaped the cluster formation. The top three most distinctive TPSs are listed, including the country of origin of the company owning the service (ownership and primary functionality of the TPSs were in many cases provided as part of the Webxray tool, or were researched by us where this was not the case). Table
2 also provides the three most frequent top-level domains (e.g. .com or .de) in each cluster, and the share of marketing/advertising oriented TPSs among the 10 most distinctive TPSs. Moreover, it provides a brief characterization of each cluster based on a manual inspection of the 20 sites with the highest eigencentrality score in the cluster. These are provided to give a sense of the characteristics of the individual clusters.

The complexity of the dataset means that we cannot provide a complete characterization of all clusters, neither with respect to the sites that constitute them, nor the TPSs that those sites have in common; the 2,150 sites in the sample are in many different languages, just as there are numerous TPSs. Our analytical strategy is to provide more in-depth analysis of a) the largest clusters, and b) the clusters with the most extreme TPS use patterns (e.g. the ones with the lowest and highest levels of TPS use). In this way the analysis can provide an outline of basic characteristics and pursue some of the central elements for understanding the underlying systematic that shape TPS use.
<table>
<thead>
<tr>
<th>CLUSTER NO.</th>
<th>NUMBER OF SITES</th>
<th>SITE CHARACTERISTICS</th>
<th>TOP 3 TLDS OF SITES</th>
<th>TOP 3 MOST DISTINCTIVE TPSS</th>
<th>TPS COUNT MEDIAN</th>
<th>MARKETING TOP 10 TPS PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>218</td>
<td>Portals (news, real estate) and tabloids newspapers. Eastern europe.</td>
<td>HU(41),PL(32),BG(31)</td>
<td>Yadro(ru),adfox(ru),rhythmone(us)</td>
<td>14</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>152</td>
<td>Local branches of transnational companies (zalando, arbnb, tripadvisor)</td>
<td>COM(18),EE(10),LT(10)</td>
<td>Intentmedia(us),tripadvisor(us),zalando(de)</td>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td>3</td>
<td>41</td>
<td>News sites, information services (dictionaries, encyclopedia).</td>
<td>COM (8), FR (5), PL(4)</td>
<td>Gemius(pl), rhythmone(us),dynadmic(fr)</td>
<td>34</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>70</td>
<td>News, chat sites, forums. Many russian sites.</td>
<td>RU(28),COM(14),EU(2)</td>
<td>Yadro(ru),adfox(ru), tns-counter(ru)</td>
<td>17</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>45</td>
<td>Classifieds, portals (cooking, weather, entertainment), tabloid news.</td>
<td>COM(6),PT(6),BG(5)</td>
<td>Blink(us),dynadmic(fr),admeta(us)</td>
<td>76</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>Accuweather national portals</td>
<td>COM(3)</td>
<td>Accuweather(us),bing_maps(us)</td>
<td>33</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>Finance information, cars, entertainment. Many eastern european.</td>
<td>CZ(20),SK(4),PL(3)</td>
<td>Bbememtus(cz),ströer(de),adscale(de)</td>
<td>25</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>22</td>
<td>Electronics sales, aggregator sites (jobs, prices, delivery services).</td>
<td>COM(3),FI(2),GR(2)</td>
<td>Rambler(ru),ligatus(de),yieldlab(de)</td>
<td>36</td>
<td>80</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>New, lottery, gambling.</td>
<td>COM(3),BE(2),PL(2)</td>
<td>Netseer(us),switchads(uk),unruly(uk)</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td>10</td>
<td>45</td>
<td>Forum sites (reddit, bitcoin tracker, deviantart)</td>
<td>CZ(11),COM(10),SI(5)</td>
<td>Netmining(us),gumgum(us),skimlinks(uk)</td>
<td>37</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>Sport sites (german)</td>
<td>DE(3),COM(1)</td>
<td>Ströer(de),svemn(de),adality(de)</td>
<td>0</td>
<td>80</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>Greek news sites</td>
<td>GR(4),COM(1)</td>
<td>At internet(fr),elastic ad(),ebay(us)</td>
<td>80</td>
<td>90</td>
</tr>
<tr>
<td>13</td>
<td>244</td>
<td>Global forums / social media (linkedin, twitch, imgur, giffy); many sites are on several countries’ top listes.</td>
<td>COM(78),GR(41),IT(17)</td>
<td>Alexa(us),oracle(us),automattic(us)</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>41</td>
<td>Online booking (booking.com) and various sports sites.</td>
<td>COM(11),GR(6),DK(4)</td>
<td>Livesportmedia(mt),simplifi(us),ensighent(us)</td>
<td>29</td>
<td>60</td>
</tr>
<tr>
<td>15</td>
<td>180</td>
<td>Picture and movie sharing (pinterest, red tube), social media (twitter, whatsapp, messenger). Global sites, many on several country top lists.</td>
<td>COM(32),DK(13),SE(8)</td>
<td>Pinterest(us),siteimprove(dk),pornhub(ca)</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>Cluster Characterization</td>
<td>Country Codes</td>
<td>Service Provider Codes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook national sites.</td>
<td>NA</td>
<td>Facebook(us)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microsoft portals (msn, microsoft products), youtube. News sites.</td>
<td>COM(35), SE(7), GR(6)</td>
<td>Youtube(us), new relic(us), dynamic1001(us)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National news, also global diasporas.</td>
<td>COM(4), IE(2), NET(2)</td>
<td>Pusher(us), visualdna(uk), integral ad science(us)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online retail (amazon, ebay, ikea, asos, paypal) and streaming services (soundcloud, vimeo, spotify).</td>
<td>COM(32), FI(13), UK(10)</td>
<td>Akamai(us), live intent(us), adobe marketing(us)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German and austrian news and magazine sites + wikia fandom</td>
<td>DE(21), NET(4), AT(3)</td>
<td>Infinet(de), compact impact(de), xpsions(de)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek news sites</td>
<td>GR(2), BE(1), COM(1)</td>
<td>At inetnet(fr), elasticad(uk), ebay(us)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portals (news, jobs, mobile tech and cars).</td>
<td>COM(38), SK(18), HU(13)</td>
<td>Hotjar(mt), gemius(pl), visual website optimizer(in)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>News sites (southern and eastern europe).</td>
<td>COM(6), FR(2), LV(2)</td>
<td>Signal(us), nativo(us), wayfair(uk)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Portals (news, chat).</td>
<td>LT(2), LV(2), FI(1)</td>
<td>Infolinks(us), walmart(us), skimlinks(uk)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>National news (tabloids).</td>
<td>LT(2), COM(2), RO(2)</td>
<td>Run(us), gumgum(us), widespace(se)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish news sites</td>
<td>SE(5), COM(3), DK(1)</td>
<td>Research-ins(se), evidence(fr), linkpulse(no)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank and insurance sites (french)</td>
<td>FR(8), COM(3), AT(1)</td>
<td>Ab tasty(fr), crosssell(de), econda(de)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aliepress, news sites</td>
<td>COM(5), UK(2), ES(1)</td>
<td>Adobe(us), alibaba(cn), avocet(uk)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>French banks,</td>
<td>FR(2)</td>
<td>Eid(fr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swebank - national branches</td>
<td>EE(1), LT(1), LV(1)</td>
<td>Swebank(se)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xnxx, xvideos (porn video sites)</td>
<td>COM(2)</td>
<td>Traffic factory(cz), xvideo(pl), xnxx(pl)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish bank (caxiabank), local branches.</td>
<td>ES(1)</td>
<td>Olyx(za), mixpanel(us), akamai(us)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rarbg (illegal) filesharing</td>
<td>ORG(2), COM(1), IS(1)</td>
<td>Dyncdn(us), imagecurl(?), statsy.net(?)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weather site, local branches (hungarian)</td>
<td>HR(2)</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government sites, local branches (hungarian)</td>
<td>NA</td>
<td>Aldi-international(de), ask.com(us), etracker(de)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orphans</td>
<td>COM(34), HR(10), NET(10)</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Cluster characterization
**TPS use and site specialization**

Cluster 36 holds 239 sites, which were all decoupled from the other sites in the network when the top 50 TPSs were removed. The sites in this cluster make limited use of TPSs beyond some of the most common ones. A manual inspection of the sites reveals that about 40% of the sites are either banking sites or belong to public authorities (tax agencies, hospitals etc.) and educational institutions. There is also a low prevalence of marketing and advertising TPSs in this group compared to the other clusters. This indicates a tendency to not monetize user activity, consistent with the utilitarian objectives (providing digital forms and information) of these sites. The same pattern repeats itself in the largest cluster, cluster 35, which has a very similar TPS use pattern with a low, median TPS count (2), and a low share of marketing and advertising TPSs (10%). Although more heterogenous (the government sites are joined by Instagram and Netflix), this cluster consists of sites which would have been ‘orphaned’ if the threshold for TPS exclusion before the projection of the network had been set slightly higher.

The other extreme in terms of TPS use is cluster 9, which consists of 11 sites, with a median TPS count of 88. This cluster consists of tabloid news sites and lottery/gambling sites. In addition to the high TPS count the cluster also has 90% of its top ten TPSs in the advertising/marketing category, and represents the opposite from the orphan sites: These sites use a massive load of TPSs in order to monetize audience traffic and attention to the fullest extent. The most used TPSs in this group are almost exclusively used by the sites in the group, suggesting a degree of specialization on the part of the TPSs. However, the massive load of TPSs also means that the sites in this cluster are primarily characterized by being the ones in the sample that use any available marketing TPS. Specifically, sites with high traffic loads make money off audience attention by auctioning off the ‘eyeballs’ of visitors with relevant demographic profiles in real-time to advertisers. This works as an incentive to sites with a wide audience to participate in as many online auctions as possible, and therefore to amass the kinds of TPSs that allow them to do so.

A different approach to audience monetization is exemplified in cluster 19, composed of 129 sites. Among the members of this cluster are some of the sites that
are most popular across the 28 country top lists; the sites in this cluster have the highest average duplication for any of the larger clusters, with each site on average appearing on the top lists of 2.6 countries. This substantial duplication is in particular driven by the websites of large retailers, online payment providers and online shops (e.g. Ikea, Amazon, Paypal, Asos and Ebay). The median TPS count is 18, and the most popular TPS is from the content delivery network Akamai, which provides support for transfer of content such as images and video. The business model of some of the most prominent sites in this cluster does not depend on monetizing traffic by auctioning parts of the sites’ screen real estate to advertisers: Instead, it depends on identifying visitors to the site, and being able to link them to ads for this business, which they have been exposed to on other sites. So while sites like these do mobilize some of the potentials of the TPS infrastructure for tracking and identifying users, they are more interested in linking visitors to profiles of existing (or prospective) clients in their customer database for later re-targeting (Turow, 2017). The TPSs used by the sites in this cluster are among the most common in the dataset, yet the specific combination and relatively low use suggests that even if sites such as Ikea.com rely on sophisticated, data-driven online marketing techniques, there is a degree of specialization. Moreover, this cluster is relatively specific to countries in the North-western part of EU, perhaps indicative of the comparatively longer maturation of online marketing (and general affluence of users) in these countries.

**Regional specialization**

The pattern of regionalization is also found in two of the other large clusters. It is most pronounced in the case of cluster 4. The cluster has 70 sites, of which 48% of sites are unique to the three Baltic countries’ top lists. These sites make extensive use of a range of TPSs which are almost exclusively used by sites in this cluster, and eight of the 10 most distinctive TPSs for the cluster are Russian (e.g. Yandex and mail.ru). The nationality of the TPSs arguably indicates the characteristic of the audience that the websites provides to the TPSs: the Russian-speaking minorities of the Baltic countries. Hence, the TPS infrastructure mirrors the cultural and linguistic specificities of the particular demographic of the web users in
these countries. Other sites from the Baltic countries do not display the same pattern of TPS use.

The non-Russian language sites from the Baltic countries do not form exclusive clusters, but are part of more widely distributed formations. The most important one is cluster 1, the second largest cluster in the sample ($n = 218$), which has a substantial share of sites in countries all across the former Soviet countries in Eastern Europe. These sites are distinguished by their use of TPSs of Polish and Hungarian origin. The Polish-run Gemius service, and its subsidiaries Onet and Ocdn, are central to this cluster, and 83% of the instances of the use of Gemius in the entire sample are in this cluster. Ocdn is exclusive to sites in this cluster (12% use it), and Onet is used by 37%. Closely related to this cluster is cluster 7, which holds 36 sites from the Czech Republic, Slovenia and Poland. The cluster shares many TPSs with cluster 1, but is allocated to a separate cluster because the sites also make use of a range of specific TPSs developed by Czech internet companies, which provide for example chat plug-ins and authentication services. These services are easily found among the offers of a wide range of global TPS providers (Facebook and Google happily offer authentication services allowing users to login using their profile information), but the characteristic mobilization of TPSs in cluster 1 suggests that many Eastern webmasters prefer the local version.

Taken together, the different clusters specific to countries in Eastern Europe account for one in five sites in the sample. Figure 1 (below) is shaded according to the number of times sites from the clusters defined by TPSs from Eastern Europe are featured in national top lists and compares it to the largest cluster of western-oriented sites (cluster 19).
As indicated above, there are several factors that may contribute to the pattern of regional specialization evident in Figure 1. Some have to do with geographic and linguistic communities (Wu and Taneja, 2016), while others suggest path dependencies in the way the web business (both TPSs and sites) has developed in different parts of Europe. The latter is particularly relevant in understanding the distinctively Western cluster in the sample, which is perhaps more aptly characterized as non-Eastern. The sites in the cluster in the left panel of Figure 1 shows that a relatively large number of sites fall in a shared cluster of which very few members are found in countries in Eastern Europe. The cluster is, however, not defined by any specifically national or regional constellation of TPS use, but more closely represents what could be called the global state-of-the-art in audience management and monetization. The most common TPSs are exceptionally prevalent across the entire sample, but are even more common to this cluster; it is for example tracking and customer interaction services from Salesforce (found on many web shops), and audience tracking services by Adobe. The sites in this cluster integrate a modular website design which can scale smoothly to large traffic volumes, supplemented with a suite of audience monitoring and tracking services.
This suggests that web businesses in the western countries, which have historically seen an earlier adoption of the internet, have begun streamlining site architecture, aimed at mass production of audience data and standardized business processes.

**Discussion**

The present analysis testifies to a highly developed system of differential commodification of user activity through big data across countries in the EU28. Even if the major players dominate the commercial game of online tracking, the analysis suggests that they have not monopolized the market to such an extent that all other players have gone out of business. We see culturally and regionally specific TPSs coexisting with global, integrative giants. Our results also demonstrate site specialization through the mobilization of the TPS infrastructure in distinct ways, indicating that different kinds of value are derived from web traffic through TPSs according to business-specific and sector-specific uses. Together, this testifies to an omnipresent and many-sided system of collaboration between websites and TPS industry. The web is a coherent infrastructure that is at once standardized in its use of TPSs and contextually dependent in the sense that it develops in and reflects underlying societal forces. The analysis shows that patterns of TPS usage are linked to wider patterns of geography, demography, language, economy and internet diffusion. Despite the fact that TPSs are widely employed to derive value from audience behaviour, they are also constituent parts of an internet infrastructure which is to some extent shaped by social and cultural dynamics beyond the narrow interests of marketing and ad tech companies. While the dataset is limited to the top-150 websites, and we certainly do not claim to cover every possible player and TPS function in the market, we contend that looking further into the long tail of websites in the EU will enhance the findings of a highly complex infrastructure, where security, content delivery, marketing, re-targeting and surveillance are channelled through similar technical means.

Our findings have implications for key discussions in critical data studies, about surveillance, privacy and commodification through big data generated by way of online tracking of user behaviour. If TPSs are a core part of the infrastructure of
the web as such, the uses of TPSs as infrastructures for datafication is neither an unintended consequence of TPS use, nor an inevitable development. The ad tech sector at the centre of big data analytics has been built on top of the existing TPS infrastructure, and is not opaque, at least for those with technical knowledge of how digital infrastructures operate. Yet, it has grown and standardized in front of our noses, while the eyes of consumers, governments and companies have been fixed on business developments, efficiency, smart solutions and economic growth. The ubiquity of tracking is a key enabler in what Shoshana Zuboff has aptly dubbed ‘surveillance capitalism’ (Zuboff, 2019), which is seen as the core ideology driving digital developments. Surveillance capitalism is first and foremost represented by the ad tech sector, but, arguably, also by the integration of traffic monitoring technology in systems for security, content delivery and site analytics. Our study contributes to understanding the infrastructural nuts and bolts of the TPSs ecology through which surveillance capitalism works. And it presents a first step in probing the role of the market structure for the ad tech sector alongside TPS-based big data consolidation, brokerage and mining endeavours that rely on this infrastructure for datafication. Indeed, infrastructural insight may qualify the theoretical sensitivities in critical data studies to how datafication capitalizes on infrastructural mechanisms that are not all invisible, and that might enable different datafied futures provided that we regulate them differently.

Our results further show the merits of moving away from ideational delimitations of the publicist news industries from other business sectors in media systems analysis. As demonstrated, online tracking and big data do not come with unique characteristics or modes of operation for news websites alone, but are spread out across different constellations of sites. TPS ecologies are heavily integrated across business sectors spanning e-commerce, games, porn, etc. as well as news outlets. This, in turn, suggests an intensified and competitive battle for user attention and associated opportunities for collecting, mining and selling big data. Focused attention on TPS-based tracking as a coherent ecosystem – a communication system – helps us qualify further the myriad types of value generation taking place on the web as a key infrastructure for datafication, as linked to wider social and cultural processes, exemplified in the regional specialization of Russian TPSs aimed at the Russian minority in the Baltic countries.
Lastly, while our methodology and analysis only begins to open the doors to an analytically complex ecosystem of tracking, it devises a number of paths for deeper analytical dives, e.g., mapping TPS typologies onto the network of TPSs to better understand their possible division of labour in infrastructures for datafication, identify critical types of TPSs for consumer surveillance and commodification, use the clustering technique deployed here to identify specific TPS networks for close investigation, and for comparative analyses of TPS infrastructures – nationally and across continents. TPS infrastructures in the EU will likely look different from those in the US, China or in developing African markets, because of differences in regulatory environment, the degrees of state and commercial involvement in building digital infrastructures and so on. Also, current regulatory developments within the EU, such as the GDPR, might contribute to changing the breadth and structure of the TPS ecology in fuelling businesses and governments with big data.

**Conclusion**

In this paper, we have analysed how the TPS infrastructure is mobilised across the top-150 websites of the 28 EU member states, focusing on ownership, TPS specialization and regional variation. Our infrastructural approach contributes with important baseline information on ownership and market structures in the online tracking business to begin assessing the role and implications of big data in the attention economy.

The results of the analysis suggest several themes for further investigation, not least with respect to the relatively low degree of scholarly attention devoted to understanding the specificities and dynamics related to the topology of the current landscape of the web: The past decade has seen a substantial attention aimed at first social media, and later at privacy, surveillance and discrimination issues regarding platforms and algorithms. While both trends are immensely important, we argue that a broader understanding of the contextual factors shaping the web and the vast number of businesses that have developed around it is needed as well: A more full-fledged understanding of the ecology of tracking and TPSs, and how
they relate to wider logics of audiences and industry, can inform our understand-
ing of the risks as well as the actual influence of rapidly growing monopolies.

We hope our analyses, in turn, can inform legislative processes aimed at recasting various stakeholders, rather than individual users – going over pages of terms of service agreements and cookie notifications – as responsible for user privacy and data protection. Beyond national and international regulatory agencies, consumer organizations and NGOs, this would include private and private organizations who may need to provide stronger justifications for their TPS use. One candidate for downsizing TPS-based data collection would be websites from public agencies and organizations (e.g. national tax agencies, hospitals), whose operations are central to society as such, but who do not have stakes in the marketplace of attention. While we have documented a fairly limited TPS mobilization on such sites, they do enable other – typically giant – players to collect and monetize their web traffic through standardized solutions. We may want to ask, if the value for the second party generated through web traffic justifies tracking users at all.

**References**


Turow J. (2011) The daily you, how the new advertising industry is defining your identity and your worth, New Haven, N.J: Yale University Press.

Turow J. (2017) The aisles have eyes : how retailers track your shopping, strip your privacy, and define your power, New Haven: Yale University Press.


A Proxy for Privacy

Uncovering the surveillance ecology of mobile apps
Abstract

The article develops a methodological and empirical approach for gauging the ways big data can be collected and distributed through mobile apps. This approach focuses on the infrastructural components that condition the disclosure of smartphone users’ data – namely the permissions that apps request and the third-party corporations they work with. We explore the surveillance ecology of mobile apps and thereby the privacy implications of everyday smartphone use through three analytical perspectives: The first focuses on the ‘appscapes’ of individual smartphone users and investigates the consequences of which and how many mobile apps users download on their phones; the second compares different types of apps in order to study the app ecology and the relationships between app and third-party service providers; and the third focuses on a particular app category and discusses the functional as well as the commercial incentives for permissions and third-party collaborations. Thereby, the article advances an interdisciplinary dialogue between critical data studies, political economy, and app studies, and pushes an empirical and critical perspective on mobile communication, app ecologies, and data economies.

Keywords

mobile apps, applications, political economy, digital infrastructures, critical data studies.

Introduction

In a recent conference talk, executive vice-president of the EU-Commission, Margrethe Vestager (2019), encouraged the audience to “imagine your smartphone was your mother”. She continued, “the first thing you see each morning and before you go to sleep. It’s not that you’re doing anything she wouldn’t approve of – she raised you after all – it’s just that she would be there with you all the time”. Vestager’s analogy underlines that we would resist it if anyone were to monitor our every move to the extents that our smartphones do. It also focuses our attention to the underlying and comprehensive surveillance mechanisms of
mobile apps that go largely unnoticed as users scroll down the endless pages of terms of service (ToS) agreements only to click accept upon request. The ubiquitous problem of “informed consent” (Nissenbaum, 2011) creates a challenge for researchers and policy-makers alike, and calls for basic research into the various infrastructures and surveillance practices of mobile communication.

Despite smartphones being an integral part of everyday communication, we know relatively little about the data collection practices of different mobile apps (for exceptions see e.g. Atkinson et al., 2015; Binns et al., 2018). We address this research gap by developing a methodological and empirical approach for gauging the ways big data can be collected and distributed through mobile apps. More specifically, the article presents a study of the infrastructural components that condition the disclosure of smartphone users’ data – namely the permissions that apps request and the third-party corporations they communicate with. These components represent the two most significant gateways into understanding the ways apps collect and distribute user data. Based on a dataset consisting of 173 unique mobile apps that appear on the home screens of 10 Danish regular smartphone users, the article asks how we assess and map the intrusiveness of mobile apps.

We offer three empirical explorations that look at permissions and third-party services from the analytical perspectives of: 1) individual app repertoires (what we in this article name appscapes); 2) categories of apps (i.e. finance, sports, weather); and 3) similar apps within a category. This approach enables us to explore the surveillance ecology of mobile apps as well as the privacy implications of everyday smartphone use, and calls for empirical and critical data studies that make otherwise invisible infrastructures visible.

The article is divided into four parts: in the first, we position ourselves as part of the field of critical data studies that focus on apps, and outline the theoretical foundation for our methodological approach and empirical explorations; the second part describes the methods and discusses methodological issues surrounding the study of apps and (big) data; the third part applies the approach and presents explorative, comparative, and descriptive analyses from the perspective of individual appscapes, app categories, and particular apps; and in the fourth, we discuss the findings and future research prospects, and argue that increased public
awareness of the surveillance mechanisms of mobile apps can support policy-efforts to regulate the data economy.

State of the art and theoretical foundations

The article is placed at the intersection between distinct yet overlapping research fields that motivate an empirical and critical perspective on mobile communication, data, and privacy. When exploring the intrusiveness of mobile apps, we build on and contribute to existing research on digital privacy, critical data studies, political economy of communication, as well as the emergent field of app studies.

Infrastructures for/against privacy

Privacy studies often revolve around philosophical and ethical issues, and investigate how comprehensive and ubiquitous datafication of people’s everyday lives challenges and transforms established, yet increasingly hazy, conceptualisations of privacy (e.g. Cheney-Lippold, 2017; Mai, 2016). Other studies have focused on regular people’s understandings and strategies towards protecting their personal data and have for instance identified a disconnect between people’s concerns and their actual behaviour (commonly referred to as the privacy paradox (Barth and Jong, 2017)). Scholars have also concentrated on individuals’ experiences and imaginations of data and algorithms (Bucher, 2017). And human rights researchers (Jørgensen, 2019; MacKinnon, 2013) have raised alarm over violations by big data corporations working within a ‘governance gap’.

In this article, we address the questions raised by digital privacy research by empirically approaching the ways the extensive collection, processing, and distribution of data through internet-based services change the fundamental conditions for protecting and controlling one’s privacy. While acknowledging the influence of socio-cultural factors on how people (and apps) behave, we stress the importance of uncovering and understanding the underlying infrastructures that support the surveillance ecology of mobile apps. In doing this, the article follows the so-called ‘turn to infrastructure’ in digital media and communication research
(Musiani et al., 2016), and uncovers the ways privacy (or lack thereof) is built into the architecture of digital infrastructures. Mobile communication is, in this perspective, conditioned by the technological configurations of smartphones and operating systems, but also by the different apps that run on them and their various terms and conditions.

In exploring app infrastructures, we inevitably have to look at how app-based activities are controlled, and by whom. If we are to comprehend the pervasive economic incentives to collect ever-increasing amounts of data, we need to understand the underlying business models, operations, and ownership structures of the mobile app ecology and the data economy. Critical data studies (Dalton et al., 2016) have pushed this agenda by looking beyond the hype surrounding big data (boyd and Crawford, 2011) and discussing the implications of datafication in terms of social sorting (Kitchin and Lauriault, 2014), commodification of users (Couldry and Yu, 2018), and data (in)justices (Dencik et al., 2019). Computer scientists have worked on reverse engineering the web tracking mechanisms of the biggest data resellers in order to expose the depth and extent of datafication (Falahrastegar et al., 2016; Kalavri et al., 2016).

Building on knowledge in these fields, we follow the infrastructural connections that support the harvesting and transport of data as a digital currency, in much the same way as researchers “followed the money” in past studies of commercial media (DeFleur, 1971). We study data as “bits of power” (Mansell, 2017) that provide insights into the emerging institutionalisation of digital (and mobile) communication and enable us to understand and discuss the conditions and control mechanisms that frame mundane communication. As such, we built on and contribute to research within the field of political economy of communication (Hardy, 2014; Mosco, 2014), focusing on the commercial incentives, business models, and power struggles of large tech companies operating in a data economy where the commodification of human experience is key (Zuboff, 2019).
App ecologies

Common for most research mentioned in the previous section is a focus on the web as a particularly important component of the internet’s application layer. However, the increasing usage of mobile connections and smartphones renders the realm of mobile apps critically important for studying and discussing digital privacy, datafication, and the political economy of communication. The emergent field of app studies addresses this research gap. Studies dedicated to apps in particular are few and far apart, although more and more are emerging across disciplines. Gerlitz and colleagues (2019) differentiate between three strands of research that are largely concerned with apps, namely studies on mobile media and communications, mobile app usage and stakeholders, and studies that adhere to the infrastructural turn (e.g. Dieter et al., 2019; Gerlitz et al., 2019). The latter often feature case studies of particular apps or app categories like policing apps (Wood, 2019), dating apps (Weltevrede and Jansen, 2019), or social media apps (Nieborg and Helmond, 2018), which provide valuable insight on the app or category in question.

A fourth strand of research, which is not included in the three mentioned above, consists of large-scale studies of app infrastructures, coming out of the computer science field. These studies focus mainly on third-party services, thereby charting the mobile tracker ecosystems. While some emphasise particular types of apps, like smart TV apps (Mohajeri Moghaddam et al., 2019), several studies map the broader ecosystem of different third-party services and their parent companies across large bodies of Android apps (Binns et al., 2018; Vallina-Rodriguez et al., 2016). Similarly, but with a focus on the permissions requested by apps, Pew Research Center (Atkinson et al., 2015) has analysed the relationship between categories of apps and permissions requested by the at that time one million apps existing in the Google Play Store.

This article extends and builds on, but also diverges from the majority of previous app studies, by: 1) combining analyses of the permissions requested and the cooperating third-parties across a larger sample of disparate apps and categories; and 2) focusing on individual smartphone users’ suites of apps as situated in the greater
app ecology in order to, 3) analyse the implications of smartphone use in light of the power structures, business models, and control mechanisms that make up the political economy of apps-based communication.

While some of the case studies on apps (Forbrukerrådet, 2020; Weltevrede and Jansen, 2019) have looked at both the permissions requested by the particular apps as well as the third-party connections the same apps make, there is, to the best of our knowledge, very limited research that combines the two, and the studies that do, focus on particular apps or categories of apps rather than the broader app ecology. This article, in other words, fills a gap in the existing studies, and the following empirical explorations will testify to the potentials in combining both aspects in future large-scale analyses. Whereas the permissions relay important information on the extents of access requested by the apps, they also represent the ‘last bastion’, insofar as users are able to exercise their agency through dismissing or granting (some of) the permissions. Third-party services, on the other hand, are not explicitly requested, but are significant for understanding the extents to which users’ data is shared with corporations that are fundamental to the (commercial) app ecology at large. Had we only looked at one or the other, we would miss out on the ways apps combine permissions and third-parties in order to get to know increasingly more about their users as well as the complex power structures and market formations of the data economy. Applying an ecological perspective encourages us to look beyond particular apps or categories of apps and directs attention to the ways in which the entire app ecology frames and conditions the use of smartphones. It enables analyses of individual smartphone users’ appscapes (the constellation of apps and their associated permissions and third-party services) as well as of the larger ‘app ecologies’ (the broader compilations of apps in for instance an app store such as Google Play Store).

Taken together, we stress that data practices of mobile apps are materially and infrastructurally rooted and that varying degrees of privacy are built into the architecture of the different apps; we emphasise the underlying business models, market structures, and governance dynamics as vital for grasping and explaining mobile data harvesting, processing, and distribution; and finally, we apply an analytical perspective that cuts across specific apps and types of apps and look into
the broader app ecology while at the same time combining two empirical indicators (mobile permissions and third-party-services).

Methods

In comprising the datasets for the article, we began by creating maps charting the apps found on the smartphones of regular users sampled according to maximum variation principles. These individuals stem from an ethnographic fieldwork on the social uses of the internet in Denmark (Lai et al. 2019). The ethnography combined various methods, including walkthroughs (Light et al., 2018) of respondents’ phones (and other devices), which relay information on, among other aspects, the download and usage of various apps. The walkthroughs entailed the respondents giving ‘guided tours’ of their phones, where they would relay information on the kinds of apps they had acquired, the ways they used them, and so on. For this study, we focused on 10 respondents from the ethnography sample, and collected their individual repertoires of apps from screenshots taken by the respondents of their home screens. We find between 34 and 104 downloaded apps on the individual phones and a total of 173 unique apps – some are shared by all ten respondents, and some are unique to just one of them. Taken together, the 173 apps create a sample of common and generally used apps as well as lesser known ones down the longtail of existing apps. None of them are paid apps, although a bulk of them offer in app purchases or require subscription upon download in order to access the app functions. There is to date no public database containing information on the most used apps (although App Annie (n.d.) offers information on the most downloaded apps at present, it does not relay information on which is most used). That is, the 173 apps constitute a valuable though not representative sample of apps downloaded and used by Danish smartphone users.

The permissions required by the 173 apps, as well as the different types of third-party services they communicate with, function as empirical indicators for gauging their intrusiveness. That is, we do not analyse particular data flows from one app (the first party) to one or several third-parties (see e.g. Weltevrede and Jansen, 2019 for the use of a package inspection tool like WireShark to decipher the exact
transmission of data between a particular (dating) app and its third-party service), or make claims about what the apps, or the third-parties, actually do, but rather what they are able to do as a result of the accesses they obtain and the connections they establish.

The permissions provide insight into the types data and metadata that are available to a given app, granted by the user upon installation. This information was acquired through scraping of the Google Play Store for all information (price, rating, category, developer, etc.) on each app as well as its required accesses (e.g. ‘location’, ‘phone’, ‘calendar’) and permissions (e.g. ‘precise location (GPS & network-based)’, ‘read call log’, ‘read calendar events and confidential information’) in February 2020. In effect, the data adheres to the standards of the Google Play Store at that time as well as the standing Android operating system (OS) (the Android 10 release), which also means that the upcoming explorative analyses reflect this particular empirical constellation and cannot account for differences across, for instance, Android and iOS, different OS releases, or different app stores. The reality of the fieldwork respondents, whose app repertoires provided the basis for the dataset, is of course more complex than that: they owned devices running OSs from both Android and Apple, some of them had the latest software on their phones, and some of them had not updated for months. This type of data moves very fast, and analyses will, as a result, always reflect a particular point in time including particular OS releases, configurations of apps, and so on. Google Play Store is, to date, the standard for large scale app studies (see e.g. Binns et al., 2018) as it allows for researcher access to data whereas for instance Apple’s Appstore blocks this kind of interaction.

The third-party services associated with each of the 173 apps were harvested from the Exodus (n.d.) database, which extracts the third-parties from the source codes of the apps. More specifically, third-party services as pieces of software are often distributed by companies in ready-made toolkits, called Software Development Kits (SDKs). These SDKs come with a multitude of different purposes, including running analytics, profiling users, serving ads, establishing location and so on, and some are more intrusive than others. Although the Exodus database is updated regularly, it has shortcomings, as apps may install new third-party services at any
given time. The database is also incomplete insofar as it lacks reports for a number of apps, be it because they expired or are not available in Google Play Store. In result, by using this database, we are unable to account for the third-parties of 11 out of the 173 apps in the dataset. Exodus represents the most substantial database over app third-party services and is a valuable tool for studying mobile tracking independent of specific and microscopic protocol analyses.

**The dataset**

All permissions and third-parties that figure in the dataset are associated with one or several of the 173 apps. There are 115 unique permissions distributed across 15 overall types of access, and 107 unique third-party services, or SDKs, owned by 88 different companies.

Table 1 overviews a number of the most prominent permissions in the dataset and the specific accesses they require. For instance the permission to ‘take pictures and video’ accesses the phone’s ‘camera’. The table also explains in more detail what the specific permissions allow the apps to do.
Table 1: Commonly asked permissions by the 173 apps. Permission marked with a * are listed as ‘dangerous’ in Google’s Android Developers Guide. Source: Android Permissions (n.d.).

Some of the 115 permissions are central to the functionings of the particular apps (that is, strictly whenever the apps are used) while other permissions are requested for commodification and data re-selling (as highlighted by for instance a recent case on flashlight apps requesting more than 70 permissions that have nothing to do with the specific functionality of a flashlight (Cimpanu, 2019)). However, importantly, permissions that are central to the functioning of an app can also provide data that is sold on to advertisers (think of for instance location data harvested in Google Maps (Gundersen, 2020). That said, it is difficult to decide what potential harm to a smartphone user follows from any specific piece of data collected by an app. Let alone what potential harms follow from any specific constellations of data collected by an app. It is hard to imagine what kinds of damage can
come from an app being able to “view Wi-Fi connections” (simply what Wi-Fi networks are available), while an app accessing “precise location (GPS and network-based)” is more straightforwardly sensitive. Yet for both permissions, the data has potential (and at times critical) implications for privacy. In other words, it all depends. Rather than deciding what is and what is not sensitive data, we here consider any kind of data to be potentially harmful – in the hands of the wrong people, in case of hacking (e.g. Easyjet, 2020) or poor anonymisation practice (Gundersen, 2020), unencrypted connections (Weltevrede and Jansen, 2019), and so on.

Two of the most common permissions, present on four out of five apps in the dataset, are the “read the contents of your USB storage” and “modify or delete the contents of your USB storage” relating to accessing “storage” and “pictures/media/files” (see table 1). They allow the apps to access data stored the phone’s external storage (e.g. the SD card) and change or erase that data. The privacy implications of these permissions depend on a variety of factors, not least the kinds of data that the individual users have in their storages (e.g. photos of children, videos containing sexual content, etc.). Another example of a prominent permission is the “find accounts on the device” accessing “contacts”, which is found on one third of the apps in the dataset. By granting access, the app can access all user accounts on the phone, including the ones in other apps.

It is important to mention two ways in which users can counter (certain) permissions. First, particular permissions, for instance ‘read phone status and identity’ or ‘take pictures and videos’, are categorised as ‘dangerous’ in Google’s Android Developers guide (Google Developers, n.d.), meaning “a higher-risk permission that would give a requesting application access to private user data or control over the device that can negatively impact the user”. That is, access to user data which is considered private in the eyes of Google developers. These permissions require that the user explicitly approves them upon download – that is, for instance in a pop-up message the first time the app is opened. Second, as of 2017 for iOS (iOS11) and late 2019 for Android (Android 10), some permissions can be turned off in the smartphone settings after instalment of the particular app or they can be limited to only apply whenever the app is in use. All permissions are, however,
upon download by default set to apply all the time. These countering mechanisms are dependent on the user making informed decisions based on privacy policies and terms of service (ToS) agreements that are both transparent and accessible, which is usually not the case (bursting with information, the Facebook ToS, for instance, is a staggering 14,000 words long). Moreover, unlike web-based tracking where users are requested to agree to the cookies of a given site they visit, there are no cookie notices or other types of consent forms that apply to mobile platforms, which effectively leaves the end-user with no way of blocking or controlling third-party tracking by apps (except for in mobile browsers) (Binns et al., 2018: 2).

Table 2 lists the top ten third-party services in the dataset as well as their functions and ownership. It shows that the most recurrent ones are owned by one of two companies, namely Alphabet (Google’s parent company) or Facebook. Found on half of the 173 apps, Google Firebase Analytics, which makes performance reports based on the user’s usage of the app, is the most common tracker. Also, Alphabet’s Google Ads and Google DoubleClick feature frequently. Similar to the distinction between permissions for functionality or commodification purposes, we can distinguish between trackers that serve performance monitoring purposes or advertising purposes only. However, it is important to mention that functionality and commodification are not mutually exclusive. Just like free-of-charge apps are not free in the sense that they do not charge anything, but only insofar as data rather than cash constitute the currency by which users pay for the services, free-of-charge performance monitoring tools like Google Analytics are not for free in a strict sense. That is, Google acquire data on market developments in real-time from the myriads of apps (and websites) that employ their services. Data which, in turn, grounds their other undertakings and thereby their bottom line. Furthermore, previous research in the realm of cookie-based tracking on websites shows how Google Analytics can, if embedded by a another third-party tracker, distribute data beyond the specific service using Google Analytics (Roesner et al., 2012). More research is needed for determining whether or not the same could be the case for app SDKs.
<table>
<thead>
<tr>
<th>TPS</th>
<th>OCCURRENCE</th>
<th>% OF APPS</th>
<th>TYPE</th>
<th>OWNED BY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Firebase Analytics</td>
<td>115</td>
<td>66%</td>
<td>Analytics</td>
<td>ALPHABET</td>
</tr>
<tr>
<td>Google Crashlytics</td>
<td>88</td>
<td>51%</td>
<td>Crash reporter</td>
<td>ALPHABET</td>
</tr>
<tr>
<td>Facebook Login</td>
<td>62</td>
<td>36%</td>
<td>In app logins</td>
<td>FACEBOOK</td>
</tr>
<tr>
<td>Facebook Share</td>
<td>58</td>
<td>34%</td>
<td>In app shares</td>
<td>FACEBOOK</td>
</tr>
<tr>
<td>Google Ads</td>
<td>54</td>
<td>31%</td>
<td>Ads</td>
<td>ALPHABET</td>
</tr>
<tr>
<td>Google Analytics</td>
<td>54</td>
<td>31%</td>
<td>Analytics</td>
<td>ALPHABET</td>
</tr>
<tr>
<td>Facebook Analytics</td>
<td>46</td>
<td>27%</td>
<td>Analytics</td>
<td>FACEBOOK</td>
</tr>
<tr>
<td>Google Doubleclick</td>
<td>41</td>
<td>24%</td>
<td>Ads</td>
<td>ALPHABET</td>
</tr>
<tr>
<td>Facebook Places</td>
<td>38</td>
<td>22%</td>
<td>Location</td>
<td>FACEBOOK</td>
</tr>
<tr>
<td>Google Tag Manager</td>
<td>37</td>
<td>21%</td>
<td>Ads</td>
<td>ALPHABET</td>
</tr>
<tr>
<td>Facebook Ads</td>
<td>22</td>
<td>13%</td>
<td>Ads</td>
<td>FACEBOOK</td>
</tr>
</tbody>
</table>

Table 2: The most prominent third-party services in the dataset.

The next part of the article features three empirical explorations that traverse micro- and macroscopic levels, deriving inspiration from quali-quantitative approaches (Venturini and Latour, 2010).

**Analysis**

In order to exemplify how the intrusiveness of mobile apps can be mapped and assessed, we analyse the material through empirical explorations according to three analytical perspectives: First, we assess the apps, permissions, and third-party services of each respondent’s phone and explore how different appscapes have different implications in terms of data harvesting and privacy; we then look at the app ecology made up by the sample of apps installed on the respondents’ smartphones and explore the characteristics of different categories of apps as well as the broader market structures; and finally, we look at one particular category, ‘communication’, in order to investigate the differences between similar apps in light of their underlying business models.
Comparing appscape

The following explorative analysis exemplifies the appscape perspective on the intrusiveness of mobile apps. Table 3 overviews the ten respondents, the number of apps they have downloaded, and the accumulated number as well as the average number of permissions and third-party services associated with their individual apps. It illustrates how the number of apps vary significantly, but also how the degree of surveillance depends on the specific app repertoires of the respondents. While Marie (pseudonymous name, ed.) is placed at the top of the table with a total of 62 installed apps, the average number of permissions required for using these apps is the lowest among the 10 respondents (15,4 permissions per app) and the average number of third-parties is 5,7. At the other end of the scale, Ena has installed 11 apps, yet, in average, each of these requires 29 permissions and connects to 6 third-parties. In order to understand the relationship between the downloaded apps, and the degrees of data harvesting and third-party tracking, we chart the respondents’ individual appscape.

<table>
<thead>
<tr>
<th>RESPONDENTS</th>
<th>AGE</th>
<th>OCCUPATION</th>
<th>APPS TOTAL</th>
<th># OF PERMISSIONS</th>
<th>AVE. PERMIS-SIONS</th>
<th># OF TPSS</th>
<th>AVE. TPSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marie</td>
<td>36</td>
<td>Teacher</td>
<td>62</td>
<td>956</td>
<td>15,4</td>
<td>354</td>
<td>5,7</td>
</tr>
<tr>
<td>Kirsten</td>
<td>24</td>
<td>Bachelor student</td>
<td>46</td>
<td>981</td>
<td>21,3</td>
<td>221</td>
<td>4,8</td>
</tr>
<tr>
<td>Fatma</td>
<td>25</td>
<td>Teacher</td>
<td>42</td>
<td>845</td>
<td>20,1</td>
<td>227</td>
<td>5,4</td>
</tr>
<tr>
<td>Stine</td>
<td>23</td>
<td>Bachelor student</td>
<td>37</td>
<td>739</td>
<td>20,0</td>
<td>185</td>
<td>5,0</td>
</tr>
<tr>
<td>Louise</td>
<td>30</td>
<td>Master student</td>
<td>22</td>
<td>581</td>
<td>26,4</td>
<td>86</td>
<td>3,9</td>
</tr>
<tr>
<td>Meriam</td>
<td>45</td>
<td>Airport staff</td>
<td>21</td>
<td>382</td>
<td>18,2</td>
<td>91</td>
<td>4,3</td>
</tr>
<tr>
<td>Noah</td>
<td>21</td>
<td>Bar manager</td>
<td>18</td>
<td>378</td>
<td>21,0</td>
<td>135</td>
<td>7,5</td>
</tr>
<tr>
<td>Sofia</td>
<td>28</td>
<td>Occupational therapist</td>
<td>16</td>
<td>408</td>
<td>25,5</td>
<td>66</td>
<td>4,1</td>
</tr>
<tr>
<td>Liam</td>
<td>52</td>
<td>Factory worker</td>
<td>16</td>
<td>350</td>
<td>21,9</td>
<td>69</td>
<td>4,3</td>
</tr>
<tr>
<td>Ena</td>
<td>43</td>
<td>Day care assistant</td>
<td>11</td>
<td>319</td>
<td>29,0</td>
<td>66</td>
<td>6,0</td>
</tr>
</tbody>
</table>

Table 3: Overview of respondent and their appscape data.
Figure 1 and 2 illustrate the apps and associated permissions of two of the respondents, Liam and Louise. The inner circle outlines the different categories of apps found on their phones, followed by the specific apps, and their accesses and permissions. These visualisations function as research tools rather than results in and of themselves, and aid us in comparing and understanding the privacy implications of individual app repertoires. For instance, in figure 1 we get an impression of how the 350 permissions on Liam’s smartphone are distributed across his 16 apps. While the privacy app Adblock requires three permissions (‘view WI-FI connections’, ‘view network connections’, and ‘full network access’), Google’s search app requires no less than 72 permissions (e.g. ‘read your own contact card’, ‘precise location’, ‘read call log’, etc.). As illustrated in figure 2, Louise’s 22 apps require 581 permissions in total among which Google Search reappear as the most pervasive, measured by the number of permissions, followed by Facebook (52 permissions), and Facebook Messenger (46) as well as the communication apps Signal and Viber (both 44). As we will discuss later in the article, the communication apps in general contain a large amount of permissions compared to other categories such as for instance ‘e-government’. Louise’s preference for communication apps (9 of her 22 apps support different types of communication (e.g. text messaging, e-mail, voice calls etc.) thereby enhances the amount of permissions granted in comparison to for instance Liam (who only have two communication apps, Gmail and Viber, that require a total of 83 permissions).
Figure 1: Liam’s appscape - permissions
While the visualisations of permissions show what types of data different apps have access to, figure 3 and 4 illustrate the third-party services found in the different apps’ source codes and thereby the third-parties that can access data from the app. The figures, again, focus on Liam’s and Louise’s appscapes, and the inner circle represents the app categories, which links to the individual apps, that in turn link to the outer layer comprising the third-parties. Similar to the permissions, the prevalence of third-party services reflects the differences in Liam and Louise’s app
repertoires: Liam’s apps connect to 31 unique third-party services that appear 69 times in total across his apps, while Louise’s apps connect to a total of 86 third-party services (30 unique). Also similar to the permissions, the communication category contains a significant number of third-parties (e.g. Messenger and Viber have 10 and 11 third-parties). For both Liam and Louise, however, the most heavily tracked are the news apps: Louise’s NYTimes app and Liam’s BlicMobile app communicate with respectively 15 and 16 third-party services.

Figure 3: Liam’s appscape – third-party services
By visualising individual appscape and zooming in on the permissions and third-parties associated with the specific repertoires of apps, we can make sense of the possible privacy implications of smartphone use at an individual micro-level. The visualisations and calculations above allow us to discuss and compare how different app constellations enable different degrees of data harvesting and distribution of data to different third-parties. However, comparing individual appscape is only
one way of analysing a disperse sample of apps like the one in this study. Another is to compare app categories as broad containers of apps that serve similar purposes.

**Comparing app categories**

The explorative analysis of this section exemplifies the app ecology perspective. Google Play Store breaks apps into a number of different categories that developers then assign their apps to appear in. The 173 apps in this study appear in 29 different categories. The smallest category (‘Auto and vehicles’ and ‘Search’) comprise just one app, while there are 25 in the largest (‘Travel and Navigation’). The distribution of apps across categories, their average number of permissions and third-party services, and examples of particular apps found in each category are outlined in table 4.
<table>
<thead>
<tr>
<th>CATEGORIES</th>
<th># OF APPS</th>
<th>AVERAGE (MEAN) PERMISSIONS</th>
<th>AVERAGE (MEAN) TPSS</th>
<th>EXAMPLES OF APPS IN THE SAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>1</td>
<td>72</td>
<td>2</td>
<td>Google</td>
</tr>
<tr>
<td>Communication</td>
<td>9</td>
<td>34</td>
<td>5</td>
<td>Viber, WhatsApp</td>
</tr>
<tr>
<td>Video players &amp; editors</td>
<td>1</td>
<td>16</td>
<td>20</td>
<td>Imgur</td>
</tr>
<tr>
<td>Social</td>
<td>8</td>
<td>30</td>
<td>5</td>
<td>Facebook, Instagram</td>
</tr>
<tr>
<td>Dating</td>
<td>1</td>
<td>18</td>
<td>12</td>
<td>Tinder</td>
</tr>
<tr>
<td>Music &amp; audio</td>
<td>6</td>
<td>19</td>
<td>8</td>
<td>Spotify, Shazam</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>13</td>
<td>12</td>
<td>Forældreintra, DuoLingo</td>
</tr>
<tr>
<td>Travel &amp; navigation</td>
<td>25</td>
<td>17</td>
<td>6</td>
<td>Airbnb, Google Maps</td>
</tr>
<tr>
<td>Photography</td>
<td>12</td>
<td>14</td>
<td>8</td>
<td>Airbrush, CoCrop</td>
</tr>
<tr>
<td>Productivity</td>
<td>8</td>
<td>19</td>
<td>3</td>
<td>Outlook, Google Docs</td>
</tr>
<tr>
<td>Games</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>Angry Birds, SongPop 2</td>
</tr>
<tr>
<td>Health &amp; fitness</td>
<td>11</td>
<td>15</td>
<td>6</td>
<td>Clue, Endomondo</td>
</tr>
<tr>
<td>Religion</td>
<td>3</td>
<td>12</td>
<td>9</td>
<td>Holy Bible, Joyce Meyer</td>
</tr>
<tr>
<td>Shopping</td>
<td>9</td>
<td>13</td>
<td>8</td>
<td>H&amp;m, Asos</td>
</tr>
<tr>
<td>Entertainment</td>
<td>6</td>
<td>13</td>
<td>7</td>
<td>Kino, IMDB</td>
</tr>
<tr>
<td>Tools</td>
<td>4</td>
<td>17</td>
<td>2</td>
<td>QR Reader, Google Translate</td>
</tr>
<tr>
<td>Business</td>
<td>4</td>
<td>17</td>
<td>2</td>
<td>Planday, My Business</td>
</tr>
<tr>
<td>Sports</td>
<td>6</td>
<td>11</td>
<td>7</td>
<td>US Open, Forza Football</td>
</tr>
<tr>
<td>Privacy</td>
<td>3</td>
<td>14</td>
<td>3</td>
<td>AdBlock, Adblocker Plus</td>
</tr>
<tr>
<td>Food &amp; drink</td>
<td>4</td>
<td>11</td>
<td>6</td>
<td>5:2 Recipes, Wolt</td>
</tr>
<tr>
<td>Vod/streaming</td>
<td>5</td>
<td>12</td>
<td>5</td>
<td>Netflix, Viaplay</td>
</tr>
<tr>
<td>Weather</td>
<td>3</td>
<td>12</td>
<td>4</td>
<td>Yr, DMI Vejr</td>
</tr>
<tr>
<td>Finance</td>
<td>11</td>
<td>13</td>
<td>2</td>
<td>Mobilepay, Mobilbank</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>Dalle Valle, Opskrifter</td>
</tr>
<tr>
<td>Events</td>
<td>3</td>
<td>9</td>
<td>6</td>
<td>Vega, Royal Arena</td>
</tr>
<tr>
<td>News &amp; magazines</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>BT, NYTimes</td>
</tr>
<tr>
<td>E-government</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>E-boks</td>
</tr>
<tr>
<td>Books &amp; reference</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>Biblioteket, Sproghjælp</td>
</tr>
<tr>
<td>Auto &amp; vehicles</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>My Mazda</td>
</tr>
</tbody>
</table>

Table 4: App categories
The table shows that the app category requiring the highest number of permissions is ‘Search’, (which consists only of Google’s search app), followed by ‘Communication’ with an average of 34 permissions across nine apps. In the bottom, we find the ‘Auto & Vehicles’ category (containing only the Mazda app) that requires four permissions. ‘Video Players & Editors’ (also with one app, Imgur) has the highest number of third-party services (20), followed by ‘Social’, ‘Education’, and ‘Games’ with an average of 12 per app. The ‘E-government’ category (comprising E-boks, a dedicated state and public communication app) has the lowest, with just one third-party.

Figure 5 looks closer at the particular permissions that the different categories of apps require. The model outlines, from the left, the different app categories (‘Search’, ‘Social’, ‘Communication’, etc.), the different types of access, and the particular permissions as they relate to the different types of access and categories. The ‘Search’ category is, not surprisingly, placed at the top as it requires a total of 72 unique permissions among which 45 are categorised as ‘Other’. ‘Other’ is by far the largest type of access: it contains 83 unique permissions, and 466 permissions of the 173 apps belong to this group, just as 28 of the 29 app categories require permissions classified under ‘Other’ types of access. The second-largest access, ‘WI-FI-connections’, has 7 unique permissions that appear 31 times in the sample. Two permissions appear across all 29 app categories: ‘modify or delete the contents of your USB storage’ and ‘read the content of your USB storage’. These are placed under the access categories of both ‘Pictures/media/files’ and ‘Storage’, and the permissions enable the apps to save, alter, and access data that is saved in the device’s storage.
Figure 5: App categories, required accesses, and permissions.
There are distinct patterns in what permissions different categories of apps ask for. For instance, the categories ‘Communication’, ‘Search’, ‘Social’, and ‘Travel & Navigation’ cluster in two ways: The apps in these categories generally ask for a lot of permissions and also tend to ask for the same types of permissions that are less frequent across the remaining categories. These include access to the phone’s calendar and permission to ‘read calendar events plus confidential information’, ‘add or modify calendar events’ and ‘send email to guests without owners' knowledge’, as well as access to messages in order to read, edit, receive, and send SMS and MMS. A common denominator for the permissions that are particular for these four categories is the labelling of them as dangerous in Google’s Android Developers guide.

In figure 6 we see similar patterns with regards to the third-parties that the different categories of apps connect to. From the left, the figure outlines the app categories and connects these with the different third-party services, which connect to the parent third-party companies. As described earlier, the most intrusive app categories measured in terms of the average number of third-parties is ‘Social’, ‘Education’, and ‘Games’. However, if we look at the number of unique third-party services within the different categories, the apps found in ‘Games’ and ‘Travel & Navigation’ rank highest as they both connect to 36 different ones, followed by ‘Photography’ with 35 unique third-parties. As mentioned in the introduction to the dataset, Alphabet and Facebook dominate the ecology of third-parties, and for instance own the 10 most common third-party services in the dataset.
Figure 6: App categories, third-party services, and the companies that own the third-parties.
Figure 7 compares the average number of permissions (Y) and third-party services (X) for each app category. The colour and size of the coordinates illustrate the combination of the average number of permissions and third-parties. The ‘Search’ category stands out, as it, as mentioned, requires 72 permissions but only communicates with two third-party services. At the other end of the scale, the ‘Video players and editors’ category cooperates with a large number of third-parties but require relatively few permissions. ‘Social’ and ‘Communication’ are examples of categories that both request a lot of permissions and connect to many third-parties.

Figure 7: App categories plotted according to their average number of permissions and third-party trackers.
The results of this analysis raise important questions about the relationship between the (technological) functionality of the analysed apps, their business models, and user commodification. Are the permissions and third-party services identified in our study necessary for the apps to work or do they serve commodification purposes only (i.e. advertisement and data reselling)? If we are to understand why different apps collect different types of data and distribute it to different third-parties, we need to look closer at the particular apps as they serve different purposes and depend on access to data in different ways. Focusing on and comparing particular types of apps is also necessary if we are to understand the large differences within the app categories as they relate to the business models, economic conditions, and market positions of the various stakeholders in the data economy. In the following and last part of the analysis, we therefore focus on the ‘Communication’ category in order to discuss and explain the prevalence of permissions and third-parties across the specific communication apps.

**Comparing apps**

The last empirical exploration serves as an example of analyses at the level of individual apps. Table 5 overviews the nine apps in the ‘Communication’ category, their occurrence in the sample, the companies that own them, and their number of permissions and third-party services. We emphasise this particular category since it comprises an interesting combination of permissions and third-parties as well as a number of common apps that are less specific to the Danish context only. Furthermore, many of the communication apps (e.g. email, instant messaging, internet-based voice calls, cooperative work platforms, etc.) are increasingly essential to most people’s everyday lives – as underlined by the current Corona crisis. The apps all serve the same or similar purposes, namely to communicate personally or in networks (more dedicated social networking apps such as Facebook, Instagram, and YouTube are placed in the ‘Social’ category) making them fairly easy to compare. More specifically, the table shows that Facebook’s Messenger is the most common communication app in our sample, and, at the same time, the app that asks for the highest number of permissions, and cooperate with the most third-parties compared to the other apps in the category. Interestingly,
Facebook also owns the second-most used and most intrusive app when measured in terms of permissions, namely Whatsapp.

<table>
<thead>
<tr>
<th>APPS</th>
<th>OCCURRENCE</th>
<th>COMPANY</th>
<th>PERMISSIONS</th>
<th>TPSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Messenger</td>
<td>9</td>
<td>Facebook</td>
<td>46</td>
<td>10</td>
</tr>
<tr>
<td>Whatsapp</td>
<td>5</td>
<td>Facebook</td>
<td>43</td>
<td>1</td>
</tr>
<tr>
<td>Gmail</td>
<td>3</td>
<td>Alphabet</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>Viber</td>
<td>3</td>
<td>Rakuten</td>
<td>44</td>
<td>11</td>
</tr>
<tr>
<td>Skype</td>
<td>2</td>
<td>Microsoft</td>
<td>40</td>
<td>5</td>
</tr>
<tr>
<td>Signal</td>
<td>1</td>
<td>Signal</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Slack</td>
<td>1</td>
<td>Slack</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>Mobilporto</td>
<td>1</td>
<td>Post nord</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>Popularsticker</td>
<td>1</td>
<td>Ghostwording</td>
<td>12</td>
<td>No data</td>
</tr>
</tbody>
</table>

Table 5: apps in the ‘communication’ category.

Figure 8 further explores the permissions required by the individual communication apps by outlining, from the left, the particular apps, the types of access they require, and the particular permissions. As mentioned in the section above, the most common access category is ‘Other’, comprising a wide variety of different permissions. Reflecting the general findings across categories, the most frequent permissions are ‘modify or delete’ or ‘read the contents of your USB storage’. The more specific permissions required by the apps in this particular category speak (to some extent) to the purpose of the apps: It seems reasonable (although strictly whenever the app is actually used) for an app like Viber to ask for permission to ‘record audio’ or ‘directly call phone numbers’; it makes sense that WhatsApp requires permission to send and receive text messages, and so on. However, it is less obvious why these apps require access to ‘read calendar events plus confidential information’ or use the smartphone’s GPS to determine the user’s ‘precise location’. 
Figure 8: ‘communication’ apps, required accesses, and permissions.

Figure 9 illustrates the third-party services that apps in the communication category connect to. From the left, we see the particular apps, the third-party services installed as SDK’s on the apps, and lastly the companies owning these services.
When looking at this figure, the positions of Alphabet and Facebook is striking: Both companies supply some of the most used apps for communication (Messenger, Gmail, and WhatsApp), and at the same time dominate the third-party-market by owning some of the most frequent services, be it for performance monitoring (Google Analytics, Facebook Analytics, etc.) or advertisement (Google DoubleClick, Facebook Ads, etc.). Again, what these third-party companies actually track through the apps and what they use it for is another story. We discuss the implications of the third-party constellations shortly, but also make a call for future further research into the business models and incentives for collecting ever-increasing amounts of app data as well as for vertical integration strategies.

Figure 9: ‘communication’ apps, their third-party services, and the companies owning the third-parties.
Returning to the perspective of the individual app users, their choices, and capabilities for navigating in the complex app ecology and data economy, the analysis above shows that, in terms of users’ privacy, there are significant differences between apps that on the surface look similar. However, as others have discussed before us (e.g. Mansell, 2017: 16; Nissenbaum, 2011), few users are equipped for making informed decisions regarding consent that rest on actually assessing the ToS and privacy policies they agree to. In the following, we discuss the approach and explorative findings of the study, as well as the future prospects for analysing and increasing awareness and transparency of the data practises of mobile apps.

Discussion

We started out asking how the intrusiveness of mobile apps can be approached empirically and answered this question by focusing on permissions and third-parties exemplified through three empirical explorations that build on and inform each other. In this section, we discuss the potentials and limitations of the approach as well as the analytical perspectives and assess the results of the analyses.

The first perspective, and the explorative analysis applying it, demonstrates that mobile apps, permissions, and third-parties make up a structuring environment for the wide, and ever-growing, range of activities that smartphones support. The mappings and visualisations of individual appscapes are useful for grasping the extensive and complex ways that smartphone users’ data is harvested and distributed. The appscapes make it possible for users, as well as researchers, to see how our lives, whether we are aware of it or not, are framed and conditioned by the datafication that is built into smartphone apps, and enabled by mobile permissions and third-party agreements. As mentioned in the methods section, the empirical approach focuses on the intrusiveness and possible privacy implications rather than the actual consequences of using particular apps, which limits our ability to determine what specific bits of data is sent where and to whom. However, as the business models of digital corporations such as Facebook and Alphabet rely heavily on data harvesting, tracking, profiling, and targeted advertisement (Curran,
the permissions and third-party services that enable these undertakings are, to say the least, important to consider.

When you grant an app permission to ‘read your text messages’ or accept terms and conditions informing you on the use of third-party services, you also contribute to the (re-)production of a complex ecology (and economy) where big data is shared across applications and different corporations, and value is generated in opaque ways. In lack of official top lists and statistics, the individual apps' appscapes thereby also serve as gateways for exploring the greater app ecology and form the basis for the second perspective and empirical analysis. By mapping all apps in a given sample (in this case 173), the permissions they require, and the third-party services they communicate with, we are able to identify the most common permissions and the largest third-party services, as well as to discuss differences and similarities across app categories (e.g. streaming and VOD apps, sports apps, etc.). By comparing the different categories in terms of both permissions and third-parties, we can visualise and explore how certain types of apps tend to require more permissions or collaborate with more third-parties (or both, as in the case of the communication category).

This approach also makes it possible to identify the dominating market actors across both apps and third-party services and to explore market structures and vertical alignment strategies. For instance, the finding that Alphabet’s apps are among the ones that require the highest numbers of permissions, but distribute data to the fewest number of third-parties is interesting. If we use only the number of permissions or the associated third-parties as strict indicators for measuring intrusiveness, we miss the complex dynamics and power structures that shape the app-ecology. By looking at both permissions and third-parties as well as the underlying ownership structures, we can interpret the results in a more nuanced way. Alphabet then appears not only as an essential app provider, but, and perhaps more importantly, as a dominating third-party service that connects to most other apps. Needless to say, the Google search app does not require any other services than Alphabet’s own. Add to that, the fact that Alphabet holds and controls the largest app store in the world (Google Play Store), and the most used OS for smartphones (Android). The analytical approach and empirical explorations
presented in this article thereby contributes to existing studies of the political economy of communication by exploring an important arena where power battles are played out, market monopolies are built and sustained, and communication systems are commercialised (Manzerolle and Daubs, 2015).

The final perspective and the empirical exploration drew attention to the business models and ownership structures of similar apps within a category in order to compare their data practices. Such investigations are important, especially when discussing the degree of privacy, as apps that serve different purposes utilize data in different ways. When interpreting the patterns illustrated in for instance figure 7, it is vital to discuss how the constellations of permissions and third-parties relate to the functionality of the apps on the one hand, and to the commodification of users on the other. When is a permission necessary for the functionality of an app (e.g. a voice call app surely needs access to your microphone and speakers, although this is only the case so long as the app is in use), when is the permission used for other purposes (e.g. the training of a voice recognition algorithm), and when does it support commercial purposes only (e.g. collection of data that can be passed on to advertisers)? How dependent is an app of various third-party services, what do these services do, how are they reimbursed, and how does a ‘free’ app as well as the cooperating companies make a living? One could argue that a certain degree of commodification is unavoidable as users have gotten used to free-of-charge apps. However, it could also be argued that the price for these services should be transparent and subject to more debate – they are indeed not cheap if we consider the large sums of data taken in return for the services. In the final section, we reflect on how the article contributes to this debate and outline the potential future analyses that can build on and develop the methodological approach and the analytical perspectives presented here.

Conclusion

The aim of this article was to explore the possible privacy implications of smartphone apps through empirically investigating mobile permissions and third-party services as key indicators of intrusiveness. Operationalising this, we
developed and tested the methodological approach in three empirical explorative analyses that can be used to further understand and discuss the data practises of mobile apps:

1) Exploring and comparing different appscapes demonstrate the important and far-reaching consequences of which and how many mobile apps users download. For instance, our analysis shows that a preference for communication apps typically results in a high degree of intrusiveness when measured in terms of both the amount of permissions and third-parties. Such analyses can serve as starting points for studying individual smartphone users’ capabilities and strategies in terms of digital privacy. This, in turn, can increase public, and individual, awareness of the surveillance practises of mobile apps and can be used for developing tools for consumer empowerment.

2) Exploring and comparing the amount of data that different types of apps collect and the number of businesses that can access information, allow for researchers to explore the conditions and dynamics of the greater app ecology. Comparing permissions and third-parties on the one hand, and the providers of apps and third-party services on the other, offer valuable insights into the complex power structures and business models of the (mobile) data economy. Our analysis, for instance, illustrates Alphabet’s and Facebook’s dominant positions as both app and third-party providers and thereby enables discussions of the consequences of their multi-sided business models. Studies building on this empirical approach can thereby be an inspiration for market analyses, monitoring, and regulatory interventions.

3) And finally, analyses and comparisons of similar apps within a particular category can qualify discussions of legitimate data practises by exploring which permissions and third-party services are vital to the technological functionality of an app and which serve commercial purposes only. In the analysis, we emphasise the communication category as particularly interesting due to the high variation in the numbers of both permissions and trackers, but also the potential harmfulness of the permissions requested. Further insights into the technological configurations as well as economic motives that drive data harvesting, processing, and
distribution are necessary if researchers, policy-makers, and regulators are to challenge the power and knowledge-monopolies of the large tech companies.

The approach, the three analytical perspectives, and the application in the following explorative analyses have scholarly and societal potentials that should be pursued through in-depth as well as large-scale studies. In effect, we suggest three next steps for research on apps, critical data, and privacy: first, packet inspection of the apps analysed in this study would increase our knowledge on the implications of different permissions, by following the specific bits of data that are transported from the app to a given third-party. While some of the studies mentioned earlier have already done this on a limited number of apps (e.g. Forbrukerrådet, 2020; Weltevrede and Jansen, 2019), no large-scale studies combining the permissions and third-parties of different apps and categories of apps have been carried out so far. If we are to investigate what data the different apps actually harvest, how it is transported and stored, what it is used for, and by whom, we need publicly available tools that can be used across disciplines and institutions; Second, and relatedly, the results of our analyses leave us with a multitude of questions unanswered about the relationship between apps and third-parties, for instance: How can and do apps control what data third-parties can access – do they use encryption, restrictions on data reselling, and so on? Future studies should look into these relationships and dynamics not only for the sake of the app users but also in order to qualify debates and initiatives related to regulating the data economy; And third, future studies should use the method and empirical approach presented here to conduct large-scale analyses comprising for instance all apps in an app store or the long tail of most popular apps used within a given societal context. Larger data collections could be used for building a database containing information on app permissions and third-parties in a transparent and assessable way and thereby countering established strategies of corporate obfuscation (Brunton and Nissenbaum, 2015) and user resignation (Draper and Turow, 2019). Such a database will for instance enable users to compare different apps that serve similar purposes and choose the less intrusive one.

In conclusion, we believe that public debate, transparency, and access to the data processing of mobile apps are necessary stepping stones for developing policies
and regulatory tools that can target digital surveillance. Enhancing our knowledge, and ability to gain knowledge, on the app ecology and data economy, is a necessary first step away from the current scenario of user exploitation and alienation and towards a future of emancipation and empowerment. Returning to the article’s opening allegory, the problem is not whether you are doing something on your smartphone that your mother would not approve of. It is rather whether your smartphone is doing something that you (or your mother) would not approve of.

References


Curran D (2018) Are you ready? This is all the data Facebook and Google have on you. The Guardian, 30 March. Available at: https://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy (accessed 10 April 2020).


Publication

MedieKultur, 2020
DOI: https://doi.org/10.7146/mediekultur.v36i69.121018

Authors

Signe Sophus Lai & Sofie Flensburg
Appscapes in Everyday Life

Studying the implications of mobile datafication from an infrastructural user perspective
Abstract

It has long been acknowledged that the use of ‘free’ mobile apps comes with a price, but few empirical studies have looked into this supposed trade-off. This article combines qualitative interviews with mappings of infrastructures for datafication in order to study the implications of mobile app usage from the perspective of individual users. It analyses users’ understandings of online tracking, maps the infrastructural tenets of mobile datafication, and finds a disconnect between what users believe happen to their data, and the actual data harvesting and distribution mechanisms of their apps. We thereby argue that users’ resigned attitudes should be understood in light of the material conditions of the app economy, and, as such, that user and infrastructure studies should join forces in exploring and enhancing users agency, empowerment and emancipation.

Keywords

infrastructure; apps; datafication; digital resignation; data economy; surveillance

Introduction

Smartphones are by now an integral part of most people’s everyday lives. However, the data collection methods of mobile apps as well as the distribution and use of mobile metadata is still far from common knowledge to users and scholars (for exceptions see Atkinson et al., 2015; Binns et al., 2018). While user studies have looked into people’s ‘algorithmic imaginaries’ (Bucher, 2017) and have identified a ‘privacy paradox’ separating users’ opinions from their actions (Barth & Jong, 2017), few studies have mapped how data actually flow from individual users’ smartphones (Lai & Flensburg, 2020) and none have, to the best of our knowledge, combined inquiries of user attitudes towards online tracking with explorations of the infrastructural mechanisms of mobile datafication. This article therefore introduces a novel approach to critically interrogating datafication by combining a user-centric focus with broader macro perspectives that emphasise
material infrastructures and business models as structuring forces that shape the everyday lives of regular smartphone users.

Building on current scholarly discussions of the implications of datafication in terms of commodification of users (Couldry & Yu, 2018) and data (in)justices (Dencik et al., 2019), we argue, on the one hand, for the advantages of an infrastructural turn (Musiani et al., 2016) in user studies that focus on the political economy of data harvesting and tracking. Infrastructure studies, on the other hand, can benefit from applying a user perspective when scrutinizing the data economy’s opaque business models and obfuscation practises (Draper & Turow, 2019). By combining insights into users’ experiences and understandings of online tracking with knowledge on the infrastructural mechanisms of apps, we are able to produce more coherent and nuanced understandings of user agency and possible ways to enhance it through, for instance, regulatory interventions.

In the following, the article explores how user and infrastructure studies, as two separate strands of research, can be combined and mutually inform each other, in order to contribute to existing studies of how datafication impact the everyday lives of smartphone users. It first analyses 20 regular smartphone users’ understandings of datafication and online tracking, and then maps how the different apps installed on individual users’ phones result in different types and degrees of data harvesting and distribution. We thereby develop what we have coined the “appscape” approach (Lai & Flensburg, 2020), in order to engage empirically with the implications of commercial mobile infrastructures for individual users, and to offer a tool whereby these implications can be made transparent and accessible to researchers and users alike.

The article consists of four main sections: In the first, we present the theoretical foundations for the article and argue that the appscape approach represents a novel contribution to critical data studies and the emergent field of app studies by applying a human-centred, but technologically-informed, approach to uncovering the datafication of everyday life. The second section presents the methods and empirical material, emphasising the value of combining thick interview data with mappings of infrastructures for datafication. The third part presents the two analyses, respectively of the respondents’ understandings of datafication and online
tracking, and of their individual appscapes. And the fourth and concluding section discusses how this approach can ground future studies as well as consumer campaigns and regulatory interventions.

**Literature review & theory**

We position our research in the broad field of so-called critical data studies (Dalton & Thatcher, 2014) that, in recent years, have pushed a critical agenda on the ubiquitous datafication (Mayer-Schönberger & Cukier, 2013) of everyday life and challenged the hype surrounding big data (boyd & Crawford, 2011). The field has, in particular, contributed with theoretical insights and conceptual work on what big data is (see e.g. Kitchin, 2013), central research areas and questions (see e.g. Dalton & Thatcher, 2014), and encouragements for research to empower and mobilize users (see e.g. Iliadis & Russo, 2016). However, consistent calls have been made for more “empirical research to underpin and flesh out critical data studies” (Kitchin & Lauriault, 2018, p. 18). Two related, but separate, strands of research have answered this call through empirical analyses of the consequences of datafication: User-oriented studies with a particular focus on privacy have explored individual internet users’ capacities to understand and abilities to actively opt out of digital data harvesting (see e.g. Barth & Jong, 2017). And digital infrastructure studies (Sandvig, 2013) have uncovered the materiality (Winseck, 2019) and ownership of digital infrastructures underlying the abundant data economy (Zuboff, 2019). The following sections give a brief account of existing research in each field and conclude by defining the concept of ‘appscapes’ as an epistemological and methodological tool for combining the two traditions in empirical analyses.

**Users’ experiences with datafication – privacy concerns & resignation**

Studies of datafication and everyday life usually build on interviews or other qualitative research designs in order to uncover how digital media users understand and cope with online tracking, algorithms, and so forth (Bucher, 2017; Dubois & Ford, 2015; Kennedy & Hill, 2017). Among the most prominent and well-
documented findings is the so-called ‘privacy paradox’ (Barth & Jong, 2017) that identifies a discrepancy between users’ attitudes and behaviour towards protecting their data. The paradox implies that knowledge and awareness of privacy risks do not necessarily lead to more restrictive or critical behaviours (Joinson et al., 2010; Oomen & Leenes, 2008; Pötzsch, 2009). The paradox has been explained in different ways. Some argue that users’ strategies build on rational calculations, where the disclosure of data is seen as a reasonable price for a desired product (Acquisti & Grossklags, 2005). Others argue that users tend to underestimate the actual implications of their digital activities (Norberg et al., 2007) and instead focus their attention to deliberate data disclosures (e.g. what information they chose to register or publish), with far less concern for the harvesting of meta data (through e.g. cookies, mobile permissions, etc.) (Young & Quan-Haase, 2013). Common for these studies, is an underlying premise of a (more or less conscious) trade-off between the user and the provider of a given service. Irrespectively of whether or not the user understands the fundamental conditions and the ‘price’ required by the supplier, the service in question is assuming making it ‘worth it’ (Barth & Jong, 2017).

Draper and Turow (2019), in turn, dismiss the idea of an actual trade off. Referring to the ‘trade-off fallacy’ (p. 1825), they argue that users are caught in the conditions put up by the digital service providers and are de facto unable to opt out. The users’ attitudes are thus not all that paradoxical but rather a result of resignation: “[…] while these people feel dissatisfied with the pervasive monitoring that characterizes contemporary digital spaces, they are convinced that such surveillance is inescapable” (Ibid.). Draper and Turow (2019) further argue that resignation to datafication does not only follow from the seeming inevitability of data harvesting but also from a deliberate market strategy revolving around corporate obfuscation (Ellison & Ellison, 2009): companies that make a living from collecting, processing, or selling user data have a commercial interest in creating obstacles (like endless pages of terms of service, hardwired functionalities, etc.) that discourage users from taking action towards protecting their data and serve to further black box tracking activities. Thereby avoiding consumer complaints, political debate, and regulatory interventions. In line with this argument, user understandings should not only be seen as something that is shaped in the minds of individual
users but also as framed by the infrastructural and economic conditions that structure mobile communication.

**Apps as infrastructures**

Digital infrastructure studies seek to uncover the material foundation for datafication, and to explain the corporate practices and power structures that are built into digital technologies. This includes studies that reverse-engineer web tracking systems and expose the depth and extent of datafication (Falahrastegar et al., 2014; Kalavri et al., 2016), as well as analyses of how different material components of the internet infrastructure are controlled and governed (Galloway, 2004; Musiani et al., 2016; Sandvig, 2013; Winseck, 2019). In the field of app studies, the infrastructural approach has been applied in studies of how individual apps collect and distribute data (Nieborg & Helmond, 2018; Weltevrede & Jansen, 2019), while computer scientists have mapped out entire mobile tracker ecosystems (Binns et al., 2018; Vallina-Rodriguez et al., 2016). These studies provide valuable knowledge on app infrastructures and business models, but often neglect to ask how these infrastructure and market configurations impact ordinary users whose lives are so dependent on them.

In the following analysis, we recast the users as part of the equation by emphasising the ways the infrastructural arrangements of the app economy frame and condition users’ digital agency. From an infrastructural perspective, users’ capabilities for understanding and controlling data harvesting and distribution are closely linked to the infrastructural architecture of smartphones and apps. As more and more of what counts as social life is lived out through mobile technologies and applications, the need for studies that scrutinises the principles of this architecture only increases. As such, the infrastructural approach holds a sensitising potential in that it recalls “the simple crucial fact that each communication technology is a material resource whose distinctive features help to explain the [...] the communicative practices that have emerged, or which may emerge in the future” (Jensen, 2013, p. 216). This does not mean that it is unimportant what people say and do with these technologies, but rather that we should ground studies of user agency in the material conditions of their everyday communications. The approach
suggested in this article therefore explores the macro structures that frame individual app use from a micro perspective: namely that of appscapes.

The concept of ‘appscapes’ epitomises the merger of macro-oriented studies of infrastructures and micro-oriented studies of users’ understanding of online tracking and establishes a middle-ground for studying the implications of mobile datafication. It is motivated by a desire to map out the landscape of mobile surveillance as seen from the perspective of the individual smartphone user. We thereby take the agency of and diversity between smartphone users into account, while at the same time considering mobile data harvesting and distribution as rooted in and shaped by the greater app infrastructures and the market actors who control them. In other words, we focus on – and talk to – empirical individuals, and go on to situate their mobile communication in the greater app environment.

The appscape approach – methods and data sources

The appscape approach is developed on the basis of a multi-sited ethnographic fieldwork study (Marcus, 1995) of the role of the internet in everyday life, carried out in early 2018 in Denmark (see Lai et al. 2019). The next sections will outline the basic principles of this fieldwork and the data it produced as well as the ways it motivated the development of the appscape approach as a necessary next step towards understanding and explaining how mobile datafication impacts users’ everyday lives.

Qualitative interviews and user understandings

The ethnography study combined maximum variation and network sampling in order to produce an interview sample of 20 individuals. As a result, the respondents are diverse in terms of age, gender, place of residence and household, ethnicity, educational level, religious beliefs, and relationship and parental status. They thereby make up a broad sample of the adult Danish population: they use the internet – and smartphone in particular – in very different ways, have installed different apps, and exhibit different levels of digital knowledge, skills, and
resources. This sampling strategy enables deep descriptions of the individual cases, shows diversity between them, and highlights patterns across seemingly different groups of people (Patton, 2015).

The interviews were semi-structured (Brinkmann & Kvale, 2009), following an interview guide that featured general questions about people’s everyday lives, and more specific questions surrounding their communication activities, including aspects of datafication and tracking. All interviews were transcribed; transcripts were fully anonymised (names of respondents, places of residence and work, and names for close relations), and the respondents were given pseudonyms. The transcripts were then subjected to thematic analysis (King & Horrocks, 2010), which entailed a consensual and iterative coding by the authors of the sections of the interviews where the respondents talked about their data streams, targeted advertisements, and so forth. The initial descriptive coding returned a multitude of codes – like ‘feelings of not knowing what is going on’ – and gave rise to a continued redefinition of the codes. Next, an interpretive coding clustered the descriptive codes in relation to the research question and the interpretive codes – like ‘loss of control’ – were applied to the material. Lastly, overarching themes – like ‘control’ – were derived from the interpretive and descriptive codes by returning to the theoretical underpinnings of resignation and obfuscation.

While the interviews led us to interesting observations that to a large extend confirmed the established notions around digital resignation, they also had us wondering what the underlying premises for the respondents’ experiences and understanding were: how and with what implications were their everyday lives tracked, monitored, and commodified through their particular configurations of apps? And how did this relate to – or not relate to – their own understandings? Upon realising this, we reached the conclusion that their understandings alone were insufficient for explaining the impact that datafication had on their lives. In other words, motivated by the empirical aim of understanding the agency of individual smartphone users, we therefore decided to combine the user study described above with an infrastructural approach to datafication.
Mapping app permissions and third parties

As part of the initial interview, the respondents would guide walkthroughs (Light et al., 2018) of their individual smartphones, and some of them would take screen dumps of their home screens from which we were able to deduct the apps that were downloaded to their phones (from hereon referred to as the participants individual app repertoires). The walkthrough manoeuvre amounted to a total of 173 unique apps distributed across a total of 10 respondents. Some apps were shared by all 10 – who would each have between 11 and 62 downloaded apps on their smartphones – and some were unique to just one of them. In lack of any official or public databases containing information on the most used apps, the dataset in this article thereby creates a valuable, though not representative, sample of common and generally used apps in Denmark as well as lesser known ones down the longtail of existing apps (Lai & Flensburg, 2020).

In order to gauge the intrusiveness of the different apps, we used the number and types of accesses and permissions requested by the individual app, as well as the different third-party services it communicates data to, as empirical indicators. An app’s accesses and permissions offer insights into the types metadata that are available to the app, granted by the user upon installation. The app’s embedded third-party services, in turn, relay information on the extent to which app data can be shared as well as the kinds of actors that harvest and use the data. While some permission can be declined or changed later in the phone settings, for the purpose of this analysis, we focus on the default settings of the apps in question. That is, the respondents could in principle have altered these settings, but the interviews did not indicate that this was the case. The data was obtained through scraping of the Google Play Store in February 2020 for all information (price, rating, category, etc.) on each app as well as its requested accesses (e.g. ‘location’, ‘pictures/media/files, ‘SMS’) and permissions (e.g. ‘precise location (GPS & network-based’, ‘modify or delete the contents of your USB storage’, ‘read your text messages (SMS or MMS)’). The scrape returned a total of 115 unique permissions distributed across 15 overall types of access. While some permissions are foundational for the functionings of the particular apps (however, only whenever the apps are used), others are strictly for user commodification and data monetisation.
purposes. For instance, flashlight apps, requesting more than 70 permissions that have nothing to do with the specific functionality of a flashlight, are extreme cases of this (Cimpanu, 2019).

Much like cookies scripts in websites, apps can implement third-party services in small pieces of software, often distributed by companies in ready-made toolkits (so-called Software Development Kits (SDKs)). The third-party services connected to the 173 apps, extracted from the source codes of the apps, were obtained through the Exodus database (n.d.), also in February 2020. This returned a total of 107 unique third-party services (e.g. ‘Branc Metrics’ or ‘Umeng Analytics’), owned by 88 different parent-companies (e.g. ‘Branch’ or ‘Alibaba’), and offering different services (e.g. running ‘analytics’, serving ‘ads’). Allowing for these types of interventions, Google Play Store is the default platform for research (Binns et al., 2018; Exodus, n.d.), yet it also comes with a number of important limitations. Most significantly to this article, the dataset reflects the Google Play Store at the time of the scraping as well as of the standing Android operating system (OS) (the Android 10 release), and, as such, it cannot reflect differences across Android and iOS, app stores, or OSs. This specific methodological constellation therefore does not allow for a precise reproduction of the individual respondents’ empirical reality, which is of course a lot more complicated: They owned devices running both Android and Apple OSs, some of them had not updated their software for months, and even if they had, this was back in 2018. This is, however, a necessary evil in the emergent app studies field, where data move fast and analyses will inescapably always reflect a particular point in time.

To sum up, smartphone users’ individual appscapes are made up from their particular constellations of apps, the accesses and permissions requested by the apps, and the third-parties they cooperate with. As described above, the appscape analysis was not part of the initial research design, but served as an explorative endeavour to explain and substantiate the findings from the interviews. In other words, it represents a first effort at uncovering the underlying data infrastructures that ground individuals’ app usage and their experiences of datafication. Future studies should employ and develop the method in order to produce more coherent empirical work in the nascent field of app studies (Nieborg & Helmond, 2018).
Interview analysis

In the following, the thematic interview analysis is summarised in four common responses to online tracking. The four themes are interrelated and fluid, and can be collected under the headlines of: “It’s scary”, “I have nothing to hide”, “So long as it’s relevant”, and “Free is a good price”. As ways of understanding or coping with the increasing datafication of everyday life, the themes offer four ways in which digital resignation (Draper & Turow, 2019) is articulated and lived out in everyday life. The analysis shows the relationship between the themes and the interview material, but also how the respondents traverse and cut across the four responses over the course of their interviews.

‘It’s scary’

One of the most common statements whenever conversation touched upon targeted ads in apps like Facebook, Google Search, CandyCrush, and any other app financed through in-app advertising, can be boiled down to, as Fatma, a 25-year-old teacher, puts it: “I just think it’s scary”. Usually statements like Fatma’s were, in the conversations, followed by different examples of how this scariness had played out in mundane situations. Johnny, an employed consultant in his 30s, for instance described how Google used to send him push notifications on his Android phone, which were based on his routines and whereabouts:

> If you take the same route enough times, then it comes up and says ‘this train is delayed’, ‘this train is not delayed’, that’s Google+, but of course it’s something you can turn off somehow, it’s a settings thing, but I never really [...] I think it’s scary because it turns into a bigger and bigger overview over what we’re doing.

In the quote, Johnny explains how he is aware that this particular feature of the smartphone could perhaps be switched off somewhere in the phone settings, and even though he clearly thought about it and was bothered by the constant notifications, he never took the actual steps to reconfigure the default phone settings.
In a different conversation, with Hanne, a 22-year-old real-estate intern, she described how, even if she was determined to alter the settings so as to block, in her case, location-based tracking, she also would not know where to start:

Actually today, I think it was on Instagram, two of my colleagues were suggested, where I thought, ‘okay I was just at work yesterday’, where they were there, and I think that’s kinda scary that, I don’t know if it [the app] registered that we were close to each other and then they’re suggested [...] I think it’s creepy, but I don’t do anything about it and I actually also don’t know how I should do it, because I think it’s really difficult to shut down everything, when you have so many apps, so I wouldn’t know where to begin.

To this, Hanne was asked, what it would take for her to take action towards protecting her location data, and she replied: “I really don’t know, perhaps if you get a sick stalker or (laughs) I don’t know, as things are now, I don’t think I’ll do anything about it, because that’s just the way it is, I think”. In this quote, Hanne highlights how the threat of another person is much less abstract than the stalking carried out by an app like Instagram. She also emphasizes two other common features across the interviews, which is the absence of proper knowledge on what to do and the general naturalisation of commercial tracking, in which corporate power over data is viewed as an “inevitable and immovable feature of contemporary life” (Draper & Turow, 2019, p. 1829).

I have nothing to hide

The second common statement in the interview material focuses on how, trying to counter online tracking might give the impression that one has something to hide. Sofia, a 28-year-old occupational therapist, first articulated a fear similar to the one described above, when talking about personalised ads in her feed and her own responsibilities: “you become a bit scared, because you start thinking about how much you’re giving permission to, because it is... I’m the one giving permission to it, but how and like when did I do that?”. As the conversation continued,
she did, however, emphasise the normalcy of it, explaining why she was not all that worried about data collected on her online activities:

*Sofia:* You’re both a bit nervous but then I also think, ‘okay, what could happen’ or... if you don’t have anything to hide.

*Interviewer:* Yes, like so what if Facebook knows that I searched for jumpsuits?

*Sofia:* Yes, exactly! But it’s still not comfortable [Interviewer: No] of course you’d rather it wasn’t there, but now it’s like, at first I thought it was a bit infringing, now it’s more normal [...] I think it’s something about getting used to it.

Kai, an 85-year-old retiree, put it more bluntly than Sofia: “Personally I don’t give damn, because I don’t do anything online that I couldn’t put out there in public and I can’t, I can’t imagine that eh, that there’s anyone who could get anything out of surveilling me, at all.” Tim, a priest in his 50s, expressed at similar attitude, explaining that he was indeed afraid if maybe the large sums of data collected about people could one day be used to control them, yet he did not fear for himself:

*I’m doing all that many things that are (laughs) are worth looking for or storing [...] I mean, I don’t think that I, as a person, am that important (laughs) in that way, it’s not myself I’m scared for, it more the... the big masses’ use of it, right [...] If I’m buying something [online], then I think, it’ll probably be alright, because I want that thing, even though (laugh) it doesn’t look totally safe.*

The quote from Tim articulates a general concern for the population or the world as a whole, but a lack of concern for oneself, perhaps, as Tim also hints as, because the need or desire to obtain a given service through a certain app overrides the concerns.
As long as it’s relevant

A third common response can be summarised under a headline of relevance. For instance, Liam, a 52-year-old factory worker, explained that he hated when online ads, for something he already bought (in this case a flight to a European country) kept coming up again and again. He also, however, resignedly said that this is an unavoidable feature of online shopping. Elaborating on the view that ads can be a welcomed distraction so long as they are relevant, the 24-year-old bachelor student Kirsten described how she appreciated the suggestions found in her Facebook feed:

Eh it’s mostly concerts and eh.. clothing like fashion ehm and (pause) also a lot of university events and lectures and stuff like that, but a lot of it is something I search for myself, so it’s obvious... but it’s funny when you look at those ads, then you don’t always think that this is because I searched for that or like, you just think, ‘waww cool that they just find this for me’ (laughs) or like, totally naïve [...] I wouldn’t think it’s annoying that they [ads] come up all the time, because I often think that they hit [the target] pretty well [interviewer: And then it doesn’t matter so much?] no, exactly, it’s more if it’s something that’s not relevant at all or that you don’t think is cool.

The responses falling under this theme are interesting because they often demonstrate somewhat advanced understandings of why relevant ads are being served in social media feeds, search results, and so forth. That is, the respondents often approve of the systems enabling the relevance of the ads. In contrast and distancing herself from her family, who also welcome ads in the form of physical ad magazines that arrive in the post every Saturday, Louise, a 30-year-old master’s student, explained how the actual relevance of online ads (unlike the physical ones her family reads through) is the exact thing that makes her want to escape them:

I know what ads are and I get tempted by it, so I think it’s nicer that it’s just not there [...] sometimes they [ads] even make me feel bad, thinking like... you’ve been looking at the same pair of boots for the seventh time and then
you only get boot ads and then you just feel like, ‘God my life is meaningless’ (laughs).

**Free is a good price**

The last and fourth theme comprises responses that, more or less conscious of the underlying data business model, celebrate that because of tracking – and the targeted advertising, using profiling and so forth – these services are for free. That is, in a monetary sense. Miriam, an airport employee, explained that it was exactly the fact there would be no monetary charge that intrigued her, when she first started using social media:

> It’s really good, the resources we have, and free on top of that, really good, but they should not control us, we should control them [...] I really got to use Facebook as part of establishing [a non-for-profit enterprise] eh because I thought: ‘this is a free opportunity, there you have a free opportunity for reaching a lot of people, but also a free opportunity for reaching your family’.

Beyond the ease by which one can reach out, as articulated by Marie, a 36-year-old teacher, also emphasised another value in these services being free of monetary charge:

> I’ve heard some rumours saying that there’s going to be payment on some of the different, I don’t know if it’s platforms or things you log into, and what’s been positive until now is that if you have, if you can go online, then it’s equal for all. Because you can visit the same pages irrespective of whether your rich or poor and that’s definitely a positive thing, but if you have to pay for it, then there’s going to be a huge difference between who can afford this or that.

The quote by Marie celebrates the egalitarian principles that were foundational for the development of the world wide web, but it also demonstrates a lack of awareness about the price that is actually paid for free-to-download apps like the ones comprised in our dataset. The fact that these services are not free in a literal sense, but only insofar as data rather than cash constitute the currency by which users
pay for the services, does not figure anywhere as part of this understanding. Neither do circumstances surrounding inequality between people whose lives are heavily datafied and people whose lives are less tracked and shaped by data. Lastly, it disregards the imbalance between the commercial corporations harvesting and controlling people’s data one the one hand, and individuals that are resigned to it as a result of successful corporate obfuscation on the other.

To sum up, the interview analysis confirms and elaborates on findings from previous user studies: The respondents express a concern regarding online tracking while at the same time showing clear signs of resignation as they lack the skills or knowledge to resist the datafication and surveillance of their everyday lives. The three themes of ‘I have nothing to hide’, ‘So long as it’s relevant’, and ‘Free is a good price’ to some degree legitimise the tracking and commodification of users and indicate a normalization of the datafication mechanisms that are by now heavily embedded into the respondents’ everyday lives. ‘It’s scary’, on the other hand, implies a more critical attitude and suggests that the respondents would have an interest in countering datafication if they knew how to and had sufficient knowledge on what is actually going on. The interviews thereby also point to the so-called trade-off fallacy: the ‘prices’ for and implications of using app-based communication services are blurred and opaque, thereby making it difficult to make informed decisions or rational ‘cost-benefit’-like calculations. As a result, in the respondents’ accounts of how they experienced and understood online tracking, we found little to no information that could explain their experiences and understandings. Their accounts alone therefore proved to be insufficient for the analysis of the impact of mobile datafication. The following appscape analysis addresses the ‘scariness’ that the respondents express head on by exploring the direct and material consequences of their individual constellations of apps. By shedding light on the actual implications of mobile communication and making the underlying infrastructures visible, we hope to develop tools and resources that can show a way out of digital resignation and evoke active user engagement, empowerment, and emancipation.
Appscape analysis

Building on the interview study above, the following analysis explores how datafication impacts the everyday lives of individual smartphone users from an infrastructural perspective. We first give an overview of the 10 respondents, whose app repertoires we focus on for the analysis, and describe their apps, the number of permissions they request, and the third-party services they embed. We then zoom in on two respondents and compare their app repertoires and the types of data harvesting and distribution enabled by them – effectively their appscape. Lastly, we discuss how the appscape method can be a useful tool to enhance the transparency of app-based data collection and distribution, and thereby increase the agency of individual users.

The respondents

Table 1 overviews the ten respondents, their ages, occupations, total number of apps, total and average number of permissions requested by these apps, and the total and average number of third-parties that the apps connect with.

<table>
<thead>
<tr>
<th>RESPONDENTS</th>
<th>AGE</th>
<th>OCCUPATION</th>
<th>APPS TOTAL</th>
<th>NO. PERMISSIONS</th>
<th>AVE PERMISSIONS</th>
<th>NO. TPSS</th>
<th>AVE TPSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marie</td>
<td>36</td>
<td>Teacher</td>
<td>62</td>
<td>956</td>
<td>15,4</td>
<td>354</td>
<td>5,7</td>
</tr>
<tr>
<td>Kirsten</td>
<td>24</td>
<td>Bachelor student</td>
<td>46</td>
<td>981</td>
<td>21,3</td>
<td>221</td>
<td>4,8</td>
</tr>
<tr>
<td>Fatma</td>
<td>25</td>
<td>Teacher</td>
<td>42</td>
<td>845</td>
<td>20,1</td>
<td>227</td>
<td>5,4</td>
</tr>
<tr>
<td>Stine</td>
<td>23</td>
<td>Bachelor student</td>
<td>37</td>
<td>739</td>
<td>20</td>
<td>185</td>
<td>5</td>
</tr>
<tr>
<td>Louise</td>
<td>30</td>
<td>Master student</td>
<td>22</td>
<td>581</td>
<td>26,4</td>
<td>86</td>
<td>3,9</td>
</tr>
<tr>
<td>Meriam</td>
<td>45</td>
<td>Airport staff</td>
<td>21</td>
<td>382</td>
<td>18,2</td>
<td>91</td>
<td>4,3</td>
</tr>
<tr>
<td>Noah</td>
<td>21</td>
<td>Bar manager</td>
<td>18</td>
<td>378</td>
<td>21</td>
<td>135</td>
<td>7,5</td>
</tr>
<tr>
<td>Sofia</td>
<td>28</td>
<td>Occupational therapist</td>
<td>16</td>
<td>408</td>
<td>25,5</td>
<td>66</td>
<td>4,1</td>
</tr>
<tr>
<td>Liam</td>
<td>52</td>
<td>Factory worker</td>
<td>16</td>
<td>350</td>
<td>21,9</td>
<td>69</td>
<td>4,3</td>
</tr>
<tr>
<td>Ena</td>
<td>43</td>
<td>Day care assistant</td>
<td>11</td>
<td>319</td>
<td>29</td>
<td>66</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: Overview of respondent and their appscape data.
While the 36-year-old teacher Marie has the highest number of apps (62), these apps, interestingly, require the lowest average number of permissions (15,4). At the other end of the scale, Ena, a day-care assistant in her 40s, has just 11 apps, but these request an average of 29 permissions, which is the highest in the sample. Looking at the third-party services, Noah, a 21-year-old bar manager, has a total of 18 apps that connect, in average, to the highest number of third-party services (7,5). With also 18 apps, the master’s student Louise, who is in her late 20s, connect to an average of just 3,9 third-parties. In order to understand these differences, it is necessary to take a closer look at the individual respondents’ app repertoires. In the next section, we therefore identify the particular apps for two of the respondents, Noah and Ena, in order to explain their different constellations of permissions and third-parties. Noah and Ena are interesting and extreme cases (Yin, 2009) for a number of reasons, most notably that Ena’s very limited app repertoire is particularly intrusive when it comes to accesses and permissions, and that Noah has a significant average third-party score for his apps.

**App repertoires and permissions**

With just 11 downloaded apps, Ena uses her smartphone for a somewhat limited number of purposes compared to other respondents (e.g. Marie with 62 apps). The circular dendrogram in figure 1 illustrates Ena’s apps according to their category (the inner branches), the particular apps (the second-level branches), the accesses they request (the third-level branches), and the related permissions (the outer branches).
Three of Ena’s apps are used for communication purposes (Messenger mainly for text messaging, Skype and Viber mainly for voice and video calls), three are categorised as social media (Facebook, Pinterest, and YouTube), two are privacy protection apps (Adblocker and Adblock plus), and the remaining three are respectively a game app (Bubble Witch), a finance app (Mobilepay), a weather app (TV2 Vejr). As the figure illustrates, the most invasive app in Ena’s repertoire is Facebook that requests a total of 52 permissions including access to the calendar,
camera, contacts, device and app history (allowing the app to e.g. access information on browsing history), location, and so on. The communication app category, however, is the most invasive, requesting a total of 130 permissions across the three apps, with Facebook’s Messenger app requesting 46 permissions alone.

In other words, Ena’s preference for communication and social media apps is an important explanation for why her average number of permissions is high. Interestingly, however, the two adblocker apps also account for a relatively high number of permissions (40 in total), but while Adblock Plus only request three permissions, Adblocker request 37 permissions, including access to the phone’s camera, contacts, device and app history, location, microphone, and pictures. This is an interesting finding as the two adblocker apps serve similar purposes, but are remarkably different in terms of their degree of intrusiveness. Why does Adblocker need access to the camera, photos, calendar, browsing history and so on, if Adblock Plus does not? This question is difficult to answer without looking into the business models of the apps in question, but for the purpose of this article, we will simply note that the choice of a particular privacy tool is highly determining for the levels of intrusiveness, a user is exposed to. Put differently, an increased awareness of the actual implications of choosing one app over another, as well as concrete tools to determine the differences between two seemingly similar apps, could lead to more informed user choices.

By strengthening transparency, users would be able to rise above the ‘It’s scary’-attitude and nuance the ‘Free is a good price’-stance by being able to assess what kinds of data are paid in return for the service, and, importantly, existing alternatives. A closer look at the adblocker apps also reveal that Adblock Plus shares data with no third-parties, while Adblocker connects to ten third-party services owned by, amongst others, Facebook and Alphabet. That is, Adblocker both collects a significant amount of metadata and shares it with a relatively high number of third-parties. The next section goes into depth with the impact and prevalence of third-party-services focusing on Noah’s apps.
App repertoires and third-party connections

As listed in table 1, Noah has downloaded 18 apps in total that connect to the highest average number of third-party services (7,5). Similar to Ena, the reason for the high amount of third-parties is to be found in Noah’s app repertoire, which is illustrated in figure 2. Like in the figure above, the inner branches outline the categories of apps followed by the particular apps, while the two outer branches represent respectively the companies that own the third-party services and the particular third-parties.
There is a clear overlap between the apps downloaded on Noah and Ena’s phones. Noah also has the communication app Messenger and the social media apps Facebook and YouTube that, as described above, allow for the collection of wide amounts of data. Apart from providing a significant degree of data to these apps, Noah’s constellation of apps also allows a large number of third-parties to access data: For instance, the music app Sound hound connects to 20 third-party services, the festival app Roskilde connects to a total of 16; and the dating app Tinder connects to 12. Across the apps, Alphabet and Facebook own half or more of the third-party services, but also a long list of lesser known companies can access Noah’s data. These include for instance LeanPlum and AppsFlyer that serve ads, and Tinder Analytics for performance monitoring and optimizing. While these apps share data with a significant number of third-parties, they collect less data compared to Ena’s apps. Tinder, for instance, asks for 18 different permissions including access to the camera, location, and pictures, most of which, at least to some extent, reflect the functionality of the app and the service provided (as Tinder allows you to upload and take pictures, search for potential partners nearby etc.).

To sum up, Noah provides comparably less data to the specific apps, but (indirectly) allows the apps to distribute his data to a large number of market actors, who utilize and monetize it in various ways (Joler & Petrovski, 2016). By uncovering this rather hidden ecology of apps and data, users, as well as scholars, can be more informed, when discussing and assessing data harvesting and re-selling. Additionally, identifying the many actors involved in mobile surveillance can refocus attention away from the ‘I have nothing to hide’-position and towards questions of how and why apps and third-parties collect data in the first place. This can, in turn, qualify knowledge building on how data is processed and return to the user in more or less relevant, curated, and filtered formats grounding the ‘So long as it’s relevant’-attitude. In the following section, we will discuss how appscapes allow us to conceptualize, measure, and compare the degrees of intrusiveness that different smartphone users are subjected to.
**Appscapes**

Focusing on the app repertoires of Ena and Noah above, we have explored and explained their differences in the number of permissions and third-party connections listed in table 1. The conceptualisation of appscapes, as the specific constellation of apps, permissions, and third-party connections pertaining to an empirical user, calculates a degree of intrusiveness. Figure 3 and Figure 4 illustrate Ena and Noah’s respective appscapes as they are made up by the categories of apps installed on their phones (the inner branches), the permissions they request and the third-party services they connect to (the two outer branches).
Figure 3: Ena’s appscape. The visualisation overviews Ena’s categories of apps, the apps, and the accesses they request on the phone as well as the third-party corporations, whose third-party services they embed. For overview purposes the permissions under each access are not featured, but can be found in figure 2 (for instance: the permissions to “add or modify calendar events and send email to guests without owners’ knowledge” or the “read calendar events plus confidential information”, which both reside under the access to “calendar”). The accesses alone relay information on the extent to which the app can harvest different types of data on the phone. Likewise, the figure does not include the specific third-party service under each third-party corporation (for instance: “Google DoubleClick” and “Google Tag Manager” are both owned by the Alphabet corporation). In other words, it shows the number of potential actors that can access data irrespective of the number of third-party services they each embed in a given app.
Figure 4: Noah’s appscape. The visualisation overviews Ena’s categories of apps, the apps, and the accesses they request on the phone as well as the third-party corporations, whose third-party services they embed. For overview purposes the permissions under each access are not featured, but can be found in figure 2 (for instance: the permissions to “add or modify calendar events and send email to guests without owners’ knowledge” or the “read calendar events plus confidential information”, which both reside under the access to “calendar”). The accesses alone relay information on the extent to which the app can harvest different types of data on the phone. Likewise, the figure does not include the specific third-party service under each third-party corporation (for instance: “Google DoubleClick” and “Google Tag Manager” are both owned by the Alphabet corporation). In other words, it show the number of potential actors that can access data irrespective of the number of third-party services they each embed in a given app.
This approach allows us to measure and compare potential data disclosure on an individual level and visualise how data is harvested and distributed across the app ecology and data economy. It sheds light on datafication practices that are difficult to grasp for regular users, who then have a better foundation for choosing to interact with various types of mobile services. For instance, figure 3 illustrates the kinds of data Ena’s apps collect and the third-parties that might harvest it. In other words, it provides valuable insight into metadata collection mechanisms as well as the ways data is transported, and raises important and critical questions to the data industry. The striking example of Ena’s two adblocker app can, for instance, operate as an eyeopener that allows users like Ena to make more informed choices in the common situation where several apps offer the same services. Furthermore, examples such as this can be an important step towards asking more critical questions to market actors and potentially increasing the efficiency and impact of regulatory initiatives.

Noah’s appscape in figure 4, similarly, testifies to the potentials of the approach by illustrating how his every move – on the smartphone as well as physically – can be tracked through the multitude of apps installed on his device. It also illustrates how the data can be distributed to known as well as lesser known market actors that in various ways benefit from collecting and monetising metadata. And finally, it helps us connect the dots between the apps and the third-party services that are often owned by the same corporations, most prominently in our dataset, Alphabet and Facebook. In other words, the appscape method sheds light on the complex business models that finance what appears to be free services, and thereby also determine the conditions that regular users are forced to adhere to.

To sum up, the analysis compared ten different smartphone users’ app repertoires and the ways data are collected and distributed through these services. By zooming in on two individual appscape, we have investigated the implications of downloading different types of apps, and further, visualised the data that is collected and the actors that access it, emphasising how datafication and intrusiveness can be measured and monitored. In the final section, we discuss how the two analyses together inform the field of critical data studies by not only identifying the
attitudes of regular smartphone users living out the datafication of everyday life, but also emphasising how these attitudes are rooted in an infrastructural and commercialised environment characterised by opaque data collection, monetisation, and user commodification. A further development and implementation of the appscape approach in future research as well as consumer regulation could be a step towards denaturalising corporate obfuscation and empowering resigned users.

Implications and conclusions

This article has argued for a need to combine user and infrastructure perspectives when studying the impact of datafication on smartphone users’ everyday lives and thereby bridge the trenches that are so often dug between social constructivist studies of user understanding and materialist analyses of infrastructures. That is, datafication is both experienced by individual users and rooted in material infrastructures, and as such, research should pay attention to both. By combining the human-centred focus of user studies with an infrastructural perspective that highlights material and economic macro-structures, our study enhances knowledge on how people’s digital capabilities are framed by infrastructural conditions – and thereby how actual agency can be regained by building infrastructures for privacy.

In the interviews, the respondents demonstrate a variety of overlapping attitudes that all cluster around the feeling of digital resignation: data harvesting is understood as unavoidable and necessary, while at the same time being somewhat creepy, baffling, and frustrating. The appsapes, in turn, map and provide evidence of the implications of particular app constellations and the ways tracking is enabled through mobile apps. They show that regular smartphone users, as represented by the respondents in the study, are heavily tracked by means of their smartphones. Taken together, the analyses show how the resignation that users express is embedded in the complex, obfuscated, and inaccessible app infrastructure. Mobile datafication is, in turn, naturalised and expanded as a result of the users’ limited abilities to critically interrogate and opt out of the conditions set up by powerful market actors that are in the business of trading in ever-increasing
amounts of data. As argued by Draper and Turow (2019, p. 1833), “the corporate cultivation of digital resignation [...] turns individual concerns about surveillance and privacy inward, leading individuals toward confusion and indecision (rather than toward collective action) about whether and how to take on the burdens of privacy self-management”. This resigned, individualised, and somewhat paralysed condition results in lack of attention to and critique of “the broader surveillance ecosystems” and fails in facilitating “changes in industrial infrastructure that result in collective empowerment or systemic change” (ibid.).

We believe that the appscape approach paves the way for more effective solutions for countering digital resignation by fostering insights into the data harvesting and distribution systems that form the infrastructural foundation for individual smartphone users’ everyday lives and the data economy alike. The appscape approach serves as an empirically grounded method for developing tools that allow researcher as well as users to screen smartphones and apps and seek concrete guidance on how to choose between services and improve digital privacy. That is, without having to meticulously read through numerous pages of deliberately obfuscated terms of service agreements and privacy policies. We thereby join, and answer, Iliadis and Russo’s (2016) call for critical data studies that: “provide individuals with the necessary tools for becoming more informed and the ability to organize efforts around data justice issues” (p. 5). The prospects of the appscape method, however, rest on the premise that increased awareness of datafication practices supports the existing informed consent processes and ultimately make users more likely to resist or critically assess digital services. In effect, a critical next step for research is to engage with how people react to their individual appscapes. Only by confronting users – not only with the implications of their app usage, but also with suggestions for how to regain control of their data – will we be able to investigate the perseverance of phenomena such as the privacy paradox and digital resignation.
References


Publication

Feminist Media Studies
Status: In review

Author

Signe Sophus Lai
“She’s the communication expert”
Abstract

The article discusses the many distinct conceptions of labour at play in the digital media and communication literature. Joining a number of recent studies, it critiques the lack of research on the gendered dimension of digital labour, and suggests communication as a missing link to understanding the role of everyday interactions in the digital economy. In doing so, it specifies relational communication as a particularly important but also under-researched aspect of social reproductive labour, and argues for a reconceptualization of the sociological concept of “the second shift”, by introducing “the digital shift”. Lastly, it charts the implications of increasingly datafied digital communication, and debates if and how the commodification and monetisation of relational communication can be understood as exploitation that reproduces and reinforces existing inequalities. Thereby, the article pushes feminist critiques from the outskirts to the centre of critical data studies.

Keywords

gender roles, digital labour, relational communication, digital media economy, datafication

Introduction

In early 2018, as part of a project on the social uses of the internet (see Lai et al. 2019), I set out do ethnographic fieldwork on the role of digital communications in Danish people’s everyday life. Denmark, as a Nordic welfare state, is characterized by high levels of equality – not least in terms of gender, where the country ranks second best in the world (“Gender Inequality Index (GII)” n.d.). Early on in the fieldwork process, I was therefore surprised to find exactly how much gender mattered, whenever people (talked about the ways they) used the internet to communicate in their everyday lives – be it in how they used it for planning things in their lives, interacting with family and friends, recreation and entertainment, or getting things done. In other words, this article is not the result of a predefined research agenda, but a necessary and crucial step in making sense of Danish people’s social uses of the internet. As such, it builds on a long history of media and
communication studies (e.g. Radway 1984) that testifies to the unfulfilled prophesy of gender equality in the early days of the internet (Herring 2003) and the persistent differences in how women and men (use the internet to) communicate (e.g. Herring 1994; 1999; 2010; Pew 2005; 2019).

Studies of how the internet amplifies inequality is no stranger to any media and communication scholar – just think of for instance the vast field of digital divide research (Couldry 2009; DiMaggio and Hargittai 2001; Mansell 2002; Mihelj, Leguina, and Downey 2019; Ragnedda and Ruiu 2020; van Dijk 2005). The Autonomist Marxist concept of immaterial labour (Lazzarato 1996) has been widely used in scholarship on the economic value of online social interactions and the inherent inequalities of the capitalist system (see e.g. Andrejevic 2002; Arvidsson and Sandvik 2007; Fuchs 2014; Terranova 2000). Yet, while media and communication research has investigated the various ways online marketers profit from people’s mundane communications, only a few scholars have focused on the gendered dimension of digital labour (see Duffy and Pruchniewska 2017; Fortunati 2011; Jarrett 2014; 2018; Portwood-Stacer 2013). These studies reinvigorate the critiques posed by Marxist feminists in the 70s (Fortunati 1981), that established ‘women’s work’ – the unpaid social reproductive labour of the domestic sphere – as crucial to capitalist accumulation (Federici 2014). This article contributes to this growing research by adding a communication perspective on digital labour and digital data commodities to the feminist critique (Hill, Kennedy, and Gerrard 2016). Departing from critical data studies (Couldry and Yu 2018; Iliadis and Russo 2016; Kitchin and Lauriault 2014; Dalton, Taylor, and Thatcher 2016), it asks how and with what implications social reproductive labour is extended in the digital economy?

The article is divided into three parts. The first specifies relational communication as a particularly important but also under-researched aspect of social reproductive labour, and argues for a reconceptualization of the sociological concept of “the second shift”, by introducing “the digital shift”. This part also features a section that draws from empirical fieldwork data, consisting of self-reported communication diaries and recurrent elicitation interviews with a sample of 20 Danes, in order to provide examples of Danish individuals’ experiences with datafied relational
communication and, as such, of the initial puzzlement that triggered the article. The second part discusses the many, distinct conceptions of labour at play in the digital media and communication literature. Joining a number of recent studies, it critiques the lack of research on the gendered dimension of digital labour, and suggests communication as a missing link to understanding the role of everyday interactions in the digital economy. The last part looks at the implications of the commodification inherent in the digital shift, and debates if and how the datafication and monetisation of relational communication can be understood as exploitation that reproduces and reinforces existing inequalities.

A communication perspective: datafication and commodification

Communication is fundamental to most if not all of the immaterial qualities of social reproductive labour, which includes ‘affect, care, love, education, socialization [...] information, entertainment, organization, planning, coordination, [and] logistics’ (Fortunati 2007, 144). Hence, a communication perspective may help us to understand the implications of the digitalisation and datafication (Dijck 2014; Mayer-Schönberger and Cukier 2013) of these aspects of social reproduction, and thereby its submission to the logics of commercial digital infrastructures and platforms. In order to move forward with this perspective, I need to clarify two central concepts to this article, namely ‘relational communication’ and ‘the digital shift’.

Relational communication

The concept of relational communication originates in the 1970s “relational communication theory” (Rogers and Farace 1975), which focused on linguistic analyses of either dominant or submissive speech acts in interpersonal conversations. In later studies, relational communication was used simply as an equivalent to interpersonal communication (e.g. Ramirez Jr and Zhang 2007; Walther and Tidwell 1995) or with specific reference to group communication (e.g. Keyton 1999). In contrast, I conceptualise relational communication as the particular
communicative acts that are aimed at maintaining, nurturing, and reaffirming relationships including caring for, interacting with, and keeping relations informed as well as all types of coordination and organisation of socialising and other everyday events. These communications are often phatic gestures (Jakobson 1960) or small acts of engagement (Picone et al. 2019) that relay little other than confirming that ‘we are here for each other’ or ‘we care’, but just as often they are filled with information on matters that other people depend on. That is, like Sofia⁶, a 27-year-old occupational therapist, summarised it in one of our conversations: “the doctor, the ear doctor, activities like sports, [daughter] going to a birthday, uhm the deadline for something, eh parent-teacher meetings, holidays, everything.”.

For the purpose of this article, the relations in question include spouses, children, close family, relatives, friends, and flatmates, but exclude professional relationships. This demarcation entails staying clear of otherwise relevant discussions of emotional labour-heavy jobs (think of nurses, stewardesses, or care workers, as well as bloggers and other social media personalities), which are more adequately described as emotional or relational labour (concepts which I will return to shortly).

From the second shift to the digital shift

The notion of a digital shift takes its outset in sociologists Arlie Hochschild and Anne Machung’s (2012 [1989]) writings in the late 80s on “the second shift”: a term that describes the unpaid labour performed at home (care work, housework, etc.) in addition to people’s “first shift”, paid work carried out in the formal work sector. The digital shift follows this ‘work shift’ conception, but refers explicitly to the task of relational communication carried out through various digital infrastructures and platforms. The digital shift is, as such, today part and parcel of the second shift, as it belongs to the realm of unpaid social reproductive labour (even if the internet and mobile communication technologies are wiping out already

⁶ All names and exact ages of the respondents and their relations have been anonymised.
increasingly blurred boundaries between the first and the second shift as it is (see e.g. Gregg 2011)).

The concept also draws on the notion of a digital shift used in studies on the gradual change from using analogue or electronic technologies to digital ones. Relational communicative acts like happy-birthday greetings, which were once maintained through for instance the card as a specific medium and the infrastructure of the postal service (stamps, mail carriers, roads), are today more than often distributed through the internet infrastructure, in emails, wall postings on social networks, and so on. But also more material types of labour, housework like cleaning, grocery shopping and the like, are outsourced more and more via digital platforms and thereby transformed into paid work, which nonetheless acquire a great amount of communication (organising cleaning tasks, making and re-making digital grocery lists, etc.) from the ones responsible for the outsourcing, most of whom are still women (Huws 2019b).

In order to ground the coming discussions, I give a few examples from the fieldwork below, which emphasize the ways in which differences in the distribution of relational communication are organized and impact on everyday life.

**Communication experts**

The respondents who were interviewed as part of the fieldwork varied in terms of age, occupation, family situations, social status, and cultural background, making for a very diverse group of people (see Lai et al. 2019). Nonetheless, across life situations, be it a female student living in and caring for a student dorm or a mother of two working full time, gender played a crucial role in what and how they used the internet to communicate. One of the initial findings of the fieldwork relied on communication diaries that relayed information on the communicative purpose of what the respondents were doing over the course of 24 hours. They each generated on average 29 diary entries, which amounts to a total of 568, many of which contained instances of relational communication. On the basis of a coding of the diaries, I found that for every time a relational communication activity figured in a male respondent’s diary, it figured twice in a female’s, which amounts
to, for the women, more than 30% of their communicative acts across a normal day.

Upon realising this pattern, I started paying attention to the gendered nature of many of the understandings and explanations of the role of relational communication in the respondents’ lives. Like when Tim, a 57-year-old priest, told me “my calendar is fed by my wife”, digitally, and with any kind of family events, or when Hanne, a 21-year real-estate intern, distinguished between her mom calling frequently “to ask how things are” and “to talk about all sorts of matters” and her dad hardly ever calling unless “there’s something to say.”.

Anna and Jens-Otto, a couple in the 70s, used to own a private business together but had now retired. I visited their home several times, and during our meetings, Jens-Otto emphasised a strict division of labour when it came to communicating with their closer and more distant relations: “Anna, you probably know by now”, he said, “she’s the communication expert here.”. This is, I was to discover, an understatement. The two developed an intricate system for being available at all times: while her phone functioned as their incoming-calls phone, his was for outgoing calls – this way, they argued, the line was never busy with them making outgoing calls. They told their family and friends only to call her phone regardless of who they wanted to get into contact with. His phone was always turned off with the exception of when they needed to make outgoing calls or the rare moments when they were apart. This complicated manoeuvre inevitably meant that she did most if not all of the communication with family and friends. People called her phone and she picked up. People texted her and wrote her emails that she checked regularly on the iPad (which he never used). Hence, following a pattern of the past where Anna would answer the landline phone when it rang or write cards and letters to their relations, she now took care of all digitally mediated social networking – be it through messages, voice and video calls, emails, or social media (which he also never used). In a two-step flow of communication (Katz and Lazarsfeld 1966), Anna would show Jens-Otto the status updates of their friends or photos of their grandchildren, which in turn allowed for Jens-Otto to stay disengaged in social media, yet engaged in the personal lives of their loved ones as they were communicated on different platforms.
Anna and Jens-Otto are far from exceptional: I found many other complicated systems for organizing communication roles along a gendered division of labour similar to the one I found in their home. Instead of the ‘hardware solution’ described above, for Liam, a factory worker in his 50s, and Ena, a 43-year-old part-time day care assistant, the division was spelled out in software, as Ena explained:

It’s also kind of stressful, you know, people they write, and family and friends and all that, as I say, they write to me more, because they’re used to it being me taking care of family and all that [...] Liam moved everything on me, you can only reach him if you call, he deleted the Messenger and Facebook apps [...] I call family, his mother, his sister and such and ask about this and that and say congratulations with this and that.

The couple came to Denmark as refugees 20 years ago, and while Ena clearly loved talking to everyone “back home” regularly, she also felt the pressure of the expectations of family and friends for her to stay on top of things, not least on social network sites or through dedicated videocall apps. The quote captures a particular struggle between feelings of profound pleasure on the one hand and a gnawing fatigue on the other that would echo throughout the interviews, making me question how a fundamental part of everyday communication could be equally gratifying and tiresome to the women, I spoke to. Moreover, given the time and energy it obviously took up, I also started wondering how we might talk about these communicative tasks as something more than pastime distractions – as a specific form of digital labour that embodies specific forms of value.

**Digital labour**

The following sections look at various conceptions of labour that taken together inform a discussion of exploitation and alienation processes in relation to the datafication of relational communication. Whereas a majority of studies emphasise the ‘newness’ of different forms of labour growing out of the digital economy, a small subset of studies critiques how continuities and reinforcements of existing gender inequalities have been ignored.
**Invisible and immaterial forms of labour**

Drawing on Autonomist Marxist frameworks (e.g. Hardt and Negri 2000), digital media and communication scholars have studied different types of labour exploitation under different conceptions. Concepts including immaterial labour (Lazzarato 1996) and invisible labour (Crain, Poster, and Cherry 2016) have been used to interrogate and explain the specificities of digital labour. Scholars have also conceptualised digital media user activity as a particular type of free labour (Terranova 2000) that entails a shift from the ‘work of watching’ (for instance commercial TV) to the ‘work of being watched’ (for instance through online data collection) (Andrejevic 2002; see also Smythe 1981). Others have charted the global division of digital production and its inherent class relationships as they map out across complex value chains that distribute low income, high-risk work to the global south (see e.g. Dyer-Witheford 2015; Fuchs 2014).

What many of these studies have in common is an insistent focus on the immaterial and invisible aspects of digital labour as something new and somewhat specific to the digital economy (Jarrett 2018). Hence, these analyses of immaterial work do not recognise that much of women’s work in the private, domestic sphere was always immaterial in nature (Gregg and Andrijasevic 2019). Also, they do not emphasise the general ‘feminisation’ (Adkins 2001; Gregg 2008; Mayer 2013) of economic life, in which “all kinds of work become increasingly precarious, under-compensated and reliant on “soft” skills such as communication, affect and cognition.” (Jarrett 2015, 17). Put differently, many studies have failed to account for the existing continuities in how large corporations exploit feelings and leisure time for capital gain. Recently, feminist scholars have pointed to this omission (Cirucci 2018; Jarrett 2014; 2015; Ouellette and Wilson 2011; Portwood-Stacer 2013), and called for research into what exactly, then, is fundamentally new about the specific ways in which immaterial, invisible, and free labour figures in digital capitalism:

> We need to determine whether there is something specific about digital labour. If it is a matter of changes in the intensity, extensity or visibility of capitalism’s exploitation of living labour, then we need to provide evidence of such shifts (Jarrett 2018, 7).
Relatedly, feminist scholars emphasize the lack of research into how the gendered division of labour is extended in the digital economy (see e.g. Arcy 2016).

Hochschild’s study The Managed Heart: Commercialization of Human Feeling (1983) represents an early and foundational investigation into emotional labour. Emotional labour is defined as the management and display of emotions in order to fulfill the expectations of specific types of work that require the labourer to produce particular emotional states in others. Women are usually overrepresented in such jobs, which include Hochschild’s designated fieldsite, namely flight attendants. She argues that the commodification and commercialisation processes at play in jobs like the flight attendants’ entail that labourers eventually become alienated from their own feelings. Although the study predates any digitalisation or datafication of such processes, it points to a critical and persistent understanding and expectation of women as “experts” in emotion management, which in turn works to maintain and enforce occupational discrimination and the gender wage gap – or, in the case of the present article, everyday communication roles. With digital and especially social media, mediated emotion management, which was also historically a communicative job for women particularly (think of for instance switchboard operators (Balbi 2013)), is extended in the realm of the digital economy.

Under the headline of relational labour, Baym (2015) looks into the kinds of work necessary for creating, sustaining, and relating to online (music) fan communities through continued communication over time. In doing so, she distinguishes relational from emotional labour by emphasising the constant efforts to build an ongoing connection that go far beyond the management of other people’s emotions in single encounters. Also, and where the concept is in direct contrast to the understanding of relational communication introduced above, the monetary aspect of such communication efforts is seen as fundamental: “Relationships built through relational labor can entail all the complex rewards and costs of personal relationships [...] At the same time, the connections built through relational labor are always tied to earning money.” (Baym 2015, 16). As such, the concept of relational labour builds on the understandings and implications from previous studies of emotional labour, while also pointing to important aspects that are particular to labour in the digital economy. Also, the conception touches upon some of the
same issues at stake in terms of digital relational communication; namely that there is a particular type of work involved in creating and sustaining relationships, and that this work has a lot to do with communication. However, by focusing on the monetary value gained by and for the professional musicians, the study overlooks an important aspect of what exactly is being commodified and who exactly monetises it. Musicians and other people with a substantial following base are surely making money through community building practices and emotion management – tasks that are often not conceptualised as part of their job but nonetheless time and energy consuming labour. However, digital infrastructure and platform companies that such communication acts depend on and that facilitate this relational labour in the first place are, by all comparisons, the financial winners of this labour power. I will return to this perspective in the following.

To provide a different answer to the question of ‘what is new’ in digital labour (Jarrett 2018, 7), and to expand on what is being commodified and for whose gain, the next section focuses on the datafication of social reproductive labour, and more explicitly, of the communication inherent in that work. It also engages with the ways relational communication tasks – which were always disproportionally placed on women – are extended and intensified through the internet and on digital platforms as new arenas for the second shift (Ouellette and Wilson 2011).

**Datafication of relational communication**

Women, in Denmark and elsewhere, are more heavy users of digital social networks than men (Pew 2019; Statistics Denmark n.d.). In a recent survey study, Cirucci (2018) finds that social network sites function largely as symbolic extensions of the home: whereas the US-based women in the study tend to see social media platforms as a social utilities that extend their abilities to be emotionally supportive through interpersonal communication, the men see the same platforms as home entertainment systems for recreation and escape from work (p. 2964; see also Pew 2005 for earlier findings mirroring the same division between communication and entertainment).
With relational communicative acts moving onto digital platforms, nothing yet everything has changed: Commenting on a Friend’s Instagram picture or a status update on Facebook, following and acknowledging everyday life events as they are shared on WhatsApp, or organising the Sunday outing via the family’s Messenger group are qualitatively the same as the relational work of the past: things get done, people feel loved, and relations are sustained. A consistent paradox in my conversations with the respondents, however, was the fact that such communicative tasks that once required a lot of time and effort in terms of writing and posting mail, making elaborate phone calls, or stopping by could now be managed in the blink of an eye: a phrase posted on a wall, a heart shaped emoji in a comment, or the click on a like button. Should relational communication not have become a matter of less work, less time, and less worry?

One finding that goes against this presumption can be seen in a number of studies on the “contagiousness” of emotions and stress through online social networks (e.g. Coviello et al. 2014; Kramer, Guillory, and Hancock 2014). Based on survey results, Hampton and colleagues (2015) conclude that social media users are generally more aware of stressful events in the lives of their family and friends, and women more so than men. Another thing that goes against it, I learned from speaking to Anna, Ena, and a dozen other women, was that the sense of workload relative to relational communication had only increased with more and more communication taking place via digital technologies. The ease with which one is able to notice, pay attention, and react to all sorts of everyday incidents had amplified not only the expectations of their friends and family as to what gets attention and what does not, but also their own. That is, whereas the possibilities for reaching out to people and staying on top of their lives were once limited to what could be managed over the phone, through cards and letters, and in face to face settings; with digital, mobile, and not least social media, the visibility of everyone’s daily life events means that now there is no limit as to what one might pay tribute to through messaging, commenting, liking, and so forth (Portwood-Stacer 2014). Hence, digitally mediated relational communication make up the new digital shift of caretaking. Unlike the notion of women’s first and second shifts, the digital shift feeds of and contributes to the continued break-down of what is commonly referred to as work-life balance: one can easily blend in and out of the digital shift,
making it a multitude of small shifts throughout the day enabled by platforms that excel at notifying whenever this other job is calling for attention.

**Commodification of the digital shift**

A key contribution of Marxist feminist scholarship lies in the argument that the unpaid social reproductive labour of the domestic sphere is crucial to capitalist accumulation: women (literally) reproduce individuals (through childbirth, breastfeeding, etc.), and they socially reproduce labouring individuals (through child rearing, homemaking, education, etc.), thereby contributing value to the capitalist system. The “wages for housework” movements of the 70s took this argument to the next level by demanding pay for domestic labour. Although unsuccessful in reaching their goal, these movements were effective in advancing recognition of the indirect value of women’s work.

Drawing on materialist second-wave feminist thoughts, Huws (2019b) looks at how housework since then has been commodified, decommodified, and recommodified through digitalisation, while fundamentally keeping to its gendered division of the past. The study develops a typology of labour that distinguishes between six categories according to two variables: “paid” or “unpaid” and “produces social value only” or “produces value for capitalists (‘surplus value’).” (p. 12; see also Huws 2019a). The first category, called “subsistence labour”, is of particular relevance here. According to the typology, subsistence labour, which includes cleaning and cooking but also for instance “teaching children to speak or read”, is labelled as unpaid and produces social value only (p. 11). Again, communication is fundamental to (at least parts of) subsistence labour, even if communication is not part of the typology. The communication perspective, however, challenges the categorisation as far as the value of the labour goes: much of the communication central to subsistence labour (and to social reproductive labour in general) is now carried out through various digital technologies and platforms, and thereby subject to extensive datafication. This means that, by way of its ‘biproduc’ (Jensen and Helles 2017) – personal data bits and trails – datafied relational communication does in fact produce surplus value for capitalists: Commodified everyday interactions are sold to the highest bidder on the online ad exchange. Given the centrality
of communication to the economic proliferation of digital platforms, relational communication thereby enters the economic circuits of the commercial internet through its exchange-value.

Even though, as spelled out by feminist movements 50 years ago, social reproductive labour was always integral to the maintenance of capitalism, the datafication of relational communication turns its indirect value into a direct one: data harvesting schemes and targeted advertising put a price on relational communication acts. Like in Baym’s study above, the lack of a communication perspective means that this type of commodification is entirely omitted from Huws’ typology and from the categorisation of subsistence labour. Be it relational labour (Baym 2015) or relational communication in the everyday sense, these communicative acts contribute with data points for personalisation and profiling purposes (topics, like, comments, shares, etc.) as well as general and increased interpersonal and networked social interaction, which, at the end of the day, is what drives business for large digital corporations like for instance Facebook (see e.g. Joler and Petrovski 2016 for an elaborate overview of Facebook’s business model). As such, recent discussions regarding people being the merchants of their personal data or data autonomy following the EU General Data Protection Regulation (GDPR) mirrors the “wages for housework” debate insofar as they call for recognition of the exchange-value of personal data as well as the inherent inequalities in terms of who gets to make money of it.

The previous sections reviewed various answers to the question of ‘what is new’ in terms of digital labour, ultimately arguing in favour of a communication perspective that allows us to understand what has changed: socially - relational communication tasks have increased with personal and social consequences for especially women; as well as economically - from indirectly sustaining a capitalist system to providing direct surplus value. Taken together, the two aspects show how the datafication of relational communication reproduces and reinforces already existing inequalities and power structures – between genders as well as between the people that provide data and the ones that control, mine, employ, and monetise it. The last section then turns to a familiar yet fragmentary and incomplete
discussion of individual agency and autonomy on the one hand and exploitation and alienation on the other.

**Implications: exploitation**

Many researchers are either cautious about labelling digital social interactions as labour (Hesmondhalgh 2010) or they reject the audience commodity thesis (Smythe 1981) – that consumption can be productive – altogether. Especially the concept of exploitation has been highly contested with critics invoking the absence of force and the pleasures of participation (Andrejevic 2011, 283) in order to challenge the idea of describing the ways large corporations employ and monetize personal data as exploitative (e.g. Lee 2013; Meehan 1984). The argument goes, in simple terms, if it is not forced, it cannot be considered exploitation, since clearly: “you don’t have to join Facebook or Twitter, it is not necessary for survival in contemporary society (yet?) to get a Gmail account, or to shop online.” (Andrejevic 2011, 283) (although, I would argue, after months of interviewing, it truly takes an immense personal and social effort to refrain from doing these things, not least if you are a woman (see also Portwood-Stacer 2013)). Hesmondhalgh (2010) holds up the concept of exploitation against the backdrop of sweatshop labour, asking “are we really meant to see people who sit at their computers modifying code or typing out responses to TV shows as ‘exploited’? [...] Clearly not” (p. 271), while also arguing that “controlling or reshaping the ways in which people communicate [...] is indeed a problem [...] [b]ut this is not the same thing as exploitation.” (p. 273). Lastly, Bolin (2010; 2011) and others contend that the actual value created and sold by digital actors is statistical representations of users that are produced and commodified by paid market research agencies and data analysts.

Arguing against this last critique, Jarrett (2015) invokes an industrial analogy in order to show how such arguments rely on a limited notion of value-creation that overlooks the origins of the raw materials that for instance Facebook analysts work with on a daily basis:
Facebook can be viewed as an assembly line in which certain paid workers produce a website which serves as the materials transformed by another set of workers – the unpaid consumers involved in “liking” and generating status updates – into raw data, which then move along the value chain to another set of paid workers who transform those data into saleable commodities (p. 150).

Going back to the canonical discussion of ‘free labour’, Andrejevic (2011) also counterargues the idea that pleasure and exploitation are binary oppositions. In his discussion, user-generated content is divided into two kinds: the visible activities on screen and the data commodity that such activities cast off. Focusing on the latter kind, the monitoring aspect of networked interactivity, he emphasizes that “exploitation is not simply about profit, but also alienation” (p. 283), and that “[a]lienation subsists not just in the surrender of conscious control over productive activity, but also, consequently in its product.” (p. 284). In other words, the ‘exploitation-as-alienation’ argument contends that exploitation is not just about a loss of monetary value, but also a loss of control (p. 284). The argument rests on two interrelated developments: the privatization of digital infrastructures, networks, and applications along with the financial incentive to take advantage of the inherent surveillance capacities of digital communication media (see also Mansell 2017). Taken together, these developments form the basis for the present constant and comprehensive datafication and commodification of people’s online communications, regardless of whether people produce or consume content. That is, any kind of user-generated content or every small act of engagement is simultaneously doubled in a distinct yet intrinsic data commodity. Such commodities have market value and while they are the result of user activity, they are not controlled by the users, who have little to say in matters of the data being captured, mined, and used for commercial purposes:

we might describe the generation and use of this data as the alienated or estranged dimension of their activity. To the extent that this information can be used to predict and influence user behaviour, it is an activity that returns to users in an unrecognizable form as a means of fulfilling the imperatives of others (Andrejevic 2011, 286).
Held up against the concept of relational communication, people use digital infrastructures for being social and communicating about things that matter to them and that are close to their hearts. Concurrently, data about their communications are collected and returned to them in unrecognisable forms of targeted advertising, social profiling, personalised feeds, and so forth, which ultimately alienates them from their own communication activities.

At this point, we might talk about two degrees of exploitation. In the first degree, exploitation occurs at the level of everyday communication and social reproductive labour in general: women, more than men, continue to be the ones responsible for relational communication tasks, and these tasks have been extended but also multiplied through digital platforms, whose market values increase alongside the increased interactivity. In the second degree, quoting the survey study by Cirucci (2018) above, “women are more alienated from their digital work products than men” (p. 2964), since they spend more time on digital social networks and contribute with more social interactivity in form of relational communicative acts as well as more data points for sale. To the debate regarding the use of the exploitation concept, I argue that as data commodities with high exchange-values, people’s communications are not compensated to their full value, or even a fragment of it. That is, exploitation without force plays out between commercial infrastructures that structure access in return for personal information on the one hand and social individuals’ everyday communication needs on the other.

Conclusions

Many studies of digital labour have focused on user-generated and creative or emotional content as immaterial labour with indirect value for commercial entities, while forgetting the long history of immaterial women’s work, and without paying attention to continuities in how commercial corporations exploit feelings and leisure time for capital gain. Other studies have focused on the digital economy and the role played by data, but from the perspective of an increasingly muddled conception of privacy, which clouds over more fundamental questions concerning the reproduction and reinforcement of existing inequalities and power structures.
Scholars drawing on Feminist Marxist thoughts have highlighted the gendered dimension of digital labour, but have largely overlooked the direct exchange-value of commodified communications as one output of what I have dubbed the digital shift, a key component of social reproductive labour, which primarily, and still, rests on women. It has been a central aim of this article to push feminist critiques from the outskirts to the centre of critical data studies. By way of a communication perspective, I have answered the question of how and with what implications social reproductive labour is extended in the digital economy, as well as the subsequent question of ‘what is new’ about digital labour. The answers comprise an account of a value chain, which includes digital relational communication as a fundamental and valuable asset. Thereby, the article refocused attention to women’s social reproductive labour as crucial to capitalist accumulation in the digital economy.

References


Conclusion: where do we go from here?

The dissertation has addressed the problem of understanding what individuals are capable of in and through digital communication; how this communication is enabled, constrained, and negotiated in a datafied society; and, in turn, the implications for individuals’ everyday lives, the choices they (do not) have, and the inequalities they live with. It has generated knowledge that traverses the macro level of global and regional infrastructures, markets, and regulation, the meso level of Denmark as a particular infrastructural, economic, and political context, and the micro level of individuals’ everyday life, their agency, and understandings. The articles above deal with questions that hover around the notion of a good – digital and datafied – life by engaging with individuals’ capabilities – what they are and what they do, in and through digital communication; with the large technological systems that these beings and doings are embedded in; and with the everyday implications of these systems. Standing on the shoulders of, and contributing to, existing research in critical data studies, infrastructure studies, and political economy of communication, the articles, in other words, follow how macro structures manifest themselves in big data, harvested from micro and mundane communications, and how the everyday life is embedded in and conditioned by the same structures.

The capabilities perspective forwarded in the theoretical framework of part I outlines the contours of a research agenda that is concerned with understanding the complex, enabling and constraining, relationship between individuals and technological systems. Following the normatively individualistic, but ontologically non-individualistic outlooks of the Capability Approach, the agenda contends that particular and micro-oriented studies of digital communication cannot stand on their own but need to take into account the structures that frame individual users beyond any particular communicative event. But also, equally, macro-oriented studies of datafication, digital infrastructures, and political economies are not enough, they need studies of lived lives too, in order to grasp the implications of how digital communication is enabled, constrained, and negotiated. As a systematic approach, the framework thereby fosters analyses towards figuring out how societal structures should be organised and regulated, so as to enhance the freedom
and wellbeing of individuals. This, however, calls for further research that goes well beyond what can be accomplished within the confines of this PhD dissertation.

By way of concluding, I will return to the strong feeling of a call to action (or, perhaps, arms) that I outlined in the very introduction. My fieldwork confirmed to me how the internet, and its hardware and software, is enabling for so much of what individuals value being and doing: it connects long lost friends or strangers and helps them find love; it makes a multitude of everyday tasks just a little more convenient in an already hectic life; it relieves immigrants that are far from their loved ones; and it helps people find communities where there are none. We can also think of how it eases the work of educational institutions that are forced to teach at a distance, helps fight poverty, and confront environmental crises. There is, indeed, a lot of problems that technology can fix, yet these solutions come with their own suite of complex and unanticipated problems – problems that materialised in the fieldwork through episodes of identity theft and cybercrime, stress and anxiety, and profound feelings of powerlessness. We can also think of hate speech, polarisation, conspiracies, and other critical and contemporary phenomena. The studies in the dissertation show how digital communication is enabled by internet infrastructures – from fibre-optic submarine cables and local networks to digital applications and minute bits of content like web cookies. But also, how it is constrained by the fundamental business model behind the majority of services that individuals employ in their everyday lives – for reaching out, coordinating, discovering information, following news, and finding their way through complex city structures or difficult life situations. Business models where any service is free of monetary charge, yet paid through ever-increasing amounts of data, thereby pushing the datafication of everyday life to new limits. Business models that rely on gradually and intangibly encouraging and persuading people to live out more and more of what counts as social life through digital infrastructures. And also, business models that are, considering the multitude of arenas that these digital systems are embedded in, difficult to circumvent and even more difficult to cancel if one wants to be online. The studies show that the companies that excel at harvesting and reselling individuals’ data and their future behaviors to advertisers have grown large, immensely powerful, and difficult to rattle. They own infrastructures across
the digital value chain, and across the world, yet they remain largely unmonitored and unregulated. In a time where data excess is conceptualised as the perfect solution to countering the COVID-19 pandemic rather than a foundational tenet for the surveillance capitalism business model, we have never needed knowledge on the enabling and constraining forces of the internet more acutely.

One obvious, and final (or first, whichever way you look at it) question is then, where do we go from here? I would argue that we continue the discussion of how to enforce the good – digital and datafied – life, how to enhance entitlements and achievements that ought to lie within the freedom of any individual, and how to counter the implications that follow from not being able to make one’s own future. More than discussing these questions in theoretical terms, we also empirically interrogate how big data is made and remade and why, thereby making visible otherwise invisible surveillance infrastructures, challenging their inevitability, and imagining future alternatives. One such alternative would be to coup the contradiction between connectivity and loss of freedom. And, in turn, make a move towards a public internet that treats the millions of human beings that use it as ends rather than means towards the commercial ends of a few powerful actors. In other words, we start asking tough questions (see e.g. Klein, 2020; Zuckerman, 2020), and we begin to answer them: why is it that digital networks and data, which are obviously so essential in a time of crisis like the present, are owned and controlled by commercial actors, and not the public? Why is it that privately-owned digital infrastructures are accepted as invisible and therefore largely unmonitored? And why is it that the internet, which is evidently critical for so much of what counts as social life, is not considered a not-for-profit public utility?
References


Afhandlingen 'Human Capabilities in a Datafied Society: empirical approaches to studying the interplay between digital communication and internet infrastructures' er en undersøgelse af forholdet mellem individer og teknologiske systemer. Gennem en række teoretiske, metodiske og empiriske analyser belyser afhandlingen, hvordan digitale infrastrukture rammesætter individers muligheder for at leve et værdifuldt liv og øve indflydelse på de måder, deres data høstes, distribueres og anvendes. Undersøgelserne kredser dermed om et overordnet problem, der består i at forstå: hvilke handlemuligheder individer har for at kommunikere digitalt; hvordan deres kommunikation muliggøres, begrænses, og forhandles i et dataficeret samfund; og med hvilke implikationer.

Del I etablerer en fælles teoretisk ramme for afhandlingens artikler ved at introducere et perspektiv på kommunikation som 'capability'. Capabilities perspektivet (the Capability Approach, Sen 1980) udgør en konceptuel og normativ optik til at undersøge og forstå, hvad individer kan og ikke kan, hvorfor, og med hvilke konsekvenser. Del II indeholder i alt syv artikler, der bevæger sig fra et mikro-niveau, der fokuserer på individers handlemuligheder, når det kommer til digital kommunikation; et makro-niveau, der beskæftiger sig med de teknologiske systemer som disse handlemuligheder er indlejret i; og tilbage til implikationerne af disse systemer for individers hverdagsliv.

Artikel 1 udvikler et forskningsdesign til kvalitative og komparative feltstudier af digital kommunikation i hverdagen. Artikel 2 og artikel 3 fokuserer på de infrastrukturelle, økonomiske, og politiske forhold, der rammesætter digital kommunikation ved dels at oprette en komparativ systematik til at kortlægge digitale kommunikation systemer i forskellige nationale kontekster, og dels at aplicere denne i et komparativt, historisk studie af internettets udvikling i Danmark. Artikel 4 og artikel 5 undersøger overvågningsinfrastrukturen i politisk ekonomi og forskellige kontekster i web- og app-baseret tracking. Artikel 6 og artikel 7 vender tilbage til data fra felt-
afhandlet, introduceret i artikel 1, og undersøger implikationerne af digital infrastrukturer og dataficerering, med fokus på mobilkommunikation og samfunds-mæssige uligheder. Endelig præsenterer Del III afhandlingens konklusion.


This dissertation is about the relationship between individuals and technological systems. It contributes to emerging understandings of the impact of digital infrastructures on individuals’ abilities to live the lives they value and to make their own futures. The dissertation deals with a three-fold problem of understanding: what individuals are capable of in and through digital communication; how this communication is enabled, constrained, and negotiated in a datafied society; and, in turn, the implications for individuals’ everyday lives, the choices they (do not) have, and the inequalities they live with.

To approach this problem, I introduce, in the first part of the dissertation, a perspective on communication as capability. The capabilities perspective (the Capability Approach, Sen 1980) offers a conceptual and normative lens through which to approach and understand what individuals can and cannot do, why, and with what implications, and it sets the agenda for the articles of the dissertation and for future research. Part II comprise a total of seven articles, and follow a trajectory...
that moves from the micro level of engaging with individuals’ capabilities — what they are and what they do — in and through digital communication; to the macro level of large technological systems that these beings and doings are embedded in; and back to the implications of these systems for individuals’ everyday life. Article 1 offers a research design for qualitative comparative studies of digital communication in everyday life. Article 2 and article 3 focus on the infrastructural, economic, and political conditions that frame everyday communications in different societal contexts. The first develops a comparative framework for mapping digital communication systems, and the second applies that framework in a comparative, historical analysis of the evolution of the internet infrastructure in Denmark. Article 4 and article 5 approach the political economy of surveillance infrastructures by analysing the use of third-party services on the web and in apps. And article 6 and article 7 return to the fieldwork data introduced in article 1 and look into the digital infrastructures and datafication for individuals’ abilities to make their own future and to counter persistent and exacerbating inequalities. Finally, Part III features the dissertations’ conclusion.

The dissertation makes three key contributions: a theoretical, a methodological, and an empirical one. First, it develops theoretical frameworks that combine critical data studies, infrastructure studies, and political economy of communication, and by offering a systematic way of understanding connections between digital communication and capabilities. Second, it contributes with developing new tools and methodological frameworks for analysing and monitoring nascent and fast-moving phenomena like datafication and digital business models in comprehensive ways and at macro, meso, and micro levels of inquiry. And third, the dissertation adds to existing research through empirical and systematic studies that uncover central mechanisms of big data, make otherwise invisible surveillance infrastructures visible, and interrogate the tenets of infrastructural power.