
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Pääkkönen, Juho; Laaksonen, Salla Maaria; Jauho, Mikko

Credibility by automation

Published in:
Convergence

DOI:
[10.1177/1354856520901839](https://doi.org/10.1177/1354856520901839)

Published: 01/08/2020

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Pääkkönen, J., Laaksonen, S. M., & Jauho, M. (2020). Credibility by automation: Expectations of future knowledge production in social media analytics. *Convergence*, 26(4), 790-807. Article 1354856520901839. <https://doi.org/10.1177/1354856520901839>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Credibility by automation: Expectations of future knowledge production in social media analytics

Convergence: The International
Journal of Research into
New Media Technologies
2020, Vol. 26(4) 790–807
© The Author(s) 2020



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1354856520901839
journals.sagepub.com/home/con



Juho Pääkkönen

University of Helsinki, Finland; Aalto University, Finland

Salla-Maaria Laaksonen

University of Helsinki, Finland

Mikko Jauho

University of Helsinki, Finland

Abstract

Social media analytics is a burgeoning new field associated with high promises of societal relevance and business value but also methodological and practical problems. In this article, we build on the sociology of expectations literature and research on expertise in the interaction between humans and machines to examine how analysts and clients make their expectations about social media analytics credible in the face of recognized problems. To investigate how this happens in different contexts, we draw on thematic interviews with 10 social media analytics and client companies. In our material, social media analytics appears as a field facing both hopes and skepticism – toward data, analysis methods, or the users of analytics – from both the clients and the analysts. In this setting, the idea of automated analysis through algorithmic methods emerges as a central notion that lends credibility to expectations about social media analytics. Automation is thought to, first, extend and make expert interpretation of messy social media data more rigorous; second, eliminate subjective judgments from measurement on social media; and, third, allow for coordination of knowledge management inside organizations. Thus, ideas of automation importantly work to uphold the expectations of the value of analytics. Simultaneously, they shape what kinds of expertise, tools, and practices come to be involved in the future of analytics as knowledge production.

Corresponding author:

Juho Pääkkönen, Department of Sociology, University of Helsinki, Helsinki 00140, Finland; Department of Computer Science, Aalto University, Espoo 02150, Finland.

Email: juho.paakkonen@helsinki.fi

Keywords

Analytics as business, analytics in client companies, algorithmic knowledge production, automation, big data, credibility, data analytics, data imaginary, objectivity, qualitative methods, social media analytics, sociology of expectations

Introduction

Social media analytics is a newly emerged business practice, born in the footsteps of media monitoring and customer insights, that aims to accumulate and analyze digital traces of the online activities of organizations and their customers to produce information for guiding business. In its simplest form, social media analytics focuses on tracing the performance of corporate communications and marketing efforts online, often using ready-made analytics tools and metrics provided by social media platforms. More advanced forms draw on developments such as data mining and machine learning to analyze message content or metadata with the aim of identifying user opinions, behavior patterns, influencers, or prospective customers (see also Kennedy, 2016; Kennedy and Moss, 2015). Social media analytics is either performed by organizations themselves, who increasingly monitor their own actions through various performance metrics (Beer, 2016; Wiesenberg et al., 2017), or by specialized analytics firms, who offer a wide variety of products and services for the task.

As an area of business, social media analytics is enabled by datafication, the transformation of social action into quantified data, which allows for real-time tracking and computational analysis (Andrejevic, 2013; Couldry and Yu, 2018; Couldry et al., 2016; Van Dijck, 2014). Some influential accounts of these developments describe how ‘big data’, consisting of digital traces of human behavior and interaction, could be effectively used to access, monitor, and, most importantly, predict people’s behavior (e.g. Mayer-Schönberger and Cukier, 2013; Pentland, 2014). Such knowledge generation relies on massive data collection practices that have been discussed in the literature through concepts like data imperative (Fourcade and Healy, 2017) and data capture (Andrejevic, 2019).

Literature has connected the rise of social media and the associated analytics to a narrative that portrays the accumulating masses of social media data as providing a novel, heightened form of social knowledge (see Boyd and Crawford, 2012; Couldry, 2014; Elish and Boyd, 2018; Kennedy and Hill, 2018). Digital media technologies emerge as a mainstay of such expectations about the value of social media analytics. As Andrejevic (2019: 9) formulates: ‘The technological affordances of digital media make comprehensive data collection seem possible and the prospect of enhanced control make it seem desirable’. Previous studies of data analytics have shown that the field is premised on the expectation that analytics can provide simple and accessible tools for producing accurate and actionable insights from large data sets – a view that Beer (2017a, 2017b, 2019) labels the *data imaginary*. This expectation has its roots in media business and audience commodification (e.g. Bolin, 2011) but has experienced new twists with the introduction of algorithmic methods of knowledge production and forms of data capture and accumulation (e.g. Andrejevic, 2013; Bolin and Andersson Schwartz, 2015; Sadowski, 2019). While social media analytics is also applied in governmental and societal fields, the understanding of value in this context is largely based on economic judgments and promises of efficiency and business profit (cf. Andrejevic, et al., 2015; Bolin, 2011). In this respect, previous research has found that expectations of the future value of social media analytics persist even in situations where both analysts and

clients recognize various practical and methodological problems with data and analysis techniques (Kennedy, 2016). Thus, there seems to be a tension present in social media analytics between the expectation of unlocking valuable information inherent in data and the shortcomings associated with data and analysis methods.

In this article, we study how social media analysts and their clients negotiate their expectations about analytics with problems that they recognize in the field. Theoretically, we build on the sociology of expectations literature (Beckert, 2016; Berkhout, 2006; Borup et al., 2006). Sociology of expectations provides a view of expectations as *performative representations* that work to bring about the future scenarios that they represent (Borup et al., 2006). As such, expectations involve ideas about technological developments that can help realize the represented futures (Berkhout, 2006). Building on this framework, we draw on 10 thematic interviews with representatives of social media analytics and client companies to investigate their ideas about overcoming the problems they recognize in social media analytics. Our aim is not to provide solutions to these problems. Instead, we investigate *how different actors make their expectations about social media analytics credible*. The problems that analysts and clients recognize in social media analytics give rise to doubts about whether the field can meet its expectations. In this context, ideas about overcoming the problems enable analysts and clients to uphold their expectations; thus, the ideas make the expectations credible. With this perspective in view, our research question is: *How do analysts and clients negotiate knowledge of the limitations of social media analytics with their expectations about the future value of social media analytics?*

Based on our empirical material, we will argue that *automation* figures centrally as a *credibility-building idea* in social media analytics. By automation, in this context, we mean practices and technologies of data processing where human interpretation is bypassed. In our material, both analysts and clients propose such practices and technologies as a way to resolve problems in social media analytics. However, their positions and attitudes toward these problems differ. From the social media analysts' perspective, problems in analytics are mostly technical and resolving them is primarily a matter of demonstrating the usefulness of social media data. Clients, on the other hand, seek comprehensible ways of integrating social media data into their already-established knowledge practices. Finally, companies with a history of working with other types of data than social media data are skeptical about the novel data and seek ways to make them comprehensible in terms of familiar methodology. Ideas about automation figure in all these views but serve different roles in them. To further analyze these different roles and how they lend credibility to expectations about social media analytics, we draw on the work of Collins and Kusch (1998), which examines the relationship between automated processes and human action. At the same time, the view of expectations as performative representations allows us to discuss how the roles given to automation by different actors drive particular views of the future of analytics as knowledge production.

Our study contributes to recent discussions about the role of automation in data analytics (Andrejevic, 2019; Andrejevic, et al., 2015; Beer 2019). This work has highlighted that the value promise of analytics rests on effective automated infrastructures and the specialist expertise of data analysts. We add nuance to this discussion by showing that the different ways in which automation can lend credibility to expectations imply diverging views of the skills and expertise that are central to social media analytics.

In the next section, we introduce our theoretical approach of the sociology of expectations, which we then connect to automation as a credibility-building idea in social media analytics. Together, these two perspectives constitute our approach to analyzing our material, presented in

the ‘Data and method’ section, followed by our analysis (fourth section to seventh section). In the ‘Discussion and conclusion’ section, we discuss our results in relation to previous literature about the role of automation in analytics.

Expectations and automation as a credibility-building idea

According to the sociology of expectations, technological expectations are ‘real-time representations of future technological capabilities’ (Borup et al., 2006: 286). Expectations are performative in the sense that they can work to bring about the imagined states of the world they represent. Expectations do this by helping actors momentarily overlook uncertainty in the future by portraying some future scenarios as more plausible or attractive than others (Beckert, 2016: 35–60). As such, expectations imply a commitment to a set of particular future possibilities (Berkhout, 2006: 302). When shared, they can work to mobilize actors and thus incentivize technological development (Flichy, 2007).

As Berkhout (2006: 302) argues, to be distinguished from mere future possibilities or objectives, expectations *need to be supplemented with an idea of how the represented future is to be achieved*. Thus, the function of expectations is to map the space of possible future scenarios within a domain and identify the salient problems that need to be resolved for the imagined futures to be realized (Berkhout, 2006: 305). In this view, the credibility of an expectation is the product of the ‘material and social structures’ (Berkhout, 2006: 306) within which it figures, dependent on the judgments of actors embedded within those structures. Thus, judgments of credibility can vary among actors in different positions (Brown and Michael, 2003). In this sense, expectations involve three characteristic features: (1) *objectives* or the represented future scenarios; (2) *order* or a set of social and institutional relationships within which the objectives can be met; and (3) *technologies*, which are the means for achieving the objectives (Berkhout, 2006: 302). Expectations can conflict with each other and enter into a contest where the credibility of different possible futures are evaluated against each other (Brown et al., 2000).

From this perspective, expectations about social media analytics can be viewed as performative representations of the capabilities of analytics technologies to uncover valuable information data. Insofar as these expectations are held by social media analysts and their clients, they involve ideas about technologies necessary to bring about the futures they represent. These ideas about technologies thus work to *support the credibility of expectations about social media analytics* by serving to ‘trigger imaginaries of successful future business’ (Beckert, 2016: 68).

As we will show later in this article, *ideas of automation have an important role in supporting the credibility of expectations about social media analytics*. Given perceived problems pertaining to the use of social media data, ideas of automated technologies that solve those problems can be drawn upon to lend credibility to compromised expectations. Here, automation is the technological means through which analysts and clients think they can reach the objectives set for social media analytics. In accordance with what Passi and Jackson (2018: 20) have observed in the context of data science, uncertainties pertaining to social media analytics emerge as ‘sites for justifying the “worth” of data, models, and results through actionable strategies’. We argue that in social media analytics, ideas of automation are a central element in such actionable strategies.

Other recent work has emphasized the importance of automation in analytics. As Beer (2019) has argued, data-led thinking in companies is supported by a powerful imaginary that frames analytics as the key to unlocking value in large data sets. Automated methodology is central here because this ‘unlocking requires an infrastructure that allows for automation and a space in which

Table 1. Interviewed companies.

Company acronym	Role	Size and stage	Number of interviewees
A1	Analytics	Middle-sized start-up and over 10 employees	1
A2	Analytics	Small start-up and several employees	1
A3	Analytics	Small start-up and several employees	2
A4	Analytics	Small start-up and several employees	1
AC5	Analytics/client	Large and established company and over 50 employees	2
AC6	Analytics/client	Large and established company and over 100 employees	3
C7	Client	Large and established company and hundreds of employees	1
C8	Client	Large and established company and thousands of employees	4
C9	Client	Large and established company and thousands of employees	1
C10	Client	Large and established company and thousands of employees	1

the data analyst's cognitive skills are enhanced by these automated systems' (Beer, 2019: 119). Similarly, Andrejevic (2019) has argued that automated infrastructure for comprehensive data captures figures as an enabling condition in the business logic of data analytics.

To analyze the roles given to automation in social media analytics, we draw on Collins and Kusch's (1998) account of automation. This account accords with our view of automation in analytics as data processing practices where human interpretation is bypassed. For Collins and Kuhn, actions can be automated when they do not depend on context-dependent interpretive work to be carried out. For instance, the operation of a ready-trained supervised machine learning classifier on a data set is automated in this sense. Importantly, automated processes can also involve humans and organizational practices, insofar as their performance does not depend on interpretative context-dependent judgments because such processes *might as well be performed by machines*. According to Collins and Kusch (1998: 119–120), automated processes can be used as 'tools' to amplify human capabilities of action in certain tasks, as 'proxies' to replace human action, or as 'novelties' to do things that humans could not possibly do. Thus, ideas of automation involve a demarcation between the expertise and capabilities of humans and the capabilities of machines.

Data and method

This study investigates social media analytics as knowledge production that is guided by certain expectations and technological conceptions. We focus on interview data from 10 companies (see Table 1), conducted both with professionals working in analytics companies ($n = 6$) and their clients ($n = 4$). Three interviews were group interviews of two to four people, and the rest had only one interviewee. All interviews were conducted by the second author, lasted from 1.5 h to 2 h each, and were transcribed verbatim. These data were collected within the framework of a wider project that studied the practices and methods of social media analytics in the Finnish context.

Social media analytics, when focused on mining text data, becomes heavily language-dependent. This means that big international companies working with text mining rarely enter smaller countries such as Finland, where the language generates barriers for business. Accordingly, of the companies we interviewed, the four that were primarily engaged in collecting and/or

analyzing social media data were all small- or middle-sized start-ups with 2–15 employees (A1, A2, A3, and A4). One of these smaller companies focused on network analysis and text mining as their main business. The main products of the other two companies included a brand measurement platform, based on a custom-built data collection and analysis infrastructure, and social media content profiling and user segmentation, respectively. Finally, one company focused on collecting and refining data from multiple platforms and providing an efficient query infrastructure for easy access.

Two of the interviewed analytics companies were large and established firms. One was engaged in survey research as their main line of business but was aiming to incorporate social media analytics as part of their products (AC5). The other was a media company that offered content recommendation, moderation, and creation services based on social media analytics but was largely based on externally bought analytics tools (AC6). These companies differed from each other in that the media company based their products more extensively on social media analytics, building on a large proprietary data set. Nevertheless, both of the established analytics firms partly occupied a client position in our analysis because they used tools, infrastructure, and services provided by smaller social media analytics companies.

The interviewees in all of the analytics companies were mainly in positions of management or middle management. However, in small analytics start-ups, the management personnel in practice also often engages in operative work, performing tasks that involve data collection, analysis, marketing, and customer consulting.

The client companies were all large and established firms in different fields of business, including insurance, retail, telecommunications, and food production (C7, C8, C9, and C10). These companies mainly engaged with social media data using externally bought data collection and analytics tools and consulting services. However, all the companies also employed personnel responsible for data analytics, which were either already engaged or were starting to engage with the use of social media data. The interviewees in these companies included mainly manager-level personnel responsible for developing customer engagement, company marketing, and communications processes.

The interviews we draw on were semi-structured and theme-centered, with analysts and clients discussing their companies' practices in relation to the project's goal of developing methods for advancing business uses of social media analytics. Thus, thematically the interviews revolved around the use of social media data and analytics for producing business insights and problems pertaining to their use and to social media analytics more broadly as a field of business. However, in addition to these themes, the interviews contained extensive discussion of what both analysts and clients expect from social media analytics and how those expectations could be realized. Thus, the interviews provide ample material for investigating what constitutes credible social media analytics for the different actors.

Our analysis of the interviews was guided by the theoretical framework discussed in the previous section, focusing on expectations for social media analytics and in particular on ways in which they could be realized despite recognized problems. We performed two rounds of analysis, coding the material with Atlas.ti software (version 8.3.1 for macOS). In the first round, the first author read through and coded all the interview material, focusing on passages that discussed ideas related to social media data and other kinds of data, methods, and aims of analyzing data, automation as part of analytics, the use of analytics information in companies, and the context within which the companies operated. The aim was to identify excerpts pertaining to what companies hoped to achieve with social media analytics and what problems were associated with these aims.

During this process, automation emerged as a key concept around which both the expectations related to social media analytics and the problems involved in reaching these goals revolved. While the term was not necessarily used in all accounts by the interviewees, the issues they talked about could be framed through it. In the second round of analysis, we focused on excerpts related to ideas about automation to examine how they were connected to different problems and aims in analyzing social media data. Each author individually read the excerpts related to automation and coded them to identify how ideas of automation relate to the aim of producing valuable information with social media data. After this stage, we met and discussed the codes to check that our interpretations agreed with each other. On the basis of this discussion, we subsequently focused our analysis on three issues: (1) current uses of automation in collecting, processing, and analyzing social media data; (2) problems that automation is thought to solve; and (3) ideas about what automation is expected to provide. We reread the excerpts focusing on these issues and jointly collated a document that described our interpretations.

On the basis of this analysis, we identified three roles allocated to automation in overcoming problems in social media analytics, which are discussed in the fifth section to seventh section, respectively. Before examining these roles more closely, however, we will first take a look at the current status, expectations, and problems of social media analytics among our case companies more generally.

The status, expectations, and problems of social media analytics

The business offering of the social media analytics companies in our material consisted of providing clients with access to information that is valuable for guiding actions and decisions. The companies' products included tools and consulting to guide clients through the entire data collection and analysis process, starting from iterative specification of keyword queries – a standard method for collecting social media data – to more extensive analytical work, such as network visualizations or topic identification with text mining and consulting with result interpretation. Most companies also provided easy-to-use versions of their analysis tools and data collection infrastructure as a ready-made pipeline to access data and produce simple representations (e.g. timeline plots of given keywords). Some companies also offered more advanced analytics in ready-made packages, such as content profiling based on natural language processing.

The client companies were largely reliant on external tools and services for their current uses of social media data, which mainly focused on monitoring social media discussions to (1) track opinions of company brand and products in relation to competitors; (2) measure the performance of product campaigns; and (3) anticipate and react to customer needs. The typical analysis flow consisted of querying social media for discussions pertaining to a given interest, filtering the produced data to identify relevant material, and applying the chosen analysis methods (e.g. topic identification algorithms or sentiment analysis) to derive metrics describing the discussion contents and collating the results to a report, which was subsequently distributed within the company.

The clients were at differing stages in integrating these procedures. While some companies had more established analytics pipelines, others struggled to incorporate the tools and services provided by social media analytics companies. One company mainly relied on external consulting services for data collection and reporting. Others had developed steady pipelines for collecting and analyzing social media data in combination with other data sources. This was the situation in two client companies, and the established analytics company more extensively engaged with social

media data. However, the remaining companies – one client company and one established analytics company – reported less success in their efforts to integrate social media analytics.

Thus, among the clients, the current status of social media analytics was still rather ad hoc, requiring investments in manual labor and craftsmanship-like expertise in data collection and processing. In all cases, the current status was thought to leave room for improvement. In the case of the client companies with already-developed pipelines, the hope was that analytics could provide more accurate measures of discussion content and user behavior in a continuous manner. Ready-made tools for sentiment analysis and content classification were commonly thought to be inaccurate and to require labor-intensive checks of result reliability and interpretation. Accordingly, one expectation both clients and analysts expressed for social media analytics was that of improved efficiency and accuracy in continuously measuring company and customer behavior online.

Another expectation was producing understanding of phenomena *offline*, such as customer opinions, brand reputation, or consumer trends. This expectation was most prevalently present in the interviews of one client and one established analytics company, which relied on demographic and representative surveys or financial data. However, it was also expressed by another client company and the analytics company that specialized in collating and refining data from different social media platforms.

The difficulties in fulfilling both of the above expectations were connected to a set of methodological and practical problems concerning social media data use. First, social media data were characterized by both clients and analysts as *messy*, *voluminous*, and *unstructured*. The data content was described as diverse, pertaining to widely varying topics that are often discussed using esoteric terminology. Consequently, queries on social media often return high volumes of material from diverse and potentially unconnected contexts. This makes distinguishing between relevant and irrelevant material difficult and complicates evaluating whether the collected samples adequately capture phenomena of interest. Messiness makes cleaning, refining, and classifying collected samples a time- and resource-consuming task, which conflicts with the aim of efficient, continuous, and accurate measurement.

Second, in addition to being messy, social media data were argued to *lack important contextual information* about the users of social media and their motivations for their actions. This issue was exacerbated by the worry that social media data are *not representative* of phenomena and populations offline. Taken together with data messiness, client companies remained uncertain about the extent to which social media data can be integrated as part of their established practices – an issue most pressing for the two companies accustomed to rigorous survey methodology with clearly delineated practices for sample evaluation.

Finally, two of the clients and both of the established analytics companies held that integrating out-of-the-box analysis methods into already-established practices of producing and utilizing information is challenging. The results produced by externally bought tools such as interfaces for querying data or machine learning and media tracking software for topic detection and sentiment analysis were thought hard to combine with heterogeneous practices and informational needs, especially in large companies.

The social media analytics companies in our material also recognized methodological problems in analytics. However, the messiness, lack of contextual information, and unrepresentativeness – which clients regarded as impediments for using social media data – were considered by analysts as largely technical challenges to be solved by developing more effective techniques for cleaning and validating data. Once resolved, the value of social media analytics could be relatively easily

demonstrated. The analysts generally held that a more pressing impediment for marketing really advanced analytics related to the clients' poor technical understanding. Thus, the different companies in our material varied in their attitudes toward the recognized problems. The most skeptical attitudes were expressed by the established analytics company, which focused on representative survey research. Client companies that did not engage in analytics as their main line of business lamented the problems they recognized but were looking for easily accessible ways to incorporate social media analytics into their knowledge practices. While all companies held on to the expectation that social media analytics could produce valuable information, they drew on different ideas about how this could be achieved.

In the following sections, we look at solutions proposed by our interviewees to the above-discussed problems. As we will see, the notion of *increasing the automation of the analytics pipeline* in various ways was taken up as a response to both the methodological problems in social media analytics and the issue of integrating analytics. In the fifth section to seventh section, we discuss three central roles given to automation by our interviewees: extending and making expert interpretation of social media data more rigorous, enabling data-driven objective measurement, and providing flexible access to analytics for nonexperts.

Automated interpretation

One expectation of both analytics and client companies was that social media data could be used in combination with more established methods to provide additional information about offline phenomena, such as consumer opinions, customer experiences, or brand reputation. This expectation was compromised by the messiness, lack of contextual information, and unrepresentativeness of social media data, which make evaluating the adequacy of the collected samples difficult.

The starting point is that, from the perspective of the researcher, social media discussions are never reliable – not even close. They are very biased. They do not represent the Finnish population's opinions about anything. (C10, I1)

In relation to these problems, social media analysts were generally optimistic about the possibility of demonstrating the value of social media data once the practices of circumscribing data collection become established. Two of the clients largely agreed with this view, but others were more hesitant. The most skeptical views were expressed by the two companies accustomed to working with representative survey methodology, which had not yet developed stable pipelines for working with social media data, as in the quote above. A solution proposed by one of these companies was the idea of *hybrid systems* that *identify relevant material* among messy data and *enable more rigorous expert interpretation* of unrepresentative samples.

This role was depicted as involving two separate processes. First, supervised protocols for data collection were proposed to enable *contextual querying of social media data* through an iterative process of training the query algorithm to recognize relevant documents.

I could in principle build some sort of an algorithm that goes through the results and tries to infer the context. And possibly search, by random, some matching articles that could be tapped as good or bad. And then use the good ones as training data for broader contextual queries. (AC5, I2)

Such algorithms would alleviate the problem of data messiness by drawing on *the user's expertise in defining which documents are relevant for a given set of specified keywords*. Thus, supervised protocols would replace the currently standard process of consulting, which takes place

between social media analysts and clients when trying to define appropriate keyword queries for capturing phenomena of interest. This imputation of contextual information was thought to enable researchers to relate the results of their particular queries to a more general picture of social media discussions.

However, for the human interpreter to be able to assess the relevance and meaning of the results, another automated process was required. Interpreting the meaning of social media discussions is difficult if no information is available about the background of the discussants. User profiling with qualitative analysis was proposed as a potential remedy but was deemed impossible because of data volume and the complexity of discussions. Thus, automated methods were required to *condense* data and represent them for the researcher in a form that is easier to interpret.

We would need something programmatic to browse through that stuff and condense it and to categorize, classify, and present it. And then the researcher could sort of see that infographic . . . and note the point they want to take up . . . that this knowledge could have value. And this is in our job queue at the moment – how to solve this thing. (AC5, I1)

Thus, in the context where social media analytics was compared to more established methods in market and consumer research, automation was given a twofold role. First, contextual querying methods were thought to enable drawing on expert interpretation to distinguish between relevant and irrelevant material and, moreover, impute lacking contextual information on the data. Here, the importance of human expertise was emphasized, and automated methods were portrayed as ‘tools’ (Collins and Kusch, 1998) that extend human capabilities for interpretation.

Second, for this extension of expertise to be possible, automated methods were needed to *guide the human in interpreting the data* by representing them so that valuable information can be detected. Here, the role of automated methods was to work as *interpretative devices* between human experts and unreliable masses of social media data *by restricting the space within which expert interpretation operates*. In doing so, automation serves as a ‘novelty’ (Collins and Kusch, 1998), producing an interpretable representation of data where qualitative analyses fail.

This hybrid process lends credibility to the expectation that social media data could be combined with the methodology that is ultimately geared toward studying offline populations with representative samples. Automated processes were portrayed as capable of circumscribing and extending human interpretation, which could not otherwise be judged to be credible. Thus, credibility is built on *the capabilities of both humans and machines* – relying on the former for context-sensitive interpretation, while the latter establishes explicit procedures that circumscribe interpretive judgments. This role of automation contrasts with the view discussed in the next section, where credibility is thought to rely on full automation of the data analytics pipeline, designed for measuring online behavior.

Automated objectivity

In the previous section, hybrid automation was proposed as a way to enable more rigorous practices of interpretation in a context marked by the established use of representative survey methods. In this section, we look at a more extensive role given to automation by analytics and client companies that expected social media analytics to be able to *generate metrics of discussion and company behavior online*. In this context, credibility was thought to depend on automating the measurement process so as to make metrics more *data-driven by precluding the effect of subjective human judgment* in the analysis process.

On the face of the problem of unrepresentativeness of social media data, the analytics companies commonly held that social media data could still provide important information about what kinds of online material people engage with and how companies should act online. However, from the perspective of many clients and one analytics company, data messiness, lack of contextual information, and representativeness also posed problems for this expectation, albeit more indirectly. Importantly, the challenge was not the measurement itself. Rather, difficulties emerged when companies attempted to turn metrics into relevant information for company actions and decision-making. For instance, evaluating the size of a given discussion topic on social media is difficult when knowledge of its relationship to other online discussions and offline phenomena is lacking. Similarly, evaluating whether a given sentiment score is ‘good’ or ‘bad’ for the company is difficult in the absence of a *standard against which social media metrics are to be assessed*.

... The first question is: How big is this? Yeah, there’s no answer ... because you cannot. What is the metric for “bigness” in these things? ... How do we measure its bigness? My standard answer in this case is that it’s as big as we make it. (C10, I1)

One solution to this problem, as recognized by both clients and analysts, would be to validate social media metrics against already-familiar data sources. As Espeland and Stevens (1998: 315–316) have argued, such a process of *commensuration* seeks to establish a common standard, through which measures normally treated as distinct can be evaluated against each other. Validation of social media metrics with, for instance, company sales data could establish such commensuration by showing that changes in company social media metrics can be used as a proxy for changes in sales. However, many client companies and one analytics company were skeptical of this solution and lamented that data sets necessary for validity testing are often proprietary or reported failures in their attempts to establish correlations among different data.

This had led to a situation where it was difficult to decipher whether social media metrics measured anything real or were just artifacts produced by measurement system design. In this context, the idea of an automated procedure that would *constantly check the produced measurements against data* was proposed as a solution by one analytics company.

Ideally, it could check itself constantly and even automatically, so it could sort of machine learn toward some reality. But what that reality is, we first thought that it is the company’s reputation or sales, but those correlations are ridiculously low still ... Maybe the trustworthiness is that, for instance, post volumes and scales should come directly through our data. If we are focused on the intensity of engagement, then it would check those [companies] that get the best score and determine how many times you have to post. We should come up with some kind of a formula, so that it would constantly calibrate itself. ... (A3, I1)

As this quote demonstrates, given the failure of validating social media metrics against more established data sources, the next best option was thought to be to tie their credibility to an automated process that calibrates the metrics against patterns detected in social media data proper. The idea behind this *data-driven measurement* enabled by automation is that the measurement system should be able to evaluate the metrics given for one company against those given for others to derive the meaning of each by comparisons against the rest. This way, by using *social media data as a standard for themselves*, the problem of validation is avoided.

In this view, automation has the crucial role of establishing credibility by making the assessment of measurements data-driven and *not dependent on human interpretation or judgment*. Given the difficulty of evaluating the results of social media metrics, automation was thought to enable

credible measurement by calibrating the metrics against a ground truth contained in social media data. This role of automation as making measurement objective was emphasized also in cases where social media metrics had been successfully validated. One client company reported having established a correlation between social media sentiment metrics and customer satisfaction surveys. However, they still maintained that having to manually check the results of their automated classifiers introduced an element of uncertainty into the pipeline. Ideally, then, even validated social media metrics would be the products of a full-fledged automated protocol.

... Having a human in between always creates the possibility that their preconceptions have an effect on the results. When the machine tells you that 17% [of the discussion] is about pricing or is negative, then it is more credible and truthful than if I myself would have done it. (C9, I1)

Thus, in a context marked by the expectation of measuring and evaluating online behavior, the credibility of social media analytics was thought to depend on removing the effect of human judgment on the produced metrics. The idea was to remove the human from the loop, using machines as a more accurate and truthful 'proxy' (Collins and Kusch, 1998) in gleaning out information inherent in data. This way, automation was thought to make measurement more data-driven and objective by enabling the use of social media data as standards for themselves.

Automated analytics interfaces

The previous two sections have focused on solutions to problems in producing information using social media data. In this section, we will turn to a different problem – namely that of putting the information produced by analytics to practice in client companies. For social media analytics to fulfill expectations, clients not only need to overcome doubts concerning analytics. Additionally, the information produced has to guide actions and decisions in the organizations. In this context, the idea of *automated interfaces*, which enable analytics to be flexibly used as part of heterogeneous organizational practices, was taken up by many clients as a way to achieve this integration.

The main issue in usefully integrating information produced by social media analytics, as explained by the interviewees in a large retail and service sector company, was that social media analytics is hard to coordinate with already-existing organizational practices. No clear procedures exist for combining divergent data sources, which make adopting new analytics slow and rigid. In the case of social media data, this problem is aggravated by reliance on externally bought analytics tools, which often are difficult to adapt to established informational practices. This latter issue was also pressing for the established analytics company that was more extensively engaged in social media analytics. Large organizations have many branches, with different practices of producing and using data for different goals. As a result, internal teams responsible for analytics are forced to stretch themselves to come up with ad hoc solutions in response to heterogeneous service calls. This creates the need to develop better coordinated ways of data use so that the resources allocated to analytics could serve as many company needs as possible.

A central way of achieving this flexibility, proposed by our interviewees, was an *automated interface* for accessing the information produced by analytics. Such an interface would ideally collate information from all divergent data sources used by the company in one location, making them comparable and easy to use together. Importantly, the interface would not necessarily work to commensurate divergent data sources – a problem that our interviewees emphasized would require

standardizing company practices of data production. However, from the users' perspective, interface access to analytics information would remove a 'senseless' phase of manually combining data sources for joint analysis.

One would hope that, if we had such a dashboard, then the data would be as uniform as possible . . . so that its not like a text file, a PowerPoint, PDF, or Excel. And then you go through those and try to figure out what can be combined . . . it would remove that completely senseless manual phase in it. (C8, I3)

The benefit associated with such an interface for social media analytics was that it would enable companies to monitor and react to phenomena such as discussion trends on social media. The clients hoped that social media data could provide them with a relatively effortless way of keeping up to date about trends on social media without having to make large resource investments on monitoring procedures.

Ideally, it would give us a list of trendy hot topics and keep a list up to date, in a way. And our aim is that it would be within our internal information system used by the commercial people . . . where we collate the basic view of everything we produce. So if we could have there a list of topics that are being discussed, then that would add to our, in a way, should automate our understanding of the surrounding world and, in this case, Finnish consumers in particular. (C10, I1)

In the context of this expectation, the analytics interface would ideally serve information in a form that is easy to access for different kinds of employees. Thus, analytics interfaces would provide a personalized view into data, enabling access to meaningful information from the perspective of the end user. This idea was further connected to real-time or on-demand access, which some interviewees contrasted with rigid periodic reporting. This way, automated interfaces were thought to help heterogeneous practices in large organizations to become relatively autonomous. The upshot is that *automation was thought to allow nonexpert end users of analytics tailor information according to their own needs*, enabling analytics processes to become more generic and less tied to specific requirements. Again, we see the idea that automation could serve as a 'novelty' (Collins and Kusch, 1998), which enables nonexperts to use data that otherwise would be inaccessible for them.

A crucial condition for the success of such an interface, pinpointed by the clients, is simplicity. Given the volume and unstructured nature of the data and the complexity of computational methods necessary for their analysis, clients maintained that the role of easy-to-use interfaces for accessing analytics was becoming increasingly important. Several companies envisioned that this accessibility is reached by using visualizations.

. . . We should invest in visual analytics . . . so that others could understand the data, and at best we can build systems that wider crowds of people can use . . . so that it does not always have to be the expert, who looks at them . . . (C9, I1)

Thus, expectations about easy access to information made credible by ideas of automated, preferably real-time interfaces work to drive the proliferation of simple visual tools for collating and personalizing information, while simultaneously emphasizing the importance of critical accounts of data visualization practices (cf. Kennedy and Hill, 2018; Kennedy et al., 2016; Laaksonen and Pääkkönen, 2020).

Discussion and conclusion

Our findings highlighted three roles given to automation as a credibility-building idea in social media analytics. First, the idea of hybrid systems for extending expert interpretation was proposed by a company accustomed to clearly delineated methodology. We saw automation lending credibility to expectations in the areas of data analytics skeptical toward the methodological capabilities of novel data and methods. Second, in the context of measurement on social media – where no clear methodological standards existed for evaluating metrics – the idea of fully automated protocols and self-calibrating metrics was proposed to remove the need for subjective judgments altogether. This idea lent credibility to the expectation that social media data could produce useful information about discussions and company behavior online. Finally, with respect to implementing analytics in companies, automated interfaces were thought to promote more efficient and seamless coordination of knowledge management by enabling end users to access analytics information according to their needs.

These findings show that ideas of automation have a central role in negotiating the future of social media analytics as knowledge production. In answer to our research question, we have shown that in the face of recognized problems, both analysts and clients draw on ideas of automation as potential solutions. Simultaneously, these ideas lend credibility to the shared expectation that social media analytics could produce valuable business insights. As recognized in the sociology of expectations literature, such shared expectations serve to mobilize resources around objectives, thus working to push toward their fulfillment (Borup et al., 2006). However, as Berkhout (2006) has argued, for shared expectations to have this performative or mobilizing power for different actors, they must lend themselves for different interpretations. In other words, shared expectations must be *interpretatively flexible* (Pinch and Bijker, 1984) to be able to mobilize various actors with divergent aims, resources, and levels of expertise (cf. Brown and Michael, 2003; Flichy, 2007). Our analysis has highlighted how automation emerges as a solution that allows both analytics and client companies to interpret the shared expectation of the value of social media analytics as credible.

In doing so, our findings both accord with and add to previous work on the importance of automation for expectations about analytics. This work has emphasized that the business promises of data analytics rest on an intermeshing of automated infrastructure and tools with the expertise of data engineers, who build pipelines for amassing and ‘sanitizing’ (Beer, 2019: 112) large volumes of ‘raw’ data (cf. Gitelman, 2013), and data analysts, whose expertise consists of puzzling together the results of analytics with business needs (Gehl, 2014). The promises made by the data imaginary involve an intertwining of the analysts’ and engineers’ expertise with automated systems, representing a ‘human-machine hybrid solution’ (Beer, 2019: 101) to the data deluge. Similarly, Andrejevic (2019: 7) has argued that the business logic and promises of analytics depend on ‘digital infrastructure and platforms on an increasingly comprehensive scale’ that essentially stand on the bedrock of automated data collection and processing. In these views, the technological practices and tools that underpin analytics act as a ‘cluster of promises’ (Mackenzie, 2013: 402), constituting a solution to the problems involved in making sense of the accumulating masses of digital data (cf. Stieglitz et al., 2018).

Our findings correspond to this picture but add richness and nuance to it. Our analysis shows that expectations for the value of social media analytics span across a host of contexts and actors, including social media analysts, clients in different domain areas, and analysts working in settings apart from social media analytics. While both analysts and clients shared the expectation of the

business value of social media analytics, the problems recognized and the corresponding solutions proposed varied according to different actors. For the clients and established analytics companies who depended on external tools for social media analytics, crucial issues concerned the interpretation of the results of those tools and the ability to incorporate social media analytics into heterogeneous organizational practices. For companies that were accustomed to a more traditional methodology, the central difficulty was to combine them with novel data. For social media analytics companies, by contrast, skepticism toward social media data appeared primarily as a business impediment related to the clients' poor understanding of the field's potential. In cases where analysts from social media analytics companies were not convinced by extant methods of measurement on social media, ideas of automation were drawn on to lend credibility to analytics. However, these ideas differed from those put forward by companies accustomed to established and clearly delineated methodology. For these companies, the intertwining of human agency with automated systems working as 'tools' and 'novelties' (Collins and Kusch, 1998) was a key to credible analytics. By contrast, in the case of social-media-focused analytics companies, credibility was grounded in the idea of automation as a 'proxy' in the form of fully automated measurement protocols.

Therefore, automation – rather than being a single concept or a mere technical necessity – emerged as an idea that actors in different contexts can adapt to lend their expectations with credibility. Automation could simultaneously cater to the requirements of the different contexts, thus enabling the nascent and heterogeneous field of social media analytics to uphold the shared expectation of value. Recognizing differences among contexts in social media analytics is important, we maintain, because the associated technological solutions have implications for how the future of the field unfolds. Thus, we argue that *automation works in social media analytics as a credibility-building idea that simultaneously shapes how analytics as knowledge production is envisioned.*

In our analysis, we could see this most clearly in the notion that automated processes could replace human interpretation. Although previous work (e.g. Beer, 2019) has emphasized the importance of the analysts' and engineers' expertise in realizing the value expectations toward analytics, we observed aspirations toward fully automated measurement protocols that downplay this expertise. Such ideas resemble the desire for numbers documented by Kennedy (2016) in data mining practices. Accordingly, a myth-like idea of fully automated knowledge production was presented in our material as a potential solution to methodological problems (cf. Couldry, 2014). In addition, we identified expectations where *nonexperts* were also included as legitimate interpreters of social media data and analytics through the development of flexible automated interfaces. The idea of automation as a 'novelty', intermeshing with the agency of users, directly lends credibility to the expectation of accessing valuable information in data through easy-to-use tools (cf. Beer 2017a). This idea conflicts with the business promise of analytics companies, which relies on the expertise and craftsmanship of the data analyst. Thus, we see how conflicts can arise between actors approaching the commonly shared expectation from different positions because of diverging needs and problems associated with its fulfillment (cf. Brown et al., 2000).

The different ways in which expectations about social media analytics can be made credible imply particular views of what kinds of expertise are relevant for the field – and which parts of analytics can be automated. In particular, ideas of fully automated and easy-to-use analytics tools foster expectations that are strongly reminiscent of technological solutionism (Morozov, 2013) – that technological development will eventually solve problems with social media data and extensivity of manual work to unleash the promises of marketing hype (Beer, 2017a). This idea contrasts with the view of expert analysts working *with* automated processes to glean information

from social media data, applying their trained vision to interpret the discovered patterns as business knowledge (Beer, 2019; Gehl, 2014). These two opposing views were present in our material. In our analysis, solutions based on hybrid automation tended more toward emphasizing the importance of expertise, while fully automated measurement strives toward removing the analyst ‘from the loop’. Automation is central to both views, yet the specific role given to it varies.

Investigating conceptions of automation is thus key to grasping how the negotiation of different expertise, tools, and practices comes to constitute credibility in social media analytics. The expected futures implicated in this negotiation contribute to the ways in which analytic practices will become realized within the field (see Flichy, 2007). Thus, conceptions of automation can work to buttress expectations that excite ubiquitous, authoritative, data-led knowledge management in organizations using social media data (cf. Beer 2016). Our work has taken a step toward developing an understanding of how the credibility of social media analytics is negotiated. However, given the pervasive role of social media analytics in organizational and everyday life (Kennedy, 2016), studying how ideas of automated solutions lend credibility across different contexts becomes increasingly important. For instance, although we could not focus extensively on conflicts between different expectations in our analysis, investigating the competing ideas that drive social media analytics calls for thorough empirical research. Likewise, extending our analysis to contexts outside of business and into different business domains is a potential avenue for future work.

Acknowledgements

We thank the anonymous reviewers for their helpful comments. This article is based on ideas presented at the Moral Machines Symposium and the NordicSTS conference in 2019. We thank the audiences of these events for their comments. We would also like to thank Matti Nelimarkka, Essi Pöyry, Arto Kekkonen, and Veikko Isotalo for assisting in the data collection.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by The Finnish Funding Agency for Technology and Innovation Tekes. The first author also received support from The Finnish Foundation for Economic Education, and the second author received support from The Finnish Science Foundation for Technology and Economics KAUTE.

References

- Andrejevic M (2013) *Infoglut: How too Much Information is Changing the Way we Think and Know*. New York: Routledge.
- Andrejevic M (2019) Automating surveillance. *Surveillance & Society* 17(1/2): 7–13.
- Andrejevic M, Hearn A, and Kennedy H (2015) Cultural studies of data mining: Introduction. *European Journal of Cultural Studies* 18(4–5): 379–394.
- Beckett J (2016) *Imagined Futures: Fictional Expectations and Capitalist Dynamics*. Cambridge: Harvard University Press.
- Beer D (2016) *Metric Power*. Basingstoke: Palgrave Macmillan.
- Beer D (2017a) Envisioning the power of data analytics. *Information, Communication & Society* 21(3): 465–479.
- Beer D (2017b) The data analytics industry and the promises of real-time knowing: Perpetuating and deploying a rationality of speed. *Journal of Cultural Economy* 10(1): 21–33.
- Beer D (2019) *The Data gaze: Capitalism, Power and Perception*. London: SAGE Publications.

- Berkhout F (2006) Normative expectations in systems innovation. *Technology Analysis & Strategic Management* 18(3–4): 299–311.
- Bolin G (2011) *Value and the Media. Cultural Production and Consumption in Digital Markets*. Farnham: Ashgate.
- Bolin G and Andersson Schwarz J (2015) Heuristics of the algorithm: Big data, user interpretation and institutional translation. *Big Data & Society* 2(2): 205395171560840.
- Borup M, Brown N, Konrad K, et al. (2006) The sociology of expectations in science and technology. *Technology Analysis & Strategic Management* 18(3–4): 285–298.
- Boyd D and Crawford K (2012) Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- Brown N and Michael M (2003) A sociology of expectations: Retrospecting prospects and prospecting retrospects. *Technology Analysis & Strategic Management* 15(1): 3–18.
- Brown N, Rappert B, and Webster A (2000) *Contested Futures: A Sociology of Prospective Techno-Science*. Aldershot: Ashgate.
- Collins H and Kusch M (1998) *The Shape of Actions: What Humans and Machines can do*. Cambridge: MIT Press.
- Couldry N (2014) Inaugural: A necessary disenchantment: Myth, agency and injustice in a digital world. *Sociological Review* 62(4): 880–897.
- Couldry N, Fotopoulou A, and Dickens L (2016) Real social analytics: A contribution towards a phenomenology of a digital world. *British Journal of Sociology* 67(1): 118–137.
- Couldry N and Yu J (2018) Deconstructing datafication's brave new world. *New Media & Society* 20(12): 4473–4491.
- Elish MC and Boyd D (2018) Situating methods in the magic of big data and AI. *Communication Monographs* 85(1): 57–80.
- Espeland WN and Stevens ML (1998) Commensuration as a social process. *Annual Review of Sociology* 24(1): 313–343.
- Flichy P (2007) *The internet imaginaire*. Cambridge: MIT Press.
- Fourcade M and Healy K (2017) Seeing like a market. *Socio-Economic Review* 15(1): 9–29.
- Gehl RW (2014) Sharing, knowledge management and big data: A partial genealogy of the data scientist. *European Journal of Cultural Studies* 18(4–5): 413–428.
- Gitelman L (2013) *“Raw Data” is an Oxymoron*. Cambridge: MIT Press.
- Kennedy H (2016) *Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary*. Basingstoke: Palgrave Macmillan.
- Kennedy H and Hill RL (2018) The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology* 52(4): 830–848.
- Kennedy H, Hill RL, Aiello G, et al. (2016) The work that visualisation conventions do. *Information, Communication & Society* 19(6): 715–735.
- Kennedy H and Moss G (2015) Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society* 2(2): 205395171561114.
- Laaksonen S and Pääkkönen J (2020) Between automation and interpretation: Using data visualization in social media analytics companies. In: Engebretsen M and Kennedy H (eds) *Data Visualization in Society*. Amsterdam: Amsterdam University Press.
- Mackenzie A (2013) Programming subjects in the regime of anticipation: Software studies and subjectivity. *Subjectivity* 6(4): 391–405.
- Mayer-Schönberger V and Cukier K (2013) *Big data: A Revolution That Will Transform how we live, Work and Think*. London: John Murray.
- Morozov E (2013) *To Save Everything, Click here: The Folly of Technological Solutionism*. Public Affairs.
- Passi S and Jackson SJ (2018) Trust in data science: Collaboration, translation, and accountability in corporate data science projects. In: *Proceedings of the ACM on Human-Computer Interaction* Vol. 2, CSCW, Article 136, November 28. New York: ACM.

- Pentland A (2014) *Social Physics: How Good Ideas Spread – The Lessons From a new Science*. London: Penguin Press.
- Pinch TJ and Bijker WE (1984) The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science* 14(3): 399–441.
- Sadowski J (2019) When data is capital: Datafication, accumulation, and extraction. *Big Data & Society* 6(1): 205395171882054.
- Stieglitz S, Mirbabaie M, Ross B, et al. (2018) Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management* 39: 156–168.
- Van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- Wiesenberg M, Zerfass A, and Moreno A (2017) Big data and automation in strategic communication. *International Journal of Strategic Communication* 11(2): 95–114.

Author biographies

Juho Pääkkönen is a doctoral candidate in sociology at the University of Helsinki. His thesis work investigates the use of digital big data sources and computational methods in the social sciences and data analytics. He is also affiliated with the Aalto University department of computer science.

Salla-Maaria Laaksonen is a postdoctoral researcher at the Centre for Consumer Society Research at the University of Helsinki. Her research intersects with ICT studies, organizational studies and consumer studies, focusing on topics such as social media analytics in organizations, organizing on digital platforms and organizational reputation narratives.

Mikko Jauho is a senior researcher at the Centre for Consumer Society Research at the University of Helsinki. His background is in sociology, with science and technology studies as a main area of focus. He has published on expertise, public understanding of science, and the sociotechnological construction of risk knowledge.