Big data observations made upon social media and policy making

Vydra, S.

DOI
10.4233/uuid:09be0fbc-fdb5-4097-a4d0-34cc205322c4

Publication date
2023

Document Version
Final published version

Citation (APA)
Vydra, S. (2023). *Big data observations made upon social media and policy making*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:09be0fbc-fdb5-4097-a4d0-34cc205322c4

Big Data Observations

Made upon

Social Media and Policy Making

Dissertation

for the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,

chair of the Board for Doctorates

to be defended publicly on

Monday 5th June 2023 at 17:30 o'clock

by

Simon VYDRA

Master of Science in Political Science, Vrije Universiteit Amsterdam, the Netherlands

born in Prague, Czech Republic

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof.mr.dr. J.A. de Bruijn	Delft University of Technology, promotor
Prof.dr.ing. A.J. Klievink	Leiden University, promotor
Prof.dr. M.O. Hosli	Leiden University, promotor

Independent members:

Prof.dr. M.E. Warnier	Delft University of Technology
Prof.dr.ir. N. Bharosa	Delft University of Technology
Prof.dr. S.W. Cunningham	University of Strathclyde, Glasgow, Scotland
Dr. A.F. van Veenstra	TNO, The Hague, Netherlands
Prof.dr. M.J.G. van Eeten	Delft University of Technology, reserve member

Other members:

Dr. J.J. Kantorowicz	Leiden University

This research was funded by the European Research Centre for Economic and Financial Governance (EURO – CEFG) – a cooperation between Leiden University, Delft University of Technology, and Erasmus University Rotterdam.

Cover by: Maggie Mulley

Available at the TU Delft open access repository

Table of Contents

Summary

This dissertation focuses on the research problem of divergence between the theoretical promise of big data use for public policymaking and the empirical support for that promise. To address this problem the dissertation asks three sequential research questions: Firstly, why does this divergence exist. Secondly, how can it be improved. And lastly, how to design and carry out research capable of these improvements. These questions form the theoretical core of the dissertation and effectively propose a research approach to studying big data use in public policymaking. However, merits of a research approach can only be demonstrated by research that adopts it, which his why this dissertation also researches a more practical problem: The difficulty of measuring and evaluating 'social investment' policies. The dissertation asks "Can social media data be used to operationalize and measure social investment?", which involves a set of research sub-questions focusing on how can requisite information be extracted, what such extraction implies for policymaking, whether the requisite information is present in Twitter data, and whether it changes between a period of normalcy and a period of crisis. Answering this research question in a more design-oriented proof-of-concept way, but while utilizing the proposed research approach, is relevant both for the research problem of operationalizing social investment as well as demonstrating the merits and pitfalls of the research approach proposed as a solution to the primary research problem.

Methodologically, the theoretical core of this dissertation relies primarily on literature review and systematization. The proof-of-concept research is more design-oriented in terms of methods and designs software artefacts capable of summarizing policy relevant information under two definitions of policy relevance. The data set this dissertation utilizes is a year-long dataset of tweets focused on early childhood education and care and labor market policy, as well as a three-month dataset of tweets from the first wave of the COVID-19 pandemic matched to the same three-moth period from the previous year to account for seasonality. The artefacts themselves are built primarily using Natural Language Processing (NLP) methods, such as topic modeling and latent semantic scaling (LSS), and aim to summarize Twitter data in ways that allow identification and exploration of policy relevant information. This allows the artefact to answer the descriptive question about whether policy relevant information exists and what it entails, but also to arrive at scientifically relevant design principles.

With regards to the primary research problem – the divergence between the theoretical promise and empirical support – the dissertation argues that it exists because the two archetypical narratives about big data in the private sector talk past one another too often. The two narratives (techno-optimism and policy-pessimism) are based on a focus on the underlying technology of big data analysis and on a focus on policy/political decision making respectively, individually explaining either successful adoption or lack thereof, but neither explaining the uneven adoption observable in the real world. With respect to the proof-of-concept part of this dissertation, it finds that Twitter does not contain a meaningful amount of policy-relevant information under either definition of policy relevance. The commentary people provide is often focused on a given policy or a given life course transition, but the commentary tends to be more political in nature rather than a disclosure of personal situation. The content on Twitter does change between periods of normalcy and crisis in policy-relevant ways, but the amount of content remains far too small and 'noisy' for this change to have practical policymaking ramifications. In exploring policy relevant information the dissertation also makes relevant observations about the methods employed, such as the interpretability of a range of topic models for short microblogging data, or the feasibility of utilizing LSS for uses outside of language independent sentiment analysis (its original use). Interestingly, the dissertation also finds that many of the analytical decisions one has to make to analyse microblogging data, even though they can be made in an informed way, have a subjective or a normative dimension. These decisions range from cleaning the data to visualising results.

With regards to techno-optimism and policy-pessimism not being integrated well to explain the uneven adoption of big data in policymaking this dissertation argues that focusing on individual combinations of policy questions, data sources, and methods (rather than a 'general theory' of big data and policymaking) is helpful. In other words, not asking **if** big data changes policymaking and how, but rather **when** (under what conditions) does big data change policymaking and how. At that more specific level both narratives can contribute relevant promises and pitfalls that can be traded-off between one another. Using this research approach (and the logic that motivates it) in the design oriented parts of this dissertation also shows how this approach can play out in practice: The research is more likely to reach a 'null' finding due to a more comprehensive notion of policy relevance and requires a thorough justification of the alignment between policy problems, data, and methods. In turn, it seems capable of answering the 'when' question of big data use; In the particular case of using Twitter data for economic policymaking the dissertation suggests (together with the existing literature) that Twitter data is used most successfully when utilized for information other than what users disclose in the texts of their tweets. Working 'bottom-up' from specific cases towards a more generalized understanding of big data in policymaking also seems feasible, with some findings applying only to the specific data and method, some applying to microblogging data and NLP more generally, and few applying to big data analysis in general.

With regards to the design-oriented part of this dissertation the lack of policy-relevant information in Twitter data tempers some of the optimism found in the existing literature. In a case where there is a good alignment between the data, methods, and policy questions (such as the one studied) we would expect the design exercise to yield more success than it does if we are to draw on the existing literature. The finding of many decisions having both a technical as well as subjective dimensions has important implications for policymaking practice, where technical experts are often not politically accountable and accountable decision makers often lack the technical expertise. The solutions one can find in the literature (such as assessment of the stability of model outputs) is suited well for a research context, but not for the policymaking context. The solution, at least as far as this dissertation is concerned, is unlikely to be purely analytical and lies in the interaction between decision makers and analysts and in designing analytical systems in ways that fosters this interaction in a manner that preserves both the democratic accountability of decision-making as well as the technical soundness of underlying evidence.

Chapter One: Introduction

Anyone interested in big data who made it past the cover page of this dissertation is now rightfully confused. Why the dated title and imagery? The expectation here is for words like 'revolution', 'new age', or 'paradigm' accompanied by imagery of the entire globe visualized as vertices and edges or ones and zeros – something conveying that we are digitizing the whole world. For bonus points the cover image would also include a human hand to convey how 'at our fingertips' all of this is. Not here. The title page is a nod to John Graunt's *'Natural and Political Observations Mentioned in a following index, and made upon the Bills of Mortality'* (Graunt, 1662) and the images on the front page of this dissertation are inspired from those included in Graunt's (1662) book.

John Graunt's book documents him gathering and using records of christenings and deaths from parishes, which were known as bills of mortality. He gathered "as much matter of that kind, even as the Hall of the Parish- Clerks could afford me" (Graunt, 1662: 1-2), constituting an overwhelming amount of records for that time, especially given that the records were not originally intended for his purposes: They were for sale and primarily used to gauge the spread of plague "so the Rich might judge of the necessity of their removal, and Trades-men might conjecture what doings they were like to have in their respective dealings" (Graunt, 1662: 1). Graunt himself praised the practice of keeping these records despite the lack of other explicit use for them: "Now, I thought that the Wisdom of our City had certainly designed the laudable practice of takeing, and distributing these Accompts, for other, and greater uses then those above-mentioned, or at least, that some other uses might be made of them" (Graunt, 1662: 1). He structured this data into tables and combined data from multiple parishes to allow him to eventually estimate the population of London, population movements, or study associations and time-series using statistical inference (Sutherland, 2013). By doing so, Graunt (arguably) founded the discipline of Statistics (Sutherland, 2013). In other words, he collected overwhelming amounts of already existing records to repurpose them for providing a new insight. He did so by structuring, cleaning, and combining data from various parishes and then analyzing it in novel ways. If we omit that Graunt did not utilize any real-time data streams, he was not only the first statistician, but arguably also the first big data analyst. Even though John Graunt's work was 360 years old last year, I will spare the reader the obvious 'coming full circle' comment.

I draw this parallel between the very inception of statistical reasoning and big data analysis not to downplay any and all differences between the two, but to immediately outline the perspective from which this dissertation approaches the topic; Big data does change things and it does spawn new questions about machine learning, artificial intelligence (AI), computing infrastructures, and many other subjects. But it also does not fundamentally change some questions - questions about inference, objectivity, fairness, power, or decision making. Many such questions, especially about governance, are rephrased or reframed by the advent of big data, but they are not answered or made obsolete. These are precisely the questions this dissertation focuses on. This is an important disclaimer as parts of the dissertation are design-oriented and they utilize and even propose approaches that fall under the umbrella of big data analysis, natural language processing, or machine learning. As much as these approaches are technical and produce technical artefacts, they are not used here to answer technical questions about the underlying models, their training, or visualization of their results. Rather, they are used to answer questions about how such models and approaches interact with the process of policymaking.

The starting point of this dissertation, defended in chapter 2, is that both novel technical questions as well as established governance questions are important for understanding the adoption of big data in policymaking practice. Currently the literature is leaning too much towards a technical approach to big data, which is also argued in chapter 2, but can be understood more intuitively by the reader's own reaction to the title page of this dissertation: How unexpected it is to associate the study of big data with 'age-old' research problems, despite the undeniable similarities.

This results in a very optimist perception of the transformative potential of big data (sometimes labeled as 'hype'). Some even argue that we are entering a second machine age, implying that computers and big-data-enabled analysis remove mental power constraints much like the invention of the steam engine removed physical power constraints (Brynjolfsson & McAfee, 2014). For social science specifically, the impact of big data can arguably "be compared with the impact of the invention of the telescope for astronomy and the invention of the microscope for biology (providing an unprecedented level of fine-grained detail)" (Hilbert, 2015: 136). Study of public administrations and governance, disciplines that this dissertation is most closely related to, share a portion of this 'hype' with some authors arguing that not only do "public bodies using big data achieve significantly more positive outcomes and benefits" (Maciejewski, 2016: 127), but also that big data "will profoundly change how governments work and alter the nature of politics" (Cukier & Mayer-Schoenberger, 2013: 35).

Despite these high hopes, the adoption of big data appears to be a slow and uneven process that takes different forms and happens at different speeds based on the institutional and policy context (Klievink et al., 2017). This is observable globally as certain policy areas see much more big data use than others, but also in regional case-studies that often conclude that "there is still little knowledge of the conditions and determinants for its [big data's] application, especially in public policy domain" (Misuraca, Mureddu, & Osimo, 2014: 176), or that "we cannot fully account for the lack of widespread diffusion of the innovative localized [big data] use practices" (Chatfield & Reddick, 2017: 346). Too much emphasis on technical factors is likely at the root of our inability to explain the diffusion of big data analytics since adopting IT solutions in public administrations resembles a "mixture of political behaviour, intuition and the exploitation of emerging opportunities, whereas technical rationality plays a minor role" (Nielsen & Pedersen, 2014: 419). As a result, the existing literature struggles to explain the uneven adoption of big data analytics for policymaking and the (lack of) change to policymaking practice this entails.

This constitutes the primary research problem of this dissertation: The divergence between what existing public administration literature theorizes the impact of big data to be and what it actually is in policymaking practice. This dissertation of course cannot 'resolve' this problem, as it concerns the existing body of literature, but it can chart a course capable of avoiding this problem in the future. To do so the dissertation asks a set of sequential questions: Firstly, why does the research problem exist? This is a rather theoretical question that requires systematizing of the existing literature. Secondly, how can this divergence be improved? This is another theoretical question, albeit a bit more pragmatic, and essentially asks what type of research is required to eventually reach an understanding more aligned with policymaking practice. Lastly, it is the question of how to design and carry out such research? What focus and methods should it include and how can it best be related back to our overarching understanding of big data use in public administrations and policymaking.

This last question is what most of this dissertation focuses on and in that regard it 'walks the walk': It adopts a design research approach and attempts to pilot a method for utilizing a particular big data source to inform policymaking in a salient policy area. It essentially does what a lot of the current 'proof-of-concept' literature does, but does it in light of the research problem of divergence between literature and practice as well as the proposed solution to that problem: Focusing not only on technical features of data and analysis, but also on policy and decision-making implications of those technical features. This essentially results in a design-oriented research project conducted from the perspective of public administrations and policymaking, following a set of prescriptive arguments aimed at resolving the divergence between literature and practice. This makes the dissertation relevant in two ways: Firstly, there is the 'proof-of-concept' design research that attempts to utilize a particular big data source for a particular policy puzzle that could benefit from it. This comes with descriptive findings about whether the requisite information is contained in the data source and methodological findings about how that information can or cannot be extracted. This contributes to our understanding of using a particular big data source and associated methods, potentially fuelling or tempering the 'hype' about their utility for policymaking. Secondly, since this research is carried out in light of the primary research problem and follows an approach that this dissertation argues can remedy it, it illustrates how the proposed approach plays out in practice, what it adds to our understanding, and even evaluates (to some extent) whether the proposed approach 'holds up'. It is normal for design research to have multiple relevant outputs, such as the artefact itself, descriptive findings, or 'lessons learned' along the way that constitute design principles. This dissertation is no different, but the lessons learned are not limited to whether the artefact demonstrably meets its goal and instead encompass many of the 'age-old' questions referred to above - questions about inference, power, or accountability.

To achieve this goal the dissertation starts by addressing the first and second questions in **chapter two**. In this chapter the dissertation systematizes the existing literature by constructing two archetypical narratives – 'techno-optimist' and 'policy-pessimist' – and arguing that even though most contributions situate themselves somewhere in the middle of those extremes, the disciplinary foundations of these two narratives result in them often 'talking past one another' and only paying lip service to each other: 'Techno-optimism' is built on a focus on the data and realizing that data, and our ability to process data, are developing rapidly and opening up new opportunities. 'Policy-pessimism' on the other hand is built on a focus on politics and public administration and realizing that, despite technological change, a lot of the interests and processes involved in policy making remain entrenched and unchanged. Each of these narratives can provide part of the explanation for the uneven diffusion of big data in public administrations, but individually they do not ask the 'when question' – under what conditions are big data initiatives transformative and under which conditions are they not? Asking such a question concedes some of the ambition for a 'general theory' of big data and policymaking, but in doing so it proposes a level of analysis at which both narratives become specific enough to be synthesized and to contribute 'promises' and 'concerns' for a specific big data use case.

In doing so chapter two answers the first question by diagnosing why the primary research problem exists – the insufficient integration of the two archetypical paradigms in the extant literature. It also provides a general answer to the second question of what can be done about this by proposing a research approach that can combine these two narratives in a more meaningful way. This approach is rather simple and proposes an appropriate level of 'aggregation' for talking about big data in policymaking: As tripartite combinations of data, method, and a policy question. Generalizations towards 'big data' and 'policymaking' should stem from a careful empirical study of individual tripartite combinations. Furthermore, it also sets a relatively high level of interdisciplinarity such an approach requires: Data is primarily studied by statisticians, methods of big data analysis by computer scientists, political institutions by political scientists, policy questions and policymaking processes by public administration scholars and domain experts. To study individual cases more thoroughly all these disciplines need to be drawn from, as features of big data use are not constrained by the neatly outlined disciplinary lines we work in. For example, it is impossible to evaluate the ethics of big data use without understanding how the data is actually being transformed and used. And it is equally impossible to evaluate a big data algorithm fully without understanding the policy question it is meant to help answer and the actors utilizing it. The ethical conundrums can be hidden in the details of data transformation and model parametrization the same way that performance of an algorithm can be driven by (mis)understanding of what policymakers want to know and domain-specific knowledge.

With the first two questions preliminarily answered and the research approach set in chapter two, the dissertation switches focus to the third question of how to design and carry out a research project following the lessons from chapter two. There is of no course no one correct answer here, as there are various combinations of data, methods, and policy questions that can be studied using various research approaches. Owing to its design-focused research approach, this dissertation must outline a potential big data use case with substantial promise (in the sense of data sources, methods, and policy questions aligning well) and try to design a system for that particular tripartite alignment. The upside of this approach is that it allows the dissertation to make a contribution to the domain the specific policy question comes from by producing a proof-of-concept approach to answering a salient policy question with appropriate data and method. In other words, contributing to answering a salient policy question is not only significant as an illustration of the research approach proposed in chapter two; It is also significant more directly – it can solve a practical problem as well as advance the domain from which a particular policy question comes from.

Chapter three specifies the domain of this dissertation to social investment (as a particular school of thought in comparative welfare state studies). It advances a theoretical argument about social investment and its analytical implications, which in and of itself contributes to social investment by explicitly outlining its analytical requirements, assumptions, and difficulties. Outside of the theoretical importance of this argument, chapter three also established practical relevance by arguing that the social investment approach is demonstrably adopted by policymaking institutions as salient, while the need for better methodological tools to translate this approach to practice is simultaneously acknowledged (and worked on) by the same institutions. But, most importantly for this dissertation, chapter three starts creating the tripartite alignment necessary for a design-oriented proof-of-concept research by arguing that the difficulties associated with social investment policy analysis are a good match for what big data analysis offers. This establishes the alignment between certain policy questions (and the logic underlying them) and use of big data.

Chapter four then further refined this tripartite alignment by specifying the general alignment between social investment policy questions and big data to specific policy questions (key labor market transitions), big data source (social media data), and methods (topic modelling and supporting NLP techniques). This tripartite alignment delivers on the third research question of this dissertation by essentially proposing (and carrying out) an instance of the type of research this dissertation argues is necessary to resolve the divergence between literature and practice. This chapter primarily asks whether Twitter data contains the type of information necessary to provide insight into certain life course transitions. In some ways this is a precursor question to more serious (action) design research – research involving aggregation and presentation of such insights, or even involving policymakers to gauge their reaction to and utilization of such insight. However, if the results are interpreted honestly, the amount of policy relevant information proposed methods can identify is extremely low, making it of no real value to policymaking. This descriptive finding, albeit simple, puts a stop to more ambitious design exercises utilizing this data. In terms of design principles, the chapter is much more successful, providing observations about the subjective/normative nature of many 'analytical' decisions that are necessary to conduct this type of analysis.

Chapter five builds on the conclusions of chapter four – conclusions that spawn more questions than they provide answers to. With regards to the descriptive findings, two avenues are possible: Change the data in order to further assess the extent to which big data do not contain social investment insight, or change the policy relevance criteria to further assess the extent to which Twitter data does not contain policy relevant information. This dissertation opts for the second option with an important expansion of the data set to include the first wave of lockdown policy response to COVID-19. This expansion, together with loosening of the policy relevance criteria, creates a 'most likely to succeed scenario' as the two primary policy areas of interest are employment (and job search more generally) and early childhood education and care. Both of these policies were very heavily affected by COVID-19 lockdowns, creating a scenario where the amount of information shared on social media is likely to be comparatively very high as these changes create difficulty in people's lives. This essentially allows the dissertation to rule out the possibility that the lack of information identified earlier is caused by the 'normalcy' of the selected time period and institutionalization of existing policies: Very consistent policy provisions and an understanding of what they are and how can be utilized could result in those policies simply not being a large topic in the public discourse. Confirming the findings from chapter four even in this 'crisis' context verifies that the lack of identified information is not simply due to the time period for which data was collected. More importantly than that, this chapter also expands on chapter four in terms of methods by proposing and utilizing a more robust approach to extracting insight and measuring the difference between two time periods. Some of these methodological changes here are not motivated only by the technical metrics - they also consider the role of the previously identified subjective decisions in the process and design the new method to facilitate input and understanding at very specific points in the process without the necessity for technical understanding. This would allow policymakers to be integrated more and to exercise more control over some of the more subjective decisions and overall focus. The method itself (the design artefact) seems to function as intended, but the negligible amount of relevant data found and not including policymakers in the research process does not allow for a more thorough evaluation. In concluding the dissertation discusses the limitations of its various findings and demonstrates the relevance of its findings with respect to the existing literature and the overall ambition to improve the alignment between theoretical promise and policymaking practice.

Chapter Two: Techno-optimism and policy-pessimism in the public sector big data debate[1]

This chapter addresses the research problem of divergence between academic literature and policymaking practice primarily by diagnosing the cause of this divergence. As alluded to in the introduction, this chapter argues that this divergence is caused by the two major narratives often talking 'past one another'. The two narratives we construct for this purpose are the following: First, a narrative focused on the study of big data analytics as a technological phenomenon, focusing on its comparative (dis)advantages to how 'traditional' data is created, handled, and analysed, often rooted in engineering and computer science disciplines (see for example Dong et al., 2017; Dumbacher & Hutchinson, 2016; Ku & Leroy, 2014; Misuraca, Mureddu, & Osimo, 2014). Second, a narrative focusing on decision-making and the study of how quantitative evidence and the advent of big data interacts with political and bureaucratic decision-making, often rooted in public administration and organisational decision-making disciplines (see for example Desouza & Jacob, 2014; Dunleavy, Margetts, Bastow, & Tinkler, 2005; Giest, 2017; Janssen & Kuk, 2016; Klievink, Romijn, Cunningham, & de Bruijn, 2016). If we would put these two narratives to the extreme – by limiting our focus purely to technology or political decision-making and accepting the underlying assumptions of these narratives as axioms - we could argue that the first narrative is optimist and the latter is pessimist with regards to the impact of big data on policymaking. We attribute this difference to the fact that technology evolves and is adopted very rapidly compared to how slowly political and governance practices change, making the technical narrative optimistic and the policy and decision-making narrative pessimistic about the magnitude of change big data will have on public sector and governance in general. We term these two extremes 'techno-optimism' and 'policy-pessimism'.

Even though these two narratives differ primarily in focus and optimism, this difference translates to important aspects of talking about big data, including something so fundamental as how we define it: The most common big data definition uses a set of 'Vs' – attributes along which big data differs from 'normal' data. Most commonly these V's are volume, variety, velocity, and veracity (IBM, 2012; Ward & Barker, 2013), but sometimes also include variability, visualisation, and value (for review of definitions see Ylijoki & Porras, 2016). This way of defining big data itself seems to be rather techno-optimist, as the attributes are primarily technical and describe the nature of the data itself (except visualisation and value, which are not commonly used). The policy-pessimist definitions of big data revolve around the social change big data motivates, especially in terms of changes to decision-making processes necessary to make use of big data (Kim et al., 2014). These definitions refer to the usage of structured and unstructured data (potentially in combination) from multiple sources both

[1] This chapter is originally published as a research article: Vydra, S., & Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly*, 36(4), 101383. More details are provided in the 'Authorship Contribution' section.

internal and external to an institution, the use of high-frequency data streams, and the use of data for radically different purposes than it was originally intended for (if there was an intent to begin with) (Klievink et al., 2017). Such definitions immediately emphasize the challenges of deriving relevant insight from data and how this insight is used by individuals in making decisions. This makes the two narratives differ not just in their focus but in terms of the fundamental ‘unit of analysis’: Techno-optimism focuses on data and analytical output whereas policy-pessimism focuses on humans turning data into insight and humans making decisions in bureaucratic structures (with the help of that insight).

Much of the literature on big data in the public sector has ingredients of both narratives, yet whether consciously or not, tends to emphasise or be based on one of them. As alluded to in the introduction, the emphasis currently seems to be on the techno-optimist side. Yet, we do acknowledge that an unequivocal distinction is very hard to make, as even rather techno-optimist accounts pay lip service to decision-making and politics (Höchtl et al., 2016; Maciejewski, 2016). In fact, even the more policy-pessimist accounts pay lip service to the big data promise and do not dismiss it outright (Iacus, 2015; Lavertu, 2016). Thus, despite our diagnosis of a techno-optimist bias, it is important to note that majority of the existing contributions are not openly and unequivocally techno-optimist and they do address relevant shortcomings, but do not do so systematically or comprehensively (Bertot & Choi, 2013; Chatfield & Reddick, 2018; Einav & Levin, 2013; Katal et al., 2013; Ku & Leroy, 2014; Misuraca et al., 2014; Sagiroglu & Sinanc, 2013). The result of offering only lip service to (as opposed to systematically addressing) the ‘opposing’ perspective and cherry-picking easy-to-address concerns is that many contributions talk past one another, rendering the existing literature incapable of explaining why is the diffusion of big data analytics in the public sector uneven. We aim to help this predicament in two ways: Firstly, we challenge key techno-optimist arguments from a more policy-pessimist lens, thus illustrating its value for interrogating big data use in the public sector. Secondly, we structure the current debate by articulating these two archetypical narratives and making them meet ‘eye-to-eye’ with the ambition of helping scholars to interrogate their work more systematically and to determine the research approach adopted in the remainder of the dissertation.

To do so we need to first disentangle the techno-optimist narrative into key arguments and assumptions, which in itself is a difficult task for two reasons: Firstly, because the benefits and shortcomings of big data articulated in the literature are numerous and how these should be aggregated into ‘key arguments and assumptions’ is not obvious. Secondly, since existing literature situates itself between the two extremes but not directly on them, it is not possible to directly extract the archetypical techno-optimist narrative from a specific contribution. In other words, we construct techno-optimism as the logical extreme of arguments we identify in the literature, but our construction of techno-optimism and policy-pessimism remains a heuristic fit for the purpose of this chapter rather than a robust categorization to sort the current literature by.

That said, to provide a structure for his chapter we disentangle these two archetypical narratives into four aspects of big data analysis they fundamentally differ on: Firstly, the quality of the data insight and subsequent decision-making. Secondly, the speed of data analysis and subsequent decision-making. Thirdly, the epistemological foundation for the analytics process. Fourthly, overcoming some of the fundamental concerns relevant to big data analytics (in this chapter we focus on privacy as an exemplary concern). These four key arguments and assumption are selected because of how foundational they are to the big data in public sector debate (conceptually and in terms of being covered by existing literature), but the two opposing narratives can be constructed using less aggregate and more context-specific set of key arguments and assumptions. These four key arguments and assumptions will be addressed in **sections 2.1 to 2.4** in the order listed above, with each section first briefly outlining the techno-optimist argument for that aspect followed by highlighting shortcomings of that argument. In **section 2.5** we conclude by summarizing the techno-optimist and policy-pessimist narratives for the four key arguments and assumptions that we deal with in this chapter, making the two narratives meet 'eye-to-eye' and highlighting some crucial questions to interrogate research with based on these narratives. In concluding we also offer our take on reconciling the two narratives in a 'best-of-both-worlds' fashion by adopting a more granular approach and focusing on a tripartite alignment between specific big data sources, methods, and policy questions - a level of analysis at which trade-offs can be meaningfully made.

2.1 How bigger does not always mean better in public decision making

A fundamental argument of the techno-optimist narrative is that big data will provide better information and that this better information will in turn facilitate better decisions. The argument essentially claims that "[t]he more quality and accurate information is available, the better the decisions will be." (Höchtl, Parycek, & Schöllhammer, 2016, p. 152). This notion is based on understanding policy decisions as based largely on empirical input and improving this input then resulting in better regulatory policy (Maciejewski, 2016). This input can of course be (and often is) an estimate, leading some to argue that "[i]f we improve the basis of prior information on which to base our estimates, our uncertainty will be reduced on average. The better the prior, the better the estimate, the better the decision" (Hilbert, 2015, p. 135).

How exactly will data (and subsequently decision making) be "better" is often left unexplained, but some authors provide a bit of elaboration: Maciejewski (2016) argues that using big data methods results in more accurate decision-making due to expansion of databases, more extensive analytics, and better data visualization and presentation (Maciejewski, 2016). Other authors focus on the overall efficiency gains in the private sector triggered by big data analytics, arguing that it is reasonable to expect similar developments in the public sector (Y.-C. Chen & Hsieh, 2014). In other words, the notion of 'better' can be related to an increase in accuracy (Höchtl et al., 2016), a reduction in uncertainty (Hilbert, 2016), or efficiency gains (Y.-C. Chen & Hsieh, 2014) and is applied to both the insight we can derive from data as well as the decision we make based on this insight.

In this section, we tackle both of the assumptions this argument rests on: That big data provides better insight and that better insight translates into better policy decisions. In sub-section 2.1.1 we point to the various important aspects of data quality that make it impossible for a big data source to be 'better' for policymaking in general. In sub-section 2.1.2 we point to factors other than data that influence the quality of public decision-making, thus complicating the link between better data and better decisions.

2.1.1 The myth of 'better' information

Taking accuracy and uncertainty as two aspects of data quality highlighted by the techno-optimist argument, it is important to point out that not all big data sets by default allow for more accurate insights: Firstly, big data sources often struggle with substantial representativeness problems that have been described both empirically and conceptually (Hargittai, 2015; Keith et al., 2016; Liu et al., 2016; Ruths & Pfeffer, 2014; Samarajiva & Lokanathan, 2016), making the resulting insight skewed and thus inaccurate in that sense. Secondly, big data often contain much more 'noise' than 'signal' and this noise has to be removed to arrive at reliable conclusions (Iacus, 2015; Scannapieco et al., 2012; Vaccari, 2015), which in itself presents an analytical challenge that introduces inaccuracy (since it is impossible to perfectly distinguish signal and noise). Because of these issues, national statistical institutions are currently primarily focused on creating quality assurance processes for big data sources (Boettcher, 2015; Dumbacher & Hutchinson, 2016; Eurostat Big Data Task Force, 2014; Hackl, 2016), rather than actually using big data for official statistics and policymaking. In fact, when compared to traditional survey-based measures that can be crafted to accurately categorize every individual (Kitchin & Lauriault, 2015), accuracy of big data tends to become more of a concern rather than a demonstrable benefit.

More importantly, accuracy is not the only (and arguably not the most important) attribute of data for policymaking. When it comes to big data, "[Data] [q]uality is composed of several elements, such as accuracy, reliability, relevance or timeliness" (Eurostat Big Data Task Force, 2014, p. 13). When it comes to reliability, arguably the most important metric, big data often perform rather poorly. Reliability in this case refers to the trust policymakers have in a specific indicator, which is established by having a good track record of accuracy and relevance for policy questions (Kitchin & Lauriault, 2015). To amass such a track record, an indicator needs to have a good and long backrun (how far back is the data available for). Given how crucial *data backrun* is for establishing reliability (demonstrated by the Bank of England deciding not to use big data based on insufficient data backrun (McLaren & Shanbhogue, 2011)) and how overlooked the concept is in the current big data debate, it deserves more elaboration: There are broadly speaking four reasons for why data backrun is crucial for data quality. Firstly, better temporal coverage of a data set allows traditional statistical methods to generate better inferential leverage. Secondly, it provides crucial contextualization to any data insight: A 'spike' in an indicator is of little use unless we can compare it to historical data showing how these spikes play out in social reality and how they react to different policies. Thirdly, and perhaps most importantly, as crucial indicators build up a reliable backdrop they get institutionalized into domestic and international policymaking practice. This creates decades of negotiated knowledge between experts, politicians, and institutions on how to measure and adapt these concepts to assure their continuous usefulness (consider the ICLS conferences on Labour market statistics organized by the ILO (Hussmanns, 2007) as an example of such negotiated knowledge and institutionalization). Lastly, this institutionalization also achieves international comparability, which big data sources struggle with as they vary greatly from country to country and cannot really be controlled by a statistical institution since much of big data is privately owned. The example of data backdrop illustrates that the debate about data 'quality' is far more nuanced than a techno-optimist narrative sometimes conveys.

Besides the multiple dimensions of data quality, the argument has a more conceptual (but no less important) dimension: Information can be seen as a contested commodity (Peled, 2014; Ruppert et al., 2017) and the use of big data depends on context and governance practices which may be subject to shifting ideals, and the (social) concepts the data ought to represent may be "negotiated, abbreviated and contested" (Robertson & Travaglia, 2015, ¶ 6). In other words, data are objects of knowledge but also power (Ruppert et al., 2017), meaning they cannot be universally "better" in a non-partisan way (Goldston, 2008). This begs the question for whom is big data better, or alternatively for what purpose is it better for. In answering such a question it is vital to understand where and how data are produced, how data are used, and what gets lost along the way: To appreciate what big data analysis tells us, we also need to know what wasn't measured, what data got filtered out, and why (McNeely & Hahm, 2014). This is not just analytical good practice – in some cases politicians are more receptive to big data evidence if they understand how and from which specific group of individuals was the data gathered (Panagiotopoulos et al., 2017). Furthermore, the generation, collection, storage, and processing of big data are done using information systems and algorithms that are perceived to be neutral (McNeely & Hahm, 2014) , but are in fact part of bureaucratic systems and structures that are inherently political (Janssen & Kuk, 2016b). All these details make it impossible to assert that big data is somehow objectively 'better'.

2.1.2 How 'better' information translates to 'better' decisions

The transformation of insight from data into policy is by no means a straight forward process and the emergence of big data influences it as well as it does the data itself. The idea that better information leads to better decisions assumes a rather linear view of policy making, where information only enters at certain places, often represented in terms of a 'policy cycle' (Helbig et al., 2015). Yet, decisions and policy processes often do not work that way. Rather, they are the product of multiple, interacting actors, that are interdependent and are hard to commit to a common problem, solution or even the value of 'facts' (de Bruijn & Ten Heuvelhof, 2008). The result is a complex policy battle in which decision-making often takes place through small, incremental steps (Lindblom, 1959) and consist of several iterations between processes, making it a plate of spaghetti rather than a cycle (Klijn & Koppenjan, 2015). In the case of big data, the process between information and decisions is subject to politicization in at least two distinct ways.

Firstly, there is the issue described above; transforming big data into information and insights is not a politically neutral process much of which depends on who decides what data is worth, what is included, what is excluded, how data are aggregated, etc. Not to mention, these concerns can often be 'hidden' in complex algorithms and thus extremely difficult to interrogate (Janssen & Kuk, 2016b). Secondly, there are political decisions to be made not only in interpreting the data, but also in gathering it; the algorithms used to capture insights from big data reflect specific conceptions of social phenomena, including preconceptions about factors of importance, expected correlations, or contested assumptions. A telling example of these two points is the debate surrounding the COMPAS risk assessment algorithm meant to predict recidivism. The algorithm, despite not including race as an input, has been argued to label blacks who do not actually re-offend with higher risk scores than whites who do not re-offend and vice versa for those who do re-offend (Angwin et al., 2016). The company developing COMPAS as well academics have argued against this critique along technical and methodological lines (Dieterich et al., 2016; Flores et al., 2016), with authors of the original critique standing by their conclusions as a response (Angwin & Larson, 2016). The debate is largely technical, but there is an underlying disagreement about the notion of 'fairness' and whether that refers to accurate calibration between groups (a specific risk score corresponding to the same rate of recidivism across population groups), or to a correct balance between the negative and the positive classes (the average assigned scores to those who reoffend should be identical across population groups) (Kleinberg et al., 2016). Not only is this a clearly political choice, it is also a choice that is difficult to avoid as both notions of fairness cannot be satisfied simultaneously in the vast majority of real-world cases (where we cannot predict perfectly and where base rates differ between groups) (Kleinberg et al., 2016). Hence, parts of what defines the search for evidence in big data and of what we infer from data, are in fact political choices.

Big data arguably adds to this problem, as it is "easy to mistake correlation for causation and to find misleading patterns in the data"(McAfee & Brynjolfsson, 2014, p. 68). There is thus the space for exploiting this 'malleability' of big data insights by policy makers seeking to find evidence for a policy that fits their pre-existing agenda – a behavior well documented in the literature (Kogan, 1999; Marmot, 2004; Nelkin, 1975; Walker, 2000). Given the number of new data sources, methods, and the size of the data itself, it will become increasingly possible to support virtually any policy intervention as 'evidence-based', which greatly expands the room for the 'political game' one can play with data. This renders the concept of 'evidence based policy' less meaningful, but also brings to the forefront some fundamental questions: "To whom do the analytics and findings go to and for which purposes? Who is profiting the most and least from big data?" (Uprichard, 2015, ¶ 2). Given that data are not inherently objective (as addressed above) and that human design and biases affect the methodologies for dealing with data (Crawford et al., 2014), questions about actor involvement, agendas, gains, and losses remain crucial.

Given the political nature of collecting and interpreting data, the more data there are, the more political choices will have to be made by those deriving meaning from the data - the analysts. For a policy maker or politician, this presents an uncomfortable situation: Analysts create algorithms to analyze big data, but the algorithms are often very complex and self-adjust, making the (political) choices made along the way difficult to interrogate (as demonstrated by, for example, the COMPAS algorithm) by the very people that (have to) use these statements and insights as a basis for policy (Janssen & Kuk, 2016b; van der Voort et al., 2018).

This dynamic introduces more actors into the policymaking process and the necessary public-private partnerships often result in tacit endorsement of security and privacy policies of private sector analytics companies (Bertot & Choi, 2013). Furthermore, this could also change the policy process itself: Entrepreneurial data analysts, scientists and enthusiasts are empowered (by the existence of big data that they can repurpose) to proactively come up with insights and services that may call for a policy response, putting the decision maker into a reactive role (van der Voort et al., 2018). Not only does this provide substantial agenda-setting power to the analysts, it also constitutes a radical decentralization of policymaking: If analysts can provide answers and solutions to problems the decision makers do not know exist yet, the cycle of "goals → gathering information → intervention" that characterizes traditional policymaking is effectively changing to "gathering information → intervention → goals", where the boundary between gathering data, making inferences, and the intervention is increasingly permeable. This is of course not a general trend, but the fact that some interventions can happen in this manner reinforces the observation of Klijn & Koppenjan (2015) that the entire process resembles a plate of spaghetti and happens in a much less structured and less predictable fashion than a techno-optimist narrative assumes.

2.2 Faster decisions: The unattainable ideal of real-time

Outside of resulting in better information that leads to better decisions, the techno-optimist narrative also maintains that big data analytics produce faster information which in turn leads to faster decisions. Not only is it argued that automation will accelerate some of public administrations' informational tasks (Maciejewski, 2016), but that real-time data streams will reduce the time period between policy coming to effect and being evaluated, as "[d]emographic data, unemployment numbers or migration patterns could be observed in real time, enabling a much faster assessment of whether the implementation of a certain policy was a success or not" (Höchtl et al., 2016, p. 162). In other words, big data will enable policy interventions to happen in real-time or near real-time.

Much like with better data leading to better decisions, this argument rests on two assumptions: That it is possible to generate relevant real-time data to inform policy decisions and that policymaking can adapt to the speed of this data. In this section we question both of these assumptions (in that order), pointing to the fact that many policy decisions are concerned with the long term, that many relevant indicators do not respond to policy interventions "in real-time", and that the speed of policy decision making is constrained by public administration and decision-making dynamics that are not removed by big data (van der Voort et al., 2018).

In policy areas concerned with long-term effects, improvements on how quickly data is available mean very little: What is the impact of education on employment outcomes, of pollution on environment, or of healthcare policy on health outcomes? All of these questions are extremely salient for policy and cannot be answered in real-time, as the effects they are concerned with materialize only years after the policy intervention. The benefits of faster measurement still exist, but a 'data lag' of even a few months is close to insignificant when measuring effects that take years or even decades to materialize, especially if there are other dimensions of quality of the measurement to be considered. Thus, notwithstanding the demonstrable potential of big data to speed up policymaking in multiple policy (Kitchin, 2014a; Lettieri, 2016; Wamba et al., 2012), this potential cannot be extended to policymaking in general.

Furthermore, an effect lag exists even for policies that are meant to have effect as soon as possible and that we often assume can be measured in real-time. Consider employment or unemployment indicators – indicators that many have tried to measure in real time using big data (Antenucci et al., 2014; Askitas & Zimmermann, 2009; Choi & Varian, 2009; D'Amuri, 2009; D'Amuri & Marcucci, 2010; Proserpio et al., 2016; Vicente et al., 2015) and that are of crucial importance for labour market policy. Both terminating and obtaining employment are not instantaneous (employees have to give notice and job seekers have to be selected, negotiate contracts, etc.) and thus assuming that a labour market policy intervention would yield (un)employment outcomes immediately is misleading. More timely measurement is of course valuable, but even if the indicator we are interested in can be measured in real-time and a policy has immediate effect on individual behaviour, translating that behaviour to a measurable change of an indicator is not instantaneous.

Lastly, much like with the quality of data and decision making, the 'speed' at which data can be generated does not translate directly into the 'speed' of decision making; The 'political game' described in section 2.1.1 does not happen instantly as actors have to co-ordinate, negotiate, and often bring in third party companies for their big data analytical expertise (Giest, 2017). This is a lengthy process that gets further extended by disagreements on interpretation, or by a misalignment with a policy window. The 'policy window' concept refers to the fact that for a policy action to be taken, multiple 'streams' have to align (Kingdon & Thurber, 1984), including a 'politics' stream that refers to whether policymakers have the will and opportunity to make the necessary policy (Cohen et al., 1972). In other words, often time it is not enough to identify a problem and conceive a solution, it is also important to implement this solution at the 'right time'. Faster data analytics are of course helpful in capitalizing on open policy windows, but it is also important to realize that just having a solution to a problem does notmean that the corresponding policy action can be taken. Needless to say, big data does not affect the political dynamics that determine when it is the 'right time' to create a specific policy.

2.3 New epistemological and methodological singularism in the works?

We now move to addressing an assumption we believe to underlie a substantial part of the techno-optimist narrative: The (often implicit) assumption that correlations identified in large datasets (or predictions made by models trained on such data sets) are at least a sufficient replacement for understanding causality of the relationship in question. Needless to say, not all big data analytics are based on this assumption, but a general link between big data analytics and privileging correlation over causation can be observed (Bollier, 2010; Kitchin, 2014b; Kitchin & Lauriault, 2015; Zwitter, 2014), as some argue rather explicitly that "we will need to give up our quest to discover the cause of things, in return for accepting correlations" (Cukier & Mayer-Schoenberger, 2013, p. 29). Perhaps even more importantly, this emphasis on correlation and prediction goes hand in hand with the belief that "[w]ith enough data, the numbers speak for themselves" (Anderson, 2008, ¶ 7). If this logic is applied to public policymaking, it translates into arguments such as "[t]he undeniable truth of facts [provided by big data] cannot be neglected even by the most stubborn politicians" (Höchtl, Parycek, & Schöllhammer, 2016, p. 146).

This leads some to argue that "[b]ig data helps answer what, not why, and often that's good enough" (Cukier & Mayer-Schoenberger, 2013, p. 29). It is important to acknowledge the use of 'often' by Cukier & Mayer-Schoenberger (2013), but in this section we argue that in policymaking more often than not knowing 'what' without the 'why' is not good enough. We first challenge the assumption of 'organic' data and meaningful correlations that can speak for themselves as a fundamental misunderstanding of data in the context of social science (sub-section 2.3.1), following which we also illustrate the practical limitations of purely predictive approaches when it comes to answering various policy questions (sub-section 2.3.2).

2.3.1 Death of the scientific method?

One of the most popular and forceful endorsements of the data-driven approach is Anderson's (2008) claim that the scientific method is dead. His argument is that creating models and theories can be useful, but models are never truly correct as reality is too complex to be captured by one (Anderson, 2008). Contrary to models and theories, enormous data sets collected with no specific analytical purpose in mind are argued to 'organically' reflect social reality more so than traditional statistical data (Groves, 2011; Zwitter, 2014, p. 2), making the patterns we find within them meaningful and informative in and of themselves. At face value, at least a part of this argument is true – models are always a simplification of the infinitely complex reality and as such might be useful, but are never truly accurate.

However, following the lines of Anderson's own argument, since reality cannot be captured by a model it cannot be captured by a data set either: Reality is infinitely complex and datasets are inherently finite, making it impossible to capture the 'full domain' of reality within a dataset (Kitchin, 2014b). Secondly, data do not really exist in a vacuum and cannot be meaningful without understanding and interpretation. Whether by design or due to practical limitations, data do not really exist in a "raw" and "organic" form (Gitelman, 2013) and always capture reality from a specific vantage point (Kitchin & Lauriault, 2015). Furthermore, we cannot derive any meaning from data without interpreting them and attaching them to domain-specific knowledge (Clemons & McBeth, 2009; Janssen & Kuk, 2016a; Kitchin, 2014b). Even if the process of translating numbers into data lacks a formal framework, science does notcircumvent the human perspective (Giere, 2006; Gould, 1981), making it impossible to capture reality in an 'organic' dataset. Because of this, data and the correlations contained in it cannot 'speak for themselves' (Goldston, 2008; Kettl, 2016; Liu et al., 2016) and are in fact crucially dependant on how we make sense of this data and interpret it – a process that is far from organic and objective.

In terms of meaningfulness of correlations, the fact that correlation does not imply causation requires no explanation, but we wish to take this argument even further: In big data sets, as Boyd and Crawford (2012) correctly note, correlation does not really imply much: "[E]normous quantities of data can offer connections that radiate in all directions" (Boyd & Crawford, 2012, p. 668). The reason for this is that with larger sample sizes the criterion of statistical significance is easier to satisfy, as well as the high-dimensionality of big data allowing for more potential correlations. This implies that in large data sets correlations are also less meaningful, for which a mathematical proof can be constructed, showing that there exist "ramsey-style correlations" that exist purely because of the size of a data set and in large data sets these can be the quantifiable majority of statistically significant correlations (Calude & Longo, 2016). This is not to say that we should not look for interesting correlations in big data, but that we should be acutely aware of the fact that some of the 'traditional' risks like spuriousness are even more pronounced in big data sets and that we should opt for statistical rigor over the assumption that correlations are meaningful because of the size and 'organic' nature of the data.

Lastly, even if data could 'speak for itself', using that for policymaking without interpretation could be unlawful. Big data insights often 'hide' a lot of discrimination, as historical exclusion and discrimination of certain groups reflects itself in data (and consequently in models trained on that data). Thus, adhering only to data-driven insight would further reinforce the discriminatory dynamic at play (Barocas & Selbst, 2016), which would be illegal in state-sponsored services (Samarajiva & Lokanathan, 2016).

2.3.2 Data-driven science in policymaking and public administrations

Despite all the above mentioned problems, there is an argument to be made that public administrations are not academia and could be more open to more inductive approaches, as politicians often revert to 'common sense' in policy decisions (Kettl, 2016, 2018) and the risks of spurious correlations are arguably context dependent. For example, it might be impossible to determine a detailed psychological theory of how work-related frustration translates to job loss, but the absence of such theory does not make this link particularly "risky" and the relationship between the two can arguably still be leveraged to understand labour market policy (United Nations, 2011). In other words, public administrations do not work along clearly demarcated 'inductive' and 'deductive' lines and are often open to 'doing what works' regardless of the epistemological implications. As such, it is important to translate this epistemological dilemma into practical terms.

The main practical limitation of an inductive data-driven approach is that it can be analytically suffocating (Lemire & Petersson, 2017) and even though it is useful for policy questions concerned with prediction, there are many other policy questions it is not useful for: Questions that beg causal proof, questions that beg explanations, or questions that beg comparative judgement (Lemire & Petersson, 2017). This is best illustrated by the difference between causality and prediction policy questions, both of which are extremely important but not answerable by the same methods. Prediction problems essentially require pre-existing knowledge about the causal link between policy intervention and outcomes, including how these outcomes depend on the occurrence of a specific event (Athey, 2017).

For example, consider evacuation policy aimed at minimizing casualties of a natural disaster: The causal link between evacuation and minimizing casualties is self-evident - if people are not in the affected area they will not be hurt by a natural disaster. In this case the effectiveness of the evacuation policy depends almost entirely on accurately predicting when the natural disaster takes place – evacuate too early or too late and you displace people without preventing casualties. However, for some of the most crucial policy problems the difficulty runs in reverse with the causality question taking precedence over the prediction question: Reducing poverty, increasing employment, or optimizing service delivery are all crucial policy areas where accurate predictions are secondary to understanding the underlying causal mechanisms. In other words, better prediction is extremely valuable for some policy question, but for other policy questions causal explanation (and other approaches and analyses in general) are a more important part of the answer and cannot be replaced solely by predictive methods.

Furthermore, inductive big data analytics are not irreconcilable with the scientific method, provided that they are only used at the stage of hypothesis formulation: Big data can point to interesting novel hypotheses and theories that can further be tested in a rigorous manner (Liu et al., 2016), resulting in more data-driven science, but science nevertheless (Kitchin, 2014b; Kitchin & Lauriault, 2015). Such approaches should allow for leveraging the insight big data can provide without abandoning the scientific method.

2.4 Unwarranted de-emphasizing of crucial issues: the case of privacy protection

The limitations and challenges to the use of big data in the public sector as outlined above are rarely systematically addressed in scholarly work on the topic. Yes, scholars generally do discuss potential limitations of big data use in the public sector, but these often end at the level of acknowledging the problem and leaving it for future legal or policy solutions. Crucial issues such as privacy are then left with "government is required to pursue this [big data] agenda with strong ethics" (Höchtl et al., 2016, p. 156). At times, these challenges are even omitted entirely. Of course, not every research can address every potential problem with big data use, but in addressing problems it is crucial to engage with how these problems are linked to the process of big data analytics itself in order to avoid assuming that the two are separable.

It is outside of the scope of this chapter to provide a comprehensive overview of all the big data challenges that tend to get overlooked, de-emphasized, or addressed selectively in the literature. In light of this chapter's objective, we do look at the underlying logic, demonstrating why it is problematic to de-emphasize these issues based on the belief that many of them can be solved down-the-line without altering the fundamental analytics. Using the example of privacy protection we illustrate that the ethical, societal, or other non-technical challenges are inseparable from big data analytics itself. Even though this section focuses on privacy as one of the best known and often referred to problems, an attentive reader can surely apply this more inquisitive approach to a host of other commonly known big data pitfalls.

2.4.1 Privacy: The big trade-off in using big data

In a way, a techno-optimist view of big data analytics makes it very difficult to engage with the issue of privacy: If we speak of data and the patterns they contain as something that is inherently objective and meaningful (as the techno-optimist narrative suggests), our data sets need to mirror social reality as closely as possible. However, in order to avoid privacy breaches, we need to distort our data. This dilemma is at the root of the trade-off between privacy protection and the validity of empirical inferences one can derive from a dataset (Daries et al., 2014).

This trade-off can be illustrated both conceptually and empirically. Conceptually, achieving a data set that does notpose a privacy risk is simple under partition-based privacy standards such as k-anonymity (Sweeney, 2002). We can set k to a large value and distort the data to the point where no two entries can be distinguished from one another, reaching a data set that poses absolutely no risk to privacy, but is also devoid of all meaning. In other words, "[t]o strip data from all elements pertaining to any sort of group belongingness would mean to strip it from its content" (Zwitter, 2014, p. 4). This is because in meeting a specific anonymity requirement the data needs to be manipulated by a combination of suppression of entries and generalization of entire variables (Daries et al., 2014).

The problem with those manipulations is that generalizing variables generalizes the data set as a whole and introduces a bias into the correlations and suppressing certain entries introduces a demographic bias (Angiuli et al., 2015). More research is needed in this area, but the current research has already shown that reaching k-anonymity (for k=5) can significantly distort conclusions derived from a data set (Angiuli et al., 2015; Waldo, 2016).

The question then becomes whether this trade-off between privacy and accuracy can be reconciled by technological solutions of either improving the popular privacy standards (such as k-anonymity), or creating a different privacy standard altogether. In terms of optimizing k-anonymity, Angiuli et al. (2015) show that the trade-off between distorting the means of variables and distorting the correlations between quasi identifiers is much more acceptable at certain "bin sizes" used for the generalization procedure. Other methods such as introducing "chaff" into the data instead of excessive suppression could also be a (part of the) solution (Waldo, 2016). Another solution would be a different privacy standard altogether, with non-partition-based standards such as differential privacy showing the largest promise by resisting a wider range of privacy attacks (Mohammed et al., 2011). Nevertheless, differential privacy still distorts the accuracy of the data and this trade-off is rather explicit in setting the privacy parameter: The more secure this parameter, the more noise is introduced to data with each query and less queries are allowed. Thus, despite its potential to optimize the trade-off (Ghosh et al., 2012; Mohammed et al., 2011), differential privacy can only be perfect for specific users and single count queries, but not for other types of queries (Brenner & Nissim, 2014). This is not to discredit these technological solutions, but to point out that their potential is merely to optimize rather than completely reconcile the privacy and accuracy trade-off.

Outside of technological solutions to this trade-off, the argument for policy solutions can be made. Here the debate turns even more speculative, since no alternative approaches to de-identification exist in practice. Theoretically, one of the promising concepts has to do with a shift from preventing privacy breaches to punishing them effectively. Such an approach would allow for sharing of de-identified data sets under the condition of tracking how individual users use this data in order to punish re-identification attempts and other misuse (Waldo, 2016). Such developments are extremely speculative, especially because there is no technical solution to enforce such a drastically different system: A scalable and practicable system of enforcement and audit of contracts on data use in the current legal system is difficult to even imagine, let alone implement (Daries et al., 2014). Thus, despite some signs of legislators re-thinking privacy regulation, no significant changes can be expected to happen soon (Angiuli et al., 2015). In sum, the evidence seems to suggest that not distorting data and respecting individual privacy are not (perfectly) reconcilable and that we are far removed from a good technical or policy solution.

2.5 Discussion and conclusion

As highlighted in the introduction, big data is expected to have a profound impact on the public sector. In recent years, a body of literature has emerged highlighting the possibilities of using big data for better insights, better decision making, and for significantly altering policy processes. Yet, the true challenge to these promises lies in where big data meets existing practice in the public sector. Although the literature has not completely neglected these challenges, the current debate on big data in the public sector emphasises technical-rational factors, focusing much more on data and analytical output rather than on its interaction with the decision-making process in public administrations. Throughout this chapter we have illustrated why political decision-making factors should be taken seriously by critiquing some of the core techno-optimist tenets from a more policy-pessimist angle, constructing these two archetypical narratives in the process.

We have first tackled the claim that big data provide 'better' insights and thus foster better decisions: Not only is big data not always 'better' in terms of accuracy, but there are also multiple dimensions of data quality. Furthermore, translating 'better' evidence into 'better' policy is subject to public administration dynamics much more complex than the techno-optimist narrative assumes. Secondly, we address a similar argument of faster insight resulting in faster policy decisions, which we challenge based on not all policy questions being able to benefit from near real-time measurement because of long-term concerns or natural delays in the causal chain. Furthermore, public decision-making dynamics are not removed by big data and introduce a substantial time lag in and of themselves. Thirdly, we tackled the less clearly articulated but no less important epistemological concerns with big data analytics as both a fundamental misunderstanding of data, but also as a practical limitation in terms of what policy questions can be answered. Lastly, we have argued against how a techno-optimist narrative de-emphasizes certain issues that are in fact crucial and should be an integral part of the debate – an argument that we illustrate on the trade-off between privacy protection and accuracy. In this concluding section, we first summarize the two narratives and have them meet eye-to-eye, and second provide a realist rejoinder.

2.5.1 Techno-optimism and policy-pessimism: an eye-to-eye comparison

Despite challenging techno-optimist arguments throughout this chapter, our goal is not to make a case for policy-pessimism as an alternative. The problem we see in the current literature is not an absence of a critical alternative to techno-optimism, but rather that such an alternative is complex, spans many disciplines, and only seldom makes it into individual research projects and agendas in a systematized and comprehensive way. As a result, even high quality research often subscribes to techno-optimist simplifications in approaching legislation (Bertot & Choi, 2013), privacy (Sagiroglu & Sinanc, 2013), data quality considerations (Ku & Leroy, 2014; Matheus et al., 2018), and many other aspects of big data use. Our contribution aims to remedy that by articulating the two archetypical narratives and making them meet 'eye-to-eye', allowing scholars to systematize the way in which they interrogate big data promises and shortcomings, paying sufficient attention to both technical-rational and political decision-making factors.

To provide this eye-to-eye comparison, in table 2.1 we summarize both narratives along the four dimensions addressed throughout this chapter. In this table we also include a hypothetical set of questions that one of these narratives would interrogate the opposing narrative with. These questions are derived from arguments we have presented in this chapter, which also constitutes an important limitation: Since this chapter is mainly challenging the techno-optimist narrative from a policy-pessimist lens, the policy pessimist questions are far better anchored in the existing literature. We still derive some key techno-optimist questions from our summary of the narrative, but recognise that a more thorough summary and defence of the techno-optimist narrative would certainly arrive at more informed and grounded techno-optimist questions. Despite this limitation, these questions as presented in table 2.1 illustrate the utility of understanding these two narratives as logical extremes that might not provide the best argument, but that are asking important questions.

Table 2.1: An 'eye-to-eye' summary of the techno-optimist and policy pessimist narratives, summarized based on quality of data and decisions, the speed of data and decisions, epistemological concerns, and fundamental concerns with big data.

Key issue	**The 'techno-optimist' narrative**	**The 'policy-pessimist' questions**	**The 'policy-pessimist' narrative**	**The 'techno-optimist' questions**
Quality of big data insight and how that translates into quality of decisions (section 2.1)	Big data provides more information which means better insight and better predictive capabilities, which then translates into better informed (and thus generally better) policy decisions.	Is big data better on all data quality dimensions? Can data be universally better? If not, who or what are they better for? How do data get translated to decisions?	On important quality dimensions big data is not better for policymaking than traditional data. Politicians will always cherry-pick data that suits their agenda – more data will diffuse the meaning of 'evidence based' and result in more political strategizing.	How will better estimates and predictions impact decision-making? How can analysts and new data source facilitate better insight? Can certain decisions be automated? How does measuring previously unmeasurable concepts help in policymaking?
Speed of big data analysis and how that translates to speed of decisions (section 2.2)	Real-time data streams provide more up-to-date information faster than currently available data, meaning that policy decisions can be made faster, making policy more agile.	Is faster data possible or useful in all policy areas? Can decision making adapt to the speed of data? What gets lost if we remove humans from the equation to allow for faster decisions?	Decision-making will not adapt to the speed of data, as negotiation and interrogation of the data by humans is a crucial part of the process. Faster data is not available for most policy questions.	Does reduced data-lag influence policy-relevant insights? Does better temporal resolution improve insight? Can't certain decisions be reliably automated?
Epistemology of big data analysis (section 2.3)	A more inductive approach based on correlation and prediction rather than causation as long as the dataset is of sufficient size.	What is the role of interpretation? How meaningful are correlations in big data? Is this approach appropriate for policy questions not predictive in nature?	No substantial change to scientific method, muting the effect big data analytics will have as they are tailored for inductive exploration and not deductive testing.	Why would public sector not emulate private sector for efficiency gains? Can inductive exploration contribute new and relevant insight? Why not 'do what works' if we can show that it does?
Connection between big	Privacy and other fundamental issues	Are these issues related to the type	Privacy and other fundamental big data	What technological solutions can provide

data analysis and fundamental concerns related to it (section 2.4)	with big data matter, but can be overcome down the line with more advanced technological solutions or policy interventions.	of analytics advocated? What are the trade-offs with these issues? How likely are these issues to be solvable 'down-the-line'?	issues are crucial and cannot be (fully) reconciled with big data analytics. To avoid them we should stop or limit big data analytics.	good results? Should limitations stop progress in terms of big data use? What is the balance of risks and rewards (including the risk of falling behind in data utilization)?

Despite our diagnosis that the literature as a whole is leaning towards techno-optimism and our subsequent case for the utility of policy-pessimism, systematizing the assumptions and arguments of the literature in this fashion has value even if one disagrees with our diagnosis. The techno-optimist and policy-pessimist systematization offers a tool that can be fitted to a specific research context: A specific research focus might require these two narratives to emphasize the various important legal or ethical concerns (such as intellectual property rights, data security, liability, accountability, etc.) and de-emphasize some of the points we focus on in this chapter. Regardless of the focus, this systematization will still pose important questions and expose where on the axis between the two narratives one is located. That in turn presents two options: Either defend a specific position as the most appropriate trade-off point (argue for one narrative over the other), or find a way to reconcile the two narratives in a 'best-of-both-worlds' fashion. Doing neither results in (unintentional) cherry-picking of the easiest to address problems and not tackling underlying assumptions that can, despite seeming inconsequential, influence research findings.

2.5.2 A realist rejoinder

To conclude this chapter, we offer our take on a rejoinder between techno-optimism and policy-pessimism in the form of a middle-of-the-road realist perspective (realist in the sense of dealing with the world as it is, not in reference to political, artistic, or epistemological positions). To achieve that, we propose a move away from the umbrella terms of 'big data' and 'policymaking' to talking about specific data sources and methods used for specific policy questions. It is difficult to make general conclusions about big data use because there are numerous associated benefits and pitfalls which depend on context. Some of the pitfalls are addressed by this chapter, but many are omitted, including the costs and challenges associated with developing skills and infrastructure, representativeness of big data sets, the procurement of data itself and the necessary public-private partnerships, accurately distinguishing 'signal' from 'noise' in big data sets, legal concerns, and many others. On the other hand, there are important and difficult to deny benefits of big data: The speed of data and analysis can be tremendously valuable for time-sensitive policy responses or monitoring systems, the large sample size can mean much more accurate disaggregation of data crucial for group-specific interventions, and analysing novel datasets can provide previously unmeasurable insight. Furthermore, once the infrastructure is in place and skills are developed, the marginal cost of an additional analytical inquiry is miniscule compared to traditional survey based sources (Kitchin &

Lauriault, 2015), further reinforced by the fact that response rates to surveys are declining (Bostic et al., 2016).

Given how many such shortcomings and benefits exist and the absence of a meaningful way to sort them, it might seem that the decision for or against adopting big data is arbitrary or heavily political at best. However, we believe that in looking at specific cases (a data source and a method applied to a policy question) the trade-offs between shortcomings and benefits become meaningful enough to make sound (albeit political) choices on. Consider the example of data backrun mentioned in this chapter: Data backrun is of tremendous importance for policy decisions on issues that policy makers have been wrestling with for decades, but of extremely little importance for more recent issues whose emergence coincides with the emergence of big data (such as e-commerce), because for those issues conventional data have no comparative backdrop advantage. This context-specificity applies to all possible pros and cons: Representativeness issues might not be serious in group-specific policy decisions, privacy is almost a non-issue when using aggregated search query data as opposed to individual search query data, and the speed of big data can benefit rapid response policies but does very little for long-term human capital policies. Outside of public policymaking, public administrations also have the task of public service delivery (and optimization), for which data needs can be different and thus also emphasize and de-emphasize various shortcomings and benefits of big data. Not only do these trade-offs become meaningful at the level of individual policy problems and data sources, they also show some space for generalizations: For example, many fundamental economic questions are naturally retrospective, and thus benefit from data accuracy much more than from timeliness (Einav & Levin, 2013), making it unreasonable to expect any shift towards 'relativized exactitude' in solving those policy questions. Through balancing these pitfalls and benefits is how decisions for or against the adoption of big data analytics can be most meaningfully made.

That said, here we draw on policy-pessimism to highlight that making 'meaningful' decisions on big data does not mean making them fully rationally: Public administrations are not purely rational entities and different stakeholders are not only likely to reach different conclusions with regards to whether big data is actually fit for a specific policy question, but also use these conclusions in different ways depending on broader strategic concerns and individual agendas. The process of public administration can resemble a strategic game rather than rational deliberation (Klijn & Koppenjan, 2015) and the adoption of big data is not immune to this dynamic. This means that to understand big data in the public sector, it is important to understand not only the rationality behind balancing the context-specific benefits and pitfalls of big data, but also the actors and institutions that participate in making the decision.

The realist rejoinder we propose can be summarized in three key points: Firstly, big data has multiple aspects of quality (including speed) and the importance of these is crucially dependant on the policy question, data source, and methods. As such, big data will be a 'game changer' for certain policy areas, but will continue to struggle with adoption in other policy areas. Secondly, big data is subject to public administrations and decision making dynamics when used for policy purposes, making the translation from big data insights into policy action rather complex. As such, even 'better' or 'faster' insights could be affected by this process and result in unexpectedly good or bad policy. Finally, as a consequence of these two arguments, big data adoption will remain uneven and will be determined by numerous balancing acts of big data benefits and pitfalls for a specific policy application and data source by networks of actors. These balancing acts will be subject to divergent perspectives, pre-existing agendas, will not be fully rational, and will require time.

We hope that our systematic way of addressing optimist and pessimist arguments and assumptions in the current debate will help scholars and policy makers to interrogate and challenge their own assumptions. This may lead to a better fit between the goals of big data for specific uses and the context in which it will be applied, as well as to more realistic expectations and hence more careful decisions about deploying big data in practice. The remainder of this dissertation takes these lessons to heart and designs and carries out a research project focusing on a particular instance of this tripartite alignment of data source, method, and policy question.

Chapter Three: Social investment as a policymaking paradigm[2]

The search for a good tripartite alignment of data source, method, and policy question commences with the policy question. That said, policy questions can hardly be reduced to a singular question as the context in which they are asked and the underlying assumptions informing them are crucial for understanding the overall policy problem. The type of policy questions this dissertation tackles are mainly informed by comparative welfare state research and ‘social investment’ more specifically - an emerging paradigm of not just understanding the welfare state as a whole, but also the functions of various policies that constitute it. Social investment as a concept has gained considerable traction when it comes to conceptualizing European social policy efforts in the last two decades: The relevance of some its key objectives is anchored in the Lisbon treaty (Hemerijck, 2013) and an endorsement of ‘social investment’ as a guiding principle for European social policy came with the ‘Social Investment Package’ (SIP) in 2013 (European Commission, 2013). Despite arguably losing some traction due to weak institutionalization (de la Porte & Natali, 2018), social investment remains an important framework for steering European social policy as evidenced by a recent report on Employment and Social Developments in Europe, which states that: “Given major demographic and technological shifts, there is a broad consensus on the need to invest in people. Such ‘social investment’ helps to improve individuals’ well-being and prevent and mitigate social risks, by enabling citizens to acquire new skills and become or remain active in the labour market and by providing them with support during critical life course transitions” (European Commission, 2019: 21).

That said, social investment displays the curious tendency of being endorsed by important institutions (European Commission, 2013, 2019; OECD, 2011, 2015) and generating scholarly traction, yet being weakly institutionalized and not playing a substantial role in policymaking practice (or doing so in very specific ways). The extent of this is difficult to assess systematically, but some academics have argued that “Within the political arena, social policies continue to be compartmentalised across sectoral ministries … making it difficult to speak of a coherent social investment paradigm” (Plavgo & Hemerijck, 2021: 283). EU’s Joint Research Centre acknowledges this quite explicitly by acknowledging that social investment has been ‘lost in translation’ and that “despite the agreement around the approach proposed by the social investment paradigm, the consistency between the programmatic ambitions of the SIP and the reform practice is not easy to gauge” (Maduro et al., 2018: 16). Existing case studies corroborate this by observing that “social investment is not recognised as a concept but definitely as a phenomenon” (Lopes & Dias, 2018: 147) for long term care policies in the context of Denmark (Greve, 2018), Portugal (Lopes & Dias, 2018), and Lithuania (Poškutė, 2018). For poverty relief in Austria it can be argued that “there is no evidence of a systematic consideration of social investment” (Heitzmann & Matzinger, 2021: 584). Even though some goals and assumptions of social investment make it on political agendas or even steer social policy reform, very few cases (if any) show adoption of social investment as a coherent and systematic approach to welfare state reform or design of individual policy interventions. On paper

[2] This chapter is a yet unpublished Manuscript authored by Simon Vydra and Anton Hemerijck. More details are provided in the ‘Authorship Contribution’ section.

there are exceptions to this, but those exceptions have a different conception of social investment than this chapter adopts.

How can this political endorsement and academic interest exist simultaneously with a virtual absence from social policy agendas? The dominant explanation has to do with resource constraints – more specifically with austerity (Deeming & Smyth, 2015; Hemerijck, 2017, 2018; Kuitto, 2016; Mertens, 2017; Ronchi, 2018; Streeck & Mertens, 2011; F Vandenbroucke et al., 2011; Frank Vandenbroucke & Vleminckx, 2011; Vanhercke, 2013; Zeitlin & Vanhercke, 2018). Austerity and indebtedness constrict the fiscal space for policy reform in general and especially so for discretionary spending (Breunig & Busemeyer, 2012; Streeck & Mertens, 2011), which includes social investment. This can force countries to pursue less social investment or pursue it at the cost of social protection policies (Ronchi, 2018), resulting in minimal recalibration towards social investment (Bengtsson et al., 2017). Other explanations have to do with the political difficulty of committing resources in the present and only reaping full 'return on investment' in the long term (Ferrera, 2017), or with the intellectual inertia of beliefs that underpin much of the austere logic on social spending – beliefs that are in many ways opposed to social investment (Hemerijck, 2017, 2018). These explanations primarily aim to explain change in welfare policy packages across countries and try and explain the extent and pace with which welfare states are 'reorienting' towards social investment in many geographical contexts including Europe (Bonoli, 2013; Garritzmann et al., 2016; Hemerijck, 2013; Morel et al., 2012), Latin America (Fenwick, 2017; Sandberg & Nelson, 2017), or South Asia (Fleckenstein & Lee, 2017; Peng, 2014). And they generally do a good job at mapping and explaining this 'reorientation'.

However, focusing on substantive policy change and not on the underlying logic sells social investment short as a paradigm for understanding welfare states: Social investment is not merely a policy direction emphasizing the importance of certain policy areas and interventions over other types of social spending. It is true that emphasis on policy areas such as childcare of activating labor market policies aligns well with social investment, but policy reforms focusing on such policy areas can be successfully supported without adhering to social investment. The main contribution of social investment, at least from the perspective of this chapter, is the understanding of the role and functioning of the welfare state as whole and an approach to individual policy packages that constitute it. It is not the emphasis on specific policy areas but rather the emphasis on human capital, life course transitions, policy complementarities, or other such priorities and causal links between them that form the underlying logic of social investment. That logic, as noted above, is very difficult to find in policymaking practice and the existing explanations are less useful in explaining this than they are at explaining the diffusion of substantive policy reforms.

This chapter offers a (partial) explanation as to why the logic of social investment isn't more widely adopted. An explanation drawing much more from public administration and study of evidence-based policymaking than existing accounts do. Approaching social investment from this perspective - a paradigm for policymaking rather than paradigm for understanding the welfare states more broadly – shows that more elaboration of social investment is needed. This chapter provides such elaboration, building on the most complete description of social investment by Hemerijck (2018) and fleshing out what social investment means for more 'day-to-day' policymaking. This is a theoretical contribution to social investment as it delineates it more thoroughly as a paradigm, but also because it provides an (alternative and complementary) explanation for why its underlying analytical logic remains less adopted than political and academic endorsement would suggest. Approaching social investment from this perspective is currently missing from the extant literature and the significance of its analytical implications is overlooked.

This lack of analytical elaboration is illustrated by, for example, interview responses reported by de la Porte & Natali (2018) when they note the following with regards to a methodological follow-up to the SIP: "In terms of identifying core quantitative indicators, the [Social Investment Expert] group fell short of its ambitions. Our interviewees noted that while the immediate explanation is technical – that is, there are no indicators to assess SI – the underlying reason is political" (Porte & Natali, 2018: 838). Even though that might be the case, there is no reason to believe that the technical reasons are unimportant: Social investment "indicators" are not available to this day and the analytical challenges of assessing social investment policies remains largely unresolved. European Commission's methodological follow-up acknowledges that "[w]e do know plenty about the broad contours, missions and hoped-for accomplishments of social investment reform, but we continue to know surprisingly little about how to identify and empirically track particular policy mixes and reforms that manifest a social investment approach as distinct from other features of (social) policy efforts" (Hemerijck, Burgoon, Di Pietro, & Vydra, 2016: 2-3). Almost 10 years later the EU is still exploring how to implement the actions suggested in the SIP, focusing much more on developing a methodological framework for evidence-based input for social policy innovation (European Commission, n.d.). This need for evidence-based input is also acknowledged as a policy-making challenge: "The emerging challenge for policy-makers is to better understand these complex relationships in a way that allows them to decide which interventions, delivered to which individuals, at what stage in their life-course, will do the most to boost resilience later in life … Further, the evidence needed on which to decide to terminate an ineffective programme, or to enhance an effective one, is often absent" (Gluckman, 2017: 2). Even at a theoretical level it is not obvious how paradigmatic claims of social investment should be translated into intermediate steps that can be taken in the short or medium term (Kenworthy, 2017).

Motivated by this need for more evidence-based input, this chapter also makes a more pragmatic contribution by suggesting a research direction that has promise of delivering on the key features of social investment policy analysis in an evidence-based fashion - the utilization of big data analytics. This also constitutes its main contribution to the dissertation as a whole: If research efforts are to focus on tripartite alignments of policy questions, data, and methods there ought to be an argument for why any such alignment is likely to provide useful practical tools and relevant academic finding (provided the given alignment has not been studied before). This chapter lays the groundwork for such an argument by demonstrating the 'fit' of big data analytics for the type of policy questions social investment tends to ask, allowing chapter four to match some of these questions with specific data and methods to arrive at a promising tripartite alignment.

To meet its aims this chapter first briefly introduces social investment and the analytical principles it comes with in **section 3.1**. **Section 3.2** then explains why these analytical implications make social investment policy analysis difficult by arguing that social investment principles put additional 'load' on policy feedback loops by requiring them to be faster and more complex. **Section 3.3** briefly outlines how utilization of big data can alleviate some of these issues, broadly outlining a promising research direction. **Section 3.4** then concludes the paper by summarizing its key claims, their importance and position in the literature, as well as contributions of this chapter to the dissertation.

3.1 Social investment policy analysis

Understanding the analytical side of social investment starts with understanding social investment as a policy paradigm (for a thorough summary see Hemerijck (2018)): Social investment is an intellectual assertion about the functions and outcomes of the welfare state in the face of important social and economic change. The social and economic change was the transition to knowledge-based economy and feminization of the workforce and what this meant for the labour market, most notably in terms of atypical employment contracts and how many times workers transition between employment, employers, and even sectors. European welfare states were designed for the industrial economy rather than for the knowledge economy and this mismatch fostered sub-optimal welfare outcomes for many population subgroups due to welfare states' inability to deal with the consequences of new social risks and erratic life courses, technological change placing high premium on skills, and increased feminization of the work force (Gøsta Esping-Andersen et al., 2002). This prompted European welfare states to overhaul policy provisions in employment policy (Bonoli, 2013; Schmid, 2008), labour market reintegration policy (Clasen & Clegg, 2011), retirement policy (Ebbinghaus, 2011), and family policy (Orloff, 2010).

The intellectual assertion of social investment is that these shifts are constituent parts of a new policy paradigm – a complete paradigm with normative commitments and political objectives as well as policy theory and instrumentation (Hemerijck, 2013, 2018). The normative commitment of social investment is Rawlsian in the sense of favouring those who are least well-off, but in a manner that is inspired by Sen's capabilities approach (Sen, 2001) in enhancing their capabilities to flourish in life (Hemerijck, 2018; Morel & Palme, 2017). However, some would argue that social investment instrumentalizes social policy by focusing too strongly on economic outcomes (Nolan, 2013; Saraceno, 2015), perhaps even reinforcing some of the neo-liberal principles it aims to steer clear of (Laruffa, 2018). The broad policy goals of social investment are less contested: Firstly it is to reduce the inter-generational transfer of disadvantage through investment in children and secondly it is to sustain the welfare state by increasing employment and productivity of those employed through investing in human capital and protecting that human capital from erratic life course transitions (Hemerijck, 2018). It is this building, protecting, and utilizing of human capital that social investment sees as the primary pathway to inclusive growth, better well-being outcomes, and sustainable welfare states.

To this end social investment has a 'positive' view of public policy as potentially generating private economic returns at low social cost, but it is not an argument built on a 'general' theory. Instead, the focus is on contextualizing social policies to different national contexts, realizing that different policies align well with certain institutional conditions rather than proposing a 'one-size-fits-all' theory of social investment progress (Hemerijck, 2018). This shift away from general theory towards acknowledging the crucial role of national context is key to understanding social investment's analytical implications. In some ways, this is a major change from its paradigmatic predecessors: In the post-war Keynesian-Beveridgean consensus, the emphasis on eradication of 'want' implied both the insurance against interruption or loss of earning power as well as adjustment of incomes for periods of earning and not earning (Beveridge, 1942: 7-8). This translated to counter-cyclical economic management by 'smoothing out' both extremes of the business cycle, firmly rooted in Keynesian economics. This is not to say that there was no room for nuance in this paradigm, but that the problem definition, governments' role in solving it, as well as tools used to resolve it remained rather consistent across (industrialized) contexts. For the neoliberal critique of the welfare state - a paradigm following the Keynesian-Beveridgean consensus – the underlying logic was even more streamlined: Public spending 'crowds out' private economic initiative and the increased equity it can provide necessarily comes at the expense of economic efficiency. From a welfare state perspective the diagnosis and solution here is even more context-independent, as more free-market mechanics were believed to be the solution to the moral hazards and inefficiency engendered by publicly funded social policies regardless of context.

Social investment does nothave any such underlying general theory. It certainly has an underlying logic about which variables are important and how they relate to desired welfare outcomes, but this understanding does not translate into policy prescriptions without considering the specific context. Whether this is due to a turn towards more sociological rather than economic understanding (comparatively to the two preceding paradigms) or simply a consequence of us better understanding the complexities of modern life courses, social investment is not built on a general theory that would easily translate its paradigm-level insights to policy prescriptions. In the absence of such a theory, what does the process of using social investment logic to inform evidence-based policymaking look like? There are a few conceptual innovations of social investment that dictate the broad contours of such a process: Firstly, it is designating life course transitions as a key driver of wellbeing gains or losses. Secondly it is the focus on 'preparing' rather than 'repairing' mechanisms of social policy. Thirdly, and perhaps most importantly, it is its focus on institutional complementarities for maximizing the wellbeing returns from social policy. These principles are explained in the following three sub-sections, followed by an overview in section 3.1.4 which articulates what 'social investment policy analysis' would, broadly speaking, entail.

3.1.1 The life course perspective and importance of transitions

The first such principle comes from the focus on 'life course transitions' (Hemerijck, 2013). Social investment moves away from analysing welfare 'beneficiaries' and 'contributors' at a specific point in time, arguing that due to more heterogeneous life courses even those who are generally 'contributors' are likely to rely on some form of welfare provisions at some point during their life (Hills, 2014). The focus is squarely shifted onto how individuals progress through their life, focusing on how human capital is created, maintained, and utilized (Garritzmann et al., 2016). This in itself is an important shift in focus. In terms of building human capital social investment relies largely on education and (re)training policies, emphasizing the work of Heckman (2006) on the importance and effectiveness (in terms of returns on investment) of early childhood education and care. The creation of human capital is not where social investment innovates – it is in terms of 'maintaining' this capital during inevitable life course transitions and 'mobilizing' this capital in the sense of enabling transitions (back) into the labour market.

Social investment focuses on these transitions as crucial drivers of welfare outcomes - at one extreme individuals who become unemployed can get re-training and enter the labour market in a more productive and prosperous sector (Nelson & Stephens, 2011) and at the other extreme their skills can atrophy to the point that they might never enter the labour market again, or only in a less productive sector (Leoni, 2015). It is this mechanism of changes to both human capital itself and its utilization during life course transitions that explains the foundational social investment claim – that social policy can generate positive wellbeing and economic outcomes. In a way success in these transitions is the underlying causal mechanism connecting social policy interventions and well-being outcomes: Social policies help individuals to go through these transitions 'successfully', meaning that their skills do not deteriorate (or are improved) and are utilized in a productive way post-transition. This in turn translates to better individual well-being as well as improved aggregate economic productivity. This is the mechanism that social investment policy analysis needs to focus on.

Outside of the specification of mechanism via which social policies 'work', this also influences the 'level' of analysis: Transitions are specific to what state an individual is transitioning from and not just the end state. This reflects the fact that different groups of people could need different interventions to achieve the same goal – consider for example fresh graduates, parents, and older workers all seeking employment. These effectively become three different policy dilemmas of improving the transitions from school to work, from parenthood to work, and from retirement to productive retirement or work. Group-specific analysis is of course nothing particularly new, but it is important to highlight that social investment analysis tends to consider not just micro-level analysis of individual trajectories and macro-level analysis of country or region-level trends, but also meso-level analysis for relevant population sub-groups and transitions specific to them (Hemerijck et al., 2016: 13).

3.1.2 Avoiding 'repairing'

The second analytical implication has to do with 'preparing' individuals for life course transitions rather than simply 'repairing' the damage of sub-optimal life course transitions (Hemerijck et al., 2016). This does not translate into pre-emptive interventions for individuals who are 'at risk' of going through a transition, but rather into making sure that the necessary policies are in place before individuals start needing them when they commence their transition. In other words the policies affect individuals 'during' their transitions rather than 'after' the transition has concluded. This change in the time horizon during which social policies impact individuals follows from social investment's focus on human capital and its mobilization: Firstly, having a period of 'repair' for human capital implies a limited participation in the labour market for that period, which results in human capital that is not (fully) mobilized. Secondly, human capital is very difficult to 'repair', especially when compared to things like demand. For example, in Keynesian terms the crucial issue is reduced demand during periods of economic downturn, which is a very 'repairable' problem: Even if people do not consume as much for a while their capacity to consume is not deteriorated by this period of reduced consumption. Their consumption can be restored by simply providing them with the means to consume more, which would then translate into the desired increase in aggregate demand. However, this is not the case with skills as they deteriorate if not used for a prolonged period of time (Gangl, 2004), reducing an individual's employability as well as productivity if employed again. Given skill erosion and the general difficulty or 'repairing' human capital, it is important to avoid its erosion in the first place.

Analytically, this determines 'when' social policy influences individual's transition, which differentiates social investment from the ahistorical neoliberal notion of always reducing social expenditure or the reactive Keynesian notion of compensating for reduced consumption during periods of economic downturn (after becoming unemployed for example). It is a much more ex-ante approach (Hemerijck, 2018). Does this imply anything different for policymaking? It does mean that social investment policies have to be established and implemented before individuals go through transitions, but from a practical perspective policymaking is a slow process and most policies (including 'repairing' policies like income replacement) are established before individuals need them.

However, this chapter argues that the combination of this ex-ante approach with an emphasis on 'new social risks' (Crouch & Keune, 2012; Kvist, 2015), another crucial part of social investment, imply the need for more agile policymaking: The concept of 'new social risks' refers to the fact that in modern knowledge based economy individuals face risks they did not have to face in industrialized economies. In fact, the heterogeneity of modern life courses – the fact that individuals are likely to undergo many transitions within the labour market throughout their working life – is a prime example of such a 'new' risk. Since social investment acknowledges that social risks do change it would seem counterproductive to not assume that they would not change further, especially given the pace of technological advancement. If social risks are changing and policy needs to respond to them then a focus on 'new social risks' (as defined at the inception of social investment) would leave social investment outflanked by 'newer social risks', making it a rather short-lived paradigm. Only some scholar associate social investment with 'continuously adjusting bundles of assistance' (Sabel et al., 2017: 140), and it is possible that some social investment theorists would leaver 'newer social risks' to another paradigm or school of thought in favour of focusing on already established 'new social risks', but when compared to previous paradigms it does imply a more agile and reactive approach to policymaking.

3.1.3 Institutional complementarities

The third, and perhaps most consequential, analytical principle of social investment is the emphasis on 'institutional complementarities' (Burgoon, 2017; Dräbing & Nelson, 2017; Hemerijck, 2017; Frank Vandenbroucke & Vleminckx, 2011). Institutional complementarities are a term originating from the Varieties of Capitalism literature where it is summarised as: "One set of institutions is said to be complementary to another when its presence raises the returns available from the other" (P. A. Hall & Gingerich, 2009: 451). These institutions can be located in different spheres of the political economy, having the positive effect of aggregating these complementarities into improved macro-economic performance, but also the negative effect of smaller returns if a change in one sphere is not accompanied by an appropriate change in other spheres as well (P. A. Hall & Gingerich, 2009; P. Hall & Soskice, 2001). Even though Hall & Gingerich (2009) extend the notion of complementarities to institutional practices and the appropriate level of analysis is arguably still debatable (Dräbing & Nelson, 2017), the mechanism of institutional complementarity exists even at the micro-level of individuals utilizing various policies. The most popular (and illustrative) example of such a complementarity from the social investment literature is that childcare provisions increase the returns available from policies aimed at getting parents back to work, since childcare provides parents with the necessary time to consider employment in the first place.

Analytically, this means that evaluating institutions and policies in isolation could be misleading as their performance can be tied to their alignment with other institutions and policies (Höpner, 2005). Even if a single policy intervention is the target, other aspects of the social policy suite and institutional context are also relevant in analysis. This is especially important when considering good practice transfer, as policies do not exist in isolation but rather as constituent parts of the overall welfare state. This means that different transitions are targeted by not just one policy, but a selection of policies that together engender the desired effect.

That said, the common approach of bundling policies together and evaluating a specific 'policy mix' can illustrate that complementarities exist, but it cannot illustrate the individual interactions between policies themselves. This would sell short the social investment claim of dependence and perhaps even causality between elements of the social policy suite (Burgoon, 2017). The analysis should ideally quantify, or at least explain, the interactive effects between the policy of interest and other relevant aspects of the social policy suite. To add further complexity, complementarities exist not just between policies as they also extend to other economic and institutional factor (such as wealth, economic growth, or institutional capacity). These complementarities are crucially important because they are the mechanism for optimizing policy effectiveness under social investment: Policies that are complementary to one another and suited to the particular institutional context 'maximize' the returns on investment, while policies not aligned with other policies in this way lead to 'incomplementarities' and sub-optimal policy outcomes.

3.1.4 Overview

The above mentioned shifts in focus and analytical principles indicate what social investment policy analysis should encompass, but without more structure the concept is still difficult to grasp. To provide more structure this chapter adopts a useful heuristic proposed by Hall (1993), which differentiates between three 'levels' of policymaking: "the overarching goals that guide policy in a particular field, the techniques or policy instruments used to attain those goals, and the precise settings of these instruments" (P. Hall, 1993: 278). These three levels are refer to as third order, second order, and first order policymaking respectively. Using these three levels it is also possible to compare social investment to its paragmatic predecessors. Even thought such an 'overview' necessarily simplifies each paradigm in a way that does notdo it justice, it illustrates the comparative difference in policymaking implications sufficiently. In doing so this chapter draws on work by Hemerijck (2018), which compares the three paradigms of welfare states in a more general and all-encompassing way. Recasting and rephrasing the features "policy problem and political objectives", "policy theory" described by Hemerijck (2018: 824) as third order policymaking and "policy instrumentation" as second order policymaking. First order policymaking is added. This overview is presented in table 3.1.

This table illustrates what the SI analytical principles outlined above mean for each level of policymaking: In terms of third order policy change (paradigm level) the policy problem is the precariousness of life course transitions due to new social risks (and what that means for wellbeing and the economy). Given the 'positive' view of state and social policies the belief is that well-designed social policies can aid individuals going through these transitions. In terms of second order policy change (techniques and instruments) social investment utilizes bundles of policies from various policy areas, targeting specific life course transitions in an ex-ante approach, trying to prevent sub-optimal outcomes rather than 'repair' the damage after. In terms of first order policy change (setting of instruments) optimal policy outcomes are reached by policy bundles that leverage institutional complementarities and that are agile enough to respond to change in social risks, institutional conditions, or changes to other parts of the policy suite.

Highlighting first-order policymaking is a consequence of approaching social investment from a policy analysis and evidence-based policymaking perspective and not something that plays a meaningful role in current social investment scholarship. In fact, social investment has its own understanding of the processes that result in welfare state recalibration and institutional change. This understanding breaks from the narrative of 'immovable' welfare states where reforms are politically costly and sparse (Pierson, 1994) in favour of recognizing that institutions do change in incremental but transformative ways, resulting in a set of mechanisms for both institutional continuity and change

(Streeck & Thelen, 2005). These mechanism of change are heavily influenced by discourse and political narratives (Schmidt, 2002). There is space for evidence to influence this process, as the epistemic authority of expert groups is important, but “only if it is able to muster political support” (Hemerijck, 2013: 102). The focus is thus squarely on country-level welfare state recalibration and on the ‘political discourse’ of social investment (third and second order policymaking). Since it does notplay a role in explaining welfare state recalibration, first order policymaking is somewhat outside of the scope of current social investment literature.

Table 3.1: Welfare state paradigms viewed via three orders of policymaking

		Keynesian-Beveridgean consensus	**Neoliberal critique**	**Social investment**
Third order policy change: Policy problems and causal understanding	Primary problem diagnosis	High degree of volatility in a cyclical economy (in the business cycle sense). Results in periodically depressed demand and increased unemployment.	Economic efficiency is undermined by welfare state programmes and the taxation and public spending these require. There is a 'big trade-off' between equity and efficiency.	Modern life courses present dangerous 'new social risks' that can lead to stagnation or erosion of human capital as well human capital not being utilized. This translates to sub-optimal well-being and productivity.
	Primary solution to the problem	To 'smooth out' the volatility in the economy and achieve full employment and social citizenship. Social policies can contribute to maintaining both demand and employment.	Reduce state-sponsored welfare programs in order to increase economic efficiency. Social policy is generally antithetical to this aim (contributing to equity over efficiency).	Improve the these transitions so that human capital does not erode (maybe even improves) and is utilized in a productive way post-transition. Social policies can improve these transitions
Second order policy change: Techniques and policy instruments	General techniques	Macro-economic steering through counter-cyclical demand management together with employment protection.	Deregulation and austerity with regards to social policy based on non-discretionary of fiscal discipline and hard currency monetary policy.	Implementing capacitating bundles of policies to address risky life-course transitions. Policies in these bundles can span multiple policy areas, but focus on the same transition and are aligned with one another.
	Specific instruments	Income-replacing social insurance in events of job loss to maintain demand and (industrial) job protection.	Various types of benefit curtailment, liberalization, or privatization. If not possible service-providers were to adopt market-like performance management to make efficiency gains.	A broad range of policy interventions (depending on the transition of focus) that assist in ex-ante fashion during the transition (or in some cases after).
First order policy change: precise settings of those instruments	Unknowns relevant for optimization	Location on the business cycle (by metrics such as current output, employment, income, sales, etc.) as well as a forecast of how relevant metrics will develop in future.	The amount of inefficiency introduced by policies as well as the amount of redistribution and equity generated 'in return'. Overall costs of policies.	What are the relevant institutional (in)complementarities and how the overall policy effectiveness changes based on the presence or absence of these. Covers a broad range of policies and institutional conditions that should be known quickly to enable agile policy adjustment.
	Analytical tools	Historically tools like IS-LM or AD-AS models for theorizing and timeseries on important metrics for management. Currently more complex DSGE models, but the understanding of business cycles is different.	Models like AD-AS to talk about deadweight loss and data on inequality and economic performance. Social policy not an explicit focus and the general task is 'minimization of cost', which is analytically simple.	Institutional analysis and ideally layered methods combining micro and macro analysis with institutional analysis. Not obvious beyond that.

Despite that, first order policymaking can play an important role in social investment, as it is argued that social investment is unlikely to happen as a grand recalibration of the welfare state based on some ex-ante calculation of future 'returns' on investment. This is not just due to the difficulties of calculating the 'returns' or 'discount rate' of social investment (Begg, 2017), but because of the uncertainty of what 'costs' would amount to, who should bear them, and who is eventually to reap the returns (C. Sabel et al., 2017). The progress is much more likely to be one of incremental change (Hemerijck, 2013; C. Sabel et al., 2017) and at least partially driven by bottom-up innovation by actors who have an understanding of how best to align policy complementarities in a given policy bundle for a given group of people (C. Sabel et al., 2017). This insight overlaps strongly with the notion of policy windows (Kingdon & Thurber, 1984) - given that social investment reform happens across multiple policy areas and is costly, it is unlikely that the problems, solutions, and political opportunity to make all the necessary reforms in one go will materialize. Curiously enough, even though social investment does not focus on first-order policymaking, it actually concedes that the bulk of social investment progress is likely to happen via incremental changes implemented at local and regional levels by civil society organizations. Despite the importance of this bottom-up innovation and local initiatives to social investment reform, "[v]ery little research looks at sub-national level examples of Social Investment" (O'Leary et al., 2018: 295). This oversight is somewhat surprising, as the EU supports social investment through the European Social Fund, which is generally utilized at sub-national level.

Not only do we understand the promise of social investment policy analysis - more context specific and life course based analysis that takes into account important interactive effects and that 'prepares' rather than 'repairs – we also know that the first order policymaking (calibrating and bundling policies in a way that maximizes institutional complementarity) is very important to social investment reform as a whole and that it is likely going to happen 'bottom-up' from local and regional actors who ought to have sufficient discretionary authority to bundle policies effectively (C. Sabel et al., 2017). However, there are a lot of challenges inherent in this proposition that are not solved simply by conceding that the change will be incremental and depend on sub-national actors.

3.2 Complex analytics and institutional taxation

The promise of social investment policy analysis comes at a substantial cost in terms of complexity and analytical load. The departure from a general theory in favour of context-sensitive focus on institutional complementarities comes with the "sobering conclusion is that there is no optimal policy mix, as welfare systems are always in flux, and more important, that each country needs to elaborate its own policy package … depending on prevailing social, economic, and institutional conditions." (Hemerijck, 2017: 28) When it comes to policy interventions 'the devil is in the details' (Hemerijck et al., 2016: 32) and those details are not necessarily transferable between national contexts due to the difference between them (Hemerijck, 2017; Kenworthy, 2017). Some 'lessons learned' can be transferred, largely due to social investment's effort to identify the most important interactive effects, but specifics have to be tailored to a given institutional context or policy outcomes are not optimal. This results in the unpleasant conclusion that details are the key to optimizing social policy bundles, but that those details cannot be transferred from different contexts.

How do these details get figured out then? Or, utilizing the heuristic adopted in this chapter, what is first order policymaking like? With the broad aims already outlined in preceding section, the specific answer is that the details can be tuned much like for any other policy – through iterative policymaking process of evaluation and adjustment. Since the very articulation of the policy cycle (Lasswell, 1956) – a concept that generated enormous traction in both academia and practice (Nakamura, 1987) – the policymaking process was presented as iterative in the sense of the last sequence of a policymaking process feeding back into the first sequence of the next iteration. As much as the concept has been critiqued for its many flaws since, its iterative nature is not often disputed. In fact, whether due to speeding up of policy cycles (Kitchin, 2014a; Lettieri, 2016; Maciejewski, 2016) or due to additional feedback mechanisms for individual stages of the policy cycle (Höchtl et al., 2016), the iterative nature of policymaking seems to be intensifying as a consequence of technological change rather than slowing down, causing some to place frequent experimentation and subsequent adjustment at the very core of the policymaking process (Sabel & Zeitlin, 2012).

These iterative feedback loops are even more important for social investment policy analysis: Firstly, because the uniqueness and importance of institutional contexts implies that any policy is going to be sub-optimal when implemented. This is because there is no ex-ante analytical approach to identifying and leveraging institutional complementarities and a policy windows in all the necessary areas is unlikely to be opened in unison. Secondly, because the changing nature of social risks implies that the policy will have to be changed when necessary to address these risks, which makes this iterative adjustment a part of the paradigm rather than just a feature of policymaking. In other words, some of the conceptual promise of social investment policy analysis (like optimal policy returns and addressing modern life course risks) come at the cost of more decision-making and complexity at the level of first order policymaking. This is apparent especially in comparison to the neoliberal critique, where policy prescriptions tend to remain relatively similar across context, driven by a very parsimonious (and/or dogmatic) third order and second order policy understanding. This is not the case for social investment, as considering institutional complementarities can lead to adjustment of multiple policies across multiple policy areas, identification of new problems with no existing policy solution, or potentially even shifting the focus entirely to a different set of necessary institutional pre-conditions rather than immediate social investment reform (Kazepov & Ranci, 2017). This is not to say that this level of context-specificity is inappropriate, but we ought to also be honest about the increased 'load' this places on first order policymaking and the policy feedback loops within it.

The first (and simplest) source of this increased load is the increase to speed, which is necessary due to the focus on 'preparing' over 'repairing': The mechanism through which social investment policies assist individuals might be during the transitions and 'preventative', but if we do not adjust policies often enough everyone who has been affected by less-than-optimal policy will be in need of 'repairing' interventions. Since social risks are dynamic, the optimal policy suite is as well, meaning that it needs to be re-evaluated frequently otherwise we are 'preparing **and** repairing' rather than 'preparing **instead of** repairing'. To really shift away from 'repairing' the policy feedback loops need to be very fast – ideally faster than how long it takes for a cohort to get 'stuck' in long-term unemployment and suffer from skill erosion (which would necessitate repairing). Following the method proposed by Hemerijck et al. (2016), which is the only available methodology for a general assessment of social investment, one needs to consider a methodologically broad range of evidence.

This evidence ranges from micro-level analysis of panel data to qualitative institutional analysis of a given policy in a given institutional context (Hemerijck et al., 2016). This method does produce an understanding of the various returns of social investment policies and the complementarities that play a role by 'triangulating' the insights from different methods, but the analysis is slow in terms of when it can deliver conclusions about current policies, relying not only on indicators suffering from substantial data lag but also on qualitative hindsight that takes years to obtain. Even though such a method is broader in its scope than analysis of individual transitions would require, there is no more specific method or approach proposed by the literature that would deliver sufficient insight and be available faster. Currently it seems that the reliance on detailed institutional analysis necessitates a level of hindsight that is almost antithetical to the mantra of 'preparing' over 'repairing'.

Another source of additional load is the increase in complexity of analysis. This is largely because of how multifaceted social investment is and that it considers multiple outcomes of interest at multiples points in time. Given the life course perspective the focus expands from immediate redistributive qualities of policies to medium term, long-term, and inter-generational returns in terms of human capital and labour market outcomes. These outcomes are also different for different groups of people (reflecting the meso-level focus) at different points in time. For example, early childhood education and care should immediately allow parents to increase their labour market participation, which in the medium term reduces skill erosion, improving human capital in the long term. The same early childhood education and care policy also affects children, ideally increasing socialization and cognitive development, allowing for better performance in school in the medium term and well as increased human capital in the long term. Social investment really highlights the multifaceted nature of policy returns and focuses on outcomes outside of the economic realm. This makes analysis of individual policies much more multifaceted.

Not analysing policies individually and instead focusing on individual transitions for a given group of people, which is what social investment does conceptually, ends up equally as complex due to the logic of institutional complementarities. Because of institutional complementarity a transitions can be improved by deploying various policies and this policy package can be seen as interacting with almost any other policy or institutional feature of a given context. This requires judgement with regard to which policy complementarities are the 'most important' to focus on, which is not a straightforward decision: Do we focus on complementarities well described in the literature? Those that best capture the uniqueness of the relevant institutional context? Those we have data available for? The pragmatic answer would perhaps be to focus on the 'strongest' complementarities in terms of demonstrable empirical record of affecting the outcome of interest, but that does notcircumvent the problem of selecting policies and features to test as 'complementarities' in the first place, not to mention the disputes and constraints that come with such a decision. The very idea of using existing empirical record of complementarity is somewhat at odds with the difficulty of transferring best practice across contexts, as the most relevant complementarities undoubtedly change as well. This means that that the complexity and 'sprawling' nature of social investment policy analysis is difficult to circumvent - it is a feature of the logic of institutional complementarities.

This increase in complexity also has some important political implications. Since the policy puzzle becomes very multi-objective and the various objectives can have trade-offs with one another, some of the political (or at least normative) decisions that were previously obvious from third and second level policymaking principles are now left for first order policymaking. Despite theorizing that social investment has a 'self-reinforcing' dynamic where its key goals reinforced one another and aggregate over the life course (Hemerijck et al., 2016), there are still obvious instances where different goals have (at least an immediate) trade-off. Consider for example features of unemployment assistance affecting how fast are individuals 'pushed' to re-enter the labour market: An earlier 'push' will increase employment in the short term but can come at the cost of lowering productivity and skill atrophy in the longer term (individuals might be pushed into a job not matching their skills). There is no analytical solution to this trade-off – even perfect analysis would only expose a set of pareto-optimal solutions (solutions where any gains in one goal come at the cost to another goal) and selecting one from this set remains a political decision. For policy approaches influenced by social investment, "[t]he range of possible interventions is infinite and the advocates for any intervention will use a mix of normative, political, humanitarian and rhetorical arguments with very mixed access to an evidence base" (Gluckman, 2017: 7). This is not to say that such dilemmas are exclusive only to social investment, but they are heightened by it multifaceted nature and by how much discretion is still left for first order policymaking. There is no immediately apparent solution: Involving accountable policymakers more with the details of first-order policy adjustment is likely to conflict with the goal of quick policy feedback loops, while outsourcing these decisions to unaccountable experts would make the entire process less democratically accountable. As much as social investment reconciles (to an extent) some goals that were believed to trade-off with one another (Wrenn, 2017), it by no means reconciles all familiar political divides and it is important to acknowledge that those might be increasingly decided at the level of first order policymaking.

In combination this puts a tremendous load on policy feedback loop at the first order level: Policy learning should happen faster to fit policies to institutional context, it should happen with a lot of complexity to account for as many outcomes of interest as possible, and it should ideally maintain some degree of accountability for political choices. Even though this seems like an impossible task, social investment has undergone some conceptual development, making it capable of providing insight despite this complexity and analytical load. This development comes mainly from better describing institutional complementarities as interaction between 'types' of policies or policy functions. The best example of this is Hemerijck's (2014) distinction between three policy functions: Developing and protecting of the **stock** of human capital and capabilities, enabling the **flow** of workers to and from the labour market as well as within the labour market, and providing sufficient safety-net **buffer** both as social protection and aid for economic stabilisation. Institutional complementarities are then understood as the interaction of these three policy functions for a specific life course transition, but also as reinforcement of individual functions across time (eg. investment in early education as stock-building improves returns available from investment in higher education later) (Hemerijck et al., 2016; Hemerijck & Vydra, 2016). These three policy functions and multiple types of interactions allow us to understand policy complementarities better, but it still does not help with the tuning of a given policy intervention in a given context, at least not in the evidence-based sense.

One of the reasons why such a conceptual understanding does not translate well into evidence-based tools is functional overlap and mutually reinforcing mechanics between these policy functions: Social policies tend to satisfy multiple functions, often for multiple beneficiaries, and in different time frames, making it very difficult to assign a singular policy function to any policy intervention (De Deken, 2014, 2017). For example, child care provisions aim to facilitate learning for the child (stock) but also allow the parents to (re)enter the labour market (flow), which in turn secures additional income for the household (buffer). These policy functions are also mutually reinforcing in the sense that one policy function also supports the others (De Deken, 2014; Dräbing & Nelson, 2017). For example, any 'buffering' policy that provides financial security will also free up capital for skill development and education (stock) and allow for better job-matching by allowing for longer job search (flow). This has been a thorn in the side of social investment since Nolan's (2013) argument that approaching social investment as a conceptual base for an analytical framework leaves much to be desired since it is difficult to differentiate between social 'investment' and other types of social spending. A similar argument is used to critique even accounts utilizing the distinction between stocks, flows, and buffers and argue that "the distinction between investment-oriented and consumption-oriented policies remains flawed" (Parolin & Van Lancker, 2021: 301) and that there is little consistency in how certain policy areas are labelled by different authors in the literature with no theoretical or analytical resolution (Parolin & Van Lancker, 2021). This is not to say that heuristics such as stocks, flows, and buffers are not useful, but rather than their usefulness lies in facilitating the understanding of social investment as a paradigm and political buy-in (something seen as crucial for facilitating policy change, as explained earlier), rather than in laying a foundation for an analytical framework that could support evidence-based policymaking, let alone at the speed and complexity required.

3.3 Towards an analytical framework using big data?

Despite presenting a somewhat bleak account of (first order) social investment policy analysis – a very promising analytical proposition, but one that places a very heavy load on policy feedback loops – there is room for optimism. This optimism is tied to recent trends of digitalization, datafication, and increased sophistication of big data analysis as trends that open up a promising research agenda for developing social investment policy analysis as a more analytically astute set of tools rather than conceptual heuristics. This chapter sees this agenda as promising for three primary reasons:

Firstly, in going through important life course transitions individuals need to search for information relevant to their transition (such as job openings if attempting to re-enter the labour market), search for information about available public services and their accessibility, potentially reach out to their social network for assistance and advice, or simply share their frustrations or successes with their social network. Both the search for information and interaction with one's social network are activities that are largely (and increasingly so) conducted online using search engines and various types of social networking platforms. In using these technologies, individuals are leaving a great amount of 'digital traces', which are stored as micro-level data about important parts of their transition. Changes in search behaviour or social media posts can indicate not just what a person is looking for, but also what are the issues preventing them from transitioning successfully: People can share their difficulties explicitly on social media, but a search for 'cheaper' services or services closer to one's residence might indicate a similar policy incomplementarity without a direct complaint. Conceptually, there is a very good pairing between what information is crucial for social investment policy analysis (detailed information about how individuals go through life course transitions) and the type of information captured as a 'by-product' of individuals using services such as search engines or social media. This is of course not universally or equally true for all life course transitions, but some of the most important elements of the most important transitions (such as job search and matching) are conducted mainly online (Stevenson, 2008) in many contexts.

Secondly, this data also has potential to deliver on the 'preparing rather than repairing' promise of social investment, both conceptually and practically: Conceptually, one of the drawbacks to using big data for policy analysis as mentioned in the second chapter becomes a strength: Much of the information provided by big data temporarily precedes information from traditional indicators such as (un)employment. Utilizing the same example from the second chapter, to become employed an individual needs to find fitting job vacancies, go through a selection process, negotiate terms of employment, and only then become employed. This is a drawback if big data re used to study unemployment, but an advantage if they are used to study the features of transitions that precede eventual (un)employment, which is precisely what social investment aims to understand and where policy complementarities optimize policy returns – during the transition rather than after. To understand what complementarities drive success or failure, it is necessary to look at precisely this type of data. Practically, many of these 'digital traces' are available as data near instantaneously – as soon as an individual searches for something online or makes a social media post that data is stored and can be processed. This removes an important data-lag inherent to traditional indicators of interest and would allow such evidence to potentially be used in faster policy feedback loops. Not only will this data respond to policy change faster, it is also available faster.

Thirdly, the less restricted scope of what information is collected as big data matches well with the 'sprawling' nature of social investment policy analysis, where new complementarities might become apparent throughout the process and need to be investigated. Here it is important to not succumb to techno-optimism and believe that the data captures the full domain of social reality, but it is also important to acknowledge that data such as peoples' social media posts are not restricted by research designs and survey questions and it is much easier to query such data again for different information should it be required to assess a newly identified complementarity. Furthermore, this data itself can potentially be utilized to identify such (in)complementarities by aggregating what people themselves identify as problematic (by eg. sharing frustrations on social media) or what their behaviour indicates. This allows analysis to be much less restrictive in scope and consider more complementarities in a relatively time-sensitive fashion.

In other words, features of big data analysis seem to be a very good fit for social investment policy analysis. To reinforce this argument this chapter reviews some of the existing research in social policy and economics that utilizes big data, which generally tends to be focused on labour market flows (Taylor et al., 2014). Labour market flows are of course a very good fit with social investment and its emphasis on life course transitions. However, much of this literature is focused on measuring unemployment and other crucial indicators in real-time. The pioneering effort in this area was conducted by Askitas & Zimmermann (2009), whose research showed that combination of Google search volumes for the Unemployment agency and biggest job search engines approximately mirrors the unemployment rate (Askitas & Zimmermann, 2009). Since then, Google data have been used to predict or improve existing predictive models for unemployment in the US (Choi & Varian, 2009; D'Amuri & Marcucci, 2010), in the UK (McLaren & Shanbhogue, 2011), in Norway (Anvik & Gjelstad, 2010), in France (Fondeur & Karamé, 2013), in Spain (Vicente et al., 2015), and in Italy (D'Amuri, 2009). Google data can be used in this fashion to also measure unemployment for a population sub-group of interest, such as youth unemployment in Italy (Naccarato et al., 2018) or France (Fondeur & Karamé, 2013). Outside of search query data Twitter can also be used to measure unemployment with some attempts achieving sub-optimal accuracy (Biorci et al., 2017), some outperforming conventional models in terms of accuracy (Proserpio et al., 2016), and some showing promise when originally published (Antenucci et al., 2014) but eventually diverging from unemployment figures when compiled past the research study period (University of Michigan, 2015).

Despite the focus on unemployment, some of the research shows (somewhat unintentionally) great promise for understanding life course transitions. For example, Proserpio et al. (2016) use Twitter data to quantify the psychological state of individuals and are able to not only show the impacts of loss of employment on individuals, but also show that “psychological well-being measures are leading indicators, predicting economic indices weeks in advance with higher accuracy than baseline models.” (Proserpio et al., 2016: 223). A similar conclusion is also reached by the UN Pulse initiative, claiming that job related micro-blogging data can be used to extract sentiment and information about behaviours that act as leading indicators for a macro-level unemployment spike (United Nations, 2011). These studies are explaining unemployment using the implicit understanding at the core of social investment – that employment change is a part of a transitionary period and that individual behaviour during that transition is consequential to the outcome of said transition (both in terms of immediate job loss and future employment prospects). There is also further research that explicitly moves ‘beyond’ measuring unemployment: Antenuccia et al. (2014) focus not just on unemployment but general labour market flows (job loss, job search and job posting). Baker & Fradkin (2011) use Google data and focus on job search and the drivers of job search and quantify the effect of unemployment policy on job search. Job listing data in natural language can be categorized into official categories (Colace et al., 2019; Turrell et al., 2019) and used to calculate information about labour supply and demand specific to professions and regions (Turrell et al., 2019). Other data sources have also been explored, such as Toole et al. (2015) using mobile phone geo-location data to explore the link between mobility patterns of individuals and labour market shock, allowing them to now-cast regional unemployment (similar research was also conducted by Dong et al. (2017)). Outside of research projects like this various national statistical institutions are running pilot programs trying to utilize big data to assess mobility patterns (Barcaroli & Righi, 2015; Heerschap et al., 2014; Lopez, 2016) or labour markets and unemployment (Barcaroli & Righi, 2015; McLaren & Shanbhogue, 2011).

The promise of this type of data for evaluating social investment policies has been acknowledged in the literature, as ‘citizen-based’ analytics utilizing detailed micro-level data arguably “allows the complex realities of multi-dimensional inputs and outcomes over the broad span of the life-course to be properly considered. It creates the possibility of fast feedback loops to adapt programmes as information as to their effectiveness emerges” (Gluckman, 2017: 18). That said, the utilization of such data and methods understandably raises concerns about issues discussed in chapter two, issues such as algorithmic bias or overestimating how ‘individual’ an analysis based on group characteristics can be (Gluckman, 2017: 17). This chapter echoes such a call for realism: A call for the simultaneous appreciation of the power of complex and group-specific analysis as well as reluctance to accept such analysis at ‘face value’ without probing for the errors and dangers we know it can come with.

In light of considering the risks of a big-data enabled approach the case of New Zealand is good to consider – a case where ‘social investment’ is adopted as an official approach and draws on big data enable evidence-based policymaking: It adopts the general understanding that social policy interventions can have multiple very varied impacts later in the life course and the “approach is highly innovative and at the cutting-edge of applying citizen-based analytics for social policy development” (Gluckman, 2017: 3) with a focus on “a very high standard of evidence and analysis” (Scott, 2018: 37). This approach is supported by the Social Investment Analytical Layer, which is a platform collecting detailed data about citizens' utilization of government services across a range of agencies. Such an approach is certainly in line with using big data, but the understanding of social investment here is different to the mainstream academic literature: The focus is much more on demonstrating returns of the investment, privileging spending on social policies that can do so and ideally policies that show a decreased reliance on other types of social policies down the line (in a 'doing less with more' mantra). This approach motivates a criticsm about over-surveiling the poor and governing strictly by the number, resulting in a situation where "the dominant logic of welfare reform becomes singularly focused on reducing the numbers of people "dependent on welfare"" (Staines et al., 2021: 159). Criticisms like this are illustrative of how distinct 'variations' of social investment are in comparison to the mainstream (where social policy is viewed in 'positive' terms and as capable of generating substantial returns for low public cost). The criticism to 'governing by the numbers' in this case is well aligned with the second chapter of this dissertation: It rejects the notion that big data can capture the 'full-domain' of reality and do so in an objective way and thus "big data under social investment is not equivalent to individuals'stories; nor are they neutral or complete" (Staines et al., 2021: 164). Databases enabling these types of analysis are constructed using subjective and normative assumptions about what the problem is and how it can be fixed "in ways that enable some questions to be asked and answered (and thus, some types of data to be valued and captured), while others are ignored and subsequently devalued" (Staines et al., 2021: 165). The utilization of big data in this space certainly shows great promise, but it also highlights the importance of a realist approach, one that isn't reductive of the complexities within social investment and one that takes seriously other impacts of big data use besides maximizing certain metrics and minimizing costs.

3.4 Conclusions

This chapter described what social investment means for first order policymaking, focusing on the iterative process of fine tuning policy provisions to achieve 'optimal' policy outcomes. As much as it is difficult to paint such a process with a broad brush, this chapter outlines some analytical principles that stem from key features of social investment and that should guide policy analysis. As argued above, first order policymaking is very important in social investment, as a lot of important decisions determining priorities, trade-offs, and policy efficiency are made at that level. Understanding what is expected from evidence-based tools at that level, what the currently utilized tools are, and perhaps even developing some are crucial tasks for social investment as a paradigm. Outside of the practical utility of doing so, understanding these tools better also provides an alternative (or complementary) explanations for how (in)frequently is the logic of social investment adopted in policymaking practice.

As we can see in practice there is a desire to measure social investment logic for first order policy making, but attempts have so far not been entirely consistent with one another or with the social investment literature: In the EU, the current approach assumes first order policymaking to happen via social innovation. Social investment is argued to capture the broad ideas and objectives of social policy reform and social innovation represents the driver of such a change (Maduro et al., 2018: 17). In that vein, the ICT-Enabled Social Innovation (IESI) project of the EU also acknowledges the key role ICT has in providing an understanding of how actions suggested in the SIP can be implemented (European Commission, n.d.). In New Zealand social investment itself is very explicitly focused on evidence-based policymaking and is seen as a combination of life-course approach with citizen-based analytics (Gluckman, 2017) or as a combination of "targeting, support for decision makers and agility with respect to learning, innovation and reinvestment" (Scott, 2018: 36). Even though these contexts are substantially different and neither obviously follows the analytical principles of social investment (as articulated in this chapter) very closely, both approaches put evidence-based policymaking front and center. This should convince even those who do not believe evidence-based policymaking is an important part of social investment – regardless of whether social investment 'should' contain such an analytical approach, it is an approach through which social investment is translated into policy advice in some cases. As such, even if it is to critique such approaches, one needs to understand what social investment implies for evidence based policymaking and how that plays out (or does not) in policymaking practice.

That said, this chapter maintains that developing this understanding is more important than simply a necessity imposed on social investment by policymaking reality. Social investment has famously been critiqued for being two things at once – a paradigm/strategy for reforming the welfare state as well as an academic framework for looking at and analysing social policy (Nolan, 2013). The assertion seems true, but it seems equally true for other paradigms of welfare state understanding: The underlying models and analysis are distinct from the substantive goals and advocated reforms. The strength of a paradigm, social investment included, is in whether these two layers complement and support one another. For example, at the layer of paradigm/strategy social investment scholars theorize a 'life course multiplier' effect (Hemerijck et al., 2016; Hemerijck & Plavgo, 2021): As people increase their human capital, become more used to life course transitions, and build robust dual-earner families the 'returns' of social investment compound themselves over individual life course and even across generations. However, this multiplier effect is "contingent on good institutional complementarities" (Hemerijck et al., 2016: 25), making it crucially connected to the analytical layer of social investment. Without the analytical principles such as institutional complementarities social investment is a political platform advocating some social spending over another, and without an overarching narrative and strategy for welfare state reform its analytical proposition amounts to little more than a life course perspective with appreciation for complexity. Neither one of these on their own constitutes a paradigm for understanding social policy and the welfare state, it is the combination of the two that does. As such, it is crucially important to develop the analytical layer of social investment not just to translate its paradigmatic claim to executable policy prescriptions, but also to provide support for its paradigmatic propositions.

That said, there are two caveats to what this chapter argues that should be addressed head on. Firstly, the call for a big-data research direction is not a call for methodological singularism in the sense that big data analysis ought to replace current policy analysis methods. The available research indicates that social investment, likely due to its multifaceted nature, benefits from a methodologically pluralist 'layered' approach, where different types of evidence are used to understand a particular mix of policies and their interaction (Hemerijck et al., 2016). In such a layered method big data enabled first order policymaking tools would allow for fast policy learning and adjustment based on the social investment logic, with bigger adjustments most likely reserved for the outcomes of more rigorous methods focused on longer-term outcomes. According to Gluckman "[a]lthough citizen-based analytics can become an important part of social sector policy-making, they do not displace the need for continual improvement in traditional policy analysis approaches, including rigorous mixed methods and qualitative work" (Gluckman, 2017: 18). This plurality must also extend to analyses that do not follow social investment analytical principles, such as the analysis of adverse ‘Matthew effects’ (Ghysels & Van Lancker, 2011; Lancker & Ghysels, 2012; Parolin & Van Lancker, 2021; Pavolini & Van Lancker, 2018). This type of analysis generally does not look at outcomes across the life course and instead looks at distributional outcomes of social investment policies, but it still describes an important issue that social investment must keep addressing. Secondly, the fact that institutional complementarities and other features of social investment are not yet operationalized and easily ‘measurable’ in the extant literature is not necessarily a critique of the paradigm: For comparison, when most of Keynesian economics was theorized business cycles were not easily measurable and comprehensive methods for their measurement only appeared after the theory (Burns & Mitchell, 1946) and have been substantially reworked since. As a comparatively young paradigm, it is perhaps not surprising that these methods are not yet available, but that is not a reason for leaving the understanding and development of such methods out of the research agenda.

With regards to this dissertation the conclusions are more straight forward than with regards to social investment as a whole: Firstly, the type of (first order) policy questions social investment asks align well with what information some big data sources are likely to contain. Secondly, The features of big data analysis and what they enable (such as increased speed or type of insight) align well with the general analytical principles of social investment policy analysis. And lastly, but most importantly, understanding and developing big data tools in this context is not just an arbitrary case to arrive at conclusions with regards to big data in the public sector. It is a case that is theoretically important as advancing social investment and practically important as an attempt to solve a dilemma that decision makers and experts have been facing for over a decade – how to translate social investment logic into context-specific policy reforms.

Chapter Four: Does it hold up? Testing big data's promise of novel information on labour market policymaking[3]

With the general understanding of why the extant literature diverges from policymaking practice and how this divergence can be improved (argued in chapter two), it is possible to conduct the type of empirical research hopefully capable of addressing this problem. This understanding also helps to position the entire dissertation with regards to the existing big data literature: Accounts of big data in public administration range from those seeing it as profoundly transformative (Höchtl et al., 2016; Maciejewski, 2016) to those highlighting the issues with big data and the difficulty of changing policymaking practice (Iacus, 2015; Lavertu, 2016). Situated between these two extremes are accounts seeing the transformative potential of big data as contingent on factors such as institutional readiness and capacity (Klievink et al., 2017), or, as this dissertation proposes, the specific alignment between data source, method, and policy question. This 'third way' approach understands big data in public administrations as a multi-faceted phenomenon and aims to generate a context-specific understanding of its utility, better mirroring the uneven adoption of big data methods in policymaking. This approach does not simply dismiss the 'big data promise' in its entirety - it aims to specify when (under what conditions) is this promise maximized and minimized.

Given the breadth of this 'big data promise' (as summarized in chapter two), it is not possible to tackle all aspects of this promise at once. The part of the promise this chapter focuses on is the promise of 'novel information' – the claim that because big data quantifies human interaction with 'higher resolution' and quantifies previously unquantified aspects of human interaction it can provide novel information. This part of the promise has been articulated at multiple levels: At the level of general research efforts Boyd & Crawford (2012) identify, somewhat critically, a mythological element of big data resting on 'the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible' (Boyd & Crawford, 2012: 663). At the level of public administrations one can argue that 'big data methods can uncover knowledge that was previously impossible to reveal. In turn, this new knowledge allows new tasks (previously impossible or even unimaginable) to be successfully carried out' (Maciejewski, 2016: 123). At the level of economic analysis and social indicators (a level closest to the empirical test this chapter conducts) one can argue that big data will 'allow economic researchers to test theories of behaviour that were previously untestable, creating a new set of metrics for issues of economic interest which were previously in the realm of theory' (Taylor, Schroeder, & Meyer, 2014: 5), or that 'Big Data have a number of relevant pros that make them very interesting also for the definition of new social indicators' (Di Bella et al., 2018: 871).

Despite the fact that this 'novel information' promise is articulated in the literature it remains empirically under-explored, creating an obvious instance of divergence between the theoretical

[3] This chapter is a yet unpublished Manuscript authored by Simon Vydra. More details are provided in the 'Authorship Contribution' section.

promise and empirical proof-of-concept efforts or policymaking practice. This chapter studies the promise of 'novel information' in the context of labour markets, which is very illustrative of this mismatch: Study of labour market flows are a data-rich and methodologically developed field of study in general, but also the most studied area of economics when it comes to big data research (Taylor et al., 2014). Despite the wealth of research most contributions attempts to predict or 'nowcast' key economic indicators, to predict marketing influence, or to substitute existing data such as census or labour market data by being cheaper and/or more accurate (Taylor et al., 2014). As mentioned in chapter three, now-casting unemployment is the prime example of replicating an existing indicator and is usually done using Google search query data (Anvik & Gjelstad, 2010; Askitas & Zimmermann, 2009; Choi & Varian, 2009, 2012; D'Amuri, 2009; D'Amuri & Marcucci, 2010; Fondeur & Karamé, 2013; McLaren & Shanbhogue, 2011; Naccarato et al., 2018; Vicente et al., 2015) or micro-blogging and social media data (Antenucci et al., 2014; Biorci et al., 2017; Proserpio et al., 2016). Other relevant research utilizes job listing data (Colace et al., 2019; Turrell et al., 2019) or mobile phone geo-location data (Dong et al., 2017; Toole et al., 2015). As much as some of the existing research provides additional 'depth' by looking at drivers of job search (Baker & Fradkin, 2011) or at psychological variables as leading indicators for a change in conventional economic indicators (Proserpio et al., 2016), the overarching ambition remains to replicate 'traditional' economic indicators. Even from a broader social indicators perspective big data is generally used to create proxies for social indicators or to substitute traditional surveys (Di Bella et al., 2018: 876). This allows us to understand (at least partially) whether big data sources can replace some existing indicators and with what potential benefits or drawbacks, but it does not really test the potential to provide novel information.

This chapter aims to provide the missing empirical test with regards to labor market flows by answering the research question **"Can social media data be used to operationalize and measure social investment?"** This is where chapter three of this dissertation becomes crucial, as it provides a comprehensive argument for (amongst other things) what type of information one would need to truly measure social investment as an analytical tool and why big data could be a good source for this type of information. In the context of this chapter, the focus on life course transitions and various (in)complementarities that affect these transitions are pieces of information not traditionally measured by any economic statistic. Some relevant (in)complementarities, or their features, might be captured in statistics such as take-up rates of various policies, but some features involving interaction between crucial policy areas such as ECEC or ALMPs are not measured. Speaking of indicators for social investment a quote from chapter three bears re-iterating: "In terms of identifying core quantitative indicators, the [Social Investment Expert] group fell short of its ambitions. Our interviewees noted that while the immediate explanation is technical – that is, there are no indicators to assess SI – the underlying reason is political" (Porte & Natali, 2018: 838). Given the theoretical novelty and lack of existing empirical methods and indicators, social investment is a fitting 'testing ground' to see whether big data will, as stated above, 'allow economic researchers to test theories of behaviour that were previously untestable, creating a new set of metrics for issues of economic interest which were previously in the realm of theory' (Taylor, Schroeder, & Meyer, 2014: 5).

To provide this test the chapter as a whole focuses on a specific policy area(s) from the lens of social investment and a particular data source. The research scope of this chapter is motivated by the understanding that terms like 'big data' and 'policymaking' are too sprawling and trying to understand their interaction at such aggregate level is bound to be misleading (chapter two)(Vydra & Klievink, 2019). The social media data source in this dissertation is Twitter – a selection defended in this chapter. This results in three research sub-questions this chapter answers in order to answer its primary research question:

1. How can required information be extracted from social media data?
2. What are the implications of this process for policymaking?
3. Do Twitter data contain the requisite information?

To avoid some of the limitations of this research with regards to external validity this chapter presents its methodology and reflection on its implications 'on par' with the empirical findings in this way. This is done to highlight the importance of these implications for practice – in this case identifying these implications is a more consequential finding than answering sub-question three and it is generalizable beyond the platform and policy selection of this chapter.

The chapter proceeds in the following steps: **Section 4.1** specifies the pairing of social investment and social media data (argued for in chapter three) to a platform, national context, and life course transition of interest. **Section 4.2** answers the first research sub-question by outlining the proposed method to identify and summarize relevant information in a corpus of tweets. **Section 4.3** then focuses on the implications of such a method in the context of policymaking. **Section 4.4** then answers the empirical question by testing whether a year-long corpus of Dutch tweets contains relevant information. **Section 4.5** then concludes and highlights some shortcomings of the chapter.

4.1 Dataset and life course transition

Following the focus on alignment between policy question, data, and methods it is crucial to specify all three more. Chapter three provides a conceptual argument for why social media is likely to contain the information necessary for social investment policymaking, but not all combinations of life course transitions or social media platforms make a good pairing for empirical research. In terms of life course transitions, and the policy areas those are most relevant to, this chapter focuses on the transition from parenthood to employment (or staying in employment while becoming a parent).This is because it is considered the 'flagship' complementarity in social investment (Hemerijck et al., 2016), because it is conceptually very clear why toddlers being cared for is necessary to free up both parents to pursue full-time employment, and because the complementarity between ECEC and (primarily maternal) employment can be observed empirically (Esping-Andersen, 2015; Simonsen, 2010; Verbist, 2017). In simplest terms: Since toddlers cannot be left to their own devices ECEC provisions are crucial to allow parents to pursue employment or remain employed, increasing the returns from policies aiming to integrate people into the labour market. Outside of the clarity of this complementarity and its political salience, there is also a practical consideration of the age of individuals going through these transitions – individuals of peak child rearing age (who are likely to go through this transition) are also likely to be active on social media.

The parenthood to employment transition can be affected by a range of policies from multiple policy areas, but the key two policy areas in this respect are labour market policies and early childhood education and care (ECEC) policies. This under-delivers on the social investment promise to take into account a large array of complementarities, but it is a valid and more focused test of the potential for novel information. What constitutes 'relevant' information in this case is determined by three simple criteria - specificity to an individual's experience, specificity to the labour or ECEC situation, and inclusion of commentary about what factors are making a transition easier or more difficult. This is a rather broad criteria for distinguishing relevant information, but it is important to acknowledge that individuals are not necessarily (or even likely) going to attach difficulties with a transition to a particular policy intervention. Since commentary on what makes a transition difficult could be contained in querying one's social network, asking for help, or simply sharing experiences it is better to 'cast the net wide' to make sure relevant information is not disregarded because users themselves do not link it to a policy. In filtering social media posts the focus is thus on posts that are specific to ECEC and labour market legislation, but also on posts that mention any of the commonly available childcare options, types of employment, and phrases related to employment status and job search. A full list of all the concepts (and keywords that capture those concepts) is available in appendix A.

With regards to finding this type of information on social media, chapter three provides a general conceptual argument, but more can be said about the fit of social media data for this specific transition. On Twitter posting updates about one's life is the most common use of the platform (Java et al., 2007), likely driven by the gratification individuals get from satisfying the need to belong and the need for self-presentation (Java et al., 2007). By the very definition of life course transitions, they have to do with very important events in one's life (such as childbirth or new employment), making them something worthy of sharing both to indicate a belonging to a certain group of people as well as self-presenting. For new mothers specifically, blogging and social media can improve various wellbeing indicators, as they feel more connected to the world outside of their home and they receive social support from friends and extended family (McDaniel et al., 2012). As such, it is not surprising that new mothers experience increased Facebook use when transitioning into parenthood (Bartholomew et al., 2012). It is also possible to identify new mothers on Twitter and observe changes in their behaviour and emotions prior to and after giving birth (De Choudhury et al., 2013).

This chapter focuses on the country of Netherlands, motivated by both practical and theoretical reasons. From a practical perspective the Netherlands has very high internet penetration and high social media use, especially for Twitter, which is the social media platform this chapter uses (discussed in paragraph below). From a theoretical perspective Netherlands is one of the earlier adopters of social investment policies (Nikolai, 2011) and boasts a relatively well developed social investment agenda, but one that has faced some challenge in the forms of substantial budget cuts, especially following the 2008 financial crisis (Soentken et al., 2017). This means that Netherlands does have the main policy elements that together support individuals in their transition from parenthood to employment, but also that those elements are not perfect and are changing. This should strike the ideal balance between social investment states where these policies are so institutionalized that people are unlikely to comment on them and states with no social investment where these policies are a very small part of the public debate.

As a social media platform this chapter utilizes Twitter – a choice driven largely by practical reasons. The ideal platform would likely be Facebook, be it due to its extremely high usage in European countries, or due to the symmetry in the 'friendship' relationship that Facebook is built on. In the Netherlands specifically the internet penetration is around 96% with close to 60% of those users having a Facebook account (Internet World Stats, 2017). However, Facebook API no longer supports searching through public posts and obtaining the data via web-crawlers is against the terms of service. Twitter on the other hand allows access to a fraction of the full stream of tweets (maximum

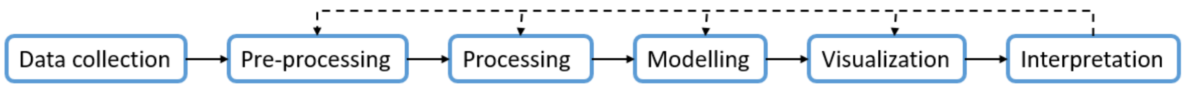
1% of total volume of tweets), which is sufficient given the specification to Dutch language and a set of relevant keywords this chapter adopts (the volume of these Tweets does not surpass 1% of overall Twitter traffic). Despite not being the ideal platform, Netherlands maintains a good amount of Twitter users at 2.8 million for 2018 and 2.5 million for 2019 (Statista, 2019), which is approximately 14.5% of the total population in 2019. However, given the focus on individuals of child rearing age Twitter penetration is likely to be substantially larger, with a 2018 poll reporting 26% of 20 to 39 year olds in the Netherlands using Twitter (Statista, 2018), which is precisely the age group this chapter focuses on.

This dataset was gathered using Twitters Filter API for one year spanning from August 2018 to (and including) July 2019. The tweets have been collected based on being written in the Dutch language (to control for national context) and containing a word or a phrase from a set of keywords included in appendix A (including English translations). This does exclude English speakers living in the Netherlands, but given how few Tweets are geo-tagged using geo-location to control for national context would have resulted in an extremely small dataset. Other more sophisticated processing methods to control for national context could certainly be designed, but given the 1% of total tweets restriction on data collection it is impossible to process all tweets so either location or language filter is necessary. The second filter included a set of 139 relevant keywords, many of which refer to the same concept but are spelled with and without a space or using English terms that have become popularly used in Dutch (appendix A). The total number of tweets collected throughout this period is just over 4.15 million tweets.

4.2 Method

With a dataset in place and relevant information outlined, the focus turns to the first research sub-question of this chapter: **How can required information be extracted from social media data?** The method this chapter proposes (and subsequently adopts) is iterative topic modelling. Topic modelling allows one to summarize very large corpus of texts in terms of the topics addressed in those texts. Topic modelling rests on the intuition that when talking about a particular topic individuals will use words reflecting that choice – some words are more related to some topics than others. This intuition is formalized into statistical models that are capable of describing which topics are contained in a document, as well as which terms are contained in each topic. Before a topic model can be trained tweets need to be collected, irrelevant tweets needs to be excluded, and relevant tweets need to be pre-processed to a format suitable for training a topic model. The entire process is also very iterative due to the fact that topic models can be constructed in many different ways and the quality of the output is judged primarily by the 'interpretability' of topics to human observers. The basic trajectory of this iterative approach is outlined in figure 4.1.

Figure 4.1: Method summary



With the data collection step outlined in previous section (using Twitter's streaming API and a set of keywords to control for content of tweets and a language restriction to control for national context) this section turns to the other analytical steps and how they apply to the dataset at hand.

4.2.1 Pre-processing

As part of pre-processing all re-tweets were removed (1.36 million tweets) followed by removing duplicate tweets, referring to tweets that have identical text to another tweet in the dataset but that are neither officially re-tweets and do not start with the 'RT' tag indicating a re-tweet (91 thousand tweets). Quoted tweets have been pre-processed by joining the text of the tweet at hand and the text of tweet it was responding to (222 thousand tweets). This manner of pre-processing quoted tweets strikes a balance of excluding the automated tweets and simple re-tweets of them, but also preserving the context of what users are responding to.

Another part of pre-processing – bot removal – is more conceptually and technically challenging. Conceptually because the distinction between 'bots' and 'humans' is fuzzy. Some accounts are 'cyborg accounts' subjected to automation but also human intervention (Nimmo, 2019), some only going as far as scheduling the dissemination of human-created posts (Radziwill & Benton, 2016). Technically this is a challenge often tackled as a supervised learning problem (Andriotis & Takasu, 2019; Inuwa-Dutse et al., 2018; Kantepe & Ganiz, 2017; Lee et al., 2011; Lundberg et al., 2019; Mazza et al., 2019), but such an approach necessitates ground truth about which accounts are bot accounts and which are not. Most available methods are too resource-intensive for a dataset of this size or not applicable to the dataset. One attempted approach was to use Twitter's own bot detection algorithm by checking whether all accounts in our corpus are still active accounts a month after data collection concluded, following the logic of Kantepe & Ganiz (2017). However, this approach assumes that irretrievable accounts have been removed by Twitter (as opposed to users disabling the account) and that the removed users are removed for bot-like behaviour rather than other breaches of the terms of service. These assumptions are not correct for our corpus as most removed users do not exhibit bot like behaviour (verified by exploring removed user's content and tweeting behaviour). This chapter thus adopts a simpler and 'cruder' approach of removing tweets that are authored by an account with a very high tweet frequency that a human is unlikely to exhibit. This is judged based on two constructed metrics: Firstly, the overall number of monthly tweets calculated by total posts divided by account age. Secondly by the monthly number of tweets captured in our dataset (tweets concerned with jobs and ECEC). Even though this approach is likely not as accurate as training a machine learning algorithm on a manually annotated data set, it is parsimonious, cost-effective, language independent, and very easily interpretable. The thresholds adopted are 1500 for general tweeting frequency per month and 450 for tweets captured in our dataset. This removes additional 1.01 million tweets from the dataset, resulting in a final dataset of 1.69 million tweets.

4.2.2 Processing

The text of tweets now needs to be processed into a format suitable for topic modelling. The most substantial step in this process is lemmatization, which tackles the problem that words often have inflectional and derivationally related forms that can make what is effectively the same word appear in many different forms (for example 'to work', 'works', 'working'), which is solved by extracting the 'lemma' of a word (its base dictionary form). The 'Frog' advanced natural language processing suite (Van Den Bosch et al., 2007) is used to lemmatize the corpus. The lemmatized tweets are then converted into a format that models can easily train on, which in this chapter is a simple 'bag of words' representation. This represents each tweet as a collection of individual unigram tokens whose occurrence in a tweet is summed.

4.2.3 Modelling

For the topic modelling step this chapter makes use of four popular topic models: Latent Dirichlet Allocation (LDA), Non-negative Matrix Factorization (NMF), Correlated Topic Model (CTM), and Structural Topic Model (STM). Use of multiple models is motivated by the fact that different topic models, despite capturing generally similar information, can produce different topics or capture the 'same' topics differently (Contreras-Piña & Ríos, 2016). Given the popularity, interpretability, and relatively low computational requirements of the LDA model we use it as a 'baseline' model for much of the iterative process, using other models to provide additional detail once we have identified a good number of topics using LDA models. We do inspect additional NMF, CTM, and STM models, but in our interpretation of the results we only include one version for these non-baseline models. This is a pragmatic choice as presenting a larger range of models for each model type would result in much longer analysis without any important additional findings due to the relative similarity of those models.

With regards to the models themselves: **NMF** is deterministic (and likely the most straight-forward) model following the simple notion that a term-document matrix can be decomposed into two matrices that when multiplied approximate the original term-document matrix: $\boldsymbol{A_{m\times n}} \approx \boldsymbol{W_{m\times k}H_{k\times n}}$ where n is the number of documents, m is the number of terms (words), and k is the number of topics. The NMF models were constructed using the scikit-learn python library. **LDA** for topic-modelling was proposed by Blei, Ng, & Jordan (2003) and it is a generative probabilistic model of a corpus that assumes the topic distribution to have a Dirichlet prior. The LDA models in this chapter are constructed using the Gensim python library which uses online variational Bayes algorithm to train the model (Hoffman et al., 2010). CTM and STM models are inspected to provide additional nuance as LDA models do not take into account correlation between topics or influence of document-level metadata on the prevalence and/or content of topics. **CTM** models assume logistic normal distribution (rather than a dirichlet one) and capture covariance amongst components, allowing for the assumption that the presence of a topic in a document is correlated with the presence or absence of other topics in that document (Blei & Lafferty, 2007). **STM** models incorporate document-level metadata and account for its relationship with topics both in terms of the prevalence of topics across documents and words across topics (Roberts et al., 2019). CTM and STM models were constructed using the stm package for R (Roberts et al., 2019).

4.2.4 Visualisation and interpretation

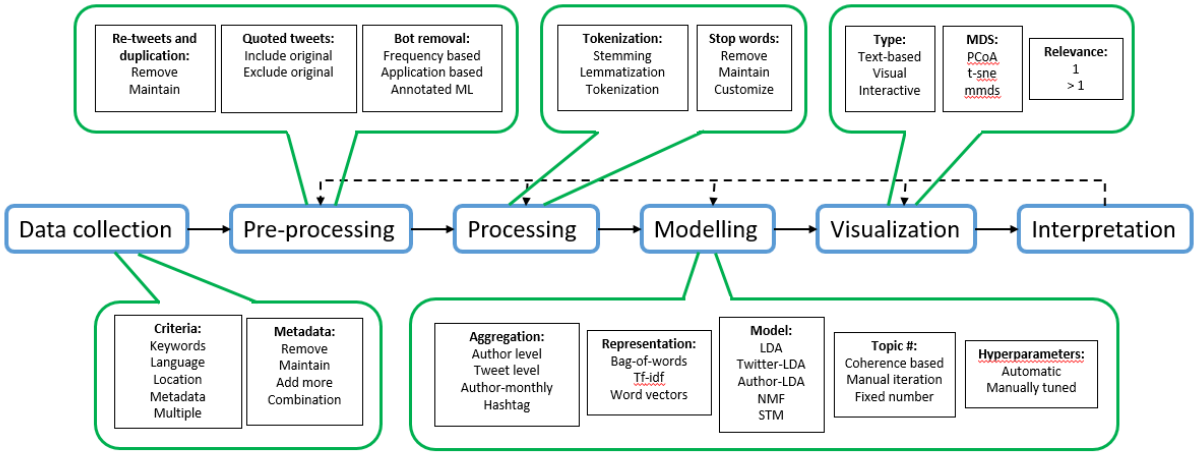
All models are visualised using primarily the LDAvis (for R) and pyLDAvis (a port to python) libraries to produce comparable visualisations across different models (Sievert & Shirley, 2014). This visualisation is selected because it provides the crucial baseline information (the words a topic model associates with each topic), but also because it provides additional relevant information like size of topics, semantic proximity of topics, or the volume of tokens captured in each topic. It is also an interactive visualisation which has two benefits: Firstly it reduces information overload and makes the model presentable - for models using a high number of topics a non-interactive summary can induce information overload and cannot be visualised on one screen without scrolling. Secondly, it also provides a 'relevance' metric that can be interactively changed to provide a deeper insight into the topics, which is important for topics that contain a lot of 'general' tokens that are plentiful in the corpus. These visualisations are used to identify and inspect topics, with relevant topics being summarized (by their top words) in appendix B.

4.3 Implications of topic modelling for policy practice

The proposed method is to be used in the context of policymaking, which comes with an additional set of expectations and challenges. In policymaking it is not just the model performance that makes a model 'fit for the job', but also its adherence to relevant public values and standards. This is an important procedural point often overlooked in the literature, but individual analytical steps can reflect assumptions about how systems in question work and what ought to be included or excluded from the analysis. There is no such thing as 'raw' data (Gitelman, 2013) and no such thing as 'value-free' analysis (Kettl, 2018), which makes it crucial to understand which parts of a method can reflect subjective assumptions about how systems do (or should) function. This understanding motivates the second research sub-question "**What are the implications of this process for policymaking?**" The process referred to in this question is the method proposed in section 4.2.

To answer this question this chapter iteratively evaluated decisions made in the analytical process in terms of whether they have an analytically 'right' and 'wrong' answer. Decisions that cannot be decided purely on analytical grounds but change the information a models provides are summarized in figure 4.2, which expands on the summary of the research process provided in figure 4.1. These decisions are a reflection of our assumptions about how individuals utilize social media and which content is of interest. As much as we believe our choices at these decision points are informed and not arbitrary, they do involve trade-offs and different analysts and policymakers might genuinely value what is being traded with differently. Furthermore, what we perceive to be the best decision often results in longer and more iterative analysis, which is something analysts cannot always afford to do in the policymaking context where human resources, computational resources, and time might all be severely constrained. In such a context one has to balance not just which decision is the 'best', but which 'good enough' decision best converts resources expended into improvements in the overall analysis. Furthermore, and perhaps most importantly, what is 'better' analysis or even how 'interpretable' a topic is are all subjective metrics, the importance of which can hardly be understated as interpretation is the step that starts a next iterative cycle and thus affects most other steps. Some analytical decisions might not even be a reflection of any assumptions but still need to be made and are potentially consequential for the outcome. In this section we provide a more critical look at some of these analytical steps.

Figure 4.2: 'Subjective' decisions consequential for analytical output



At the stage of data collection and pre-processing many of the decisions are obviously subjective and politically informed, such as deciding on keywords and additional criteria to filter by. In the case of this chapter this applies to constraining the data to the country of study: Using language of the tweet to do that is relatively straightforward, but it also includes Dutch speakers living abroad and excludes English (or other) speakers living in the Netherlands. An alternative approach is using the geo-tags of tweets, which runs into the problem of vast majority of tweets not being geo-tagged, resulting in large information loss. Some more advanced pre-processing approaches could be adopted, such as named entity recognition focusing on areas and cities in a given country, but such an approach is analytically challenging as it requires access to a much larger volume of tweets (since the language or geographical constraints are missing) and the analytical capabilities to process them in near real-time. Regardless of which approach one adopts there are trade-offs to be made between removing 'noise' from the data and information loss due to the fact that 'data' and 'noise' cannot be perfectly distinguished. This is true for many other steps at this stage of the analytical process. For example, in terms of bot removal, the thresholds we select are essentially a reflection of what we consider to be 'normal' tweeting behaviour and what deviation from this 'normal' is significant enough. Such decisions reflect not only our assumptions about user behaviour but also our preference on minimizing information loss versus removing 'noise' from the data. This dilemma is the same for our representation of quoted tweets, which, despite being reasoned, is a subjective one and consequential to what content gets represented in the analysis and what gets left out.

In the modelling stage many of these decisions are even more explicit, starting with one already mentioned in previous section – the number of topics. The topic coherence measure used has an important caveat: Since topic coherence aggregates the coherence of all topics in a given topic model, low coherence score does not necessarily mean that *all* the topics are uninterpretable. A model with a few topics can have all of its topics interpretable whereas a model with many topics can have only half of its topics interpretable but still provide a higher total of interpretable topics and better overall insight. This presents a trade-off between models that present simplistic and aggregate topics and models that contain more detailed topics but do so at the expense of also containing less interpretable topics. This is a trade-off that persists regardless of whether quantitative coherence metrics are used, the models are inspected manually, or both. Another explicit trade-off in the modelling stage are model hyper-parameters. These differ per model, but for our 'baseline' LDA models those hyper-parameters reflect our expectations with regards to sparsity of words per topic and topics per documents (very simply put). These hyper-parameters can be left to the model to 'learn' them from the data at hand (which is what this chapter does for most models), but especially for short micro-blogging data concerned with multiple policy areas there can be value in tweaking these hyper-parameters to reflect the assumption that a tweet likely contains relatively few topics (from the total number of modelled topics). Similar parameters that require a researcher to make assumptions about the corpus and topics within it are also present for other models like STM: What function of metadata variables determines topic prevalence is up to the researcher. In this chapter we opt for a rather simple function $prevalence = frequency + followers + spline(days)$ and we assume that the content does not change based on metadata. In choosing this function we effectively assume a linear relationship between topic prevalence and frequency or followers and a non-linear relationship between topic prevalence and the date a tweet was posted on. In the modelling steps there is also a decision that is consequential for model output but is more arbitrary. This decision is about the randomness involved in training the model. Different libraries offer different solutions to initializing the training of a topic model (in the case of gensim a 'seed' can be controlled). Sufficient training of the models helps with convergence of models generating based on different seeds, but some differences still remain.

The visualisation step contains a similarly arbitrary decision due to the necessity for dimensionality reduction. Because of the dimensionality reduction the two axes used to capture the 'semantic space' do not have an inherent meaning to them. The main purpose of this visualisation is to illustrate semantic similarity – essentially maintaining the distance between topics that exists in the topic model (which is n-dimensional where n is the number of topics) in a two-dimensional representation. This results in some identifiable 'clusters' of topics, which can help with interpretation, but the dimensionality reduction algorithm used drastically effects this visualisation. Principle coordinates analysis is utilized in this chapter as it seemingly provides most informative clustering, but that judgement is subjective and there are other algorithms that generate a substantially different semantic space and that are not analytically 'incorrect' to use.

Utilizing some of these analytical steps as illustrative examples reveals an important implication of this method: Many of the decisions that have to be made in topic modelling are consequential for the insight a model provides but also based on a subjective understanding of user behaviour and the broader social system, or even made arbitrarily and then iterated over in order to obtain the most interpretable (or desirable) outcome. This has some important implications for policymaking practice. Firstly, it highlights our lack of understanding of how some of these choices influence the result. Unfortunately, sensitivity analysis is difficult in this case because of the difficulty of objectively measuring 'interpretability' or 'value' different topic models provide. Secondly, there is the practical issue of politicians generally not being analytical experts and having to outsource this type of analysis to experts, which risks the outsourcing important subjective/normative decisions to experts who are not democratically accountable and who might not be aware of the importance of the assumptions they are making in the analytical process.

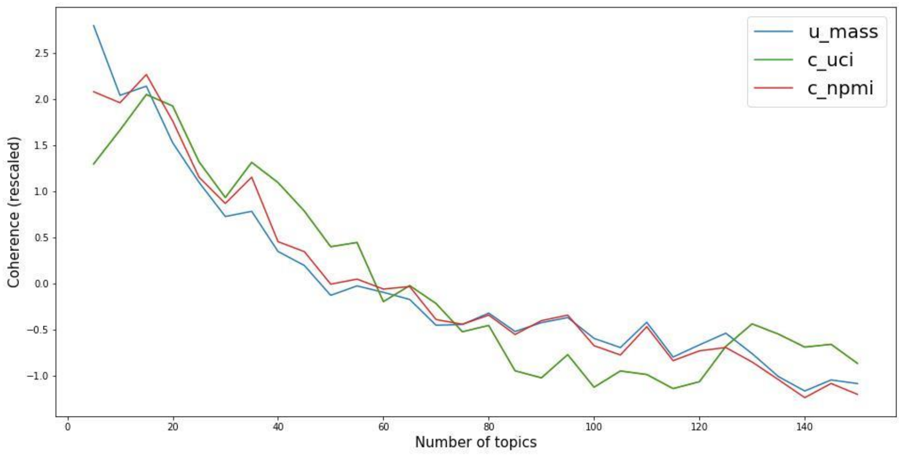
4.4 Empirical findings

The chapter now turns its attention to the final research sub-question "**do Twitter data contain the requisite information?**" By requisite information we refer to information that is a) personal to an individual's experience, b) concerned with a specific life-course transitions and the policies of interest related to it, and c) explicit about what factors are making that transition difficult or easy.

The topic modelling method proposed in section 4.2 is then applied across the entire processed corpus to generate an overall summary. This has the benefits of not just providing a necessary overview, but also eliminates an important risk inherent in more supervised methods: It is impossible to a-priori identify all the potential factors that can genuinely influence the parenthood to employment transition. Because of this, a topic could be relevant if inspected, but would not be searched for.

As mentioned in section 4.3, deciding for a specific number of topics to model will not be 'right' or 'wrong' and there is no substitute for careful validation of the actual topic models by making sense of individual topics (Grimmer & Stewart, 2013). To aid with selecting models for manual inspection we draw on the topic coherence approach of Röder, Both, & Hinneburg (2015), which aims at approximating human interpretability and the various interpretability metrics are plotted in figure 4.3 for LDA models from 5 to 150 topics (increment of 5 topics). Worth noting here is that the downward trend in coherence with the increase in topic number is unsurprising due to the fact that topics containing very general words tend to score well on coherence metrics (Roberts et al., 2014) – our aim here is to balance the detail of a topic model (which increases with number of topics) and the coherence of the model (which generally decreases with number of topics). As baseline we select models with 20, 35, and 65 topics due to their good performance in terms of topic coherence comparatively to models with a similar number of topics (an upward 'spike' of coherence in figure 4.3) and, more importantly, due to them capturing the range of interpretable models. Models with less than 20 topics tend to provide an overly crude summarization and models with more than 65 topics tend to start losing too much interpretability with no obvious gains in the insight provided. As such, we present and interpret LDA models for 20, 35, and 65 topics and NMF, CTM, and STM models for 35 topics (which we find to be the most informative topics number).

Figure 4.3: Coherence metrics for LDA topic models based on number of topics



Note: The three metrics are described by Röder et al. (2015) and implemented in gensim. There are multiple differences between them but fundamentally u_mass uses document co-occurrence counts to estimate coherence and c_uci and c_npmi use sliding windows and pointwise mutual information (which is normalized in c_npmi).

Given that this chapter presents multiple models we focus the overall summary of our data on topics that are robust across these models (identifiably present in each model) and relevant to the parenthood to employment transition in some way. We identify six such topics and in appendix B a full summary of these topics across the presented models is provided. The six topics robust across models are the following: First topic (table B1 in appendix) is concerned with employment openings advertised by employers and generally contains tokens like 'you', 'we', 'search', 'job offer', 'colleague', 'new', or 'team'. In some models this topic also includes the qualities a prospective employee should have (eg. 'enthusiastic'), seniority of the position (eg. 'manager' or 'assistant'), or location. Second topic (table B2 in appendix) is about training and education and includes tokens like 'training', 'day', 'education', 'internship', 'year', 'school', or 'teaching'. The topics are generally about education and following various programmes, with the exception of the 20 topic LDA model which combines education and athletic training making the topic more generic (35 topic models are detailed enough to disaggregate these two topics). Third topic (table B3 in appendix) is about unemployment and general social assistance, generally concerned about who is contributing to the welfare system and who receives the benefits (and if that ought to be so). It includes tokens like 'unemployment', 'employee insurance agency', 'payment', 'euro', 'assistance', 'to get', 'year', or 'right'. This topic is rather consistent across models but in the 65 topic LDA model it is less interpretable. Fourth topic (table B4 in appendix) is focused on the number of hours worked per month including tokens like 'hour', 'per', 'week', '24', '32', '36', '40'. This topic is focused mainly on job offers (and often co-occurs with topics concerned with job offers), but sometimes also includes education level and integers for years. Fifth topic (table B5 in appendix) is concerned with self-employment and freelance work including tokens like 'self-employment', 'the self-employed', 'entrepreneur', 'freelance', and 'interim' and the topic is both commentary on the situation of the self-employed as well as offers for freelance contracts. In some models the topic is more concerned with offers of contracts (mentioning the location for example) and in some more with the commentary (mentioning issues such as 'pension', 'obliged', or 'economy'). Sixth topic (table B6 in appendix) is concerned with childcare and includes tokens like 'childcare', 'daycare', 'kindergarden', 'pedagogical', or 'out of school care (bso)'. In some models this topic is more focused on the practicalities (including tokens like 'job application', 'company', or 'industry').

Of course, there are other interpretable topics, but some of these are irrelevant (such as a topic containing English words), extremely general, or not robust across enough models. The number of topics modelled makes a difference in expected ways – as the topic number increases the topics get more granular and detailed but the fraction of uninterpretable topics increases. A good illustrative example of this dynamic is the topic concerned with training and education: In the 20-topic model topic 14 combines training programmes with general education bundling 'training', 'internship', and 'year' together. In the 35-topic model these themes get disaggregated into topics concerned with formal education and topics concerned with athletic training. In some models this gets disaggregated even further with a topic about internships. It seems that the 35 topic model provides an appropriate level of aggregation for most topics, but not for all – some topics might be 'most understandable' in a 65 topic model and some in 35 topic model. This is partially why the multi-model approach of this chapter has substantial benefits, as we cannot expect all relevant topics to be most informative modelling only for one topic number.

The type of model also influences the summary we are able to get. NMF models generally capture similar insight but highlight different topics and are arguably less interpretable: Tokens tend to be included in more topics, giving many topics the appearance of 'mixed' topics and thus having multiple topics concerned with job offers that tend to differ mainly in terms of region and industry – which is a valid summary but not useful for the purposes of this chapter. The CTM models are a bit more similar to the baseline LDA models capturing many similar topics. The differences that exist do sometimes result in a novel topic: For example a 65 topic LDA model includes topics that are about general economic situation (topic 10 including tokens like 'joblessness', 'high', or 'government') or about pensions (topic 29 including tokens like 'everyone', 'obligate', 'pension', or 'expensive'). More specific versions of relevant topics also exist as, for example, the 65 topic LDA model contains a topic about parents, income, and childcare subsidy (topic 33) and the 35 topic CTM model contains a topic about internships (topic 8). Lastly, STM models present topics that are close to identical to CTM topics showing only a minor variation in how tokens are distributed across topics, which is understandable given that we model for a change in topic prevalence.

The main benefit of CTM and STM models is that they generate additional information with regards to correlations (co-occurrence in one tweet) between topics and correlation between topic prevalence and metadata (such as tweeting behaviour). However, in this case this is of little value as the primary finding is that it is impossible to identify a single topic that would satisfy our criteria for relevance with regards to social investment – information that is a) personal to an individual's experience, b) concerned with a specific life course transitions and the policies of interest related to it, and c) explicit about what factors are making that transition difficult or easy. Many topics are concerned with relevant policies or aspects of transition (this is expected due to keyword selection), but the specificity to an individual's experience is something we cannot really ascertain at a topic level. The answer to the third research sub-question thus appears to be a simple 'no, there is no relevant information'.

In order to get a better evidence for the potential absence of this information the chapter focuses on two much more specific sub-sections of the overall dataset: One focusing on affordability of childcare and one the (in)sufficiency of unemployment insurance. We focus on these policy features due to their relative importance as policy features as well as due to mentions of affordability in some topics or the amount of benefits received. The two sub-sections are constructed to already satisfy the three criteria that make information ideal for social investment – being specific to **an individual's experience**, specific to a **life course transition**, and would include commentary about what **factors are making a transition easier or more difficult**:

1. To assure focus on individual's experience tweets need to include the pronouns 'I' and 'we' or their possessive forms.
2. Here we make the focus on parenthood to employment transition more specific by focusing exclusively on childcare (by including one of the ECEC keywords listed in table A1 in the appendix) or unemployment support (by including one of the following keywords: 'uwv' unemployment insurance agency, 'payment', 'assistance', 'unemployment insurance', 'joblesness')
3. To assure focus on factors that make a transition easier or more difficult we focus on the affordability of childcare (by including one of the following keywords: 'cheap', 'expensive', 'costly', 'cost', 'price', 'affordable', 'unaffordable') and by focusing on the (in)sufficiency of unemployment assistance (by including one of the following keywords: 'little', 'less', 'low', 'lower', 'sufficient', 'enough').

This approach is rather crude and results in an extreme information loss, leaving a corpus sub-section of only 501 tweets for ecec affordability and 2077 tweets for unemployment insurance sufficiency. However, the purpose here is to obtain corpus sub-sections densely filled with relevant tweets to find out whether relevant tweets exist at all and what are their features. The two corpus sections are then topic modelled using LDA and after iterating through a few versions a 6-topic model and an 8-topic model are selected for the childcare affordability sub-section and the unemployment insurance sub-section respectively. Here the topics are understandably general and do not provide a useful summary, but some topics certainly capture relevant information better than others and by identifying them we can sort tweets according to the degree to which they represent these relevant topics. We inspect 50 most representative tweets for two topics for each sub-section: For ecec affordability we inspect topic 2 (consisting of tokens like 'daycare', 'cost', 'want', 'we', 'expensive', 'child benefits') and a relatively similar topic 3 (consisting of similar tokens but also including 'cheap', 'really', or 'money'). For unemployment insurance sufficiency we inspect topic 1 (consisting of tokens like '@', 'payment', 'less', 'assistance', 'enough', 'should', 'to work', 'to get') and topic 6 (consisting of tokens like 'low', 'little', 'unemployment', 'we', 'uwv' - unemployment insurance agency, 'income').

In terms of childcare affordability the majority of tweets are about things (childcare being one of them) getting more expensive and what policymakers are (not) doing about it, about the distributive aspects of policies like child benefits, or even news reports about childcare costs (including phrases like 'our correspondent' or 'we talked to parents' and thus meeting the personal criteria). The tweets generally meet the three relevance criteria but offer a personal opinion about a political issue rather than personal anecdotes or requests for assistance (such as, for example, asking if anyone in one's network knows about more affordable childcare options). Only three inspected tweets included meaningful personal commentary mentioning that an individual relies on grandparents due to costs of childcare, that childcare feels too expensive, and that one cannot make ends meet without utilizing childcare and the associated benefits. In terms of sufficiency of unemployment insurance tweets mainly include political statements about the welfare state, especially with regards to migrants receiving unemployment benefits. Much like with the affordability of childcare the tweets express a personal opinion but about broader social issues and not about an individual situation – in the inspected tweets there is no instance of personal (rather than political) statements. Iterating through different topic numbers and thus different topics does not meaningfully alter the findings for either corpus sub-section. This finding suggests that Twitter data does in fact not contain the requisite data, as the relevant content is heavily politicised – personal commentary is almost non-existent and sometimes personal commentary is also highly politicised (by @ mentioning various political parties for example).

4.5 Conclusions and discussion

The goal of this chapter is to address the lack of empirical work with regards to the potential of big data to provide novel information in policymaking. It does so by trying to answer the question "**Can social media data be used to operationalize and measure social investment?**" via answering three sub-questions that can be crudely summarized as: what is the method, what does the method imply for practice, and what are the results of this method. Answering the first research question does not necessarily constitute an academic contribution in and of itself, as topic modelling is understood relatively well as research method and it is often iterative when used in an exploratory way. The iterative process is of course always different depending on the dataset and research questions, but the articulation of the method itself, especially given the sobering answer to sub-question three, is not an important academic contribution. Why the method is articulated in such detail and is highlighted by dedicating a research sub-question to it is because the second research sub-question builds on it in a very direct way.

Answering the second research sub-question provides an overview of the complexities of the proposed method and illustrates that some of the decisions made along the way can be subjective but also consequential for the model output, bringing to the forefront some important questions about deploying this type of analysis in policymaking practice (such as the interaction between domain experts and modelling experts). This finding is also somewhat generalizable, as the plethora of analytical choices made is not dependent on the corpus used in this chapter or even Twitter as a platform – same or similar choices will have to be made in modelling discussion board posts, Facebook updates, or Instagram posts as long as the topic model remains largely unsupervised. More supervised approaches have a distinct set of advantages, but in this case unsupervised models are a good fit due to the difficulty of a-priori defining all potentially important factors as well as the practical concern that policymakers are not analytical experts and might lack the understanding of what questions a particular method can answer (unsupervised topic modelling approaches can start with rather broad question). Most importantly, this cautionary finding should not be interpreted as topic models being arbitrary statistical artefacts – they do capture underlying structures in a corpus of text, but those structures are so multifaceted that they can be constructed in multiple ways. For example, should a topic model of the corpus differentiate between a childcare topic and a self-employment topic each of which mentions multiple regions, or should it differentiate between regional topics each including both childcare and self-employment mentions? Neither is wrong or less interpretable, but selecting one representation of the corpus over another is consequential for the insights a model carries. This of course does not need to be a conscious or deliberate choice, as the more normative impacts of various analytical decisions are not immediately obvious or quantifiable.

Answering the third research sub-question provides an even more sobering conclusion. As much as the multiple topic models presented in this chapter provide a summary of the information that the corpus contains, the information considered relevant in the scope of this chapter appears to be virtually absent from the corpus. We provide an illustrative example of zooming in on two crucial and politically salient policy aspects – the affordability of childcare and sufficiency of unemployment insurance. These examples show that even though we can reasonably well filter tweets that are personal and focused on a specific aspect of policy the overwhelming majority of these tweets constitute a political commentary rather than sharing of personal experience or debating others using personal anecdotes. The hypothesis that people will share information relevant to their personal life-course transitions is not correct for the examined case. Despite the limited external validity of this finding, it does provide a relevant instance where we would reasonably expect big data to provide new and better insight, but where despite this theoretical promise almost no useful insight can be generated. Should only 'successful' (in the sense of generating requisite insight and affirming the potential of big data) research efforts be reported an elevated level of optimism about the transformative potential of big data is inevitable, making even conclusions such as this one relevant to understanding when big data is transformative and when less so.

Lastly it is important to acknowledge some of the limitations of this chapter. Firstly, there are problems with external validity when it comes to the third research sub-question: The empirical results presented in section 4.4 are specific to Twitter as platform, specific to the Netherlands, specific to parenthood to employment as a life course transitions, and specific to certain factors within that transition. These findings could be very different when using Facebook data and looking at school to work transitions for example. Limitations of scope also apply to modelling, as there are other possible topic models and approaches that are likely to illuminate different aspects of the information contained in our corpus. Even the presented method could effectively be iterated over endlessly to discover other topics carrying different (and perhaps more relevant) information.

There are also some issues that would benefit from further elaboration, such as key social investment concepts like life course transitions and policy interactions and how those could be traced in social media data. This chapter uses these concepts in a simplistic way to a) provide an intuitive rather than rigorous sense of what information is needed, and b) to expand the scope from a single policy intervention to multiple interventions and general context. Inquiries such as what users who are seeking ECEC are also saying about labour markets would illuminate the concept of policy interactions much more but are outside the scope of the chapter and the data is unlikely to yield different conclusion if queried for even more specific information.

There are other issues that this chapter does not address, which might seem counterintuitive given the strong emphasis on a realist rejoinder between techno-optimism and policy-pessimism, including ethics and public value concerns. Needless to say, there are some important ethical implications of governments collecting, analysing, and utilizing this type of personal data in designing welfare state policies. These challenges would involve balancing the potential insight with privacy protection: We know that different groups experience different life course transitions differently, meaning that the individual data should be anonymized to such a degree to allow at least a reliable categorization of population sub-groups, but the more identifiable data the more privacy threats it poses for individuals whose data are collected. Even if the analysis would preserve individual privacy, group privacy might be a serious issue, especially given the sensitive nature of some of this information. A lot of important and subjective decisions would also have to be made in aggregating the information extracted into indicators or other forms of digestible insight – even this paper has to be selective in what models it presents and how thoroughly it describes them, but it is undoubtedly still too technical and dense to provide clear recommendations and directions that policymakers might be looking for. In aggregating any of this information choices about which information is important and which information is less important are unavoidable, which is where the subjective nature of some of the 'analytical' choices really start having important normative and ethical implications. These and other ethical questions are sidestepped in this dissertation simply because the research never progressed far enough to meaningfully engage with them. With almost no relevant information found, it is not informative to assess how existing analytical decisions influenced the final finding or how the found information can potentially be aggregated.

That said, ethical considerations did play a role in designing the data collection and anonymization process. Other collection approaches are possible, such as identifying accounts of interest (eg. new parents) and collecting all of their Twitter traffic to be analysed. These approaches have been used and had some success, but they are a bit more contentious in light of data minimization requirements enshrined in GDPR and what 'scaling up' such a system would imply in terms of ethics and surveillance, especially given the fact that sampling accounts would require much more 'identification' of said accounts. Here the breadth of social investment policy analysis becomes a double edged sword, as it is true that many different policies and factors can influence individual transitions, but casting the net too wide essentially justifies collecting all (some potentially sensitive) information about parenthood and its challenges. Utilizing changes in mood and behaviour patterns to identify new mothers might be possible, but it is not difficult to see why many mothers would object to their social media data being used this way. It is possible that the ethically 'safe' and thus very 'aggregate' approach of this chapter is partly to blame for the empirical conclusion of almost no relevant information being present, but it is firmly outside the research scope of this paper to push the legal and ethical boundaries to see if better information can be found then. If anything, the empirical findings of this paper suggest that such attempts are unlikely to be worth the ethical cost.

Chapter Five: Tracing policy-relevant information in social media: The case of Twitter before and during the COVID-19 crisis[4]

Despite the sobering conclusions from chapter four with regards to the promise of novel information (at least with respect to Twitter data and social investment), the 'big data promise' is multifaceted and by no means limited to novelty of information (as argued in Chapter two). Another part of this promise, one that has seen more practical success when leveraged for policymaking, is utilizing the increased velocity of data distinguishing big data from more conventional data sources (Emmanuel & Stanier, 2016; Ward & Barker, 2013; Perspectives to Definition of Big Data: A Mapping Study and Discussion, 2016). Regardless of whether we keep using the 'big data' label, a growing part of the data used in research and policymaking is re-used data created by users interacting with various online services, such as search engines or social media, resulting in data streams that are 'always on' (Salganik, 2018). The promise associated with this increase in velocity is mainly providing existing indicators faster (Antenucci et al., 2014; Biorci et al., 2017; Di Bella et al., 2018), hopefully allowing for faster decision making (Chapter two). Even though this dissertation takes issue with the notion of faster data automatically equating to faster decisions, it is true that under certain conditions this data can speed up decision making, or provide evidence for policy decisions that have to be made in a timeframe for which conventional data is simply not available – such as when policymakers need to respond to a crisis.

Even though in such scenarios big data is essentially competing with a 'no data' scenario, there are still important challenges related to the comparatively little control researchers and policymakers have over the data generation process and to the difficulty of integrating processes to guarantee reliability, accuracy, or representativeness of data into a 'near real-time' timeframe. Nowhere is this tension between grand promises and substantial challenges more apparent than in the case of social media data: Conceptually, this data is a gold-mine of information about important events in peoples' lives and their perception of them. In the case of Twitter, the social media platform this chapter focuses on, the two primary motivations for using the platform are to connect with others and to seek or share information and advice (G. M. Chen, 2011; Johnson & Yang, 2009), which results in users posting updates about their life (Java et al., 2007) or sharing their beliefs and concerns with regards to current (crisis) events (Gilardi et al., 2020; McNeill et al., 2016; Signorini et al., 2011). Such tweets provide very detailed information at micro-level and in near real-time, lending them well to being utilized in the policymaking process, especially in situations requiring a rapid policy response.

The policy areas of employment and early childhood education and care (ECEC) policies, which chapter four focused on, are comparatively static and generally do not need to be changed rapidly

[4] This chapter is originally published as a research article: Vydra, S., & Kantorowicz, J. (2021). Tracing Policy-relevant Information in Social Media: The Case of Twitter before and during the COVID-19 Crisis. *Statistics, Politics and Policy*, 12(1), 87-127. More details are provided in the 'Authorship Contribution' section.

based on current events. This makes these policies interesting for studying aspects of big data such as provision of novel information to improve understanding (much like chapter four does), but not for real-time insight and crisis management. Or at least that was the situation before the COVID-19 pandemic, which greatly disrupted both the employment and ECEC situation of many. Given the data itself and the data collection infrastructure developed for this dissertation as part of chapter four, this dissertation was in the unique position to evaluate the change in information contained in social media as a result of the COVID-19 outbreak. To study this, this chapter needs to first address a familiar research gap - the lack of knowledge with regards to what relevant information for social and economic policymaking do real-time social media data streams contain. Even though this research gap was addressed in chapter four and the data source and pre-processing remain the same, the inclusion of data covering the first wave of COVID-19 pandemic makes it more likely that the policies of interest will be discussed since they became a more prominent part of the public discourse and more people struggled with employment as well as ECEC provisions during this time period. This means that as much as the research gap is the same, the updated data creates a scenario in which it is (even more) likely that social media will contain policy relevant information.

The second research gap (enabled by filling the first one) this chapter explores is whether the relevant information in real-time social media data changes in periods of abrupt social, economic, and policy change. This is a research gap the literature pays little attention to despite the fact that real-time social media data (and Twitter specifically) have been used to map and understand public reaction to crises ranging from short-term crises like natural disasters (Acar & Muraki, 2011; Terpstra et al., 2012) or shootings (Heverin & Zach, 2010) to longer term crises like the refugee crisis (Gualda & Rebollo, 2016; Öztürk & Ayvaz, 2018). These cases include the 2009 H1N1 outbreak (Ahmed et al., 2019; Chew & Eysenbach, 2010; McNeill et al., 2016; Signorini et al., 2011; Szomszor et al., 2011) as well as the COVID-19 outbreak (Gilardi et al., 2020). We thus know that people use social media like Twitter to share information about the crisis and their personal experience with it, but we lack the understanding of whether there is a meaningful change in what they say about social and economic policies that are not established as a result of the crisis, but that are affected by it. Intuitively we can hypothesize some degree of change, but the empirical work testing such a hypothesis is currently not available – existing research focuses on direct response and reaction to a crisis event, making any comparison with the pre-crisis period trivial and not informative.

Understanding this change (or lack thereof) is important for our understanding of the value of big data in policymaking: The key advantage of real-time data streams is the ability to support a faster policy response, but there is little incentive to make fast decisions and abruptly change a policy suite in a time period of normalcy and stability. This can result in very little incentive to trade-off the generally higher accuracy and reliability of data sources like household samples surveys for the higher velocity of real-time data in dealing with some policy puzzles (Chapter two)(Vydra & Klievink, 2019). In other words, the potential benefits of using real-time data to support policymaking are greatest in a time of crisis, such as the 2020 COVID-19 outbreak, when policy needs to be changed in a timeframe in which traditional data will simply not be available. Yet we have little understanding of how the content of real-time social media data changes in such situations with respect to existing policies. It could be that affected policies get debated even more (since they are more salient) or that the crisis itself creates new grievances that people comment on. It could also be that the content is too narrowly focused on the crisis event itself or that generic or politicised ‘noise’ drowns out the ‘signal’ of meaningful commentary. Having an empirical understanding of the change that happens can further inform our understanding of the transformative potential of real-time data for social and economic policymaking.

This chapter attempts to (partially) fill both of these research gaps by empirically studying tweets (as an instance of real-time social media data) focused on (un)employment policies and early childhood education and care policies – two well-established policy areas heavily affected by the COVID-19 outbreak. We utilize data for two time periods – a four-month period of 'crisis' following the 2020 COVID-19 pandemic, and the same time period from 2019 as a period of 'normalcy'. This allows us to tackle two research questions: Firstly, the more exploratory question **'What policy-relevant information does Twitter contain?'** This broad question necessarily touches on whether the information exists, what insight it carries, how much of it there is, and how well are we able to extract it. Secondly, the more descriptive question **'How does this information change between a period of normalcy and a period of crisis?'** By 'policy-relevance' we refer to information where individuals express their opinion on the (in)sufficiency of relevant policies, specific aspects of those policies, or the situations those policies aim to address (e.g. joblessness). Our approach is distinct from other approaches utilizing social media data to measure policy-relevant indicators (Proserpio et al., 2016) or issues associated with the COVID-19 pandemic (Gilardi et al., 2020) in having an a-priori determined (and rather broad) focus and not relying on sampling individual users, as mentioned in concluding chapter four. Our approach here closely approximates that of chapter four and can of course result in somewhat general information, but this information is not meant to replace a thorough policy evaluation in practice and in terms of testing the theoretical promise of real-time data for policymaking we have to avoid an overly restrictive focus to maintain any ability to generalize.

We answer the two research questions for the case of the Netherlands during the 2020 COVID-19 outbreak, focusing on two policy areas to improve the external validity of our findings. We select these policy areas due to their general political salience, but also due to how severely they were impacted by the lockdown policies responding to the COVID-19 pandemic. We describe our case, policy, platform, timeframe selection, and conception of 'policy-relevance' in detail in **section 5.1**. We then analyse this data using a novel methodological approach that combines topic modelling and latent semantic scaling – an approach that we propose and justify as fitting for this particular chapter in **section 5.2**. In **section 5.3** the results are presented, and we discuss the policy-relevance of our findings. **Section 5.4** mentions limitations of this chapter and **section 5.5** concludes by answering the two research questions and commenting on our approach as a whole.

5.1 Case and data selection

To meaningfully test for presence of policy-relevant information, we constrain our dataset to a single country and to two relevant policy areas. This section justifies our country selection – the Netherlands – based on high take-up rates and salience of the two selected policy areas (section 5.1.1) as well as social media utilization. It further substantiates the policy selection of early childhood education and care (ECEC) policies and (un)employment policies based on their political salience and the degree to which they are affected by the COVID-19 crisis (section 5.1.2). It then argues (section 5.1.3) for utilizing Twitter as a social media platform due to its fit with the two selected policy areas as well as pragmatic reasons and for the two four-months data collection periods (section 5.1.4). Most importantly, we define what constitutes a 'policy-relevant' piece of information (section 5.1.5), adopting a crowdsourcing approach to creation of socioeconomic indicators.

5.1.1 Country selection

This chapter focuses on the Netherland for two primary reasons: Firstly, from a practical perspective the most important factors are a very high internet penetration which the Netherlands has, together with high rate of social media use (Internet World Stats, 2017), which is necessary to gather a sufficient amount of data representing a broad section of the population. Secondly, both of the policy areas selected for this chapter (unemployment and childcare policy) are well developed in the Netherlands: The Netherlands has a relatively low percentage of children under the age of 3 who receive no formal childcare (35.2% for 2019) (Eurostat, 2020). In 2018 the Netherlands was the 3rd lowest in Europe in this statistic followed only by Luxembourg and Denmark (Eurostat, 2020). Multiple childcare options are also comprehensively supported by the government. In terms of unemployment policies, the Dutch expenditure on out-of-work income maintenance and support is the third highest in Europe (1.38% of GDP in 2018) (European Commission, 2020), but the contributory conditions are some of the strictest in Europe (Matsaganis et al., 2014). Data on unemployment assistance take-up rates is not available for the Netherlands.

5.1.2 Policy selection

There are two groups of policies that this chapter focuses on – labour market (LM) policies related to (un)employment as well as early childhood education and care (ECEC) policies. Tackling two policy areas rather than one is simply to improve the external validity, as it is to be expected that the public assessment and overall conversation will be different for different policies. There are two main reasons for selecting these policies: Firstly, it is because of the general political salience of these two policy areas. On a European level, employment is an explicit goal of the Europe 2020 agenda, and childcare contributes to both the employment (of the parents) as well as the educational goals. This translates well to Dutch policy priorities – as mentioned, the Dutch utilize childcare a lot and unemployment policies are sufficiently generous. Secondly, and more importantly, it is because of how heavily impacted these policy areas are by the COVID-19 outbreak. The amount of global working hours has decreased by approx. 17.3% in second quarter of 2020 (compared to Q4 2019) with women affected more severely, in part due to the increased burden of unpaid labour (ILO, 2020) such as caring for children. In the Netherlands, unemployment has increased from 2.9% in March 2020 to 4.6% in August 2020 (CBS, 2020), which is despite the government's intervention providing companies with financial support to pay their employees and providing financial support, credit, and relaxed taxes to the self-employed. In terms of childcare (and schools), the access was restricted for everyone except children of crucial workers in late March, and the restriction was lifted on May 11th 2020. No official statistics are available on how this impacted the number of children formally enrolled in childcare during the restrictions or immediately after. Both labour markets and childcare options have been heavily affected globally, and the Netherlands is no exception to this. Such a period of abrupt change to the labour markets and childcare is a well-fitting instance of a 'crisis' situation, where substantial policy intervention is required, but where traditional data is not available in time and will not provide sufficient detail.

5.1.3 Platform selection

For data collection Twitter is the platform of choice for several reasons. Firstly, the demographics of Twitter are a good fit with the policies we focus on. In the Netherlands the best represented demographic group on Twitter are those between 20 and 39 years of age (Statista, 2018), which is the prime age for activity in the labour market as well as starting a family and child rearing. Furthermore, it is also reasonable to expect that these two policy areas will in some way be debated on social media. In terms of childcare, new mothers often seek support from their networks on social media (McDaniel et al., 2012) and sharing ones experiences or looking for advice about childcare options is also something we hypothesize to see in social media data. With regards to employment, one's personal network is an important tool that can be leveraged using social media platforms. In other words, we select policy areas that we expect to be discussed on social media more than other policies, and we select policy areas likely to be of interest to the demographic group most heavily represented on social media (in this case on Twitter). Despite other platforms such as Facebook having a substantially higher user-base in the Netherlands, Twitter maintains 2.5 million users in the Netherlands for 2019 (Statista, 2019) and a 2018 poll reports 26% of 20 to 39 year olds in the Netherlands using Twitter (Statista, 2018), which is good comparatively to other social media platforms. Furthermore, as much as the debate on Twitter can be influenced by 'opinion leaders', opinion leaders on Twitter do not necessarily consume a lot of traditional media (Park, 2013a) and do not share the same socio-economic characteristics of 'offline' opinion leaders (Park & Kaye, 2017). This suggests that the discussion on Twitter is unlikely to just be a reflection of the narratives found in traditional media and thus it should be capable of providing additional insight, even if some of that debate is driven by opinion leaders.

Choosing Twitter is also a pragmatic choice due to the accessibility of research data compared to other social media platforms where this access is not provided and automatically scraping the platform would violate the terms of service as well as users' expectations with regards to privacy. Secondly, this data is by default available in a real-time data stream, which is not the case for other social media data that would either have to be retroactively scraped or retroactively searched for. This allows us to explore the real-time aspect of social media data more than other platforms would allow for. Even though this chapter does not process the data itself in real-time, they are gathered from a real-time data stream and the methods adopted here are generally applicable for deploying the proposed method in a real-time context.

5.1.4 Timeframe selection

This research gathers Twitter data for two time periods – from the 11th of May until the 11th of September 2019 and 2020. This selection is of course limited by practical constraints, but the 11th of May starting point is selected to coincide with the re-opening of schools and day-care centres (to non-essential workers) during the COVID-19 outbreak. With the re-opening, these facilities can be used again but people will likely have concerns that are not captured in any conventional data source. This re-opening also means that a lot of parents could focus on their labour market situation with the children being back to school/day-care. Understanding these developments in near real-time could be crucial for designating appropriate policy action. The end point for our data collection assures we capture the 'first' wave of the COVID crisis and the start of a new academic year. With regards to unemployment assistance the start and end date are not as consequential, but the time period in 2020 covers the period of abrupt unemployment increases as well as the availability of special governmental assistance. The same four-month period from 2019 is used as a period of policy 'normalcy', minimizing the effect of seasonality.

5.1.5 Defining 'policy-relevance'

The most consequential definition adopted by this chapter is what it means for information on social media platforms to be 'policy-relevant'. This is due to the fact that any such definition carries with it assumptions about the functioning of policymaking as well as about the specific data needs of some policymaking decisions.

Starting at the broadest possible level of the general role of social media data in governance: The extant literature outlines a range of potential uses like electronic participation, engagement, transparency, communication, trust, collaboration, democracy, crowdsourcing, security, and open data practices (Dwivedi et al., 2017). In practice, the ways governments utilize social-media are also quite varied, ranging from utilization in elections, information dissemination, making processes transparent, but also sourcing information and feedback from users to be utilized in decision-making (Grubmüller et al., 2013). In this chapter we focus exclusively on using social media as a 'information and feedback source' (Grubmüller et al., 2013) for decision-making, which is also known as crowdsourcing. Crowdsourcing involves problem-solving, idea-generation, and production tasks that use IT to leverage the dispersed knowledge of a 'crowd' of individuals (Prpić et al., 2015). This knowledge is then utilized as evidence in evidence-based policymaking.

With our focus on crowdsourcing set, the next question is how should the 'evidence' we are seeking to find be utilized, as there is no one-size-fits-all answer with regards to how evidence from social media enters existing policymaking practice (Höchtl et al., 2016; Janssen & Helbig, 2018; Mergel & Bretschneider, 2013). For the purposes of this chapter we assume that the information we aim to find is relevant mainly for the latest and the earliest stages of the policy cycle: Either the information can be used as near-real time monitoring tool to evaluate the perceptions of and actual problems with a specific policy (Singh et al., 2020), or it can be leveraged in an exploratory way during agenda setting and problem definition (Panagiotopoulos et al., 2017). This makes our focus somewhat broad as we are not only looking for very policy-specific commentary in which people voice concerns about specific policy (or its aspects), but also for more general commentary where people identify problems with the situation those policies aim to address (childcare situation and employment situation). We do this for two reasons: Firstly, to avoid excluding information where users have genuine and relevant complains but simply do not link them to a policy explicitly. Secondly, to allow policymakers and researchers a degree of freedom in determining what policy can solve a given issue – even if people link a problem they are experiencing to a policy they might not do so 'correctly', especially given that many problems can be solved by different policies.

Given our policy selection and focus on crowdsourcing information to be used as a monitoring and/or agenda setting tool, our empirical contribution is most closely aligned with literature on real-time indicators for economic policymaking. In this area, indicators based on real-time social media data have some huge advantages over traditional household and business survey approaches (Antenucci et al., 2014): First, (labour market) indicators created in real-time allow for a more rapid diagnosis of an issue and a timely policy response, especially when it is most appropriate such as in times of crisis. Second, they can enable a more targeted policy response by, for instance, identifying socio-economic or geographical groups most hit by a crisis, and by pointing to particular aspects of policy which demand adaptations. Third, this type of data comes at relatively low cost as this information exists regardless of its potential utility for the policy insight. The costs stem from creating an appropriate system of retrieving relevant information and maintaining it later on; thus at arguably lower cost than running consumer surveys. These benefits have already motivated efforts to create real-time social media indicators with varied degrees of success (Antenucci et al., 2014; Biorci et al., 2017; Proserpio et al., 2016; United Nations, 2011). In terms of adding additional 'depth' to economic indicators, as already described in chapters three and four, these efforts range from simple replication of the unemployment indicator (Antenucci et al., 2014) to understanding psychological impacts of unemployment (Proserpio et al., 2016) or coping mechanism (United Nations, 2011). There is currently no research (that we would be aware of) testing the merit of such indicators in periods of crisis, but in terms of other real-time data sources, there is some promise for tracking economic activity in response to the COVID-19 crisis. A good example is research by Chetty et al. (2020) which was able to quickly identify relative ineffectiveness of state-ordered reopening to stimulate economic activity in the aftermath of the lockdown and argued that the only effective approach to mitigating the short-term economic hardship is through providing benefits to those who have lost their incomes (Chetty et al., 2020). It thus offered a clear recommendation for policy adaptation. While this research was based on economic transaction data provided by private entities, we test in this chapter Twitter's utility in providing relevant real-time information which could trigger policy adaptation, especially in periods of crisis.

Conceptually restricting 'policy relevance' to **crowdsourcing** of **real-time socioeconomic indicators** to inform the **monitoring and/or agenda setting stages** of policymaking is necessary to firmly situate this chapter in the existing literature, but it does not solve the practical issue of differentiating between 'relevant' and 'irrelevant' tweets. To do that, we adopt a simple and intuitive definition: 'Policy relevant' tweets comment on the adequacy or inadequacy of specific policies, specific aspects of those policies, or the situations relevant policies aim to address. In the analysis we operationalize 'adequacy/inadequacy' in multiple ways, but in general we are interested in people identifying issues that make them (un)happy with a policy and that render the policy somehow (in)effective or (un)desirable. This operationalization is very similar to that of chapter four, with the criteria for this information being specific to an individual relaxed (here we accept more general commentary as relevant as well). What we mean by 'aspects of policy' here are different features of a policy provision as perceived by individuals; For unemployment policy these could include the generosity, length, or eligibility criteria or unemployment benefits (Gallego & Marx, 2017). For childcare these could include affordability, capacity, or quality of care (Carta & Rizzica, 2016; Grammatikopoulos et al., 2014; Kawabata, 2012). Trying to differentiate between these individual aspects in both policy areas is important because negative comments on one such aspect of a policy convey fundamentally different feedback than those on another.

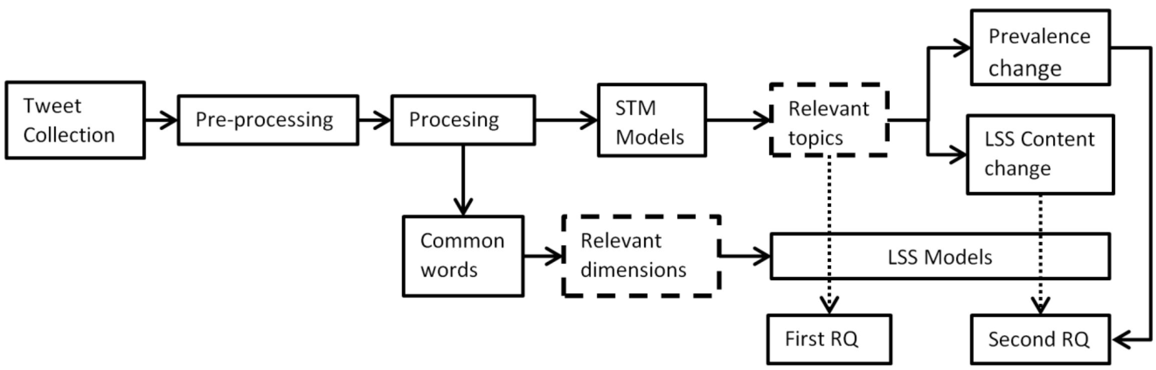
Despite defining policy relevance and the practical rules making it applicable to individual tweets, our decisions about which aspects of policies are relevant are ultimately subjective and in practice would be made by policymakers. In the absence of a real-world policy dilemma and extensive cooperation with policymakers responsible for resolving it, we can only approximate 'policy relevance' by assuming that factors that are either shown (in the literature) to influence the effectiveness of a policy, or that we hypothesize to influence the effectiveness of a policy are 'policy relevant'. For the purposes of this chapter this is not a limitation per se, but we do address this issue in articulating our methodology (section 5.2) by introducing specific steps where policy relevance is decided without necessitating a technical understanding of other stages of the method, making it relatively easy to include policymakers at appropriate points in the analysis.

5.2 Methods

To answer the research questions of this chapter we propose an approach combining topic modelling and latent semantic scaling (LSS), both of which are machine learning methods developed for automated text analysis. In this approach we first establish (using topic modelling) the substantive focus of individual tweets allowing us to extract clusters of tweets (topics) focusing on relevant policies or situations those policies aim to address. However, such a summary does not always tell us which policy aspects are being talked about or what exactly is being said about them as the topics can focus on multiple aspects of a policy or convey multiple opinions about a policy. Topic modelling alone does not provide this insight, which is why LSS is utilized to further summarize the content in these topics along policy-relevant ‘dimensions’ that we specify. These dimensions can be rather general such as positive sentiment versus negative sentiment (the main use of LSS), but also more tailored to policy such as succeeding versus failing. The two methods are used in sequence with topic modelling first identifying a group of tweets relevant to a topic of interest and LSS models then estimating whether these tweets tend to be negative or positive and concerned with success or concerned with failure (as those are the two LSS dimensions we construct). The LSS scores for groups of tweets that we report in the results section and in appendix C are simply the mean polarity score of the selected group of tweets, which also allows us to run a statistical test for the difference in mean/median between tweets posted in a period of normalcy and tweets posted in a period of crisis.

The general methodological approach and how it answers the research (sub)questions is illustrated in figure 5.1. This figure outlines what we consider major steps in the analysis process, with dashed borders denoting the two steps where the meaning of ‘policy-relevance’ is primarily established. In practice many analytical decisions made along the way can carry subjective assumptions (as argued in chapter four), but the two highlighted steps are almost purely subjective and normative and when applied to real-world policymaking would necessitate the involvement of policymakers.

Figure 5.1: Method diagram



This approach is able to answer the first (more exploratory) research question by identifying relevant topics in the corpus and then testing whether tweets representative of those topics have a polarity score significantly different from zero on LSS dimensions. It then builds on this finding in order to answer the second (more descriptive) research question by utilizing two analytical outputs: Firstly, whether there is a significant difference in the prevalence of policy relevant topics between a time period of abrupt change and a time period of normalcy. Secondly, whether there is a significant change in LSS polarities between these two time periods for a given topic. This allows us to answer both research questions of the chapter not just at a level of detail unparalleled by other methods but also by using statistical tests of difference, providing a more definitive conclusion than one based on our qualitative interpretation of topics and their change.

Outside of fitting the research focus of this chapter, this approach has multiple important and more general advantages: Firstly, it can leverage the real-time nature of social media data by providing outputs in near-real-time. There are steps of the process that cannot happen in near-real-time, such as training the full models or the two more normative steps that determine 'policy-relevance' (highlighted by dashed borders in figure 5.1), but once those steps are taken, the models we utilize can take new tweets and predict their topic memberships and their polarity on relevant dimensions in near-real time. These underlying models will need to be iteratively updated to avoid them becoming out-dated, but applying them to a stream of tweets can leverage its real-time nature. Secondly, the two instances where the models have to be 'supervised' are actually an advantage: As mentioned earlier and argued in Chapter four, decisions on which information is 'policy-relevant' are inherently subjective and to be determined by policymakers rather than technical experts. The fact that these decisions are transparent and can be made in a very intuitive way without requiring technical expertise allows policymakers to be involved, make informed decisions, and be held accountable. Thirdly, both topic modelling and LSS are entirely language-independent, making the entire approach language-independent as well. Processing the Tweets themselves can be language-dependent, in our case it is, but there are also language-independent options and any potential language-dependence is easier to overcome because it is limited purely to the 'Processing' step. Fourthly, this approach is not specific to Twitter as a platform and can be easily adapted to other user-generated text data such as other social media or comments on governmental platforms. Fifthly, it allows for good internal validation by validating the STM models themselves, the LSS models themselves, as well as the final summary by inspecting tweets and assessing whether a topic theta (belonging to a topic) and LSS polarity estimate (positive/negative or success/failure) are assigned in a way that is interpretable and agreeable to human annotators. We cover the validation of STM and LSS models themselves in this section and comment on the validity of the final output in a topic-by-topic fashion in appendix C.

The rest of this section describes the methodological steps in more technical detail in three sub-sections focusing on the collecting, pre-processing, and processing of the data (sub-section 5.2.1), on the topic models used and their parameterization (sub-section 5.2.2), and on LSS and creation of relevant 'dimensions' (sub-section 5.2.3).

5.2.1 Collection, processing, and pre-processing

The data was collected via Twitter's Stream API by gathering tweets that are a) in the Dutch language (as identified by Twitter) and b) containing some of the keywords from a list of keywords aimed to capture relevant labour market and ECEC policies, as well as situations those policies aim to address (list of keywords available in appendix A). This entire procedure is identical to that of chapter four. Pre-processing includes removing re-tweets and duplicate tweets and joining quoted tweets with the text of the tweet quoting them (including '|' between the two texts). Bot removal is very rudimentary and consists of removing tweets authored by accounts tweeting more frequently than 1500 times a month and/or authoring more than 450 tweets a month that get captured in our dataset. This issue is likely better tackled as a supervised learning problem (Andriotis & Takasu, 2019; Inuwa-Dutse et al., 2018; Kantepe & Ganiz, 2017; Lee et al., 2011), but such an approach necessitates ground-truths about which accounts are bot accounts and would sacrifice the language-independence and easy applicability of the general method we are proposing. Given that we focus our interpretation on specific topics, the existence of other topics that include some content from bot accounts is not a major drawback (any remaining irrelevant content authored by bots gets removed by omission at the stage of selecting relevant topics).

The entire resulting dataset (approx. 740 000 tweets) is used in training LSS models, but for training the topic models we create two corpus sub-sections, one concerned with ECEC (approx. 40 000 tweets) and one with labour markets (approx. 403 000 tweets). We do that utilizing groups of the same keywords used to collect the data. Tweets that are not included in either corpus are excluded because not all keywords used for collection are used to create corpus sub-sections (see appendix A for details on keywords). This is the most importance difference between the collection procedure of chapter four and chapter five. Text for all tweets is then tokenized and lemmatized – splitting it into individual tokens that get reduced to the 'lemma' of a word, which is its basic dictionary form. This resolves the issue of words that have inflectional and derivationally related forms appearing as different tokens despite being (effectively) the same word. To do this step we utilize the 'Frog' natural language processing suite (Van Den Bosch et al., 2007).

5.2.2 Topic models and parameterization

Topic models are a class of algorithms that aim to identify existing 'themes' of a corpus of documents and can organize the documents according to those themes at a scale where human annotation would be near-impossible (Blei, 2012). These models operate on the assumption that documents can be summarized as collections of various 'topics' and that those topics can be summarized as collections of various tokens. These assumptions are most famously formalized into a generative probabilistic model called Latent Dirichlet Allocation (LDA) (Blei et al., 2003), which has since become widely adopted and utilized for its simplicity and good interpretability. Since then topic models have expanded to take into account correlation between topics (Blei & Lafferty, 2007), the evolution of topics across time (Blei & Lafferty, 2006), the evolution of topics due to influential documents (Gerrish & Blei, 2010), or to be able to train in real-time from a data stream (Wang et al., 2012). The model utilized in this research is the Structural Topic Model (STM) (Roberts, Stewart, & Airoldi, 2016) and its R implementation (Roberts et al., 2019). This model includes the information about topic correlations but more importantly assumes that topic prevalence and/or topic content can be influence by a generalized linear model of document-level covariates.

The ability to include document-level covariates into the analysis makes STM the best fit for this chapter for two reasons: Firstly, Twitter is a platform that different users use in different ways and we would expect the accounts of political parties, companies, or individuals to address different topics. STM allows us to include document-level covariates describing the type of account posting a Tweet (such as follower count or tweeting frequency) as factors that influence either the prevalence or the content of topics (or both). This makes the model more conceptually fitting than models that would force us to 'assume away' the difference between users. Secondly, it allows us to answer the research sub-section about how relevant information changes in times of abrupt social, economic, and policy change by simply including a dummy covariate capturing whether a tweet was posted in a period of abrupt change or a period of normalcy. With regards to assuming change in topical prevalence and/or topical content, our baseline is to assume change only in topical prevalence as that makes topics easier to understand and provides us with more solid footing to run a statistical test of difference on mean or median polarity (if we assume a change in the topic itself between those two periods we risk comparing 'apples to oranges'). We thus assume a simple generalized linear model for prevalence change of topics: Prevalence =tweeting frequency + number of followers + crisis or normalcy. To test this difference in polarity we adopt a two-sample t-test and a Mann-Whitney U test: A two-sample t-test is generally sufficient due to the size of the data, but the polarity distribution is often bi-modal (due to the fact that we are looking for non-zero polarity), data can be small in size, and samples can be imbalanced. A potential divergence between the t-test and Mann-Whitney prompts us to investigate a test closer. We report the results of both tests in the results section.

With any topic model there is the challenge of determining an appropriate number of topics to model for. There is not a 'wrong' topic number in topic modelling (Grimmer & Stewart, 2013; Roberts et al., 2019) as the quality of topic models generally depends on their interpretability (to humans) and the insight they deliver. The quality of the insight depends entirely on why the model is used and interpretability is highly subjective and dependant on domain-specific knowledge about a corpus. That said, there are multiple metrics assessing the quality of models that can aid with the selection of a topic number. These metrics are presented for both the ECEC corpus sub-section and the LM corpus sub-section in figures 5.2 and 5.3 respectively and there are four of them:

1. Semantic coherence: This is a per-topic metric and is maximized when the top words of a topic tend to co-occur in the corpus. This metric is important because it correlates with human judgement of topic interpretability (Mimno et al., 2011). However, Roberts et al. (2014) show that this metric can be maximized by having topics dominated by relatively common words resulting in semantic coherence generally declining as number of topics increases making coherence insufficient on its own.

2. Exclusivity: This is a per-topic metric that remedies the problem of coherence by providing a measure of how exclusive assigned words are to a given topic, scoring models where general words are shared across many topics poorly. This metric should be read together with semantic coherence to select topic numbers where both metrics are relatively high (but trade-offs are unavoidable).

3. Heldout likelihood: This is a metric to be maximized and is calculated by removing a proportion of words (in our case 40%) from a proportion of documents (in our case 10%) before training the model and then measuring how well the model predicts the missing words.

4. Dispersion of residuals: Metric utilized by Taddy (2012) computes the dispersion of residuals and should be minimized as over-dispersed residuals can mean that the variance has not been sufficiently accounted for.

Figure 5.2: Model metrics for ECEC sub-section

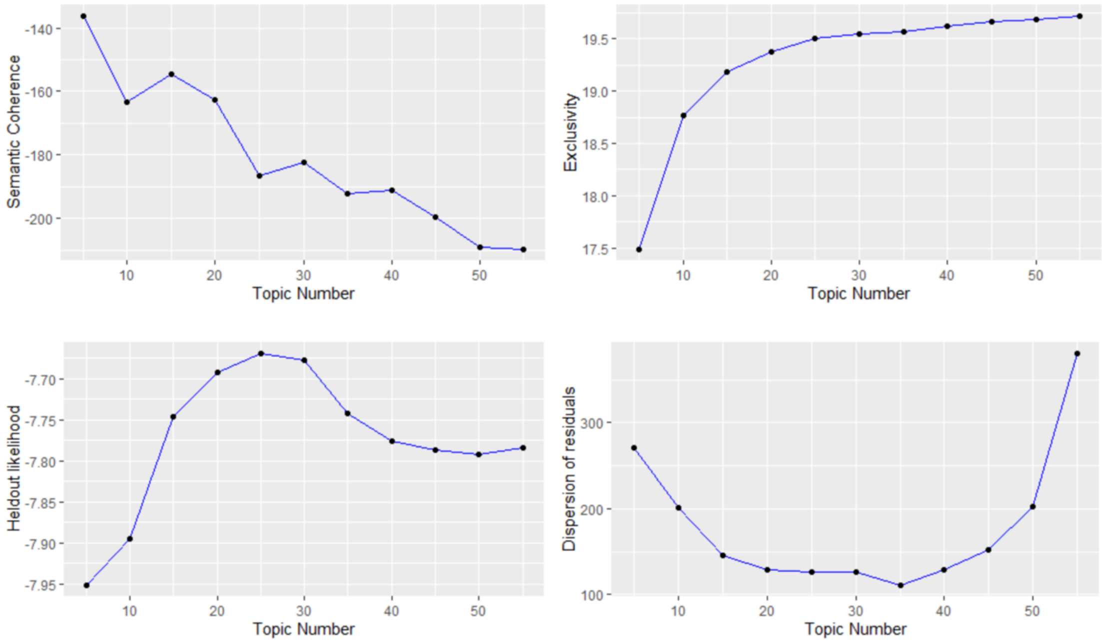
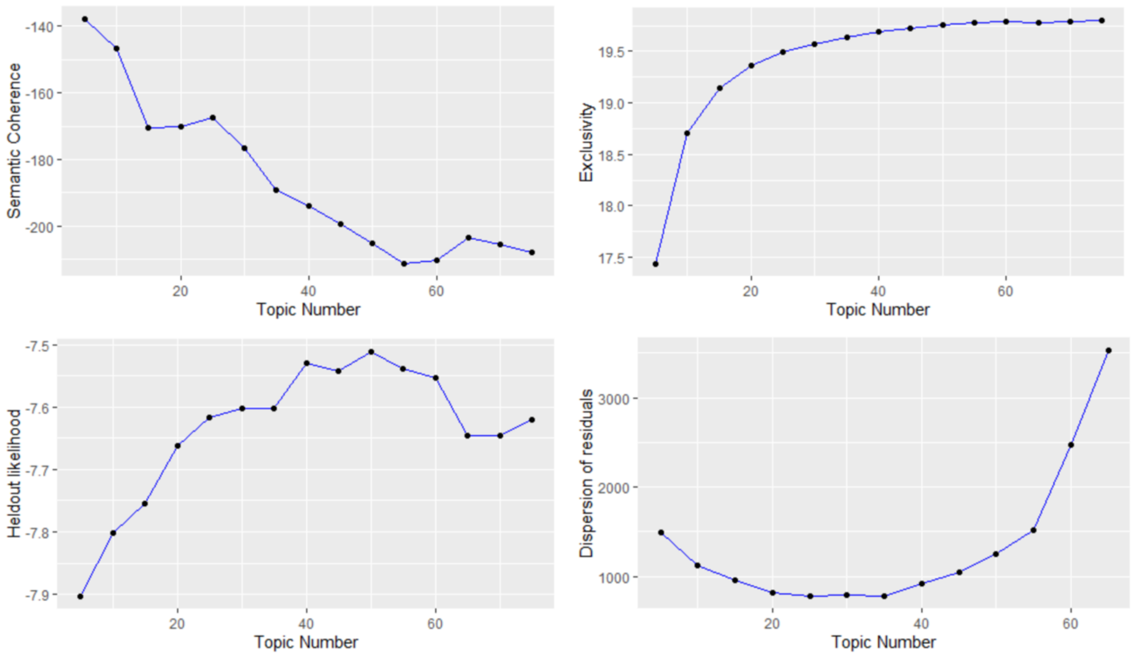


Figure 5.3: Model metrics for LM sub-section

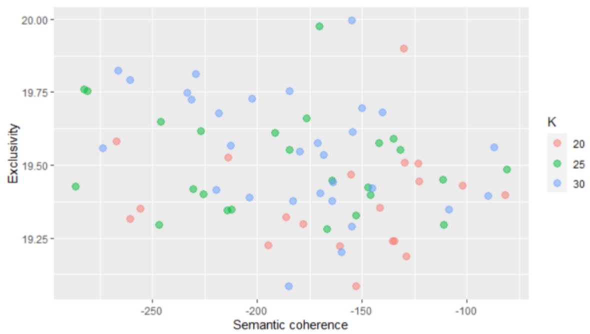


Note: For the dispersion of residuals models with 70 and 75 topics are excluded as they are outliers and including them stretches the scale thus obscuring the trend.

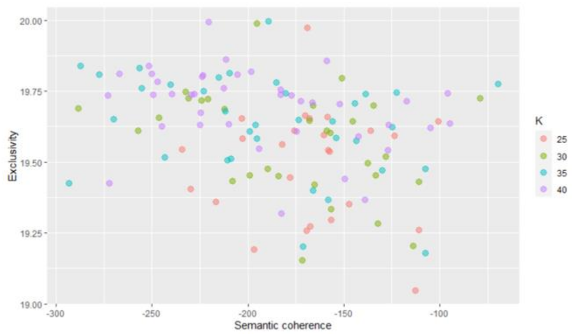
We use these four metrics to determine a reasonable range of topic numbers, but the final selection is made based on manual inspection of models in that range. This is because exclusivity and semantic coherence are computed per-topic and plotting the averages for a model discards information about variance in this metric. Figure 5.4 illustrates this on the candidate models: 20-30 topics for the ECEC sub-section and 25-40 topics for the LM sub-section. From figure 5.4 we observe that each model clearly contains topics that are relatively coherent and exclusive, meaning that any of these models can provide good insight given that we will be focusing only on 'relevant' topics and discounting the rest. We inspect models by first looking at 15 'top words' for each topic in each candidate model and for seemingly interpretable topics we inspect 30 tweets – 6 groups of five tweets along a range of theta values (maximal, 0.85, 0.8, 0.75, 0.7, 0.65), which represents documents that are 'representative' of a topic to a given degree. We do this to get a deeper insight into the topic (avoiding issues such as top tweets all belonging to the same conversation), but also to provide a layer of internal validation for our topic models: This inspection allows us to select the most 'valid' model and to also identify which topics validate and which do not, which is simply judging of whether the model assigns tweets to 'correct' topics with a 'correct' theta as assessed by human observers (in this case the authors). Using this method of manual inspection, we select 20 topics for the ECEC sub-section and 30 topics for the LM sub-section and identify topics that are potentially policy-relevant and clearly validate (in the sense of being interpretable by humans).

Figure 5.4: Exclusivity and Semantic coherence across multiple topic models.

5.4a: Candidate models for ECEC



5.4b: Candidate models for LM



5.2.3 Document scaling using LSS

The other method we utilize is Latent Semantic Scaling (LSS), which is a semi-supervised technique for document scaling that has been shown to perform comparably to lexicon-based approaches for sentiment analysis (Watanabe, n.d.) and has been used as a language-independent sentiment analysis method (Watanabe, 2017a, 2017b). It relies on a word embedding approach that expresses individual tokens as vectors that can then be compared to assess the semantic similarity between them. LSS relies on singular value decomposition of a document-term matrix representation of a corpus where each document is a singular sentence (Watanabe, n.d.). Sentences are the context in which words are considered to co-occur for the purposes of training a word embedding model (which is appropriate for our corpus of rather short texts). Researchers are then required to provide seed words that represent two extremes of a dimension of interest and LSS infers the polarity of other word vectors based on their proximity to the two defined extremes. Conceptually these dimensions do not have to be limited to sentiment and can be constructed in various ways – the limits of what such dimensions can meaningfully capture are to still be tested by the literature, including by this chapter.

LSS as a method is appropriate for this chapter for a multiplicity of reasons. The Dutch language does not have as many options for sentiment analysis as English, and existing lexicon-based approaches do not perform well on our corpus (language on Twitter tends to be highly informal and as a result majority of tweets cannot be annotated reliably). Secondly, the ability to control exactly what a dimension captures is crucial for this research, as sentiment alone, despite carrying valuable information, is not perfect. Identifying that tweets about a specific policy tend to be negative could be users expressing displeasure with the policy, but it could also be users condemning the problem that the policy is addressing (and voicing support for intervention). We try to partially resolve this problem by creating a more policy-specific dimension of success (at one extreme) and failure (on the other extreme). This dimension bears some similarity to the sentiment dimension but also limits the issue of negative sentiment being associated with a situation rather than intervetion mentioned above. Many other dimensions of interest can be constructed, but the selection of those is ad hoc without the involvement of policymakers who would articulate a particular interest.

However, the ability to define the relevant dimensions comes with a drawback – these dimensions need to be validated and interpretable (beyond the provided seed words). To train LSS models for individual dimensions we utilize the full corpus of tweets and remove all '@' mentions to remove unwanted bias (without doing so, certain political parties and public figures would be strongly associated with certain extremes, constituting an undue bias). We include a form of internal validation into constructing these dimensions. We base our keyword selection on a manual inspection of top 500 most commonly occurring adjectives, adverbs, verbs, and nouns (assuming that conjunctions, prepositions, etc. do not convey strong polarity). We then pick the most relevant words from that list that convey polarity with respect to a given dimension. We provide these tables for the dimension positive sentiment versus negative sentiment in table 5.1a and or the dimension success versus failure in table 5.1b. For each dimension we seed only a part of the selected words into the model and the rest are later displayed along the given dimension to validate whether they are placed 'correctly' as judged by human annotators (in this case the authors). The resulting plots are displayed in figures 5.5 (for positive vs. negative sentiment) and 5.6 (success vs. failure) – the words that were not seeded into the models are those highlighted in blue in table 5.1. Rather than randomizing which words are seeded and which are heldout for validation we select them manually to include word pairs that are clearly antonyms ('good' & 'bad' or 'positive' & 'negative') and words that are not ambiguous in terms of polarity. This is especially important for Dutch as many generally positive words can simply be preceded by 'niet' (translates as 'not') to capture the opposite, making those words very ambiguous as they can be used in both negative and positive contexts. In general the sentiment dimension validates well as the heldout words are placed towards the 'correct' polarity to an understandable degree (figures 5.5 and 5.6).

Table 5.1: Selected words for LSS Dimensions

5.1a: Positive vs. negative sentiment

Positive sentiment			Negative sentiment		
count	token	translation	count	token	translation
127810	goed	good	9262	slecht	bad
2834	positief	positive	1290	negatief	negative
5490	Blij	happy	1375	boos	angry
4913	gelukkig	happy	968	zielig	pathetic
930	tevreden	satisfied	472	kwaad	pissed off
3203	prachtig	magnificent	616	idioot	idiot
21721	mooi	beautiful	1039	belachelijk	ridiculous
1076	lief	sweet	2972	dom	stupid

5.1b: Success vs. failure

Success			Failure		
count	token	translation	count	token	translation
1917	succesvol	successful	634	onnodig	redundant
1766	handig	useful	511	mislukken	fail
2409	lukken	succeed	804	falen	to fail
1495	behalen	achieve	2660	verliezen	to lose
1866	bereiken	to achieve	473	nutteloos	useless
1159	afmaken	finish up	1930	tekort	shortage
1100	realiseren	realize			

Figure 5.5: Words on a positive vs. negative sentiment dimension

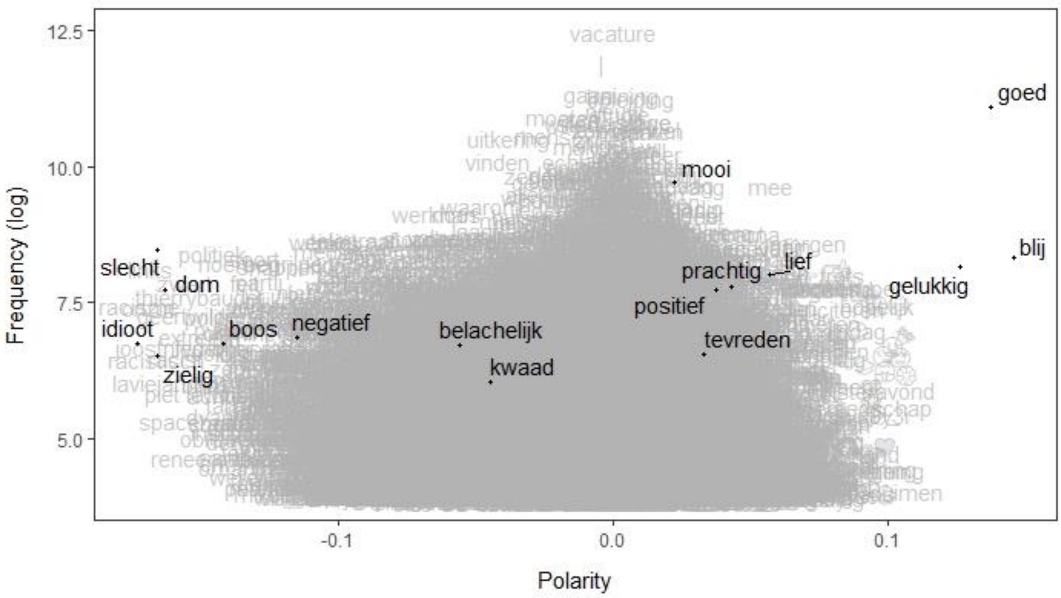
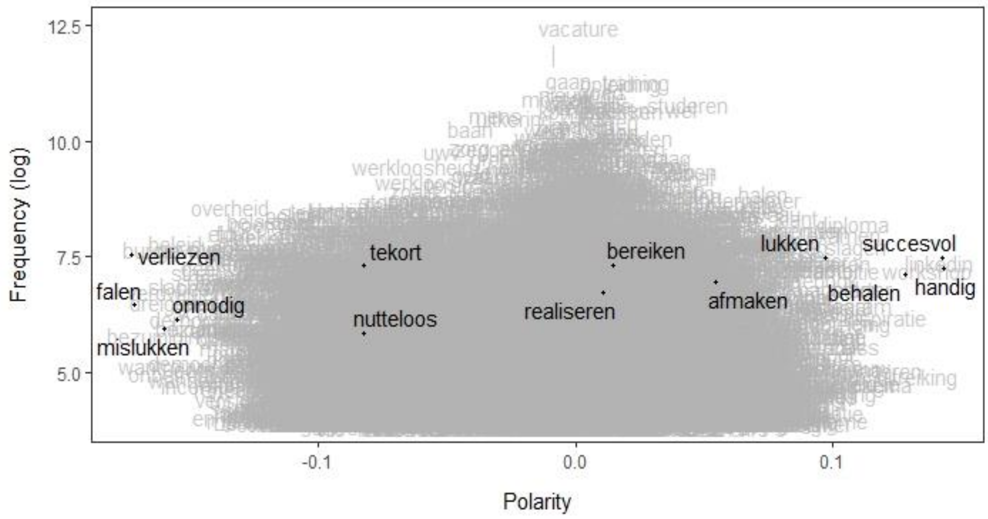


Figure 5.6: Words on a success vs. failure dimension



Another form of validation is inspecting the words assigned the strongest polarity in either direction. Doing so for the example dimension validates it further but also reveals that the dimension captures positive and negative sentiment specific to different topics. In this example the most 'positive' words include words like summer, sunny, vacation, and various happy emojis. The most negative words include racist, racism, nonsense, politics, the names of right-wing political parties and commentators, and Black Pete (captured by tokens 'zwart' and 'piet', referring to a controversial Dutch tradition that many consider racist). In other words, the negative extreme is related to political sentiment and the positive extreme is related to sentiment about weather. This is not (necessarily) a wrong representation of the data, but might not be desired for certain topics. LSS includes an approach that resolves this issue by restricting the model terms to only terms that tend to co-occur with selected tokens (effectively restricting model terms to a given theme or topic), but we do not always find meaningful improvements in utilizing it and it sacrifices the generality of our dimensions. Furthermore, our approach utilizes STM modelling to identify tweets representative of a certain topic prior to applying interpreting LSS estimates. This means that even if the dimension is somewhat general it is applied to a rather specific collection of tweets, reducing the risk of providing a misleading summary.

The last and most important form of validation comes from predicting a polarity score of individual tweets, which is done for every dimension applied to every topic of interest (manually inspecting 20 tweets with the highest and the lowest polarities that 'contain' at least 50% of a given topic). This is important because even a dimension that seems interpretable might not be interpretable when applied to tweets from a specific topic, either due to technical reasons or due to the nature of the topic cluster.

5.3 Results

Applying the proposed method we focus on a 20-topic model for the ECEC sub-section and a 30-topic model for the LM sub-section of our corpus. These models generally do not deliver a fundamentally different insight to other candidate models, but they tend to deliver more 'focused' topics. We present summaries of the full models in Appendix C. From these models we mainly focus on 4 topics from the ECEC sub-section and 5 topics from the LM sub-section as those topics interpretable and potentially policy-relevant: From the ECEC sub-section we focus on **topic 5** concerned with ECEC benefits and their beneficiaries, **topic 11** concerned with playgroups and babysitters, **topic 17** concerned with working mothers, and **topic 18** concerned with health. From the LM sub-section we focus on **Topic 7** concerned with labour market benefits and their beneficiaries, **topic 12** concerned with work contracts, **topic 16** concerned with the Dutch unemployment insurance agency, **topic 18** concerned with the general state of the economy and joblessness, and **topic 28** concerned with self-employment. We provide a more thorough topic-by-topic summary and validation of these in appendix C. The remainder of the topics are either not easily interpretable or are interpretable but not policy-relevant – this would include topics such as (automated and manual) advertising of job vacancies, advertising ECEC services, or responding to other content only with a generic emotional response (eg. emojis denoting laughter or phrases like 'lol'). These topics are an important part of the summary of the overall information in the corpus, but are not further analysed in this chapter due to their obvious lack of policy-relevance.

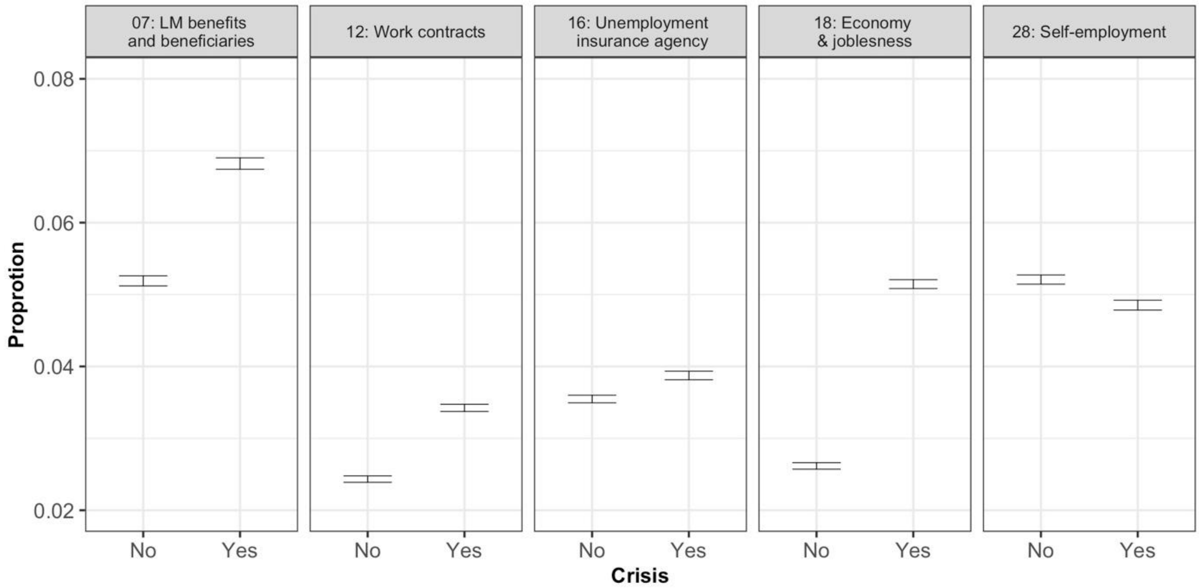
From the above mentioned interpretable and potentially relevant topics we highlight a few successful examples: ECEC Topic 17 is concerned with women balancing child rearing and formal employment, which is a crucial insight for multiple policies, especially given that both sentiment and success LSS dimensions validate well for this topic: The success vs. failure dimension successfully captures how much of this content expresses dissatisfaction with this balance or with policies aimed to help this balance. ECEC topic 18 is concerned with the health of children (and parents) at day-care and how that influences peoples' tendencies to (not) utilize day-care services. This is a very policy-relevant topic in terms of aggregating various limiting factor to take-up rate of childcare services, with the success vs. failure LSS dimension delivering a good summary of how much of this content is about policy failures and (perceived) health hazards. LM topic 28 (self-employment) is very clearly focused on various policies effecting the self-employed and the sentiment dimension provides valuable insight about how negative this content is about those policies. LM topic 12 (contracts) is a topic containing more personal commentary than other topics (generally specific to one's own employment contract), but the prevalence of both typical and a-typical working arrangements is policy-relevant.

However, there are two key issues with the policy-relevance of the identified topics. Firstly, some topics are, despite their seemingly high policy-relevance, rendered irrelevant due to factors such as strong politicisation. The prime examples of this are ECEC topic 5 and LM topic 7. Both of these topics are concerned with who are the 'beneficiaries' and 'contributors' of the Dutch welfare system, but both topics contain primarily strong anti-immigrant rhetoric. This means that despite being concerned with the redistributive effects of the two policy areas (which would be policy-relevant) the individual tweets would often bundle many different welfare policies (including but not limited to ECEC and LM) and convey that they find it unjust for immigrants to have access to these policies and for benefits to be of the size they are. Secondly, some topics remain too broad and cover multiple distinct aspects of policies in one topic, such as LM topic 28 (self-employment), LM topic 16 (unemployment insurance agency), or ECEC topic 17 (working mothers). This does not render the topics irrelevant, but limits them to a more 'agenda setting' use as they do not distinguish between individual policies (or their aspects) sufficiently. In general the summary provided here and in appendix C offers an overview of what policy-relevant information exists in our corpus, but it also points to the relative sparsity of this information among irrelevant and generic tweets and to the difficulty of summarizing this information in topics that are specific (enough) to individual policies or their aspects.

With regards to quantifying any change in these topics between the period of normalcy and of crisis we rely a) on the change in topic prevalence and b) on the change in the mean polarity score on the two LSS dimensions. The prevalence change is illustrated in figure 5.7 for the 4 ECEC topics and in figure 5.8 for the 5 LM topics. The confidence interval plotted on these figures is 95%. For the ECEC corpus sub-section we get insignificant change for topics 11 and 17. For topic 18 (health) we get a very large increase in topic prevalence and for topic 5 (ECEC benefits and beneficiaries) we see a statistically significant decrease in prevalence. For the LM corpus sub-section the only topic whose prevalence decreases is topic 28 (self-employment) with topics 7 (LM benefits and beneficiaries), 12 (work contracts), 16 (unemployment insurance agency), and 18 (Economy and joblessness) increasing in prevalence.

Figure 5.7: Change in topic-prevalence in ECEC sub-section

Figure 5.8: Change in topic prevalence between normalcy and crisis



With regards to change in LSS dimensions there are two significant changes in the ECEC corpus sub-section: The sentiment of topic 17 (working mothers) becomes more negative (mean change from -0.024 to -0.031) and significantly so (two-sample t-test p-value is 0.022 and Mann-Whitney p-value is 0.003). The success/failure dimension for topic 11 (playgroups and babysitters) becomes less associated with success (mean moving from 0.02 to 0.01) and significantly so (p-value for two-sample t-test is 0.004 and p-value for Mann-Whitney is 0.04). For the LM corpus sub-section the changes are the following: Topic 16 (unemployment insurance agency) experiences a very slight increase on the sentiment dimension (from -0.023 to -0.022) that is statistically significant (t-test p-value of 0.017 and Mann-Whitney p-value of 0.017). However, the success dimension actually becomes less concerned with failure (mean polarity changes from -0.037 to -0.033) and this change is very statistically significant (t-test p-value of 0.001 and Mann-Whitney p-value of 0.001). This shows that the topic becomes slightly less negative and concerned with failure, but that this change is much more prominent in its focus on failure rather than in its sentiment. Topic 12 (work contracts) actually changes substantially with regards to its sentiment (mean polarity changes from 0.001 to -0.014) and the change is statistically significant (both two-sample t-test and Mann-Whitney have p-value of 0.001). This shift is also reflected in the success dimension (a change in mean polarity from -0.004 to -0.018) at a statistically significant level (both two-sample t-test and Mann-Whitney have p-value of 0.001).

The changes in these two metrics can be read together to makes some policy-relevant observations. For example, the decreased prevalence of ECEC topic 11 (playgroups and babysitters), given that the topic is largely about existing playgroups and activities in those groups, illustrates the decline of those activities during the COVID crisis. ECEC topic 17 (working mothers) does not change in prevalence but the sentiment dimension becomes more negative during COVID crisis which illustrates that the topics itself is not 'covered' more but that its content is reflective of the increased difficulty of combining child rearing and a career with lockdown policies in place. LM topic 28 (self-employment) decreases in prevalence, potentially illustrating that lockdown policies affect the self-employed less obviously, which makes sense due to workplace shutdowns affecting mainly those with an employer. LM topic 16 (unemployment insurance agency) becomes bo1th more prevalent and more concerned with failure, which reflects displeasure users are voicing with regards to how the agency distributes COVID-specific assistance. LM topic 12 (contracts) experiences the largest shift towards negative sentiment and towards failure observed in the data, which captures the decrease of permanent contracts being awarded (tweets celebrating obtaining such contracts are a large portion of the 'positive' tweets for that topic), which is relevant information about employment that does not get captured by existing indicators – people do not necessarily lose their jobs, but they seemingly experience less career advancement during the COVID-19 outbreak. That said, some of the observed change is not extremely relevant, such as the increase in prevalence of ECEC topic 18 (health) and LM topic 18 (joblessness) which is entirely expected and serve as a sanity check more so than as a finding. Some change is not as expected but still not very policy relevant, such as the decrease in prevalence of ECEC topic 5 (childcare benefits and beneficiaries) but an increase in prevalence of LM topic 7 (LM benefits and beneficiaries) despite a substantial overlap between those two topics, indicating a shift of the discussion more towards labour market issues.

Despite our ability to identify statistically significant and policy-relevant change, there are important caveats to the practical policy-relevance of these observations. For the ECEC sub-section the main caveat is a lack of data: After focusing only on tweets sufficiently representative of a topic whose sentiment is also significantly non-neutral the sample size for some topics becomes extremely small – for ECEC topic 11 we are comparing 104 tweets with 66 and for ECEC topic 17 we are comparing 48 tweets with 41. This obviously erodes the utility these observations can have in practice as the sample is extremely small. For the LM sub-section the main caveat is the generality of topics in the sense that they contain multiple sub-topics: LM topic 28 (self-employment) captures an important debate about occupational disability insurance that was very relevant in our period of 'normalcy', but also issues related to COVID. This is also the case for LM topic 16 (unemployment insurance agency) where the negative content in 'normalcy' is about discrimination and data leaks and in 'crisis' about decisions on distributing COVID-specific assistance, or ECEC topic 18 (health) that focuses on vaccinations in period of 'normalcy' and COVID testing and policies during 'crisis'. This is not an unexpected finding, but one that highlights the importance of obtaining even more disaggregated topics.

5.4 Limitations

There are some important limitations of our research approach and data worth noting. In terms of our data the two corpus sub-sections are very different in size and the amount of noise, but neither is able to avoid issues with both: The ECEC sub-section is smaller (by a factor of 10) and thus runs into the problem of insufficient amount of data, especially when it comes to less prevalent topics. However, this comes with the benefit of less noise (such as automated tweets or tangentially relevant commentary). The labour market sub-section is the inverse of that, as it contains much more data but at the cost of also containing a lot of noise, mainly in the form of automated tweets. In short, both sub-sections could benefit from more or better data. Some of the quality limitations of our data seem inherent to social media, and Twitter specifically, with regards to how opinions get formed on the platform: Moderate users tend to change their opinion over time to fit the average opinion of their friend group (Kozitsin, 2020) and over time one opinion can start dominating the discourse (Xiong & Liu, 2014). This adjustment of opinions is heavily influenced by 'opinion leaders' who tend to post or re-post frequently and tend to have a substantial following and be more engaged in politics (Park, 2013b; Park & Kaye, 2017) and the strength of this influence is mediated by variables such as trust (Xiong et al., 2017) or emotion (Mansouri et al., 2019). As much as users still prefer to voice their opinion rather than to change it (Xiong & Liu, 2014), these dynamics are likely to over-emphasize certain opinions over others. This is not a dynamic our research design can tackle sufficiently, but there are variations of topic modelling that can potentially do so in future work (Gerrish & Blei, 2010).

Outside of concerns related to data quality and opinion formation, there are also some methodological limitations of our approach. With regards to topic modelling, our approach relies on a conceptually inaccurate assumption that topics remain constant across time. This can be fixed by allowing STM models to also infer a change in the content of topics based on the crisis or normalcy variable. We do not do so for two reasons: Firstly, because models with changing topics are, in our case, less interpretable and have more diffused topics. Secondly, because we are running statistical tests of difference on the mean and median polarity, we want to keep the topic construct constant to avoid an 'apples to oranges' comparison. That said, the assumption of fixed topics remains a conceptually sub-optimal assumption.

With regards to LSS the two dimensions do not validate for every topic and some topics also contain a noticeable amount of noise that can skew the LSS polarity estimate. For example, ECEC topic 11 (playgroups and babysitters) includes some comments on unrelated things like cooking that mention that the activity is happening while children are in daycare. Tweets like that can often be much more positive or negative than the more relevant policy-focused tweets, meaning that the mean for a group of tweets can be skewed by these less relevant tweets. In some cases the LSS dimension itself can be a source of bias, as some 'neutral' tokens are associated with a strong polarity in the LSS model. This happens with, for example, ECEC topic 5 where LSS models score the token 'kinderbijslag' (child benefit) itself negatively, resulting in many short tweets that mention child benefit being labeled as negative even though the sentiment of the tweet was neutral. In general we ascribe these problems to shortcomings of our data (for LSS modeling) due to a few reasons: Firstly, the discourse on Twitter can be highly critical which means that even neutral tokens tend to occur in negative commentary and cause LSS to assign that token a negative polarity. Secondly, our data contains short sentences which impede the training of LSS models. We also attempted to train LSS models on entire tweets with little improvement. Thirdly, our documents (tweets) are also very short themselves and make it difficult for a trained LSS model to accurately predict a polarity as the model often only has a few words to work with (prediction is better for longer tweets in our estimation). Fourthly, the dimensions are trained on a very general corpus without restrictions to a particular topic, which results in aggregating multiple features of any dimension (such as sentiment about weather with sentiment about politics in our case). We rely on topic belonging to minimize the effect of this (e.g. we do not expect weather and politics tweets to occupy the same topic), but that assumption is imperfect for broad topics. In general, our data might be sub-optimal for LSS modelling.

5.5 Conclusions

This chapter, much like chapter four, answers the research question 'What policy-relevant information does Twitter contain?' The key additions here are a relaxed definition of 'policy-relevance' (due to a more general approach and focus on agenda setting) and a different temporal scope including a period of substantial 'crisis' for both policy areas. The second, and more novel, research question this chapter answers is 'How does this information change between a period of normalcy and a period of crisis?' To do so, we analyse Twitter data for two 4-month periods – one in a period of relative normalcy and one during the first wave of the COVID-19 crisis in the Netherlands. We propose and utilize a novel method combining topic modelling with latent semantic scaling, which we design to work for real-time data streams and to easily include stakeholders without technical skills by giving them a comparatively large degree of control in specifying what information is 'policy-relevant' in an intuitive way. As such, we also offer concluding remarks on the merits of this novel method.

With regards to the first research question – what policy-relevant information Twitter contains in the context of socio-economic policies – the answer is, much like in chapter four, somewhat sobering: There are relevant 'topics' that include policy-relevant tweets, but these topics are relatively scarce amidst content such as job vacancy postings or responding to other tweets, which is not policy-relevant information for the purposes of this chapter. Even within the relevant clusters of tweets there are substantial issues as the content can be heavily politicised, too broad in coverage, contain noise, or not validate (without a bias) on relevant LSS dimensions. Most of these issues are identified in chapter four and persist even for the new dataset used in this chapter. Some of these problems are methodological and can likely be solved with more sophisticated topic modelling or data cleaning methods, but some of these problems (like the heavy politicisation of some of the relevant content) are not a methodological flaw and are simply a feature of our data. This results in most of the identified relevant topics likely only having utility for agenda setting and not for monitoring performance of specific policies. Furthermore, some of these tweet clusters are policy-relevant according to our definition, but their extremely small size makes them of no practical utility for policymaking, precisely as was the case in chapter four. This does extend the cautionary conclusion from chapter four by showing that even when focusing on more general commentary during a time period 'most likely' to engender such commentary the information is still lacking (albeit to a lesser degree).

The issues we run into with our data and method for extracting insight are seemingly related to features of Twitter data that necessitate us to make trade-offs that we would ideally avoid entirely. The central issue for this chapter is that it is impossible to perfectly distinguish between signal and noise: Despite removing re-tweets, accounts posting with high frequency, and duplicate tweets, near-duplicate tweets still exist due to various sharing features and differences in '@' mentions. If we pre-process data more aggressively, we start losing a noticeable amount of 'signal'. Our approach of conceding an amount of 'noise' in the corpus and letting it cluster into topics that can be excluded works well, but certainly not perfectly. Similar trade-offs exist for our treatment of '@' mentions (we do not remove them and in doing so minimize information loss but it has implications for our topic models), re-tweets (we exclude them to prioritize original and personal content but lose the associated information), quoted tweets (we join the tweet text with the text of the tweet it quotes which introduces noise but maintains the context of those tweets), and others. This is, once again, strongly analogous to the 'subjective' nature of some analytical decisions identified in chapter four. For any of these decisions our choice is informed, but at the same time far from perfect, which aggregates into meaningful problems for our analysis. We consider this as a part of our findings relevant to understanding the limitations of using Twitter data in policymaking and an extension of the findings of chapter four.

With regards to the second research question the answer is, despite being plagued by the same issues mentioned above, much more straightforward. We show that the prevalence (how much is something being talked about) as well as what is being said (in terms of LSS dimensions) changes for some topics and remains consistent for others and that this (lack of) change is statistically significant at a 95% confidence level. Some of these changes are entirely expected, such as the increase in prevalence of topics concerned with the impact of corona on health of children or on the economy. Some findings are a bit more interesting and can carry a great deal of policy relevance, such as the changes to ECEC topic 11 (playgroups and babysitters), ECEC topic 17 (working mothers), LM topic 28 (self-employment), or LM topic 12 (contracts) that we highlight in sub-section 5.3. Some of these insights, such as the decrease in tweets celebrating the obtaining of permanent (and generally better) contracts, capture information that is both policy-relevant and not easily obtainable from other data sources. That said, these changes are only as informative as the topics themselves, which constitutes an important limitation given the general difficulty of identifying relevant topics and their size.

Even though our findings point to serious limitations of utilizing Twitter data for policymaking, our conclusion should not be viewed as contradictory to other research that finds Twitter data useful in the context of labour market indicators (e.g. Proserpio et al. (2016), or in the context of the COVID-19 crisis (e.g. Gilardi, Gessler, Kubli, & Müller (2020)). Our conclusion is complementary to those findings due to differences in approach and sampling strategies. Our approach is distinct in that we a-priori select two rather broad policy domains and collect all relevant tweets without controlling for users or a strict policy focus, unlike research efforts collecting data for specific users and gathering a more ‘complete’ profile for those users (Gilardi et al., 2020; Proserpio et al., 2016). There are of course also important methodological differences but taken together with the existing literature our findings point to a potential limit to the utility of Twitter data.

Despite this cautionary conclusion, the proposed method performed relatively well, especially given the relative brevity of tweets, the fact that some dimensions are not applicable to all topics (some topics are simply not concerned with, for example, failure or success), or the skew of topics themselves (some topics seemingly do not contain a lot of positive sentiment or talk of success). As a general method it has a number of comparative advantages in the level of detail it provides, and the degree of control researchers or policymakers have when it comes to defining dimensions of interest. Technically the method is also language independent and able to utilize streaming data in a real-time fashion with periodical re-training of the underlying models. Needless to say, more future work is needed to validate our approach as it is currently not validated to ground-truth data about public perception such as general opinion surveys or surveys of users whose tweets are included in our corpus. Beyond validation on a more fitting data substantial work also remains with regards to seeing how policymakers would utilize such a system and how well it would perform compared to existing data and decision support systems.

Chapter Six: Conclusions and the realist rejoinder in practice

This dissertation addresses an unconventional research problem – a divergence between the theoretical promise of big data use in policymaking found in the literature and the empirical support for that promise. Despite not being able to (strictly speaking) 'resolve' this problem, the dissertation answers three research questions that together chart a course towards resolving it. This starts with the first research question concerned with the cause of this divergence: **Why does this divergence between theoretical promise and empirical testing of that promise exist?** The answer to this question is provided in chapter two, where my co-author and I argue that the cause of this divergence is the existence of two archetypical narratives in the literature – techno-optimism and policy-pessimism – and their tendency to talk past one another. This is because techno-optimism focuses on the underlying technology of big data analysis, which evolves rapidly and is adopted rapidly in the private sector, whereas policy-pessimism focuses on policy/political decision making, which is comparatively much more resistant to change. These narratives can individually provide arguments for why different aspects of decision making remain largely unchanged despite a change in the insight experts can generate (policy-pessimism), or how dramatically the insight we can get changes and encompasses new areas (techno-optimism), but they struggle to combine the two and talk about how the change in insight can change decision making. Because of this, they fail to answer the question of when (under what conditions) is big data in public administration successful, which is precisely the question needed to better align the practical success (both in the sense of big data use cases and proof-of-concept design research) of big data analysis with the theoretical promise: Success is varied with some policy areas and some data sources showing a lot of promise (and some adoption) and some far less so, which is not sufficiently explained by either narrative.

Following this answer to the first research question, the dissertation turns to the second question: **How can this divergence be improved?** Our answer, also presented in chapter two, is primarily to stop looking for a general understanding of big data and policymaking; It is not a question of *whether* big data changes policymaking, but rather *when* (under what conditions) it changes policymaking and *how*. This shift in necessitates studying more specific cases where big data is/can be used and leveraging insight from both techno-optimism and policy-pessimism: The two narratives might be talking past one another when we talk about the umbrella term of 'big data', but if our inquiry gets more specific so do these two narratives: They become capable of highlighting promises and pitfalls *for a specific instance* of big data use. It is neither the case that big data universally provides 'better' insight that translates itself without friction into 'better' policy decisions, nor is it the case that policymakers are universally going to ignore and/or cherry-pick all evidence and analysis equally. We argue that because the individual promises and pitfalls of big data analysis are extremely different for various data sources and answering different policy questions has very different information requirement, the quest for a 'general theory' of big data (which large portion of the literature seems to be on) ought to be abandoned. The success of big data use, we argue, is determined at a much more specific level – at the level of combinations of sources of big data, associated methods, and particular policy questions. This is not to say that we should not pursue a more generalized understanding of big data in public administrations, but that such an understanding should progress from individual instances of (possible) big data use to generalizing across similar data sources (eg. social media data, search query data, transaction data, etc.) and across either type or domain of policy questions (eg. predictive policy questions, labour market policy questions, etc.).

The answers to both research questions carry important lessons for both researchers and practitioners. In answering these two questions we make these two archetypical narratives meet 'eye-to-eye' in table 1.1, where we take four exemplary features of big data use for policymaking (quality of data and decisions, the speed of data and decisions, epistemological concerns, and fundamental concerns such as privacy) and show how the two narratives approach these features and how one narrative would interrogate the opposing one. We do this to offer a tool for what we argue researchers ought to do: Interrogate their research from both perspectives to realize how techno-optimist or policy-pessimist their assumptions are. Following such a realization researchers can either prefer one paradigm over the other and argue why that is fitting for their research, or they can try to reconcile the two paradigms by taking them both seriously. The questions we provide in table 1.1 can provide a starting point for either approach. In that sense this dissertation provides not just the conceptual argument, but also a broad-strokes method for how interrogating one's research from both perspectives could look like and multiple examples of how various features of big data analysis apply to various policy questions (eg. importance of data backdrop for some policy questions and not for others).

The very same reasoning and broad-strokes method is also of relevance for practitioners: Policymakers encountering expert advise or recommendation based on big data can utilize this approach to interrogate the work of advisors and researchers that gets presented to them. Should any policy recommendations or reasonings fit one of the narratives over the other one, it might be valuable to interrogate it from the opposing perspective, whether that is internally or directly with those issuing a policy advice or constructing a system. Policymakers often have to consider multiple public values, which technical experts do not, and asking probing questions relevant to those values can expose oversight and better communicate the insight policymakers need and the values that need to be respected while obtaining it. This can help avoid analysis that is primarily technical and data-focused (and pays only lip-service to more decision-making and public values challenges), as well as analysis that is primarily political and public-administration focused (paying only lip-service to capabilities of big data analysis). To avoid inefficient analysis or unintended erosion of public values lip-service is not enough.

Even though the answers to research questions 1 and 2 are of direct relevance to both scholars and practitioners, there is an important limitation that needs to be explored before proceeding because it informs the third research question. The first two research questions are both very theoretical and the answers to them are not 'provably' correct. Chapter two is explicit about the techno-optimist and policy-pessimist narratives being *constructed* as a heuristic rather than *discovered* as a distinction clearly traceable to individual papers. This does not only mean that there can be alternative diagnoses and prescriptions for future research that other scholars might arrive at, but also that (methodologically speaking) there is little I can do to demonstrate that the diagnosis and prescriptions presented in this dissertation are 'correct' and that alternative explanations are 'wrong'. The dissertation essentially suggests a research approach. In that respect I can return to where this dissertation started, to John Graunt. Despite the enormous significance of his work, Graunt was very modest with regards to what his contributions amount to (Sutherland, 2013), stating that “How far I have succeeded in the Premisses, I now offer to the World's censure. Who, I hope, … will take it well, that I should offer at a new thing” (Graunt, 1662: 2). As modest as that statement is, it is not overly so: Any new approach that cannot be immediately ‘supported’ or ‘disproven’ using currently available tools and understanding is inevitably subject to ‘world’s censure’. As much as I can argue for the merits of the approach proposed research approach, I cannot demonstrate its ability to fix the divergence between theoretical promise and policymaking reality. Only its adoption and eventual success or failure can do that.

Since the only way to truly demonstrate the merits of the proposed research approach is to adopt it and see if it yields the hoped for results, the third research question does exactly that. The question itself - “**How to design and carry out such research?**” - does not have a universal answer: Given the answer to the second research question, how to actually construct and carry out research would be conditional on the data source, method for analysing it, and policy question we are trying to answer. As such, this dissertation offers only one instance of answering this question by proposing a theoretically promising combination of data source, method, and policy questions and then carrying out proof-of-concept research for that particular tripartite alignment. That said, the ‘answer’ to this question cannot be reduced down and summarized in a few sentences, as the proposed ‘realist’ approach permeates the entire research process from pairing data with policy questions to interpreting results. Chapters three, four, and five all provide an illustrative example of how to design and carry out research that follows the research approach set in answering research questions 1 and 2.

This is the first of two ways in which the proof-of-concept research at core of this dissertation is significant – it serves as an illustrative example of the approach proposed in answering research questions 1 and 2. It is impossible to summarize this in a few sentences here, but the next two sections (6.1 and 6.2) go over the results of this proof-of-concept research and interpret them. Furthermore, this allows for a tentative evaluation of the merits of this approach. Now that the research is complete, it is possible to read its main findings in light of the extant literature and see what it can contribute (done in section 6.3.). This by no means allows me to claim the proposed approach a success or a failure, but it does show how this research approach can be integrated with existing literature. Whether such additions to the literature are desirable will inevitably be left to ‘world’s censure’.

The second way in which this proof-of-concept research is significant is much more straight forward, as it also has a stand-alone significance: It outlines an important practical shortcoming for using social investment in policymaking, designs a method utilizing big data to remedy it, and identifies design principles for similar problems. These contributions are valuable independently of its adoption of the proposed research approach. This proof-of-concept research, at its core, focuses on the research question **“Can social media data be used to operationalize and measure social investment?”**. This question is not the fourth research question and it doesn’t answer the primary research problem directly – it contributes to the primary research problem only because in answering this question the dissertation follows the research approach proposed in answering research question 1 and 2. To fully answer this proof-of-concept research question chapters four and five of this dissertation answer four research sub-questions:

1. How can required information be extracted from social media data?
2. What are the implications of this process for policymaking?

3. Do Twitter data contain the requisite information?
4. How does this information change between a period of normalcy and a period of crisis?

There is substantial overlap between chapters four and five here, as they both address versions of research sub-questions 1, 2, and 3 but with a different data set, different policy-relevance criterion, and ultimately using a different method. Because of the difference in data set chapter five is also capable of answering research sub-question 4 about how policy-relevant information changed between a period of normalcy and a period of crisis.

The first research sub-question cannot be effectively summarized here, as it asks what is the method for analyzing the data set at hand. This method and associated code is the design artefact that the proof-of-concept research proposes, the methods are described thoroughly in chapters four and five and the code is available on an open-access online repository.[5] The second research sub-question concerned with implications of the method is addressed in **section 6.1**, where the primary procedural finding about various analytical decisions being subjective or normative is summarized. The third and fourth research sub-questions are answered in **section 6.2**, where the primary descriptive finding about a lack of relevant information in Twitter data is summarized. **Section 6.3** then takes these two findings and relates them back to the existing literature and to the research approach the proof-of-concept research adopts. In doing so it illustrates how a realist rejoinder of techno-optimism and policy-pessimism can contribute to existing literature, which highlights some of the merits and shortcomings of the proposed research approach.

Even though this chapter mainly reflects on the primary research problem of divergence between theoretical promise and its empirical testing, as well as on the findings of the proof-of-concept research, it does not mean that the dissertation is only limited to those findings: This dissertation expands the description of social investment (chapter three), it identifies an issue with social investment policy analysis (chapter three), it extensively describes Twitter content (chapters four and five), it demonstrates the utility of novel methods like LSS for tasks it has not been applied to prior (chapter five), and others. Given the nature of individual chapters in this dissertation (being written as stand-alone papers), these findings are all reflected on and discussed within their respective chapters, making it unnecessarily repetitive to reflect on them again at this stage. These findings are individually important, but they are neither key findings of the proof-of-concept research or important for reflecting on the theoretical problem of divergence between theoretical promise and its empirical testing.

6.1 The procedural finding

This section discusses the answer to the second research sub-question: What are the implications of this process (referring to the proposed method) for policymaking? In this case, as argued in chapter four, the implications here are not specific to the particular method adopted and apply to using social media data and topic modelling in general. The most important implication is that many of the analytical decisions one has to make to analyse this type of data, even though they could be made in

[5] Both of these are available on github. All code and its explanation for chapter 4 can be accessed at https://github.com/SimonVydra/chapter-4-code and all code and its explanation for chapter 5 can be accessed at https://github.com/SimonVydra/STM-LSS.

an informed way, have a subjective or a normative dimension. This includes things like cleaning the data (what is 'normal enough' behaviour for bot detection, how to deal with forwarding of information, etc.), processing the data (how to translate words into tokens, whether to remove or keep punctuation, etc.), modelling the data (how many topics can we expect, what makes topic valuable, etc.), or communicating the data (type of visualisations, balancing information overload with complete information, etc.). The fact that these decisions have to be made and are both technical but also subjective in nature is an important observation for how tools like this get deployed in practice, especially since technical experts are often not politically accountable and decision makers are not technical experts.

A part of this finding is by no means novel: It is understood that topic models are multimodal and present a non-convex optimization problem (Roberts, Stewart, & Tingley, 2016) – They do not have one globally optimal solution, instead they have multiple locally optimal solutions. To make matters worse, even if there is a globally optimal solution, it might not actually be the 'best' solution: Given that topic models are primarily judged on how they are interpretable by human observers and whether they provide a meaningful insight, it is very difficult to construct quantitative metrics that would approximate such judgement. It is possible to, for example, train models to maximize their predictive performance when it comes to categorizing texts, but metrics like predictive performance do not necessarily correspond to how domain-expert annotators evaluate topic models (Mimno et al., 2011). Not only is there no single analytically 'correct' solution in topic modelling, even if there was one it might not be the preferred solution.

If this is a known problem, what exactly can be done about it? The answer to that question is twofold. Firstly, it is to not only focus on the objective function the topic model is optimizing. Selecting a topic model can involve automated approaches such as computing metrics like semantic coherence and exclusivity, semi-automated approaches such as reading and interpreting the model summaries (such as top-words for topics), or even manual approaches (such as reading through documents to validate the topics assigned to them) (Roberts, Stewart, & Tingley, 2016). The selection between these tools is then pragmatic: For example, automated approaches can identify a broad range of candidate models, semi-automated approaches narrow the selection down, and manual approaches validate the selected model or chose from a small set of candidate models.

Secondly, it is to consider the stability of topics across a range of parameters. In general this is seen as important for parameters such as model initialization as a way to demonstrate that the output is not merely a result of arbitrary choices or randomness (Roberts et al., 2016: 62). As much as there are cutting edge techniques that provide consistent initialization and maximize relevant metrics, such as spectral learning which this dissertation uses for STM models, these approaches do not 'resolve' the multimodality of the problem (Roberts et al., 2016: 82). They simply provide a 'generally good' and consistent solution. This logic of assessing stability can then be extended from just initialization to other parameters that our findings should either be robust to or that influence the outcome and we want to know how. As should be apparent by now, chapters four and five of this dissertation adopt these approaches to dealing with multimodality: Chapter four utilizes semantic coherence for its improved correlation to human judgement as an automated metric (Mimno et al., 2011) and chapter five utilizes a range of metrics (Roberts, Stewart, & Tingley, 2016). Both chapters also discuss the stability of relevant findings across various parametrizations of the models. To make matters worse, results change not only based on features of topic modelling but also based on pre-processing decisions (Denny & Spirling, 2017), which multiplicatively expands the number of 'possible' or 'reasonable' outputs. In the research context looking at stability remains a valid (albeit computationally expensive) option as differences across models or individual topics across those models can be quantified (Roberts, Stewart, & Tingley, 2016) and it is possible to assess how robust results are to pre-processing choices (Denny & Spirling, 2017).

But stability also has its limits, which is where this dissertation makes a contribution: If we approach this problem not purely as a research problem but also as a policymaking one - one where the results inform decision making and policy – we run into two issues: Firstly, should pre-processing and processing of our texts change, new assumptions about what constitutes 'similarity' or 'equivalence' will have to be made and they are also likely to be (in part) subjective and influence the results. Secondly, and more importantly, stability does not actually make models or the topics they contain more or less valuable in this context. In other words, "an unstable topic is not inferior or less substantively interesting" (Roberts et al., 2016: 69). Stable topics tend to be more 'specialized', which can make them more interpretable and less 'mixed', but that does not automatically mean that the representation they provide is the most useful one. Unstable topics will not provide the most robust summary of the entire corpus of texts, but that is not always what topic models are used for. This dissertation is an example of that – it is looking for very specific information and that information can be framed in different way. For example, we can tease out topics about employment, motherhood, the economic situation, or a various combination and 'overlap' of these. The most useful topics here are not going to be the most stable ones, but the ones that best fit to the question we want answered (provided we can interpret and validate them). The very concept of stability, as useful as it is in research context, does not really suffice in policymaking context.

Furthermore, even if stable topics would be 'better', we still do not circumvent the subjective and normative nature of these decisions: A 'full' sensitivity analysis essentially tells us how the results differ across all 'reasonable' decisions across all decision points in the process. But in the policy making context this still does not give us the necessary tools to select the most 'appropriate' parametrization of the analysis. The 'analytical' decisions made along the way can impact various public values and it would be contentious to tailor them to maximize some criterion like interpretability or value (metrics that are very difficult to measure in an automated way anyway). To do so would amount to something like 'analytics washing' or 'big data washing': Presenting a collection of subjective assumptions about what matters and what we ought to look at as 'analytically correct'. Researchers and practitioners cannot claim policymaking relevance of a big data system purely based on the insight it provides. In this context relevance is not limited to the final accuracy improvement or even provision of an entirely novel insight. It includes decisions made along the way that impact relevant public values not just in what the resulting insight captures (or doesn't), but also how it collects and processes the digital footprints of citizens. These decisions can sometimes be made by design (in the sense of designing for a particular public value), but given the number and complexity of such decisions they are just as likely to simply reflect the value positions of those making them.

This of course raises questions around how to resolve this conundrum, and as nihilistic as the description here might seem, this dissertation does not argue that there is no solution. It just argues that there is no **analytical** solution. The very issue is that analytical and political decisions are inseparable in an analysis like this. Should there be conflicting results, or conflicts in making some of the decisions, no amount of analysis can show one course of action to be 'correct' and the other one 'incorrect'. One solution is to adopt the same standards applicable to research and assume that assessing stability (and other technical approaches) do arrive at the 'best' final insight. The other solution, one that this dissertation sees as promising, takes the opposite approach: Admitting that these decisions can be political and, as a consequence of that admission, include democratically accountable decisionmakers in (some of) these decisions and hold them accountable for value judgements and assumptions carried forward by big data systems. Alternatively, some form of democratic accountability needs to reside with those actually making these decisions (the analysts) to avoid a situation where the value-laden nature of these decisions is used to knowingly privilege some public values over others without accountability, or to unknowingly outsource or omit important value-based decisions.

This translates into lessons for both researchers and practitioners. For researchers, it suggests a range of research questions related to how policymakers understand and perceive this type of analysis. Primarily, are policymakers aware of this issue? And if they are, do they see this issue as problematic? Answering such questions can open the doors to more action-design research, where policymakers are actively involved in design and adjustment of a particular analysis, allowing researchers to study how relevant decisions can be made jointly by policymakers and technical experts, or what is missing for such a cooperation to materialize. Based on the literature we can already hypothesize a few limiting factors, such as the technical skill requirements this would place on policymakers, or the time requirements (given how iterative the analysis is). Such research direction also reinforces the need for interdisciplinarity articulated at the start of this dissertation as 'analytical' decisions need to be understood both analytically but also in terms of their impact on ethics, public values, or legal standards.

For practitioners (or design research that aims to be useful in policymaking practice), there are lessons about design goals and limits of technical solutions. In policymaking context, purely technical analysis can only takes us so far: For example, it can narrow down the decisions that actually matter for the model output and it can present those decisions in an understandable way, which can foster more agreement, but it cannot privilege one decision over another should they both meet a certain standard of technical soundness. More importantly, this lesson also suggests that the **integration of non-technical users** and the **transparency and explainability of how their decisions effect analytical output** are crucial design goals and should be designed for, much like technical performance metrics are. In that respect chapter five of this dissertation already makes some progress by designing its method in a way that concentrates a lot of the most important normative decisions into particular points where even non-technical users can easily understand them and adjust the analysis in intuitive rather than technical ways. Future research that aims to have policymaking relevance can expand on this logic, as well as evaluate whether such design steps actually translate to better integration of policymakers in analytical processes.

6.2 The descriptive finding

This section tackles the third and fourth research sub-questions concerned with whether Twitter data contains policy-relevant information and whether this information changes between periods of normalcy and crisis. The answer is that Twitter does not contain a meaningful amount of relevant information for either of the two data set (chapters four and five) and under either of the two definitions of policy-relevance. The information does change between periods of normalcy and crisis, but the amount of information remains so small this change is of no substantive policy relevance. What does this mean at face value? The simple answer is that it tempers some of the optimism with regards to big data use in policymaking. This is because the dissertation focuses on an alignment of data, methods, and policy questions that is 'designed for success': Methodologically the net is cast wide and the process is highly iterative. In terms of policy questions the focus is, broadly speaking, on information we can expect to be posted on social media. In terms of country and timeframe selection we are looking at a country with high internet penetration, high use of social media and Twitter specifically (especially for the demographic of interest), and time period of disruptive change to relevant policy areas and associated life course transitions. Following the theoretical promise articulated in the literature, big data should provide meaningful value here. The fact that this is not the case and that there is very little relevant information contradicts the existing literature that hypotheses that this type of data could be of immense practical (supporting policymaking) and theoretical (providing novel information to test novel theories) value. That is the simple answer.

However, to fully evaluate the importance of this finding we must consider it in light of its limitations. Given the difficulty of obtaining additional data sources at the time of writing this leaves the dissertation in an uncomfortable position of almost finding an absence of evidence (finding only very small amounts of it), which is generally a conclusion with predictable limitations. In this case, there are three main issues limiting the validity of the finding: Firstly, the evidence could be absent because of the platform selection. Twitter could simply be a social media platform where the requisite type of information is not shared. This greatly limits the external validity of the finding as it is, strictly speaking, a valid finding only for Twitter.

Secondly, it is the context and timeframe of the research. External validity is further limited to the Netherlands (or at least Tweets labeled as Dutch), the policy context, and to a specific time frame used to collect this data. The specification to the Netherlands and a set of policies is undoubtedly limiting external validity as in other countries and/or for other policy areas the discourse on Twitter could contain more relevant information. The timeframe concern is at least partially addressed by the combination of chapter four and five: Chapter four utilizes a comparatively broad timespan to gather data and chapter five includes a crisis period that greatly affects the policy areas of interest. It is, of course, still possible that 'early Twitter' or 'future Twitter' will contain more or even less policy relevant information, but for the time period this research was conducted in the temporal coverage is substantial, reducing the external validity concerns in this respect. That said, due to its context-specific approach, this dissertation explicitly sacrifices external validity of its findings in favor of a more nuanced understanding of a specific tripartite alignment of data, method, and policy question. This is a theoretical choice, one justified in chapter two, and will always result in lower external validity of findings.

Thirdly, and perhaps most importantly, there is the opacity of sourcing data from Twitter, which could be a challenge to internal validity. Twitter provides a Streaming API access to developers, which can be either sampled (Sample API) to receive a random sample of 1% of all tweets or filtered (Filter API) to receive specific tweets up to the 1% of total traffic threshold. The Sample API has been shown to not be a truly random sample by experimental designs (Morstatter et al., 2013) and the sampling procedure used, one based on the millisecond signature of when tweets are received by Twitter, can be exploited to make some tweets over-represented (Pfeffer et al., 2018). As much as this is not an issue for the Filter API (Pfeffer et al., 2018), which is used throughout the dissertation, there are two limitations that remain relevant for it: Firstly, that 'corporate spammers' can be over-represented in both Sample API and Filter API due to their ability to exceed the rate limits (Pfeffer et al., 2018: 15), which is relevant to the findings of chapters four and five that identify a large amount of corporate posts. Secondly, it is the opacity of the data collection process. The scripts developed to collect data throughout the dissertation were not getting timed out due to rate limitations, suggesting that they were not crossing the 1% threshold. The number of tweets collected throughout the dissertation was roughly 11 thousand per day, which should be very firmly under the 1% threshold of total tweets. That said, such a statement cannot be easily verified. Twitter does not publish metrics such as daily volume of tweets, so they can only be roughly extrapolated from the volumes of the 1% sample, which would equate to between 300 million and 400 million tweets per day (Leetaru, 2019). Earlier experimental research shows that the Filter API collected approximately 43% of all relevant Tweets, collecting between 10 and 40 thousand tweets per day, but never collecting all tweets for a given day (Morstatter et al., 2013). Later research successfully used the Filter API to collect all tweets posted during an experiment, but the volume here was substantially lower at approximately 15 thousand tweets in total (Pfeffer et al., 2018: 6). As much as this dissertation 'should' have a majority of relevant tweets, without an access to full Twitter data it is impossible to quantify how much data is missing and whether the sample is biased in some meaningful way. This is by no means a shortcoming specific to this paper and much of research utilizing Twitter data suffers from the same problem, but it is a relevant shortcoming nevertheless, especially in light of recent changes to accessing Twitter data.

These recent changes to Twitter's API allow academic researchers to search the full historical archive of Tweets, provided it is done for research purposes (Twitter, n.d.). This is of great value for future research utilizing Twitter data, but less great for this dissertation, as this change was only fully implemented after all the data was collected and analyzed. This makes the above mentioned threat to internal validity (opacity about how much of the total data is captured and what biases exist in its sampling) something almost universally present in research utilizing Twitter data before functionalities of API 2.0 were released, but also something unlikely to be present, or at least to a far lesser degree, in research that is to come. Even if this dissertation has majority of the relevant data, arguing for the absence of (meaningful amount of) evidence is necessarily limited by the uncertainty about what data, if any, is missing.

Given these limitations, the simple answer to what this finding means for the literature bears re-interpreting. Even though this is an unforeseen development rather than a methodological oversight, it is an important limitation for taking this finding at face value, especially given the fact that it is 'negative' or a 'null' finding. In principle this should not be an issue, not if the research is methodologically rigorous, but the bias against negative results is a well-documented feature of the scientific discourse: The proportion of papers reporting a null finding in academic journals is decreasing over the years - a trend known as positive-outcome bias - and is especially prominent in social sciences (Fanelli, 2012). The positive-outcome bias can even be identified in randomized control trials, showing that even when a manuscript differs only in the direction of the finding and is otherwise identical reviewers will recommend the positive result for publishing more, they will rate its method as better (even though they are completely identical), and they will pick up on less errors comparatively to the negative result (Emerson et al., 2010). It is important to note that there has been a meaningful amount of progress made with respect to publishing negative findings in the past decade, but the finding still generates some skepticism. For example, chapter four was originally written as a stand-alone paper and was reviewed in a prominent journal, where one reviewer noted: "The relevance of this result is questionable, considering that other work based on a Twitter corpus (also using simple word matching) has been used to construct …" and listed research works utilizing

Twitter to provide insight. A negative finding in a sea of positive ones is not seen as 'interesting' but rather as 'questionable', which is illustrative of the diagnosis of a techno-optimist bias argued for in chapter two.

This is by no means to complain about reviewers, and certainly not about this particular reviewer, since they also provide a very good reason for why the finding is questionable: "However, it is unclear if this result is an artefact of the coarse filtering applied … and not whether the dataset could really be used for this type of research." In that argument they are absolutely right. In fact, that critique could be made even more forcefully: There is a myriad of other methodological decisions that could cause the findings, many of them are explicitly identified in chapter four as subjective but potentially consequential for the results and their interpretation. The issue gets more complicated here, as this very critique can be made against virtually any of the existing research that presents positive findings. In general, research utilizing big data for social indicators or policymaking is a relatively novel field and it is not filled with randomized control trials or natural experiments – it is largely filled with observations of interesting correlations or features of data and outlining what could be a promising research direction.

In this context (big data analysis and more exploratory research), not only are there many more associations that can be discovered, but there are also multiple methods and multiple ways to parametrize those methods to answer a particular research question. With this deluge of research approaches and their increasing complexity, most findings presented in the literature (whether positive or negative), can be critiqued for presenting only a sub-section (or only a singular) approach and parametrization(s) of that approach. The difference between positive and negative findings is that for positive findings this criticism does not erode their relevance: Even a single parametrization is indeed sufficient to show potential, since the limits of that potential can easily be offloaded onto future research. But for research that reaches a negative finding, the inability to demonstrate that this negative finding is consistent across an exhaustive range of parametrizations is a problem, as we cannot be certain that the absence of evidence is due to the method or due to its actual absence. In that sense, it becomes increasingly easier to support a positive finding and more difficult to support a negative one in the context of big data analysis.

There are multiple reasons for why this can be seen as problematic for academia in general, whether that has to do with discouraging pioneering (and thus risky) research, incentivizing unethical behaviour, or a myriad of other problems, but for this dissertation it is more directly relevant: The comparative ease of supporting positive as opposed to negative findings almost inevitably supports a techno-optimist bias. As argued in chapter two, it is crucial to ask the when question with regards to big data – under what conditions are big data projects successful and under what conditions do they fail. But those conditions can only be discovered at the boundary between projects that 'succeed' to use big data and those that 'fail'. In that sense, negative findings, or in general projects that 'fail' to utilize big data effectively, are as important as those that 'succeed' in utilizing big data. This is why the answer to what this descriptive finding actually adds to the literature gets more complicated: From a perspective of methodological rigor with no concern for a techno-optimist bias in the literature, the finding holds very little significance as it (self-admittedly) could be somewhat different under different parametrizations, failing to provide scientific proof that the lack of evidence it finds is because it actually does not exist. However, from the perspective of the dissertation, one that emphasizes context-specific research and focuses on conditions enabling big data success, this descriptive finding is essential for answering the when question, which is argued to be the primary goal of the literature. The complicated answer is thus the following: It depends on what we value about the literature and how we are willing to trade-off between methodological rigor and 'balance' in the literature (given that positive findings are more easily rigorous, as argued above). It is difficult to suggest how this trade-off ought to be made, especially since there are many ways to reconcile rigor and balance, such as forcing studies that arrive at a 'positive' finding to more thoroughly test the 'limits' of their findings.

This is thought provoking for researchers, as this dissertation is calling for more 'failed' proof-of-concept efforts and arguing that such 'failure' is necessary for complete understanding of big data use in public administrations (argument developed further in the next section as well). For researchers who have reached a 'null' finding or failed to design a functioning system this offers some encouragement, as, provided that the research is theoretically informed and rigorous, there is still immense value in the design principles one can derive from their efforts. For research that reaches a positive finding or successfully designs a system the suggestion is to interrogate such findings further and push them 'to failure' to illustrate their limits. Doing so will result in more thorough research and allow us to have a more nuanced understanding of big data use in public administrations.

6.3 The realist approach

Even though these two findings have significance in and of themselves, in this section the chapter returns to the primary research problem – the divergence between theoretical promise of big data for policymaking and the empirical testing of that promise. As outlined at the start of this chapter, the proof-of-concept research at the core of this dissertation also serves as an illustrative example of the proposed research approach, eventually resulting in our ability to answer the **when** question of big data and understand under what conditions is big data use successful in policymaking. This dissertation of course does not give us a complete answer, but it does move us forward in a few ways. If, as section 6.2 claims, 'failures' are important to understand the conditions for big data use success, what are those conditions in this particular case? It is important to specify here that 'failure' does not refer to quality of the research as such, but simply to whether the proof-of-concept ambitions of a research effort yield a 'positive' finding or a 'null' finding. In other words, when speaking about failure in proof-of-concept research, I use it to refer to research that successfully identifies where big data fails, rather than to research I believe does not meet requisite academic standards.

That said, keeping the specific when question close to empirical work in chapters four and five, when can Twitter data be successfully used to answer policy questions related to the labor market? The most important condition for success here seems to be what approach to information extraction one has. One such approach completely disregards the content of Tweets and primarily relies on when (and where) a tweet was posted. This includes efforts of Llorente et al. (2015), who utilize this data to identify patterns of mobility and times during the day when users post to approximate regional unemployment rates. Bokányi et al. (2017) adopt a similar approach for the US context and support the findings of Llorente et al. (2015) by showing that the timing of people's tweets can indicate their employment status (eg. starting tweeting later in the day or tweeting more during work hours can correspond to unemployment), at least when aggregated to the county level. Both of the research efforts are successful in this respect and outperform baseline predictive approaches.

Another approach disregards the contents of tweets only partially, by first using it to identify relevant users but then collecting all of their tweets to quantify their psychological state regardless of which topic they are addressing, essentially focusing on “psychological micro-foundations of macroeconomics” (Proserpio et al., 2016: 223). Here Proserpio et al. (2016) identify individuals who lost a job or gained a job and gathers their five-year tweeting history. They then use Linguistic Inquiry Word Count method to quantify the psychological state of people and to identify correlations between psychological state and disclosed change of employment status. Some of the identified variables, such as anxiety, are leading or lagged indicators of employment shocks (eg. anxiety decreases for those who gained a job and increases for those who lost a job following the shock). Proserpio et al. (2016) then use these findings to create an autoregressive model that utilizes some of these leading psychological variables to predict monthly unemployment rates. It outperforms a baseline autoregressive model with its predictions (Proserpio et al., 2016).

The last, and perhaps most intuitive use of Twitter data is to focus on the content of the tweets themselves and any self-disclosure they contain. This is what this dissertation does, and it fails: Chapter four and five show that in the Dutch context there is insufficient information and the information that is there is generally not personal enough to amount to disclosure of personal employment situation or perceived difficulties with employment-related life course transitions. Biorci et al. (2017) also focus on content of tweets and count the frequency of keywords like ‘unemployment’ or ‘work’ (in Italian, due to the context) and explore whether this data is correlated to regional unemployment information, using the ‘location’ information users can provide about themselves in their profile. This effort finds no obvious association and, much like this dissertation, points to a large volume of ‘noise’ in the data (Biorci et al., 2017). Here it is also important to note that Proserpio et al. (2016), given that their dataset contains seemingly reliable identification of tweets that indicate job loss and job gain, also investigate whether the frequency of users mentioning job loss or gain on Twitter can improve unemployment prediction. Autoregressive models utilizing this information fail to outperform baseline autoregressive models in all samples (despite coming close) (Proserpio et al., 2016). In a similar fashion, Llorente et al. (2015) also explores the value of the contents of tweets by counting number of unemployment related keywords and seeing whether it aids in prediction. In the proposed model this variable turns out to not be statistically significant (Llorente et al., 2015).

Despite these failures of Twitter data to predict unemployment under this approach to information extraction, there are some cases that conclude with a ‘success’. One of them is research by Ryu (2018), which crowdsources 662 unique keywords from 100 people and considers three timeseries for each of these keywords: The Google Trends timeseries, its frequency in Twitter data (and other media), and the sentiment associated with this word in social media. It cross-correlates these time series with the official unemployment statistic timeseries to look for correlations and then tests the predictive power of a series of autoregressive models utilizing some of these keywords. The model utilizing frequency of keywords on Twitter has the best predictive performance, outperforming the baseline autoregressive model. However, out of the 662 keywords, this model utilizes only two: “price” and “inflation”, lagged by one and two months respectively (Ryu, 2018: 911-912). These are not keywords associated directly with labor market situation, but rather with consumer price expectations and inflation. It is thus impossible to parse out whether the improved predictive performance is due to utilizing Twitter data or due to including inflation and price information in the model, especially since Twitter data has successfully been used to measure inflation expectations

(Angelico et al., 2022). A more informative comparison would be between a baseline autoregressive model and a model utilizing only employment-related keywords, or between a model including conventional measures of inflation (and/or consumer price indices) and one including the two selected keywords. The comparison, as it currently stands, cannot support a conclusion with regards to Twitter data containing information about employment situation of individuals.

The last piece of research, and one that concludes with a 'success', is by Antenucci et al. (2014). The work of Antenucci et al. (2014) is one of the most widely cited works from this research area and one largely inspiring this dissertation. Antenucci et al. (2014) utilize a set of n-grams (keyword combinations) to detect disclosure of information such as job loss in Twitter data and utilize this information to now-cast initial claims for unemployment insurance (analogue for job loss). The effort is very successful, showing that it is feasible to gather this information and that it can be used as "both substitutes and complements to data generated from surveys and administrative records by statistical agencies and the private sector" (Antenucci et al., 2014: 27). In fact, recalling the reviewer who found the findings of chapter four questionable because other research efforts have successfully used Twitter data, the research of Antenucci et al. (2014) was the first example provided. The reviewer is by no means alone in reading the work of Antenucci et al. (2014) in this way - it is generally cited for showing the possibility of utilizing social media data for economic indicators. However, given that the design ambition was to now-cast initial claims data, Antenucci et al. (2014) let the model run after the paper was published to further verify their findings and establish how useful they are. These predictions start diverging strongly from the initial claims data rather quickly and the model ceases to make predictions in 2017 (University of Michigan, 2015). Thus, despite showing that Twitter data can capture relevant information, the conclusion of the research is not one of 'success', as the resulting indicator cannot be relied on and was eventually 'discontinued'.

Here it is important to state that, should there be any blame to go around for the interpretation of the findings of Antenucci et al. (2014), none of it lays with the authors. The authors explicitly state that longer time series and further research is needed to verify the usefulness of this data and that "[i]n practice, the rapid evolution of the use of social media could make the relationship between the measurement and the underlying fundamental being measured unstable" (Antenucci et al., 2014: 28). The decision to continue issuing predictions is exemplary research practice, especially from the perspective of this dissertation, as it provides a more realist notion of usefulness rather than a techno-optimist one: The utility of the indicator does not lie just in its technical performance at the time of construction, but also in the data backdrop it can amass as a demonstration of its reliability, as well as the exploration of the value of its features (eg. what value can a better temporal resolution provide for decisions). Returning back to the discussion of 'failure' and publishing of 'negative' results, Antenucci et al. (2014) clearly demonstrate how more 'failure' with regards to its proof-of-concept ambition can equate to more rigorous and better research efforts. It also engenders much more useful questions about why this divergence happened: Was it due to training of the model? Change in what data was supplied to it by Twitter? Or simply an evolution of how users utilize Twitter? In this case failure is important and it is not at the expense of research quality or effort, the very opposite in fact - it takes more effort to interrogate one's research to the point of failure. In the realm of big data finding something that works at face value is often not difficult, it is the 'failure' or 'limits' of that discovery that takes effort and rigor to discover.

As a general commentary on 'failure' in this field I can say that many of my peers are apprehensive about my labelling of the proof-of-concept ambitions of this dissertation as 'failure', fearing that I am being too 'harsh' on myself and that I am looking past the successes. In some ways they are right and there is more 'success' in the dissertation than it lets on. Chapter five, for example, can be framed as a success: Under a more technical and narrow definition of success it does identify interpretable and policy relevant topics, extracts novel insight about them, successfully shows how they change in policy relevant ways, and validates this insight down to individual tweets. However, such a definition of success fails to consider how such an approach would apply to policymaking practice and what its use would mean for relevant public values. In other words, claiming success here would be techno-optimist. The extent to which we claim 'failure' is largely a choice and given how the work of Antenucci et al. (2014) is referred to in the literature, despite it demonstrating crucial limitations to using Twitter data, this dissertation is as explicit as possible about the failure of its proof-of-concept ambition. In fact, this dissertation considers its failure as a point of pride and a testament to empirical honesty and adherence to a 'realist' perspective.

The importance of accepting the value of failure is certainly a part of the research approach this dissertation proposes and illustrates, but before reflecting on it fully it is important to answer the **when** question with regard to Twitter and labour market policymaking: The most intuitive use of Twitter data – utilizing the content and trying to identify disclosure of personal situation relevant to policymaking - generally fails: This dissertation finds a lack of information, Biorci et al. (2017) find no association, Proserpio et al. (2016) do not improve a baseline model utilizing this information, Llorente et al. (2015) find this information statistically insignificant in combination with other variables, Ryu (2018) does not provide an appropriate baseline, and Antenucci et al. (2014) show that the information is prone to changes. Even though this is by no means a common reading of the literature, I believe it to be the most empirically honest one. Interestingly enough, it shows that using Twitter Data tends to be **unsuccessful when we are using the information users knowingly provide**, such as disclosures of one's employment situation or struggles with a life course transition. However, we tend to be **more successful when utilizing Twitter data for information that users do not knowingly provide**, such as what is their 'daily rhythm' or their general mental state.

What does this seemingly simple answer mean for researchers? Firstly, it brings into question emphasis on national context, which does not seem to be important according to available evidence – the summarized research includes Netherlands, Italy, Spain, and the US and the context does not seem to change the finding about when use of Twitter data for labor market assessment succeeds or fails. The evidence is of course not sufficient to write of national context as unimportant, but it does open up future research to more systematic study of how exactly national context matters for the content in social media data. Most directly it also indicates which approaches to information extraction tend to be more successful for Twitter data, but that is not to say that utilizing Tweets (or other social media data) to identify disclosure of important events is research that should be abandoned. Identifying things like disclosure of job loss are technically difficult in this data because only a miniscule fraction of total tweets actually do that, making it difficult to train machines to recognize it. Some of the work in this space points to problems analogous to the ones experienced in this dissertation and explores much more advanced methods for identifying such disclosure, like active learning of pre-trained language models (Tonneau et al., 2022). This is another promising research direction, also because methods such as active learning can integrate non-technical users quite well, since annotation does not require technical knowledge and the amount of it that needs to be done is drastically reduced compared to other (semi)supervised approaches.

In terms of demonstrating the proposed research approach (one argued to potentially resolve the primary research problem), this dissertation shows that it is possible for more design oriented proof-of-concept research. Most importantly, the research approach is seemingly capable of helping answer the when question associated with big data, as it does above for Twitter data and labor market assessment. In doing so, it is more prone to 'failing' its proof-of-concept ambitions, as it establishes more 'realist' criteria for success that are not limited only to technical performance. This also makes it more demanding in terms of interdisciplinarity. As much as the research approach seems feasible and promising for eventually providing a more nuanced understanding of big data use in public administrations, it is certainly not an easy sell for existing researchers due to its interdisciplinary requirements, more thorough interrogation of results, and resulting proneness failing its design goals.

Outside of researchers, this finding has important lessons for practitioners as well. Despite the 'simple' answer of approach to information extraction being the key, differences in that approach can have important knock-on effects: For example, it necessitates a type of sampling where all tweets for relevant accounts are collected and analyzed, which aligns with an observation made in chapter five: That other research efforts yielding more success differ in their sampling strategy and collect a much broader range of tweets (for relevant accounts). Knock-on effects like this can be very relevant in policymaking context as, in this case, the successful uses seem to collect more data less discriminately and analyze it to get insights that Twitter users are not necessarily intending to disclose. Working with such insight could be contentious in terms of public values, but it could also be legally contentious due to requirements such as data minimization; Things like the mental state of citizens is likely 'relevant' for many things an agency could be trying to measure, potentially justifying an almost unlimited collection of Tweets (or other social media content). Deciding whether to use social media data (and for what type of information) requires careful appraisal of the potential benefits for policymaking practice so that decisionmakers can appropriately balance the upside with the costs (both financial and in terms of public values).

Acknowledgements

It has been close to a year since I finished my dissertation, and I am relieved I did not write my acknowledgements back then. They would have been very cookie-cutter, quickly put together, and pretty much the acknowledgements everyone expects. Some things are no different (I had to put these together even quicker than I would have had to back then), but this time away from my work has given me perspective only distance can give.

Given that I am no longer in academia and it is uncertain whether I will return, I am starting from a very broad perspective; There are three people without whom I would not be at this point. Firstly, it is Marek Havrda whom I started working for during my Bachelor's and continued working for (on and off) ever since. My Bachelor's was a time of disillusionment with political science and "theory-heavy" approached to social sciences, but Marek showed me how individual people - ones with drive and genuine interest in better policies - can actually change policies for the better. That you don't have to comment on and theorize things or play the political game, that you can just "do" and by "doing" make a part of the world a better place. Marek was also one of the first people to suggest I pursue a PhD focused on big data. If I hadn't met Marek then, my career would be very different. Secondly, it is Anton Hemerijck who supervised my master's thesis and for whom I worked as a research assistant after. During my Master's I had developed a strong interest in "big data" and hoped to get a PhD position, but I could not find a fitting one (or one that would accept me) for a time. During that time Anton convinced me that I have enough work ethic and talent to succeed in academia and that my own arguments have value. My search for a PhD, which was discouraging and long, turned into a period where I learned that I could have a voice in academia. If I hadn't met Anton then, my career would be very different. Thirdly, it is Bram Klievink, who was first my PhD supervisor and later my promotor and boss. My PhD (like any other PhD) had its fair share of challenges and Bram allowed me tackle them in my way, but supported me and gave advice like a great friend when I would inevitably fail. Bram taught me how to be an independent researcher and take charge of a scientific inquiry. Not just to carry out work, but also to plan it and decide when it is finished. If Anton helped me find my voice, Bram taught me how to use it: When to whisper but also when to speak up or shout, whether that is with respect to quality of my own work, academic principles, or leadership and support. If I hadn't met Bram then, my career would be very different.

Without these three I would not be defending this dissertation (for better or worse – I don't mean to flatter them that much). Curiously enough, I don't think any one of them realises what they really taught me. It is not about evidence-based policymaking, welfare states, or digitization of governance. It is about the value of relentlessly pursuing meaningful change, finding your voice and believing in the value of that voice, and learning to use that voice. All three had faith in me when others (including myself) didn't, and I would not have finished or even started writing this book without that. I doubt I can ever repay them, but I will pay it forward.

Now after "starting small" and acknowledging those who formed my entire academic career, let me return to the topic at hand - this dissertation. Firstly, there are the rest of my promoters - Hans de Bruijn and Madeleine Hosli, who have helped to tailor the focus of my PhD and helped me answer some difficult questions about its contributions. Secondly, there is my supervisor Jaroslaw Kantorowicz, who has not only worked on a paper with me and who I taught a course with, but who also provided guidance and help throughout, going far beyond his duties as a supervisor. Speaking of going above and beyond as well as teaching great courses I must mention two other colleagues – Scott Cunningham and Jan Kwakkel. Scott has always given so much of his valuable time to me for discussing topic models and helping me understand and interpret what I was seeing better. Jan teaches one of the best Master's courses I have ever seen and allowed me the privilege to teach a part of it as well, which continues to inspire me (and everyone else I talk to about it) to this day. Then there are all my colleagues at the Hague Centre for Digital Governance, who have been some of the most welcoming people I have ever met and who always managed to take my weird ideas and eventually hammer out something useful out of them. It was a privilege to work with them – there are not many teams with their level of expertise and commitment (if any). I have already given up on naming names here, since there are so many. We are all painfully aware of the readership statistics of doctoral dissertations, so chances are that if you are one of the people reading this you are also one of the people who I am grateful to have met and worked with throughout this process. Thank you.

Outside of those who helped with the process of writing this book, there are those who helped with everything else that this process demands from a person. I am thankful to my TU Delft peer group for sharing the journey with me; to Maria for helping me stay sane; to Moty for advocating for the truth; to Viki for always sharing a dram when needed; to David, Lucas, and Merlijn for not changing in any of the important ways; to my training partners for always kicking my ass more than my PhD could; and to everyone else who was there for me. Much like previously, if you are reading this your name should likely be here too – thank you. To end it is important to say one last thing: All the lofty pursuits like finding my voice, using it for good, fixing my mistakes, or producing this body of work sound really good on paper, but they absolutely suck in real life. They are difficult, arduous, and they take time. The associated personal growth is great, but it comes with a cost and that cost is never borne just by the person doing the growing. In my case the brunt of that cost was borne by my girlfriend Maya, who bore it with more grace and kindness than I could have ever imagined.

References

Acar, A., & Muraki, Y. (2011). Twitter for crisis communication: Lessons learned from Japan's tsunami disaster. *International Journal of Web Based Communities*, *7*(3), 392–402. https://doi.org/10.1504/IJWBC.2011.041206

Ahmed, W., Bath, P. A., Sbaffi, L., & Demartini, G. (2019). Novel insights into views towards H1N1 during the 2009 Pandemic: a thematic analysis of Twitter data. *Health Information & Libraries Journal*, *36*(1), 60–72. https://doi.org/10.1111/hir.12247

Anderson, C. (2008). *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*. Wired. https://www.wired.com/2008/06/pb-theory/

Andriotis, P., & Takasu, A. (2019). Emotional bots: Content-based spammer detection on social media. *10th IEEE International Workshop on Information Forensics and Security, WIFS 2018*, 1–8. https://doi.org/10.1109/WIFS.2018.8630760

Angelico, C., Marcucci, J., Miccoli, M., & Quarta, F. (2022). Can we measure inflation expectations using Twitter? *Journal of Econometrics*, *228*(2), 259–277. https://doi.org/10.1016/j.jeconom.2021.12.008

Angiuli, O., Blitzstein, J., & Waldo, J. (2015). Balancing statistical accuracy and subject privacy in large social-science datasets. *COMMUNICATIONS OF THE ACM*, *58*(12). https://doi.org/10.1145/2814340

Angwin, J., & Larson, J. (2016). *ProPublica Responds to Company's Critique of Machine Bias Story*. ProPublica. https://www.propublica.org/article/propublica-responds-to-companys-critique-of-machine-bias-story

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks.* ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Antenucci, D., Cafarellab, M., Levensteinc, M., Red, C., & Shapiro, M. (2014). *Using Social Media to Measure Labor Market Flows* (No. 20010; NBER Working Paper). http://www-personal.umich.edu/~shapiro/papers/LaborFlowsSocialMedia.pdf

Anvik, C., & Gjelstad, K. (2010). *"Just Google it": Forecasting Norwegian unemployment figures with web queries* (11/2010; CREAM Publications). https://biopen.bi.no/bi-xmlui/handle/11250/95460

Askitas, N., & Zimmermann, K. (2009). *Google econometrics and unemployment forecasting* (No. 4201; IZA Discussion Papers).

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science*, *355*(6324).

Baker, S., & Fradkin, A. (2011). *What Drives Job Search? Evidence from Google Search Data* (No. 10–020; Working Paper). https://siepr.stanford.edu/research/publications/what-drives-job-search-evidence-google-search-data

Barcaroli, G., & Righi, P. (2015). Big data in official statistics : ongoing work at Statistics Italy. *Big Data for the Analysis of the Digital Economy and Society*, *September*.

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, *104*(671).

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

Bartholomew, M. K., Schoppe-Sullivan, S. J., Glassman, M., Kamp Dush, C. M., & Sullivan, J. M. (2012). New Parents' Facebook Use at the Transition to Parenthood. *Family Relations*, *61*(3), 455–469. https://doi.org/10.1111/j.1741-3729.2012.00708.x

Begg, I. (2017). Social Investment and its Discount Rate. In A. Hemerijck (Ed.), *The Uses of Social Investment*. Oxford University Press.

Bengtsson, M., de la Porte, C., & Jacobsson, K. (2017). Labour Market Policy under Conditions of Permanent Austerity: Any Sign of Social Investment? *Social Policy & Administration*, *51*(2), 367–388. https://doi.org/10.1111/spol.12292

Bertot, J. C., & Choi, H. (2013). Big Data and e-Government: Issues, Policies, and Recommendations. *The Proceedings of the 14th Annual International Conference on Digital Government Research (Dg.o 2013)*, 1–10. https://doi.org/https://doi.org/10.1145/2479724.2479730

Beveridge, W. (1942). *Social Insurance and Allied Services*. https://www.sochealth.co.uk/national-health-service/public-health-and-wellbeing/beveridge-report/

Biorci, G., Emina, A., Puliga, M., Sella, L., & Vivaldo, G. (2017). Tweet-Tales: Moods of Socio-Economic Crisis? *Data Science and Social Research*, 205–213. https://doi.org/10.1007/978-3-319-55477-8_19

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84. https://doi.org/10.1145/2133806.2133826

Blei, D. M., & Lafferty, J. D. (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics*, *1*(1), 17–35. https://doi.org/10.1214/07-AOAS114

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*, *148*, 113–120. https://doi.org/10.1145/1143844.1143859

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*, 993–1022. http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf

Boettcher, I. (2015). *Automatic data collection on the Internet (web scraping)*.

Bokányi, E., Lábszki, Z., & Vattay, G. (2017). Prediction of employment and unemployment rates from Twitter daily rhythms in the US. *EPJ Data Science*, *6*(1). https://doi.org/10.1140/epjds/s13688-017-0112-x

Bollier, D. (2010). *The Promise and Peril of Big Data*. The Aspen Institute. https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf

Bonoli, G. (2013). *The Origins of Active Social Policy: Labour Market and Childcare Policies in a Comparative Perspective*. Oxford University Press.

Bostic, W. G., Jarmin, R. S., & Moyer, B. (2016). Modernizing Federal Economic Statistics. *American Economic Review*, *106*(5), 161–164. https://doi.org/10.1257/aer.p20161061

Boyd, D., & Crawford, K. (2012a). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information Communication and Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Boyd, D., & Crawford, K. (2012b). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Brenner, H., & Nissim, K. (2014). Impossibility of Differentially Private Universally Optimal Mechanisms. *SIAM Journal on Computing*, *43*(5), 1513–1540. https://doi.org/10.1137/110846671

Breunig, C., & Busemeyer, M. R. (2012). Fiscal austerity and the trade-off between public investment and social spending. *Journal of European Public Policy*, *19*(6), 921–938. https://doi.org/10.1080/13501763.2011.614158

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age : work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.

Burgoon, B. (2017). Practical Pluralism in the Empirical Study of Social Investment: Examples from Active Labour-Market Policy. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 161–173). Oxford University Press.

Burns, A., & Mitchell, W. (1946). *Measuring Business Cycles*. National Bureau of Economic Research.

Calude, C. S., & Longo, G. (2016). The Deluge of Spurious Correlations in Big Data. *Foundations of Science*, *22*(3), 595–612. https://doi.org/10.1007/s10699-016-9489-4

Carta, F., & Rizzica, L. (2016). Female employment and pre-kindergarten: on the unintended e↵ects of an Italian reform. *Dondena Working Papers*, *91*. ftp://ftp.dondena.unibocconi.it/WorkingPapers/Dondena_WP091.pdf

CBS. (2020). *Monthly labour participation and unemployment*. https://www.cbs.nl/en-gb/figures/detail/80590eng

Chatfield, A. T., & Reddick, C. G. (2018). Customer agility and responsiveness through big data analytics for public value creation: A case study of Houston 311 on-demand services. *Government Information Quarterly*, *35*(2), 336–347. https://doi.org/10.1016/J.GIQ.2017.11.002

Chen, G. M. (2011). Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior*, *27*(2), 755–762. https://doi.org/10.1016/j.chb.2010.10.023

Chen, Y.-C., & Hsieh, T.-C. (2014). Big Data for Digital Government: Opportunities, Challenges, and Strategies. *International Journal of Public Administration in the Digital Age*, *1*(1), 1–14. https://doi.org/10.4018/ijpada.2014010101

Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., & OI Team. (2020). *The Economic Impacts of COVID-19: Evidence from a New Public Database Built from Private Sector Data*.

Chew, C., & Eysenbach, G. (2010). Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak. *PLoS ONE*, *5*(11), e14118. https://doi.org/10.1371/journal.pone.0014118

Choi, H., & Varian, H. (2009). *Predicting Initial Claims for Unemployment Benefits*. http://static.googleusercontent.com/media/research.google.com/en//archive/papers/initialclaimsUS.pdf

Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record*, *88*(1), 2–9.

Clasen, J., & Clegg, D. (2011). *Regulating the Risk of Unemployment: National Adaptations to Post-*

Industrial Labour Markets in Europe. Oxford University Press.

Clemons, R. S., & McBeth, M. K. (2009). *Public policy praxis : a case approach for understanding policy and analysis*. Routledge.

Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A Garbage Can Model of Organizational Choice. *Administrative Science Quarterly*, *17*(1), 1. https://doi.org/10.2307/2392088

Colace, F., De Santo, M., Lombardi, M., Mercorio, F., Mezzanzanica, M., & Pascale, F. (2019). Towards Labour Market Intelligence through Topic Modelling. *Proceedings of the 52nd Hawaii International Conference on System Sciences* , 5256–5265. https://hdl.handle.net/10125/59962

Contreras-Piña, C., & Ríos, S. A. (2016). An empirical comparison of latent sematic models for applications in industry. *Neurocomputing*, *179*, 176–185. https://doi.org/10.1016/j.neucom.2015.11.080

Crawford, K., Gray, M., & Miltner, K. (2014). Big Data| Critiquing Big Data: Politics, Ethics, Epistemology| Special Section Introduction. *International Journal of Communication*, *8*(10), 1663–1672. http://ijoc.org/index.php/ijoc/article/download/2167/1164

Crouch, C., & Keune, M. (2012). The Governance of Economic Uncertainty: Beyond the "New Social Risks" Analysis. In G. Bonoli & D. Natali (Eds.), *The Politics of the New Welfare State* (pp. 45–67). Oxford University Press.

Cukier, K., & Mayer-Schoenberger, V. (2013). The Rise of Big Data: How It's Changing the Way We Think About the World. *Foreign Affairs*, *92*(3), 28–40. https://www.jstor.org/stable/23526834?seq=1#page_scan_tab_contents

D'Amuri, F. (2009). *Predicting unemployment in short samples with internet job search query data* (No. 18403; MPRA Paper). https://ideas.repec.org/p/pra/mprapa/18403.html

D'Amuri, F., & Marcucci, J. (2010). "Google it!" Forecasting the US unemployment rate with a Google job search index. *FEEM Working Paper No. 31.* https://doi.org/http://dx.doi.org/10.2139/ssrn.1594132

Daries, J. P., Reich, J., Waldo, J., Young, E. M., Whittinghill, J., Ho, A. D., Seaton, D. T., & Chuang, I. (2014). Privacy, anonymity, and big data in the social sciences. *Communications of the ACM*, *57*(9), 56–63. https://doi.org/10.1145/2643132

de Bruijn, H., & Ten Heuvelhof, E. (2008). *Management in networks : on multi-actor decision making*. Routledge.

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Major life changes and behavioral markers in social media: case of childbirth. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13*, 1431–1442. https://doi.org/10.1145/2441776.2441937

De Deken, J. (2014). Identifying the Skeleton of the Social Investment State. In B. Cantillon & F. Vandenbroucke (Eds.), *Reconciling Work and Poverty Reduction: How Successful Are European Welfare States?* (pp. 260–285). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199926589.003.0009

De Deken, J. (2017). Conceptualizing and Measuring Social Investment. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 184–193). Oxford University Press.

de la Porte, C., & Natali, D. (2018). Agents of institutional change in EU policy : the social investment moment. *Journal of European Public Policy*, *25*(6), 828–843. https://doi.org/10.1080/13501763.2017.1401110

Deeming, C., & Smyth, C. (2015). Social Investment after Neoliberalism: Policy Paradigms and Political Platforms. *Journal of Social Policy*, *44*(2), 297–318. https://doi.org/10.1017/S0047279414000828

Denny, M. J., & Spirling, A. (2017). *Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It*. https://doi.org/10.1017/pan.2017.44

Desouza, K. C., & Jacob, B. (2014). Big Data in the Public Sector: Lessons for Practitioners and Scholars. *Administration & Society*, *49*(7), 1043–1064. https://doi.org/10.1177/0095399714555751

Di Bella, E., Leporatti, L., & Maggino, F. (2018). Big Data and Social Indicators: Actual Trends and New Perspectives. *Social Indicators Research*, *135*, 869–878. https://doi.org/10.1007/s11205-016-1495-y

Dieterich, W., Mendoza, C., & Tim Brennan, M. (2016). *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*. http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf

Dong, L., Chen, S., Cheng, Y., Wu, Z., Li, C., & Wu, H. (2017). Measuring economic activity in China with mobile big data. *EPJ Data Science*, *6*(1), 6–29. https://doi.org/10.1140/epjds/s13688-017-0125-5

Dräbing, V., & Nelson, M. (2017). Addressing Human Capital Risks and the Role of Institutional Complementarities. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 128–139). Oxford University Press.

Dumbacher, B., & Hutchinson, R. (2016). Enhancing the Foundation of Official Economic Statistics with Big Data. *European Conference on Quality in Official Statistics*.

Dunleavy, P., Margetts, H., Bastow, S., & Tinkler, J. (2005). New Public Management Is Dead--Long Live Digital-Era Governance. *Journal of Public Administration Research and Theory*, *16*(3), 467–494. https://doi.org/10.1093/jopart/mui057

Dwivedi, Y. K., Rana, N. P., Tajvidi, M., Lal, B., Sahu, G. P., & Gupta, A. (2017). Exploring the role of social media in e-government: An analysis of emerging literature. *ACM International Conference Proceeding Series*, *Part F1280*, 97–106. https://doi.org/10.1145/3047273.3047374

Ebbinghaus, B. (2011). *The Varieties of Pension Governance: Pension Privatization in Europe.* Oxford University Press.

Einav, L., & Levin, J. D. (2013). The Data Revolution and Economic Analysis. *NBER Working Paper*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Emerson, G. B., Warme, W. J., Wolf, F. M., Heckman, J. D., Brand, R. A., & Leopold, S. S. (2010). Testing for the Presence of Positive-Outcome Bias in Peer Review: A Randomized Controlled Trial. *Arch Intern Med*, *170*(21), 1934–1939.

Emmanuel, I., & Stanier, C. (2016). Defining Big Data. *Proceedings of the International Conference on Big Data and Advanced Wireless Technologies - BDAW '16*, 1–6. https://doi.org/10.1145/3010089.3010090

Esping-Andersen, Gosta. (2015). Investing in Early Childhood. *Revue Belge de Sécurité Sociale*, *1*, 99–112. https://socialsecurity.belgium.be/sites/default/files/rbss-1-2015-espingandersen-fr.pdf

Esping-Andersen, Gøsta, Gallie, D., Myles, J., & Hemerijck, A. (2002). *Why We Need a New Welfare State* (Gosta Esping-Andersen (ed.)). Oxford University Press.

European Commission. (n.d.). *IESI (ICT-Enabled Social Innovation)*. Retrieved March 2, 2022, from https://joint-research-centre.ec.europa.eu/iesi-ict-enabled-social-innovation_en

European Commission. (2013). *Towards Social Investment for Growth and Cohesion – including implementing the European Social Fund 2014-2020*. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2013:0083:FIN:en:PDF

European Commission. (2019). Employment and social developments in Europe 2019. In *Employment and Social Developments in Europe*. https://ec.europa.eu/commission/presscorner/detail/en/IP_19_3412

European Commission. (2020). *LMP expenditure by type of action*. https://webgate.ec.europa.eu/empl/redisstat/databrowser/view/LMP_EXPSUMM/default/table?lang=en&category=lmp_expend

Eurostat. (2020). *Children in formal childcare or education by age group and duration*. https://ec.europa.eu/eurostat/databrowser/view/ILC_CAINDFORMAL__custom_129524/default/table?lang=en

Eurostat Big Data Task Force. (2014). *22 nd Meeting of the European Statistical System Committee*. https://ec.europa.eu/eurostat/cros/system/files/ESSC doc 22_8_2014_EN_Final with ESSC opinion.pdf

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Fenwick, T. B. (2017). From CCTs to a social investment welfare state? Brazil's 'new' pro-poor strategy. *Development Policy Review*, *35*(5), 659–674. https://doi.org/10.1111/dpr.12247

Ferrera, M. (2017). Impatient politics and social investment: the EU as 'policy facilitator''.' *Journal of European Public Policy*, *24*(8), 1233–1251. https://doi.org/10.1080/13501763.2016.1189451

Fleckenstein, T., & Lee, S. C. (2017). A Social Investment Turn in East Asia? South Korea in Comparative Perspective. In A. Hemerijck (Ed.), *The Uses of Social Investment*. Oxford University Press.

Flores, A., Bechtel, K., & Lowenkamp, C. (2016). False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks." | United States Courts. *Federal Probation*, *80*(2). https://www.uscourts.gov/federal-probation-journal/2016/09/false-positives-false-negatives-and-false-analyses-rejoinder

Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, *30*, 117–125. https://doi.org/10.1016/J.ECONMOD.2012.07.017

Gallego, A., & Marx, P. (2017). Multi-dimensional preferences for labour market reforms: a conjoint experiment. *Journal of European Public Policy*, *24*(7), 1027–1047. https://doi.org/10.1080/13501763.2016.1170191

Gangl, M. (2004). Welfare States and the Scar Effects of Unemployment: A Comparative Analysis of the United States and West Germany. *American Journal of Sociology*, *109*(6), 1319–1364. https://doi.org/10.1086/381902

Garritzmann, J., Häusermann, S., Palier, B., & Zollinger, C. (2016). *WoPSI-the World Politics of Social Investment*.

Gerrish, S. M., & Blei, D. M. (2010). A language-based approach to measuring scholarly impact. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 375–382. https://dl.acm.org/doi/abs/10.5555/3104322.3104371

Ghosh, A., Roughgarden, T., & Sundararajan, M. (2012). Universally Utility-maximizing Privacy Mechanisms. *SIAM Journal on Computing*, *41*(6), 1673–1693. https://doi.org/10.1137/09076828X

Ghysels, J., & Van Lancker, W. (2011). The unequal benefits of activation: an analysis of the social distribution of family policy among families with young children. *Journal of European Social Policy*, *21*(5), 472–485. https://doi.org/10.1177/0958928711418853

Giere, R. (2006). *Scientific Perspectivism*. The University of Chicago Press. https://books.google.nl/books?id=nxJnpU8ePocC&lpg=PP1&hl=nl&pg=PP1#v=onepage&q&f=false

Giest, S. (2017). Big data for policymaking: fad or fasttrack? *Policy Sciences*, *50*(3). https://doi.org/10.1007/s11077-017-9293-1

Gilardi, F., Gessler, T., Kubli, M., & Müller, S. (2020). *Social Media and Policy Responses to the COVID-19 Pandemic in Switzerland*. https://fabriziogilardi.org/

Gitelman, L. (Ed.). (2013). *"Raw Data" Is an Oxymoron*. MIT Press.

Gluckman, P. (2017). *Using evidence to inform social policy: The role of citizen-based analytics*. https://www.nbr.co.nz/sites/default/files/17-06-19-Citizen-based-analytics.pdf

Goldston, D. (2008). Big data: Data wrangling. *Nature*, *455*(7209), 15–15. https://doi.org/10.1038/455015a

Gould, P. (1981). Letting the Data Speak for Themselves. *Annals of the Association of American Geographers*, *71*, 166–176. https://doi.org/10.2307/2562790

Grammatikopoulos, V., Gregoriadis, A., Tsigilis, N., & Zachopoulou, E. (2014). Parental conceptions of quality in Greek early childhood education. *European Early Childhood Education Research Journal*, *22*(1), 134–148. https://doi.org/10.1080/1350293X.2012.738868

Graunt, J. (1662). *Natural and Political Observations Mentioned in a following index, and made upon the Bills of Mortality*. Roycroft.

Greve, B. (2018). Do stakeholders in Denmark know about social investment within long-term care? *Journal of International and Comparative Social Policy*, *34*(2), 127–139. https://doi.org/10.1080/21699763.2018.1465447

Grimmer, J., & Stewart, B. M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, *21*(3), 1–31. https://doi.org/10.1093/pan/mps028

Groves, R. M. (2011). Three Eras of Survey Research. *Public Opinion Quarterly*, *75*(5), 861–871.

https://doi.org/10.1093/poq/nfr057

Grubmüller, V., Götsch, K., & Krieger, B. (2013). Social media analytics for future oriented policy making. *European Journal of Futures Research*, *1*. https://doi.org/10.1007/s40309-013-0020-7

Gualda, E., & Rebollo, C. (2016). The Refugee Crisis On Twitter: A Diversity Of Discourses At A European Crossroads. *Journal of Spatial and Organizational Dynamics*, *4*(3), 199–212. https://www.jsod-cieo.net/journal/index.php/jsod/article/view/72

Hackl, P. (2016). Big Data: What can official statistics expect? *Statistical Journal of the IAOS*, *32*, 43–52. https://doi.org/10.3233/SJI-160965

Hall, P. (1993). Policy Paradigms, Social Learning, and the State: The Case of Economic Policymaking in Britain. *Comparative Politics*, *25*(3), 275–296.

Hall, P. A., & Gingerich, D. W. (2009). Varieties of Capitalism and Institutional Complementarities in the Political Economy: An Empirical Analysis. In *British Journal of Political Science* (Vol. 39, pp. 449–482). Cambridge University Press. https://doi.org/10.2307/27742754

Hall, P., & Soskice, D. (Eds.). (2001). *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*. Oxford University Press.

Hargittai, E. (2015). Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 63–76. https://doi.org/10.1177/0002716215570866

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902. https://doi.org/10.1126/science.1128898

Heerschap, N., Ortega, S., Priem, A., & Offermans, M. (2014). Innovation of tourism statistics through the use of new big data sources. *12th World Telecommunication/ICT Indicators Symposium*. https://www.itu.int/en/ITU-D/Statistics/Documents/events/wtis2014/002INF-E.pdf

Heitzmann, K., & Matzinger, S. (2021). The (changing) consideration of social investment in the design of welfare benefits: The case of poverty relief in Austria. *Social Policy and Administration*, *55*(4), 573–588. https://doi.org/10.1111/spol.12643

Helbig, N., Dawes, S., Dzhusupova, Z., Klievink, B., & Mkude, C. G. (2015). Stakeholder Engagement in Policy Development: Observations and Lessons from International Experience. In *Policy Practice and Digital Science* (pp. 177–204). Springer International Publishing. https://doi.org/10.1007/978-3-319-12784-2_9

Hemerijck, A. (2013). *Changing Welfare States*. Oxford University Press.

Hemerijck, A. (2017). Social Investment and its Critics. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 3–42). Oxford University Press.

Hemerijck, A. (2018). Social investment as a policy paradigm. *Journal of European Public Policy*, *25*(6), 810–827. https://doi.org/10.1080/13501763.2017.1401111

Hemerijck, A., Burgoon, B., Di Pietro, A., & Vydra, S. (2016). *Assessing Social Investment Synergies*. Publications Office of the European Union.

Hemerijck, A., & Plavgo, I. (2021). Measuring returns on social investment beyond here-and-now redistribution: A commentary on Parolin and Van Lancker’s response article. *Journal of European Social Policy*, *31*(3), 309–320. https://doi.org/10.1177/09589287211018144

Hemerijck, A., & Vydra, S. (2016). Le champ d'analyse de la politique d'investissement social. *Informations Sociales*, *192*, 8–20. https://www.cairn.info/revue-informations-sociales-2016-1-page-8.htm

Heverin, T., & Zach, L. (2010). Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma, Washington Area. *Proceedings of the 7 Th International ISCRAM Conference*.

Hilbert, M. (2016). Big Data for Development: A Review of Promises and Challenges. *Development Policy Review*, *34*(1), 135–174. https://doi.org/10.1111/dpr.12142

Hills, J. (2014). *Good times, bad times: The welfare myth of them and us*. Bristol Policy Press.

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 147–169. https://doi.org/10.1080/10919392.2015.1125187

Hoffman, M. D., Blei, D. M., & Bach, F. (2010). Online Learning for Latent Dirichlet Allocation. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culotta (Eds.), *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1 (NIPS'10)* (pp. 856–864). Curran Associates Inc. https://papers.nips.cc/paper/3902-online-learning-for-latent-dirichlet-allocation.pdf

Höpner, M. (2005). What connects industrial relations and corporate governance? Explaining institutional complementarity. *Socio-Economic Review*, *3*, 331–358. https://doi.org/10.1093/SER/mwi014

Hussmanns, R. (2007). *Measurement of employment, unemployment and underemployment – Current international standards and issues in their application*. International Labour Organization. http://www.ilo.org/global/statistics-and-databases/WCMS_088394/lang--en/index.htm

Iacus, S. M. (2015). Big Data or Big Fail? The Good, the Bad and the Ugly and the missing role of Statistics. *Electronic Journal of Applied Statistical Analysis*, *5*(1), 4–11. https://doi.org/10.1285/I2037-3627V5N1P4

IBM. (2012). *4 Vs*. http://www.ibmbigdatahub.com/tag/587

ILO. (2020). *ILO Monitor: COVID-19 and the world of work. Sixth edition* .

Internet World Stats. (2017). *Europe Internet Usage Stats Facebook Subscribers and Population Statistics*. http://www.internetworldstats.com/stats4.htm

Inuwa-Dutse, I., Liptrott, M., & Korkontzelos, I. (2018). Detection of spam-posting accounts on Twitter. *Neurocomputing*, *315*, 496–511. https://doi.org/10.1016/J.NEUCOM.2018.07.044

Janssen, M., & Helbig, N. (2018). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*, *35*(4), S99–S105. https://doi.org/10.1016/j.giq.2015.11.009

Janssen, M., & Kuk, G. (2016a). Big and Open Linked Data (BOLD) in research, policy, and practice. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 3–13. http://www.tandfonline.com/doi/pdf/10.1080/10919392.2015.1124005?needAccess=true

Janssen, M., & Kuk, G. (2016b). *The challenges and limits of big data algorithms in technocratic* *governance*. https://doi.org/10.1016/j.giq.2016.08.011

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. *Procedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*. http://aisl.umbc.edu/resources/369.pdf

Johnson, P. R., & Yang, S.-U. (2009). Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use. *Paper Presented at the Communication Technology Division of the Annual Convention of the Association for Education in Journalism and Mass Communication*. https://www.researchgate.net/publication/228959109

Kantepe, M., & Ganiz, M. C. (2017). Preprocessing framework for Twitter bot detection. *2017 International Conference on Computer Science and Engineering (UBMK)*, 630–634. https://doi.org/10.1109/UBMK.2017.8093483

Katal, A., Wazid, M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good Practices. *In Contemporary Computing (IC3), 2013 Sixth International Conference*, 404–409. https://pdfs.semanticscholar.org/34ad/09cda075101dc4ce3c04006ff804aca3ebf8.pdf

Kawabata, M. (2012). Access to childcare and the employment of women with preschool-aged children in Tokyo. *CSIS Discussion Paper*, *114*. http://www.csis.u-tokyo.ac.jp/dp/114.pdf

Kazepov, Y., & Ranci, C. (2017). Is every country fit for social investment? Italy as an adverse case. *Journal of European Social Policy*, *27*(1), 90–104. https://doi.org/10.1177/0958928716673314

Keith, J., Ginnis, S., & Miller, C. (2016). Addressing quality in social media research: the question of representivity. *Social Research Practice*, *2*, 21–30. http://the-sra.org.uk/wp-content/uploads/social-research-practice-journal-issue-02-summer-2016-v2.pdf

Kenworthy, L. (2017). Enabling Social Policy. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 89–96). Oxford University Press.

Kettl, D. F. (2016). Making Data Speak: Lessons for Using Numbers for Solving Public Policy Puzzles. *Governance*, *29*(4), 573–579. https://doi.org/10.1111/gove.12211

Kettl, D. F. (2018). *Little Bites of Big Data for Public Policy*. CQ Press.

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-Data Applications in the Government Sector. *Communications of the ACM*, *57*(3), 78–85.

Kingdon, J., & Thurber, J. (1984). *Agendas, alternatives, and public policies*. Little Brown.

Kitchin, R. (2014a). The real-time city? Big data and smart urbanism. *GeoJournal*, *79*(1), 1–14. https://doi.org/10.1007/s10708-013-9516-8

Kitchin, R. (2014b). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, *1*(1), 205395171452848. https://doi.org/10.1177/2053951714528481

Kitchin, R., & Lauriault, T. P. (2015). Small data in the era of big data. *GeoJournal*, *80*(4). https://doi.org/10.1007/s10708-014-9601-7

Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). *Inherent Trade-Offs in the Fair Determination of Risk Scores*. https://arxiv.org/pdf/1609.05807.pdf

Klievink, B., Romijn, B. J., Cunningham, S., & de Bruijn, H. (2017). Big data in the public sector: Uncertainties and readiness. *Information Systems Frontiers*, *19*(2), 267–283.

https://doi.org/10.1007/s10796-016-9686-2

Klijn, E. H., & Koppenjan, J. (2015). Strategic complexity in governance networks: strategies, games, rounds, and arenas. In *Governance Networks in the Public Sector* (pp. 66–98). Routledge.

Kogan, M. (1999). The impact of research on policy. In F. Coffield (Ed.), *Speaking truth to power: Research and policy on lifelong learning*. Policy Press.

Kozitsin, I. V. (2020). Formal models of opinion formation and their application to real data: evidence from online social networks. *The Journal of Mathematical Sociology*, *00*, 1–28. https://doi.org/10.1080/0022250X.2020.1835894

Ku, C.-H., & Leroy, G. (2014). A decision support system: Automated crime report analysis and classification for e-government. *Government Information Quarterly*, *31*(4), 534–544. https://doi.org/10.1016/J.GIQ.2014.08.003

Kuitto, K. (2016). From social security to social investment? Compensating and social investment welfare policies in a life-course perspective. *Journal of European Social Policy*, *26*(5), 442–459. https://doi.org/10.1177/0958928716664297

Kvist, J. (2015). A framework for social investment strategies: Integrating generational, life course and gender perspectives in the EU social investment strategy. *Comparative European Politics*, *13*(1), 131–149. https://doi.org/10.1057/cep.2014.45

Lancker, W. Van, & Ghysels, J. (2012). Who benefits? The social distribution of subsidized childcare in Sweden and Flanders. *Acta Sociologica*, *55*(2), 125–142. https://doi.org/10.1177/0001699311433428

Laruffa, F. (2018). Social investment: Diffusing ideas for redesigning citizenship after neo-liberalism? *Critical Social Policy*, *38*(4), 688–706. https://journals.sagepub.com/doi/pdf/10.1177/0261018317749438

Lasswell, H. D. (1956). *The decision process: Seven categories of functional analysis*. Bureau of Governmental Research, College of Business and Public Administration, University of Maryland.

Lavertu, S. (2016). We All Need Help: “Big Data” and the Mismeasure of Public Administration. *Public Administration Review*, *76*(6), 864–872. https://doi.org/10.1111/puar.12436

Lee, K., Eoff, B., & Caverlee, J. (2011). Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter. *ICWSM*. https://www.semanticscholar.org/paper/Seven-Months-with-the-Devils%3A-A-Long-Term-Study-of-Lee-Eoff/1dd5355e62b9fc37a355e135d5909ed28128d653

Leetaru, K. (2019). *Visualizing Seven Years Of Twitter’s Evolution: 2012-2018*. Forbes. https://www.forbes.com/sites/kalevleetaru/2019/03/04/visualizing-seven-years-of-Twitters-evolution-2012-2018/

Lemire, S., & Petersson, G. J. (2017). Big Bang or Big Bust? The Role and Implications of Big Data in Evaluation. In G. J. Petersson & J. D. Breul (Eds.), *Cyber Society, Big Data, and Evaluation: Comparative Policy Evaluation*. Transaction Publishers. https://books.google.nl/books?id=mb3lDQAAQBAJ&printsec=frontcover&hl=nl&source=gbs_ge_summary_r&cad=0#v=onepage&q=singular&f=false

Leoni, T. (2015). Welfare state adjustment to new social risks in the post-crisis scenario. A review with focus on the social investment perspective. *WWWforEurope Working Papers Series*.

https://ideas.repec.org/p/feu/wfewop/y2015m3d0i89.html

Lettieri, N. (2016). Computational Social Science, the Evolution of Policy Design and Rule Making in Smart Societies. *Future Internet*, *8*(2), 19. https://doi.org/10.3390/fi8020019

Lindblom, C. (1959). The Science of Muddling Through. *Public Administration Review*, *19*(2), 79. https://doi.org/10.2307/973677

Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 134–142. https://doi.org/10.1016/j.isprsjprs.2015.11.006

Llorente, A., Garcia-Herranz, M., Cebrian, M., & Moro, E. (2015). Social Media Fingerprints of Unemployment. *PLoS ONE*, *10*(5), 1–13. https://doi.org/10.1371/journal.pone.0128692

Lopes, A., & Dias, I. (2018). The Portuguese stakeholders' perspective on social investment and quality assessment in LTC. *Journal of International and Comparative Social Policy*, *34*(2), 140–150. https://doi.org/10.1080/21699763.2018.1472135

Lopez, J. (2016). Sandbox Project on Social Media Data for Sentiment Analysis/Mobility. *Conference on Big Data for Official Statistics*.

Lundberg, J., Nordqvist, J., & Laitinen, M. (2019). Towards a language independent Twitter bot detector. *Proceedings of 4th Conference of The Association Digital Humanities in the Nordic Countries*. http://lnu.diva-portal.org/smash/record.jsf?pid=diva2%3A1302270&dswid=5681

Maciejewski, M. (2016). To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences*, *83*(1), 120–135. https://doi.org/10.1177/0020852316640058

Maduro, M., Pasi, G., & Misuraca, G. (2018). *Social impact investment in the EU. Financing strategies and outcome oriented approaches for social policy innovation: narratives, experiences, and recommendations*. Joint Research Center of the European Commission. https://doi.org/10.2760/159402

Mansouri, A., Taghiyareh, F., & Hatami, J. (2019). Improving Opinion Formation Models on Social Media Through Emotions. *5th International Conference on Web Research (ICWR)*, 6–11. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8765288

Marmot, M. G. (2004). Evidence based policy or policy based evidence? *BMJ (Clinical Research Ed.)*, *328*(7445), 906–907. https://doi.org/10.1136/bmj.328.7445.906

Matheus, R., Janssen, M., & Maheshwari, D. (2018). Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly*. https://doi.org/10.1016/J.GIQ.2018.01.006

Matsaganis, M., Ozdemir, E., & Ward, T. (2014). *The coverage rate of social benefits*.

Mazza, M., Cresci, S., Avvenuti, M., Quattrociocchi, W., & Tesconi, M. (2019). *RTbust: Exploiting Temporal Patterns for Botnet Detection on Twitter*. http://arxiv.org/abs/1902.04506

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harward Business Review*, *90*(10), 60–66. https://hbr.org/2012/10/big-data-the-management-revolution

McDaniel, B. T., Coyne, S. M., & Holmes, E. K. (2012). New Mothers and Media Use: Associations Between Blogging, Social Networking, and Maternal Well-Being. *Maternal and Child Health*

Journal, *16*(7), 1509–1517. https://doi.org/10.1007/s10995-011-0918-2

McLaren, N., & Shanbhogue, R. (2011). *Using Internet Search Data as Economic Indicators*. Bank of England. http://www.bankofengland.co.uk/publications/Documents/quarterlybulletin/qb110206.pdf

McNeely, C. L., & Hahm, J. (2014). The Big (Data) Bang: Policy, Prospects, and Challenges. *Review of Policy Research*, *31*(4), 304–310. https://doi.org/10.1111/ropr.12082

McNeill, A., Harris, P. R., & Briggs, P. (2016). Twitter Influence on UK Vaccination and Antiviral Uptake during the 2009 H1N1 Pandemic. *Frontiers in Public Health*, *4*, 22. https://doi.org/10.3389/fpubh.2016.00026

Mergel, I., & Bretschneider, S. I. (2013). A Three-Stage Adoption Process for Social Media Use in Government. *Public Administration Review*, *73*(3), 390–400. https://doi.org/10.1111/puar.12021

Mertens, D. (2017). The “New Welfare State” Under Fiscal Strain: Austerity Gridlocks and the Privatization of Risk. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 77–88). Oxford University Press.

Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & Mccallum, A. (2011). Optimizing Semantic Coherence in Topic Models. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272. https://doi.org/10.5555/2145432.2145462

Misuraca, G., Mureddu, F., & Osimo, D. (2014). Policy-Making 2.0: Unleashing the Power of Big Data for Public Governance. In M. Gascó-Hernández (Ed.), *Open Government* (pp. 171–188). Springer New York. https://doi.org/10.1007/978-1-4614-9563-5_11

Mohammed, N., Chen, R., Fung, B. C. M., & Yu, P. S. (2011). Differentially private data release for data mining. *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’11*, 493. https://doi.org/10.1145/2020408.2020487

Morel, N., Palier, B., & Palme, J. (Eds.). (2012). *Towards a Social Investment Welfare State?* Birstol Policy Press.

Morel, N., & Palme, J. (2017). A Normative Foundation for the Social Investment Approach? In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 150–157). Oxford University Press.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from Twitter’s streaming API with Twitter’s firehose. *Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013*, 400–408.

Naccarato, A., Falorsi, S., Loriga, S., & Pierini, A. (2018). Combining official and Google Trends data to forecast the Italian youth unemployment rate. *Technological Forecasting and Social Change*, *130*, 114–122. https://doi.org/10.1016/J.TECHFORE.2017.11.022

Nakamura, R. T. (1987). the Textbook Policy Process and Implementation Research. In *Review of Policy Research* (Vol. 7, Issue 1, pp. 142–154). https://doi.org/10.1111/j.1541-1338.1987.tb00034.x

Nelkin, D. (1975). The Political Impact of Technical Expertise. *Social Studies of Science*, *5*. http://journals.sagepub.com/doi/pdf/10.1177/030631277500500103

Nelson, M., & Stephens, J. (2011). Do social investment policies produce more and better jobs? In N.

Morel, B. Palier, & J. Palme (Eds.), *Towards a Social Investment Welfare State?: Ideas, Policies and Challenges*. Bristol Policy Press.

Nielsen, J. A., & Pedersen, K. (2014). IT portfolio decision-making in local governments: Rationality, politics, intuition and coincidences. *Government Information Quarterly*, *31*(3), 411–420. https://doi.org/10.1016/J.GIQ.2014.04.002

Nikolai, R. (2011). Towards social investment? Patterns of public policy in the OECD world. In N. Morel, B. Palier, & J. Palme (Eds.), *Towards A Social Investment Welfare State?: Ideas, Policies and Challenges* (pp. 91–116). The Policy Press.

Nimmo, B. (2019). *Measuring Traffic Manipulation on Twitter*. http://archive.is/G1Hjj.

Nolan, B. (2013). What use is "social investment"? *Journal of European Social Policy*, *23*(5), 459–468. https://doi.org/10.1177/0958928713499177

O'Leary, C., Baines, S., Bailey, G., McNeil, T., Csoba, J., & Sipos, F. (2018). Innovation and Social Investment Programs in Europe. *European Policy Analysis*, *4*(2), 294–312. https://doi.org/10.1002/epa2.1052

OECD. (2011). *Doing Better for Families*. OECD Publishing. https://www.oecd.org/social/soc/doingbetterforfamilies.htm

OECD. (2015). *In It Together - Why Less Inequality Benefits All*. OECD Publishing. https://doi.org/10.1787/9789264235120-en

Orloff, A. (2010). Gender. In F. Castles, S. Leibfried, J. Lewis, H. Obinger, & C. Pierson (Eds.), *The Oxford Handbook of the Welfare State* (pp. 252–264). Oxford University Press.

Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, *35*(1), 136–147. https://doi.org/10.1016/j.tele.2017.10.006

Panagiotopoulos, P., Bowen, F., & Brooker, P. (2017). The value of social media data: Integrating crowd capabilities in evidence-based policy. *Government Information Quarterly*, *34*(4), 601–612. https://doi.org/10.1016/J.GIQ.2017.10.009

Park, C. S. (2013a). Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2013.01.044

Park, C. S. (2013b). Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Computers in Human Behavior*, *29*(4), 1641–1648. https://doi.org/10.1016/j.chb.2013.01.044

Park, C. S., & Kaye, B. K. (2017). The tweet goes on: Interconnection of Twitter opinion leadership, network size, and civic engagement. *Computers in Human Behavior*, *69*, 174–180. https://doi.org/10.1016/j.chb.2016.12.021

Parolin, Z., & Van Lancker, W. (2021). What a social investment 'litmus test' must address: A response to Plavgo and Hemerijck. *Journal of European Social Policy*, *31*(3), 297–308. https://doi.org/10.1177/09589287211012974

Pavolini, E., & Van Lancker, W. (2018). The Matthew effect in childcare use: a matter of policies or preferences? *Journal of European Public Policy*, *25*(6), 878–893.

https://doi.org/10.1080/13501763.2017.1401108

Peled, A. (2014). *Traversing Digital babel: Information, E-Government, and Exchange*. The MIT Press.

Peng, I. (2014). The Social Protection Floor and the 'New' social investment policies in Japan and South Korea. *Global Social Policy: An Interdisciplinary Journal of Public Policy and Social Development*, *14*(3), 389–405. https://doi.org/10.1177/1468018114533711

Pfeffer, J., Mayer, K., & Morstatter, F. (2018). Tampering with Twitter's Sample API. *EPJ Data Science*, *7*(1). https://doi.org/10.1140/epjds/s13688-018-0178-0

Pierson, P. (1994). *Dismantling the Welfare State? Reagan, Thatcher, and the Politics of Retrenchment*. Cambridge University Press.

Plavgo, I., & Hemerijck, A. (2021). The social investment litmus test: Family formation, employment and poverty. *Journal of European Social Policy*, *31*(3), 282–296. https://doi.org/10.1177/0958928720950627

Poškutė, V. (2018). Are the concepts of 'social investment' and 'long-term care for the elderly' compatible? Perceptions of long-term care stakeholders in Lithuania. *Journal of International and Comparative Social Policy*, *34*(2), 151–167. https://doi.org/10.1080/21699763.2018.1479652

Proserpio, D., Counts, S., & Jain, A. (2016). The psychology of job loss: using social media data to characterize and predict unemployment. *Proceedings of the 8th ACM Conference on Web Science*, 223–232. https://doi.org/10.1145/2908131.2913008

Prpić, J., Taeihagh, A., & Melton, J. (2015). The Fundamentals of Policy Crowdsourcing. *Policy & Internet*, *7*(3), 340–361. https://doi.org/10.1002/poi3.102

Radziwill, N. M., & Benton, M. C. (2016). *Bot or Not? Deciphering Time Maps for Tweet Interarrivals*. http://arxiv.org/abs/1605.06555

Roberts, M., Stewart, B., & Airoldi, E. (2016). A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association*, *111*(515), 988–1003. https://doi.org/10.1080/01621459.2016.1141684

Roberts, M., Stewart, B., & Tingley, D. (2016). Navigating the Local Modes of Big Data: The Case of Topic Models. In M. Alvarez (Ed.), *Computational Social Science: Discovery and Prediction*. Cambridge University Press.

Roberts, M., Stewart, B., & Tingley, D. (2019). Journal of Statistical Software stm: R Package for Structural Topic Models. *Journal of Statistical Software*, *91*(2). https://doi.org/10.18637/jss.v000.i00

Roberts, M., Stewart, B., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Robertson, H., & Travaglia, J. (2015). *A Politics of Counting – Putting People Back into Big Data*. Discover Society. http://discoversociety.org/2015/07/30/a-politics-of-counting-putting-people-back-into-big-data/

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining -*

WSDM '15, 399–408. https://doi.org/10.1145/2684822.2685324

Ronchi, S. (2018). Which Roads (if any) to Social Investment? The Recalibration of EU Welfare States at the Crisis Crossroads (2000-2014). *Journal of Social Policy*, *47*(3), 459–478. https://doi.org/10.1017/S0047279417000782

Ruppert, E., Isin, E., & Bigo, D. (2017). Data politics. *Big Data & Society*, *4*(2), 205395171771774. https://doi.org/10.1177/2053951717717749

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), 1063–1064. https://doi.org/10.1371/journal.pmed.0020124

Ryu, P.-M. (2018). Predicting the Unemployment Rate Using Social Media Analysis. *Journal of Information Processing Systems*, *14*(4), 904–915.

Sabel, C. F., & Zeitlin, J. (2012). Experimentalist Governance. In D. Levi-Faur (Ed.), *The Oxford Handbook of Governance*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199560530.013.0012

Sabel, C., Zeitlin, J., & Quack, S. (2017). Capacitating Services and the Bottom-Up Approach to Social Investment. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 140–149). Oxford University Press.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47. https://doi.org/10.1109/CTS.2013.6567202

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Pricneton Unviersity Press.

Samarajiva, R., & Lokanathan, S. (2016). *Using Behavioral Big Data for Public Purposes: Exploring Frontier Issues of an Emerging Policy Arena*. http://lirneasia.net/wp-content/uploads/2013/09/NVF-LIRNEasia-report-v8-160201.pdf

Sandberg, J., & Nelson, M. (2017). Social Investment in Latin America. In A. Hemerijck (Ed.), *The Uses of Social Investment*. Oxford University Press.

Saraceno, C. (2015). A Critical Look to the Social Investment Approach from a Gender Perspective. *Social Politics: International Studies in Gender, State & Society*, *22*(2), 257–269. https://doi.org/10.1093/sp/jxv008

Scannapieco, M., Virgillito, A., & Zardetto, D. (2012). *Placing Big Data in Official Statistics: A Big Challenge?*

Schmid, G. (2008). *Full Employment In Europe: Managing Labour Market Transitions and Risks*. Edward Elgar.

Schmidt, V. A. (2002). Does discourse matter in the politics of welfare state adjustment? *Comparative Political Studies*, *35*(2), 168–193. https://doi.org/10.1177/0010414002035002002

Scott, G. (2018). *From output budgeting to social investment : reflections on thirty five years of evolution in public sector management in New Zealand*.

Sen, A. (2001). *Development as Freedom*. Oxford University Press.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70.

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. *PLoS ONE*, *6*(5), e19467. https://doi.org/10.1371/journal.pone.0019467

Simonsen, M. (2010). Price of High-quality Daycare and Female Employment. *Scand. J. of Economics*, *112*(3), 570–594. https://doi.org/10.1111/j.1467-9442.2010.01617.x

Singh, P., Dwivedi, Y. K., Karanjeet, &, Kahlon, S., Ravinder, &, Sawhney, S., Alalwan, A. A., & Rana, N. P. (2020). Smart Monitoring and Controlling of Government Policies Using Social Media and Cloud Computing. *Information Systems Frontiers*, *22*, 315–337. https://doi.org/10.1007/s10796-019-09916-y

Soentken, M., van Hooren, F., & Rice, D. (2017). The Impact of Social Investment Reforms on Income and Activation in the Netherlands. In A. Hemerijck (Ed.), *The Uses of Social Investment* (pp. 235–243). Oxford University Press.

Staines, Z., Moore, C., Marston, G., & Humpage, L. (2021). Big data and poverty governance under Australia and Aotearoa/New Zealand's "social investment" policies. *Australian Journal of Social Issues*, *56*(2), 157–172. https://doi.org/10.1002/ajs4.129

Statista. (2018). *Netherlands: Twitter users, by age group 2017-2018*. https://www.statista.com/statistics/828876/Twitter-penetration-rate-in-the-netherlands-by-age-group/

Statista. (2019). *Netherlands: number of Twitter users 2013-2019*. https://www.statista.com/statistics/880865/number-of-Twitter-users-in-the-netherlands/

Stevenson, B. (2008). The Internet and Job Search. *National Bureau of Economic Research Working Paper Series*, *13886*. http://www.nber.org/papers/w13886

Streeck, W., & Mertens, D. (2011). Fiscal Austerity and Public Investment: Is the Possible the Enemy of the Necessary? *PMlfG Discussion Paper 11/12*, 33. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1894657

Streeck, W., & Thelen, K. (Eds.). (2005). *Beyond continuity: Institutional change in advanced political economies*. Oxford University Press.

Sutherland, I. (2013). John Graunt : A Tercentenary Tribute. *Journal of the Royal Statistical Society*, *126*(4), 537–556.

Sweeney, L. (2002). K-anonymity: A model for protecting privacy. *International Journal on Uncertainty*, *10*(5), 557–570. http://www.cs.pomona.edu/~sara/classes/cs190-fall12/k-anonymity.pdf

Szomszor, M., Kostkova, P., & De Quincey, E. (2011). #Swineflu: Twitter predicts swine flu outbreak in 2009. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, *69 LNICST*, 18–26. https://doi.org/10.1007/978-3-642-23635-8_3

Taddy, M. A. (2012). On Estimation and Selection for Topic Models. *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, 1184–1193. http://proceedings.mlr.press/v22/taddy12.html

Taylor, L., Schroeder, R., & Meyer, E. (2014). *Emerging practices and perspectives on Big Data analysis in economics: Bigger and better or more of the same? 1*(10), 1–10.

https://doi.org/10.1177/2053951714536877

Terpstra, T., de Vries, A., Stronkman, R., & Paradies, G. L. (2012). Towards a realtime Twitter analysis during crises for operational crisis management. *Proceedings of the 9th International ISCRAM Conference*. http://www.nrc.nl/nieuws/2011/08/18/pukkelpop-stilgelegd-na-noodweer/

Tonneau, M., Adjodah, D., Palotti, J., Grinberg, N., & Fraiberger, S. (2022). Multilingual Detection of Personal Employment Status on Twitter. In *arxiv preprint*. https://arxiv.org/abs/2203.09178

Toole, J. L., Lin, Y.-R., Muehlegger, E., Shoag, D., González, M. C., & Lazer, D. (2015). Tracking employment shocks using mobile phone data. *Journal of Royal Society Interface*, *12*. https://doi.org/10.1098/rsif.2015.0185

Turrell, A., Speigner, B. J., Djumalieva, J., Copple, D., & Thurgood, J. (2019). Transforming Naturally Occurring Text Data Into Economic Statistics: The Case of Online Job Vacancy Postings. *NBER Working Paper Series*, *25837*. http://www.nber.org/papers/w25837

Twitter. (n.d.). *Twitter API: Academic Research Access*. Retrieved April 14, 2022, from https://developer.Twitter.com/en/products/Twitter-api/academic-research

United Nations. (2011). *Unemployment Through the Lens of Social Media*. http://www.unglobalpulse.org/projects/can-social-media-mining-add-depth-unemployment-statistics

University of Michigan. (2015). *Prediction of Initial Claims for Unemployment Insurance*. http://econprediction.eecs.umich.edu/

Uprichard, E. (2015). *Most big data is social data – the analytics need serious interrogation*. LSE Impact Blog. http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/12/philosophy-of-data-science-emma-uprichard/

Vaccari, C. (2015). *Report: Twitter data*. UNECE Sandbox Big Data Projects 2015. http://www1.unece.org/stat/platform/display/bigdata/Report%3A+Twitter+data

Van Den Bosch, A., Busser, B., Canisius, S., & Daelemans, W. (2007). An efficient memory-based morphosyntactic tagger and parser for Dutch. In F. van Eynde, P. Dirix, I. Schuurman, & V. Vandeghinste (Eds.), *Selected Papers of the 17th Computational Linguistics in the Netherlands Meeting* (pp. 99–114).

van der Voort, H., Klievink, B., Arnaboldi, M., & Meijer, A. (2018). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly*. https://doi.org/10.1016/J.GIQ.2018.10.011

Vandenbroucke, F, Hemerijck, A., & Palier, B. (2011). The EU Needs a Social Investment Pact. In *OSE Paper Series* (Issue 5). https://pure.uva.nl/ws/files/2781101/179011_OpinionPaper5_Vandenbroucke_Hemerijk_Palier_2011.pdf

Vandenbroucke, Frank, & Vleminckx, K. (2011). Disappointing poverty trends: is the social investment state to blame? *Journal of European Social Policy*, *21*(5), 450–471. https://doi.org/10.1177/0958928709XXXXXX

Vanhercke, B. (2013). Under the radar? EU social policy in times of austerity. In D. Natali & B. Vanhercke (Eds.), *Social developments in the European Union 2012* (pp. 91–112). ETUI. http://www.ose.be/files/bilan2012/Bilan_social_2012_EN.pdf#page=92

Verbist, G. (2017). Measuring Social Investment Returns: Do Publicly Provided Services Enhance Social Inclusion? In A. Hemerijck (Ed.), *The Uses of Social Investment*. Oxford University Press.

Vicente, M., Lopez-Menendez, A., & Perez, R. (2015). Forecasting unemployment with internet search data: Does it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting and Social Change*, *92*, 132–139.

Vydra, S., & Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly*, *36*(4). https://doi.org/10.1016/j.giq.2019.05.010

Waldo, J. (2016). *Can Accuracy and Privacy Co-Exist?* Data for Policy. https://www.youtube.com/watch?v=Cf9F0i2ZHt8

Walker, R. (2000). Welfare policy: tendering for evidence. In S. Nutley, P. Smith, & H. Davies (Eds.), *What works? : evidence-based policy and practice in public services*. The Policy Press.

Wamba, S. F., Edwards, A., & Sharma, R. (2012). Big Data as a Strategic Enabler of Superior Emergency Service Management: Lessons from The New South Wales State Emergency Service. *Society for Information Management and MIS Quarterly Executive Pre ICIS Workshop*, 1–3. http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1153&context=buspapers

Wang, Y., Agichtein, E., & Benzi, M. (2012). TM-LDA: Efficient online modeling of latent topic transitions in social media. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 123–131. https://doi.org/10.1145/2339530.2339552

Ward, J. S., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions*. http://arxiv.org/abs/1309.5821

Watanabe, K. (n.d.). Latent Semantic Scaling: A Semisupervised Text Analysis Technique for New Domains and Languages. *Communication Methods and Measures.*

Watanabe, K. (2017a). The spread of the Kremlin's narratives by a western news agency during the Ukraine crisis. *Journal of International Communication*, *23*(1), 138–158. https://doi.org/10.1080/13216597.2017.1287750

Watanabe, K. (2017b). Measuring news bias: Russia's official news agency ITAR-TASS' coverage of the Ukraine crisis. *European Journal of Communication*, *32*(3), 224–241. https://doi.org/10.1177/0267323117695735

Wrenn, A. (2017). Social Investment and the Service Economy Trilemma. In A. Hemerijck (Ed.), *The Uses of Social Investment*. Oxford University Press.

Xiong, F., & Liu, Y. (2014). Opinion formation on social media: an empirical approach. *Chaos (Woodbury, N.Y.)*, *24*(1), 013130. https://doi.org/10.1063/1.4866011

Xiong, F., Liu, Y., & Cheng, J. (2017). Modeling and predicting opinion formation with trust propagation in online social networks. *Communications in Nonlinear Science and Numerical Simulation*, *44*, 513–524. https://doi.org/10.1016/j.cnsns.2016.09.015

Perspectives to Definition of Big Data: A Mapping Study and Discussion, 4 Journal of Innovation Management 69 (2016). https://doi.org/10.24840/2183-0606_004.001_0006

Zeitlin, J., & Vanhercke, B. (2018). Socializing the European Semester: EU social and economic policy co-ordination in crisis and beyond. *Journal of European Public Policy*, *25*(2), 149–174. https://doi.org/10.1080/13501763.2017.1363269

References

Zwitter, A. (2014). Big Data ethics. *Big Data & Society*, *1*(2), 1–6. https://doi.org/10.1177/2053951714559253

Authorship Contribution

This dissertation is based on four separate research papers, some of which are peer-reviewed and published and some of which remain as a working paper for the time being. These papers correspond to chapters 2, 3, 4 & 5 of the dissertation, with changes to those papers being concentrated mainly the introductory and concluding sections to improve clarity and avoid repetition. This makes the dissertation a combination of a paper-based model and a traditional monograph. I was the research lead on all four studies, but I also had the privilege of working with co-authors on three of the four studies. The contribution of my co-authors has been indispensable and will be described in more detail in this section for each paper.

Paper 1:

Vydra, S., & Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. *Government Information Quarterly*, 36(4), 101383.

This paper corresponds to chapter 2 of the dissertation and is co-authored by Bram Klievink. Bram and I have jointly developed the overall framing of the paper, with Bram being indispensable in steering the paper towards a contribution that is theoretically greater than the sum of the individual arguments it contains. Bram also wrote a part for the paper itself, which is part of the published paper and now a part of section 2.1 of this dissertation, as well as improved multiple versions of the draft, especially with regards to clarity of the central message. Remainder of the conceptual work as well as writing was carried out by me.

Changes to this paper are relatively minor and involve mainly cuts to its introduction and re-phrasing of its conclusions (to better tie into the dissertation).

Paper 2:

Vydra, S., & Hemerijck, A. (2022). Social Investment as Policymaking Paradigm. *Manuscript.*

This paper corresponds to chapter 3 of the dissertation and is co-authored by Anton Hemerijck. Anton and I have debated at length, and jointly developed, the central theoretical point of the paper with regards to role of big data analysis in policy assessment from a social investment perspective. Anton has been especially indispensable for discussing the contribution of even developing the paradigm in this way. Much of the groundwork this study builds on is a result of our previous and continued cooperation with respect to methodological development of social investment. The writing of the paper was carried out by me with multiple rounds of feedback on its theoretical aims and structure with Anton.

Changes to this paper are relatively minor and mainly happen in introduction to connect it to the previous chapter.

Paper 3:

Vydra, S. (2022). Does it hold up? Testing big data's promise of novel information on labour market policymaking. *Manuscript.*

This paper corresponds to chapter For chapter 4 and does not have a co-author. The design, writing, data collection, analysis, and interpretation were all carried out by me. That said, there are colleagues and supervisors who have provided important feedback and discussions, such as Scott Cunningham, Bram Klievink, or Jaroslaw Kantorowicz, who I am grateful to.

The changes to this paper are more substantial - in its stand-alone version the paper has an additional section that argues that social investment (as a school of thought) is both relevant for policymaking and in need of analytical development, making it a fitting testing ground for use of big data. Since that argument is now developed in more detail by the previous chapter, the section is missing and the paper is reworked to work without it and to tie into the dissertation.

Paper 4:

Vydra, S., & Kantorowicz, J. (2021). Tracing Policy-relevant Information in Social Media: The Case of Twitter before and during the COVID-19 Crisis. *Statistics, Politics and Policy*, 12(1), 87-127.

This paper corresponds to chapter 5 of the dissertation and is co-authored by Jaroslaw Kantorowicz. It is fair to say that Jaroslaw took the lead on outlining the contribution the paper can make and finding an outlet for publishing it. From there we have jointly developed the theoretical framework for the paper and Jaroslaw was indispensable in focusing the paper and polishing drafts of the paper. Jaroslaw also wrote a part for the paper itself, which is part of the published paper and a constituent part of section 5.1.5 of this dissertation, as well as produced figures 5.7 and 5.8 based on the selected results. The remainder of the writing, data collection, processing, analysis, and interpretation are my own work.

The changes to this paper are also somewhat substantial, as much of the justification for platform, policy area, or country are all covered in chapter 4 and, as a consequence, are not fully defended in this chapter to avoid repetition. Instead of outlining its stand-alone relevance the contribution is now framed in terms of expanding the conclusions from chapter 4, which is a substantial change to not just introducing and concluding the paper, but also how its focus and method are described.

Paper 5:

Vydra, S., Poama, A., Giest, S., Ingrams, A., & Klievink, B. (2021). Big Data Ethics: A Life Cycle Perspective. *Erasmus L. Rev*., 14, 24.

This paper does not correspond to any chapter in the dissertation and is not included in it in any other way. However, it is a research project conducted during my PhD trajectory and it is closely related to arguments made throughout the dissertation about taking ethics seriously when it comes to big data use. It is not included as a chapter in order to maximize the coherence of the dissertation, but working on it with my co-authors informs the overall dissertation and arguments made throughout. This paper is co-authored with Andrei Poama, Sarah Giest, Alex Ingrams, and Bram Klievink. All of my co-authors here have contributed an original piece of research in the form of a case study fitting into the ethical framework of the paper. This framework was jointly developed by Andrei Poama and myself, with Andrei taking the lead on the theoretical front. All authors were also crucial in providing feedback on the central arguments of the paper and improving the draft. The remainder of the writing, editorial work, and steering of the overall contribution are my own work.

Curriculum Vitae

PERSONAL INFORMATION

Simon Vydra

Vydra.simon@gmail.com

Sex Male | Date of birth 16/01/1992 | Nationality Czech

WORK EXPERIENCE

01/08/2022 - present

Head of Section for Legislative Analysis

Office of the Government of the Czech Republic – Cabinet of the Minister for Legislation – Prague, Czech Republic

- Assisting ministries with research and analysis to directly support their efforts of Regulatory Impact Assessment for upcoming legislative interventions.
- Building the Government's analytical unit in the sense of organization, processes, and personnel

01/03/2021-07/12/2021

Researcher

Leiden Universiteit – Institute of Public Administration – The Hague Centre for Digital Governance, Hague

- Teaching research methods (NLP and general research methods) as well as courses on digitization of the public administration
- Research primarily focused on utilization of NLP tools (by various US departments) in analysing publicly issued comments on proposed legislation.

01/01/2017-01/01/2021

EURO-CEFG PhD Fellow

Technische Universiteit Delft – Faculty of technology, policy and Management, Delft

Leiden Universiteit – Faculty of governance and global affairs, Hague

- PhD Fellowship under European Research Centre for Economic and Financial Governance
- Research on the interaction of big data with policymaking processes, focused (more theoretically) on the use of big data in governance and (more applied) on labour markets, social media data, and NLP.

01/01/2016-01/04/2016

Research assistant

Vrije Universiteit – Faculty of social science, Amsterdam,

Universiteit van Amsterdam – ACCESS Europe, Amsterdam,

- Editorial work on the "Uses of Social Investment" volume edited by Anton Hemerijck and published by Oxford University Press
- Research work on a methodological inquiry for European Commission focusing on measuring the returns of social policies in line with a social investment approach

01/02/2014-01/01/2015

Policy researcher

INstrategy - Institute for European and national strategies, Prague, www.instrategy.cz

- Involved in multitude of projects with thematic foci such as: Social innovation, decision architecture, pre-school education, teledermatology and others.
- Research and administrative responsibilities

EDUCATION AND TRAINING

01/09/2014 - 31/08/2015 **Master's degree in Political Science**
Vrije Universiteit , Amsterdam

▪ Specialization in Policy and Politics (focus on political science and public administration)
▪ Graduated Cum Laude

01/09/2011 - 31/06/2014 **Bachelor of Liberal Arts and Sciences**
Amsterdam University College , Amsterdam

▪ Major in Social Science (focus on political science)
▪ Graduated Cum Laude

ACADEMIC PUBLICATIONS

2021 Vydra, S., Kantorowicz, J. (2021). Tracing policy-relevant information in social media: The case of Twitter before and during the COVID-19 crisis. *Statistics, Politics and Policy.* 12(1), 87-127

2021 Vydra, S., Poama, A., Giest, S., Ingrams, A., & Klievink, B. (2021). Big Data Ethics: A Life Cycle Perspective. *Erasmus Law Review*., 14(24).

2019 Vydra, S., Klievink, B. (2019). Techno-optimism and policy-pessimism in the public sector big data debate. *Government information Quarterly. 36(14)*

2017 Burgoon, B., Hemerijck, A., De Pietro, A., Vydra, S. (2017). Assessing Social Investment Synergies. European Commission, Brussels

2016 Hemerijck, A., Vydra, S. (2016). Le champ d'analyse de la politique d'investissement social. *Informations Sociales,* 2016/1(192), 8-20

Appendix A

This appendix describes the keywords used to gather data from Twitter's API for chapters 3 and 4. There are essentially three versions of the dataset that these chapters refer to.

Firstly, it is the general **corpus**, which is used throughout the entirety of chapter 3 and for training of the LSS models in chapter 4. This corpus includes all gathered Tweets based on all keyword in tables A1 – A4, including words that are crossed out.

Secondly, there is the **ECEC sub-section**, referred to in chapter 4, which includes only tweets that contain at least one keyword from the 'ECEC' section below (Table A1) in the text of the tweet.

Thirdly, there is the **LM sub-section**, referred to in chapter 4, which includes only tweets that contain at least one keyword from the 'Labour Market' section (Tables A2 – A4), with the notable exception of all keywords that are crossed out (to focus more on unemployment as opposed to more general labour market policies including re-training or apprenticeships).

ECEC:

Table A1: All ECEC Keywords

'kinderopvang' 'kinder opvang'	General childcare term
'kinderdagverblijf' 'kdv'	Refers to daycare centers Cover children up to 4 years old and 10 hours a day during working hours. Apparently limited capacity
'gastouder' 'gastouders' 'gastouderopvang' 'gastouder opvang' 'gastouderbureau'	Available for toddlers up to pre-school by parents caring for up to 6 children in locations approved by the national childcare register. Often administered by agencies
'peuterspeelzalen' 'peuterspeelzaal' 'peuterspeelplaats'	Preschools, these are usually part of a primary school and are a preparatory program for children between two and four. These do not cover the whole week or full days.
'peutergroep' 'peutergroepen'	A more informal playgroup setting for young children
'buitenschoolseopvang' 'buitenschoolse opvang' 'naschoolseopvang' 'naschoolse opvang' 'naschoolse' 'BSO' 'voorschoolse opvang' 'voorschoolse' 'voorschoolseopvang'	Afterschool and outdoor school care. However, this is connected to primary schools and thus generally available to children from 4 years of age
'oppas' 'oppassers' 'babysitter' 'babysitters' 'nanny' 'nannies'	Babysitter options
'kinderopvangtoeslag'	Childcare subsidy in the Netherlands
'kindgebonden budget'	Automatic child benefit if your child is under 18 and your income is not high

'kinderbijslag'	Covers part of the cost of raising children and depends on their number and residence

Labour market:

Table A2: Legislation and programs

'Participatiewet' 'Participatie wet' 'Gesubsidieerde arbeid'	Overarching legislation in place to support people who can work but need some sort of assistance in order to work.
~~'opleiding' 'scholing' 'heropleiding' 'omscholing' 'training' 'retraining' 're-training' 'studie' 'studeren'~~	~~Training and re-training.~~
~~'praktijktraining' 'werkervaringsplek' 'stage' 'stage lopen'~~	~~On the job training~~
~~'werkervaringsplek' 'werkervaring plek' 'studeer en werkplek' 'studeer- en werkplek' 'traineeship'~~	~~apprenticeship~~
'Werkbedrijf' 'werk.nl' 'werkplein' 'werkpleinen' 'arbeidsadviseur' 'uwv' 'arbeidsbemiddelaar' 'arbeidsbemiddeling' 'loopbaan coach' 'werk coach'	Employment services
'WW-uitkering' 'uitkering' 'bijstand' 'bijstandsuitkering'	Unemployment (benefits)
~~'meewerkaftrek'~~	~~Subsidy for when your partner works in your business without pay~~

Table A2: Type of employment

'full-time werk' 'full time werk' 'fulltime werk' 'full-time baan' 'full time baan' 'fulltime baan' 'voltijd baan' 'voltijd werk' 'voltijdwerk' '1 fte' '1 wtf'	Full time work
'deeltijd werk' 'part-time werk' 'part time werk' 'deeltijd baan' 'part-time baan' 'part time baan'	Part time work
'vast contract' 'vaste baan' 'vaste aanstelling'	Permanent contract
'tijdelijk contract' 'tijdelijke baan' 'tijdelijke aanstelling'	Temporary contract

‘uitzendcontract’	Contract with recruitment agency
‘nul uren contract’ ‘0 uren contract’	Zero hour contract
‘zelfstandige zonder personeel’ ‘zzp’ ‘zzp’ers’ ‘zzp’er’ ‘zzper’ ‘zzpers’ ‘DBA modelovereenkomst’	freelancers
‘schijnzelfstandigheid’	Sham independence
‘loondienst’ ‘in loondienst’	Salaried employment
‘eigen baas’ ‘eigen baas zijn’	Self-employment

Table A4: Generic employment-related phrases

‘werkloosheid’ ‘werkeloosheid’ ‘werkloos’ ‘zonder baan’ ‘jobless’ ‘in between jobs’ ‘between jobs’ ‘in between two jobs’ ‘between two jobs’	Unemployment
‘onderbezetting’ ‘onderbezet’	Underemployment
‘zoek naar werk’ ‘kijken voor werk’ ‘een baan zoeken’ ‘zoeken naar een baan’ ‘banen zoeken’	Job search
‘passend werk’ ‘passende arbeid’ ‘passende baan’ ‘passende job’	Correct or fitting job
‘goed werk’ ‘slecht werk’ ‘beter werk’ ‘betere kansen op werk’ ‘beter arbeidscontract’ ‘goed arbeidscontract’ ‘slecht arbeidscontract’	A good job
‘vacature’ ‘vacatures’ ‘openstaande baan’	Job vacancies
‘vaardigheidseisen’ ‘ervaringseisen’ ‘werkervaring’ ‘werkervaringseisen’ ‘competenties’	Skill/experience requirements

Appendix B

This appendix presents topic summaries referred to in Chapter 4. Each table (B1-6) represents a topic (label assigned to each topic is in the title of each table) and lists all models that feature that topic (or some variation of it). The appendix provides 15 top-words for each topic in each model.

Table B1: Employment openings topic

LDA 20 - topic 5		LDA 35 - topic 6		LDA 35 - topic 23		LDA 65 - topic 60		CTM 35 - topic 17		NMF 35 – topic 27	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
vacature	job offer	jij	you	vacature	job offer	jij	you	jij	you	wij	we
jij	you	wij	we	#	#	wij	we	wij	we	op_zoek	looking
zoeken	search	op_zoek	looking	nieuw	new	op_zoek	looking	op_zoek	looking	team	team
#	#	jou	you	zoeken	to search	jou	you	zoeken	to search	enthousiast	enthusiastic
wij	we	zoeken	to search	medewerker	employee	collega	colleague	vacature	job offer	!	!
op_zoek	looking	willen	want	baan	job	technisch	technical	collega	colleague	collega	colleague
nieuw	new	leuk	fun	amsterdam	Amsterdam	limburg	limburg	team	team	opdrachtgever	client
jou	you	collega	colleague	amp	amp	arnhem	arnhem	we	we	kijken	look
medewerker	employee	team	team	via	through	checken	check	jou	you	graag	gladly
bekijken	see	we	we	utrecht	utrecht	uitdaging	challenge	reageren	comment	direct	straight away
willen	want	werken	to work	werk	work	senior	senior	komen	come	snel	fast
collega	colleague	snel	fast	manager	manager	!	!	snel	fast	leuk	fun
amp	amp	graag	gladly	bekijken	see	assistent	assistant	communicatie	communication	informatie	information
werken	to work	komen	come	v	v	tijdelijk	temporarily	enthousiast	enthusiastic	bekijken	see
team	team	functie	position	m	m	ervaren	to experience	bekijken	to look	interesse	interest

Table B2: Training and education

LDA 20 - topic 18		LDA 35- topic 21		LDA 65 - topic 18		CTM 35 - topic 4		NMF 35 - topic 13	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
training	training	opleiding	education	opleiding	education	@	@	opleiding	education
opleiding	education	leren	to learn	staan	stand	studie	study	volgen	to follow
#	#	onderwijs	teaching	leren	to learn	opleiding	education	starten	start
1	1	student	student	vaak	often	zullen	will	leren	to learn
jaar	year	volgen	to follow	halen	fetch	wel	well	student	student
vandaag	today	school	school	professional	professional	moeten	should	maken	to make
nieuw	new	groep	group	brengen	bring	gaan	to go	vandaag	today
2	2	tijdens	while	niveau	level	studeren	to study	2019	2019
2019	2019	jong	young	soort	kind	denken	to think	geven	to give
dag	day	starten	start	helaas	unfortunately	maken	to make	hbo	HBO education
week	week	leerling	pupil	geleden	ago	jaar	year	tijdens	while
3	3	docent	teacher	vorig	last	ander	other	onderwijs	teaching
volgen	to follow	geven	to give	elkaar	each other	weten	know	mbo	MBO education
stage	internship	kind	child	waarin	in which	eigen	own	dag	day
amp	amp	kennis	knowledge	mnd	month	zeggen	say	mooi	beautiful

Table B3: Unemployment and social assistance

LDA 20 - topic 20		LDA 35 - topic 20		LDA 65 - topic 9		CTM 35 - topic 11		NMF 35 - topic 2	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
@	@	uitkering	payment	uwv	unemployment insurance agency	@	@	@	@
uitkering	payment	uwv	unemployment insurance agency	bedrijf	company	uitkering	payment	uitkering	payment
|	|	bijstand	assistance	gemeente	municipality	uwv	unemployment insurance agency	uwv	unemployment insurance agency
uwv	unemployment insurance agency	@	@	(	(	mens	person	krijgen	to get
krijgen	to get	krijgen	to get	werkgever	employer	krijgen	to get	zullen	will
mens	person	betalen	To pay	bieden	offer	bijstand	assistance	eigen	own
jaar	year	geld	money	mogelijkheid	possibility	betalen	To pay	bijstand	assistance
moeten	should	via	through	kwaliteit	quality	moeten	should	gewoon	just
zzp	self-employed	recht	right	oplossing	solution	jaar	year	ander	other
bijstand	assistance	moeten	should	ruim	spacious	werken	to work	zitten	to sit
via	via	euro	euro	cursus	class	geld	money	kind	child
betalen	to pay	land	country	enorm	huge	via	through	nl	NL
werken	to work	huis	House	branche	industry	land	country	betalen	To pay
eigen	own	telegraaf	News outlet	bestuur	governance	nederland	The Netherlands	alleen	only
nederland	The Netherlands	kind	child	lokaal	local	gaan	to go	denken	to think

Appendix B

Table B4: Hours per week worked

LDA 20 - topic 16		LDA 35 - topic 27		LDA 65 - topic 21		CTM 35 - topic 29		NMF 35 - topic 23	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
uur	hour	uur	hour	week	week	uur	hour	uur	hour
zorg	care	1	1	per	per	week	week	per	per
per	per	per	per	uur	hour	per	per	week	week
vacature	job offer	week	week	€	€	maand	month	1	1
week	week	2	2	maand	month	hbo	HBO education	32	32
nijmegen	Nijmegen	2019	2019	10	10	1	1	dag	day
chauffeur	driver	3	3	aantal	number	2	2	maand	month
24	24	4	4	euro	euro	3	3	24	24
40	40	2018	2018	20	20	mbo	MBO education	2	2
welzijn	wellbeing	januari	January	extra	additional	4	4	36	36
32	32	5	5	16	16	40	40	aantal	number
verpleegkundig	nursing	maand	month	12	12	plaats	place	40	40
c	c	jaar	year	30	30	24	24	functie	position
36	36	€	€	lid	member	20	20	€	€
begeleider	mentor	plaats	place	organiseren	to organize	vacature	job offer	3	3

Table B5: Self-employment

LDA 20 topic 13		LDA 35 - topic 7		LDA 65 - topic 62		CTM 35 - topic 1		CTM 35 - topic 15		NMF 35 - topic 17	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
zzp	self-employed	zzp	self-employed	zzp	self-employed	|	|	vacature	job offer	zzp	self-employed
#	#	opdracht	order/contract	via	via	zzp	self-employed	#	#	freelance	freelance
opdracht	order/contract	interim	interim	opdracht	order/contract	'	'	amsterdam	Amsterdam	opdracht	order/contract
interim	interim	freelance	freelance	interim	interim	ondernemer	entrepreneur	zzp	self-employed	interim	interim
freelance	freelance	ondernemer	entrepreneur	freelance	freelance	verplichten	oblige	opdracht	order/contract	info	info
gen	gen	#	#	slag	battle	economie	economy	freelance	freelance	ondernemer	entrepreneur
vacature	job offer	info	info	sturen	send	zzp-er	self-employed person	interim	interim	lezen	read
parttime	part-time	mkb	SMEs	specialist	specialist	willen	want	info	info	maken	to make
€	€	lezen	read	manier	way	werknemer	employee	rotterdam	Rotterdam	mkb	SMEs
lezen	read	freelancer	freelancer	.	.	pensioen	retirement	project	project	loggen	logging
info	info	professional	professional	verzorgen	take care of	mkb	SMEs	locatie	Location	planetinterim	planetinterim
postbezorger	mail deliverer	ondernemen	to undertake	half	half	loondienst	salaried service	werk	work	project	project
nieuw	new	tip	tip	volkskrant	Volkskrant	zzper	zzper	zwolle	Zwolle	professional	professional
planetinterim	planetinterim	verplichten	oblige	veilig	safe	<emoji>	<emoji>	senior	senior	€	€
freelancer	freelancer	blog	blog	basis	base	petitie	petition	adviseur	adviser	verplichten	oblige

Appendix B

Table B6: Childcare

LDA 20 - topic 14		LDA 35 - topic 14		LDA 35 - topic 17		LDA 65 - topic 65		LDA 65 - topic 59		CTM 35 - topic 30	
Dutch	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**	**Dutch**	**English**
kinderopvang	childcare	kinderopvang	childcare	kinderdagverblijf	daycare	kind	child	sector	sector	kind	child
the	the	the	the	agent	agent	kinderopvang	childcare	kdv	daycare	kinderopvang	childcare
thanks	thanks	to	to	oss	oss	onderwijs	Education	verwachten	expect	ouder	older
to	to	thanks	thanks	lelystad	lelystad	school	school	kids	kids	media	media
latest	latest	latest	latest	kunst	art	bso	out-of-school care	onzin	nonsense	bso	out-of-school care
pedagogisch	pedagogical	pedagogisch	pedagogical	peuterspeelzaal	kindergarten	passen	to fit	tonen	show	kinderdagverblijf	daycare
aantal	number	<emoji>	<emoji>	den_helder	den Helder	leerling	pupil	effect	effect	pedagogisch	pedagogical
dienstverband	employment	sollicitatie	job application	hogeschool	University	gelukkig	happy	lachen	laugh	open	Open
branche	industry	bso	out-of-school care	vdab	vdab	leerkracht	teacher	rtlnieuws	rtlnews	social	social
status	status	branche	industry	sa	sa	gebruik	use	meestal	mostly	kdv	daycare
medewerker	employee	dienstverband	employment	basisonderwijs	primary education	vier	four	behandelen	to treat	aantal	number
sollicitatie	job application	open	Open	psychiater	psychiatrist	samenwerking	cooperation	nauwelijks	barely	bedrijf	company
open	open	status	status	ko	ko	betreffen	concern	oor	ear	opvang	day care
bso	out-of-school care	bedrijf	company	lg	lg	durven	to dare	collectief	collective	1	1
bedrijf	company	kdv	daycare	lelystad	Lelystad	wo	Wed	verband	bandage	branche	industry

Appendix C

In this appendix we offer a matter-of-fact summary of selected topic models for both corpus sub-sections presented in Chapter 5. Section C.1 focuses on early childhood education and care corpus sub-section and section C.2 focuses on the labour market sub-section. We provide summaries of the entire model (in English and Dutch) and a topic-by-topic description of all potentially policy-relevant and interpretable topics including insight from manual inspection of the topics and LSS dimension scores for individual tweets.

C.1 ECEC Corpus

For the ECEC sub-section of the corpus, the most interpretable model is the 20-topic model, delivering similar topics to the 25 and 30-topic models but with more interpretability and more ‘focused’ topics. All topics of this model (described by top 10 tokens) are summarized in figures C1 (English translation) and C2 (Dutch original). For figures C1-4 we sort top tokens by probability to appear for a topic, which is inferred directly from the topic-word distribution. There are alternative metrics to sort top words by, but in this case we find this to represent the topics most accurately.

Figure C1: Topics in a 20-topic model of ECEC sub-section (English translation)

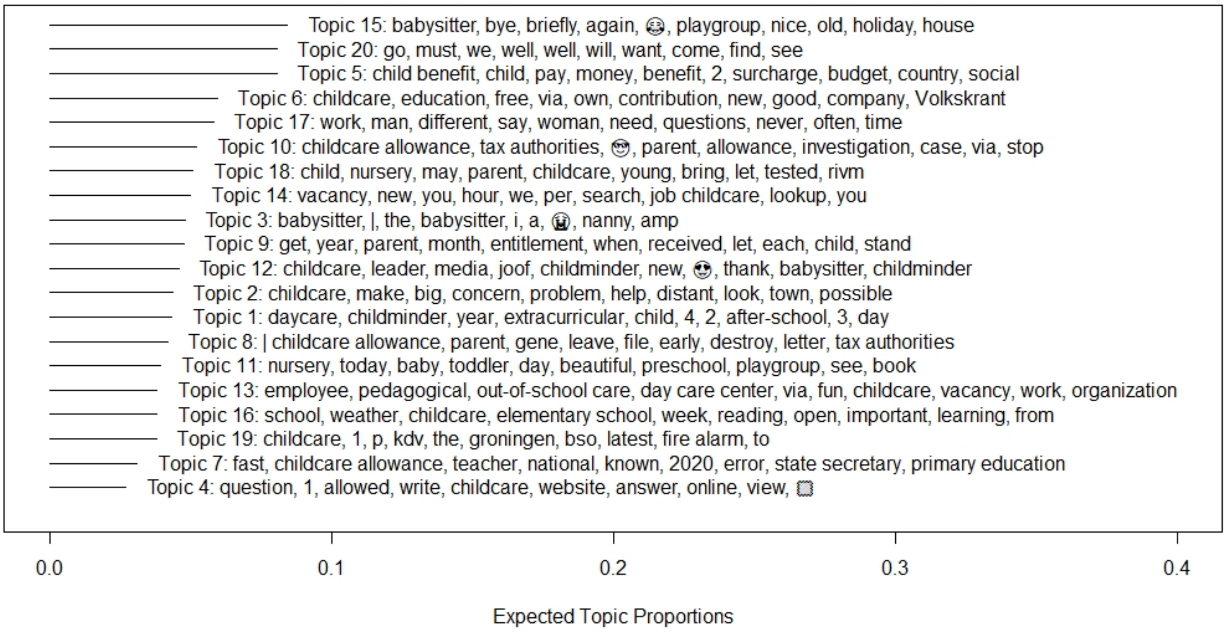
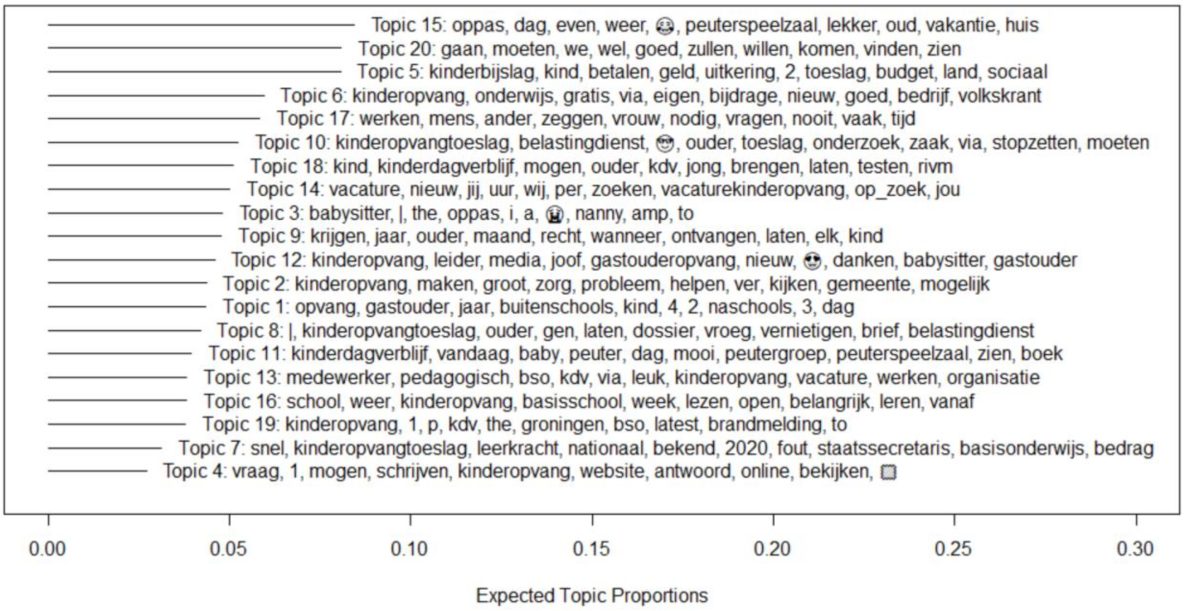


Figure C2: Topics in a 20-topic model of ECEC sub-section (Dutch original)



The interpretable topics in this case are the following: **Topic 5** is defined by tokens like 'child benefit', 'benefit' (toeslag), 'payment', 'to pay', 'money', 'budget', or 'country' and generally focuses on who receives what benefits and who is paying for them. This makes the topic not very exclusive to childcare and also concerned with political issues like the support migrants receive. The sentiment and success LSS dimensions are both negative, but mainly reveal a bias: The models score the token 'kinderbijslag' (child benefit) itself negatively, resulting in many short tweets that mention it as being highly negative and related to failure. This results in neither dimension validating.

Topic 17 is defined by tokens like 'to work', 'people', 'women', 'time', 'necessary', 'often', or 'never' and generally focuses on issues related to women's role in the labour market and (unpaid) domestic work. It includes commentary on the choice between paid or unpaid labour and family life both from the perspective of women stating and justifying their choices and from a perspective of more general commentary. For the sentiment dimension the mean polarity is negative (-0.027) and statistically significant. The dimension validates, with negative polarity corresponding to people arguing and insulting one another and positive polarity corresponding to people looking for babysitters or commenting on their choices with regards to child rearing and employment in neutral or positive terms. The success/failure dimension also validates and is negative (mean -0.027), with the failure polarity remaining 'negative' but focused more on failure of policy or individual providers rather than insults.

Topic 18 is defined by tokens like 'child', 'parent', 'to bring', 'childcare' (both as full word and 'kdv'), 'to test', or 'rivm' (National Institute for Health and Environment). In top 20 tokens words like 'sick' or 'corona' also appear. The topic generally focuses on health in childcare, with a detectable focus on COVID-19 and vaccinations. Only the success/failure dimension validates here - mean polarity on this dimension is negative (-0.024) and significant, with the negative polarity being about necessary policy adjustments and various failures of policies, including some general negative commentary.

Topic 11 is defined by tokens like ‘childcare’, ‘playgroup’ (multiple tokens), ‘today’, ‘day’, ‘nice’, or ‘to see’ and it is a somewhat general topic about childcare and playgroups that contains personal commentary on playgroups or preschools, pointing to newspaper articles about preschools, or preschools advertising themselves. The sentiment dimension validates but shows the generality of the topic as some positive tweets are only tangentially relevant. The mean polarity of the sentiment dimension is slightly positive (0.012) and statistically significant. The success/failure dimension also validates well with the failure polarity focusing on failure, insufficiency, and general hazard and the success polarity focusing on success. The mean polarity is slightly towards success (0.016).

C.2 LM Corpus sub-section

For the Labour Market sub-section, the most interpretable model is the 30-topic model. Similar to the ECEC sub-section, this model contains generally the same interpretable topics, but seemingly includes less noise for those topics than other candidate models. All topics of this model (described by top 10 tokens) are summarized in figures C3 (English translation) and C4 (Dutch original)

Figure C3: Topics in a 30-topic model of LM sub-section (English translation)

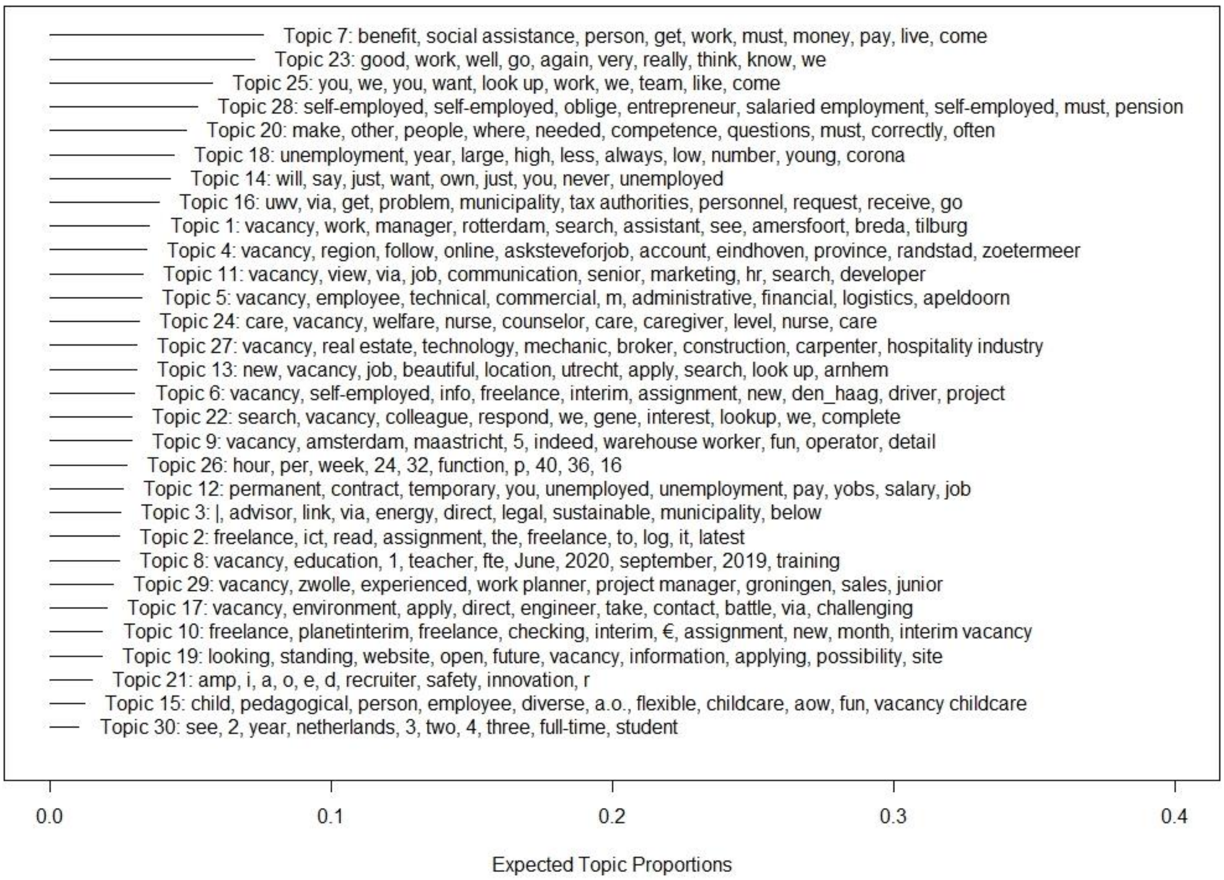
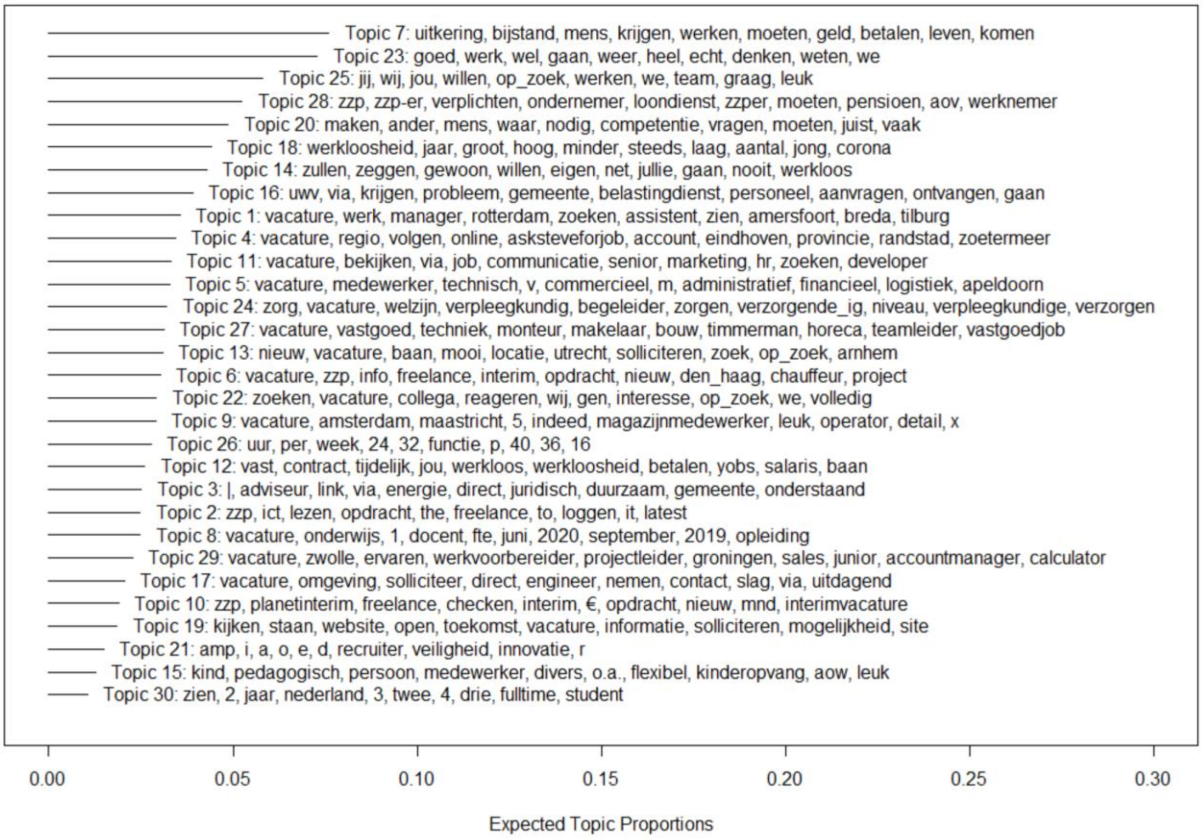


Figure C4: Topics in a 30-topic model of LM sub-section (Dutch original)



The interpretable topics in this case are the following: **Topic 7** includes tokens like ‘payment’, ‘assistance’, ‘people’, ‘receive’, ‘must’, ‘money’, or ‘work’ and describes who is receiving benefits, what type of benefits, how much they total to, and whether it is deserved. This topic is heavily concerned with immigration, and there is a strong overlap between this topic and topic 5 from the ECEC sub-section. In terms of sentiment the dimension validates, but with a noticeable bias: The positive extreme is generally slightly positive or neutral and the negative extreme shows noticeable bias due to the word ‘uitkering’ (payment) being labelled as negative on this dimension. The mean sentiment polarity is slightly negative (-0.03) and statistically significant. The success/failure dimension doesn’t validate.

Topic 28 includes tokens like ‘self-employment’, ‘self-employed (noun)’, ‘entrepreneur’, ‘oblige’, ‘pension’, ‘occupational disability insurance (aov)’ and describes various aspects of policy prescriptions for the self-employed. These include pension, the legal distinction between self-employed and entrepreneurs, but mainly whether the self-employed should have to contribute to occupational disability insurance. The sentiment dimension validates well, with the positive extreme praising policy changes or stating that policy is finally negotiated and passed, and the negative dimension mentioning that something is being critiqued or perceived negatively. The mean sentiment polarity is slightly negative (-0.02) and significant. The success dimension validates with mean polarity towards failure (-0.03) and significant.

Topic 18 includes tokens like ‘joblessness’, ‘high’, ‘big’, ‘year’, ‘less’, ‘corona’, ‘economy’ and is clearly a topic providing commentary about joblessness and the economic (and overall) impact of corona. However, it doesn’t differentiate between the situation in the Netherlands and elsewhere in the world, including commentary on the US or the Eurozone. Sentiment dimension doesn’t validate due to un-interpretable positive polarity, and success dimension doesn’t validate due to not interpretable success polarity.

Topic 16 includes tokens like ‘unemployment insurance agency’, ‘to receive’, ‘problem’, ‘municipality‘, ‘tax authority’, or ‘via’ (often used to link to a news story). It is mainly concerned with issues relevant to the UVW (unemployment insurance agency) like data leaks, miscalculations, or the misuse of corona-specific assistance. The sentiment dimension validates fine with negative polarity clearly corresponding to negative and critical comments, and positive polarity containing neutral or slightly positive commentary. The mean sentiment polarity is slightly negative (-0.02) and significant. The success dimension validates very well, with the failure polarity associated with tweets that are about failures and shortcomings of policies or the UVW, and the success polarity being much more neutral but reliably excluding comments about blatant failure. The polarity here is towards failure (-0.035) and is significant.

Topic 12 includes tokens like ‘fixed’, ‘temporary’, ‘contract’, ‘joblessness’, ‘jobless’, ‘to pay’, ‘salary’ and is rather general but maintains a focus on type of contracts. The issues this covers range from technical errors in the administration’s systems, to family postponement due to non-fixed contracts. Sentiment validates despite the negative polarity being largely dispassionate and the positive polarity including some noise. The sentiment polarity is very slightly negative (-0.006) but still significant due to large sample. The success dimension is similar with success polarity corresponding to people celebrating new employment contracts and the failure polarity corresponding to negative comments concerned primarily with unemployment. The success dimension leaning towards failure slightly (-0.01) and is significant.