Four Conceptual Lenses for Journalism Amidst Big Data: Towards an Emphasis on Epistemological Challenges

Oscar Westlund, *University of Gothenburg, Sweden*
and Seth C. Lewis, *University of Oregon, USA*

Introduction

The social world is increasingly quantified, rendered as digital trace information – geolocation, web metrics, self-tracking, social graphs, likes and shares, and much more. Such data may be collected and analysed at ever larger scale amid the growing ubiquity of mobile devices, always-on sensors, 'smart' homes, algorithms and automated systems, digital repositories and archives, and the many fragments of social activity represented by ones and zeroes (Kitchin and McArdle 2016; Mayer-Schönberger and Cukier 2013).[1] Harnessing, combining, manipulating and visualizing such data – techniques that once required supercomputers – increasingly can be accomplished via standardized software or cloud computing (Manovich 2012). In academia, for instance, such contexts have facilitated the rise of computational social science (Shah, Cappella and Neuman 2015). Importantly, big-data developments are not merely a technological transition but truly a *sociotechnical* phenomenon – a complex amalgamation of digital data abundance, emerging analytic techniques, mythology about data-driven insights, and growing critique about the overall consequences of big-data practices for democracy and society (boyd and Crawford 2012). As its opportunities and challenges come into focus, big data promises enormous potential for yielding various types of value – social, cultural and monetary – for the institutions that manage it effectively, while simultaneously posing troubling questions about privacy, accuracy and ethics (Howard 2015; O'Neil 2016; Tufekci 2014).

Like the corporate, scientific and government sectors, institutions of journalism increasingly need to work out how they are to approach big data. News organizations have long been familiar with data and databases – through computer-assisted reporting (CAR) and other forms of information visualizations – both as a form of journalistic evidence and as a type of news product (Anderson 2013a, 2015). But, as Usher (2016)

---

[1] This chapter draws in part on material previously published in Lewis and Westlund (2015a).

describes, journalism is at a constitutive moment in which changes in technology, culture and economics have enabled the rise of 'interactive journalism', which, at its core, is about the application of data, code and software to novel forms of news presentation. Moving forward, the open question 'is not whether data, computers, and algorithms can be used by journalists in the public interest, but rather how, when, where, why, and by whom' (Howard 2014, 4).

In a conceptually oriented research review article on the intersection of big data and journalism in a special issue in *Digital Journalism* guest-edited by Lewis, we argued that big data, in the context of journalism, embodies emerging ideas about, activities for, and norms connected with datasets, algorithms, computational methods, and related processes and perspectives tied to quantification as a key paradigm of information work (Lewis and Westlund 2015a). Importantly, we think of big data as neither good nor bad for journalism. Rather, it is freighted with equal parts potential and pitfall, depending on how it is imagined and implemented – and, crucially, towards what purposes and in whose interests. Scholars have begun to examine and conceptualize big data in the broad context of media and public life, but less has been accomplished with regard to journalism. Just as journalism is trying to make sense of big data, journalism studies must develop conceptual and theoretical toolkits to understand what it means for how news is perceived and practiced.

In that context, our *Digital Journalism* article presents four conceptual lenses for understanding the nexus of big data and journalism: epistemology, expertise, economics and ethics. Each of these addresses relevant questions for news media, helping both to illuminate existing research and to guide future work. Briefly summarized, the concept of epistemology addresses the legitimization of new claims about knowledge and truth. Expertise is concerned with the negotiation of occupational status, authority and skill sets as new specializations are developed and deployed. Economics encompasses the potential for and challenges with new efficiencies, resources, innovations, value creations and revenue opportunities. And, finally, ethics deal with the issues raised by these developments for the norms and values that guide human decision-making and technological systems design. In this chapter, at the editor's invitation, we reprise our discussion of the four concepts, but this time place particular emphasis on *epistemology*, addressing current and future research on big data's implications for journalism and the production of news knowledge. As such, the chapter will first revisit the three conceptual lenses of expertise, economics and ethics before synthesizing current research literature on the epistemology of big data and journalism and, finally, discussing future directions for such research.

The expertise, economics and ethics of big data and journalism

In reviewing expertise, economics and ethics, we first offer a brief concept introduction, followed by a discussion of the concept's meaning for big data and journalism and finally a series of examples or possible applications in the context of news production

and distribution. This systematic approach will be used also in the succeeding section that goes into further details about epistemology.

Expertise

The term 'expertise' comes from the Latin root *experiri*, meaning 'to try', and generally refers to 'the know-how, the capacity to get a task accomplished better and faster because one is more experienced' – hence *expertus*, or 'tried' (Eyal 2013, 869). The sociology of professions field has since long focused on how *expert* (the social actor) is connected with *expertise* (the specialized know-how). There are diverse studies into how occupations (and the experts who constitute them) work to forge and maintain 'jurisdictional control' over the boundaries around a body of abstract knowledge and the application of that knowledge through work practices. Consequently, this allows professionals to claim autonomy, authority and other benefits (Abbott 1988, 60). Expertise thus distinguishes those possessing specialized knowledge *and* experience from those who do not.

The expertise of journalists is being contested by increasingly active and prolific 'produsers', people who both produce and consume media content (Bruns, 2012), as well as new forms of digital aggregation and curation (Anderson 2013b). Moreover, as media users spend more and more time on social media platforms, and as those platforms – Facebook especially – assume greater control over information distribution writ large (Bell 2016), the role of journalist-as-expert-gatekeeper falls into question (cf. Braun 2015). And, unlike law and medicine, journalism has a more tenuous link to formalized expertise, lacking the protective trappings of credentials bestowed via examination. Altogether, journalism's claim to social expertise through its 'professional logic' (Lewis 2012) – a bargain to control the production and distribution of news on society's behalf – is rather beset by challenges to authority (Carlson 2017). Against that backdrop, and alongside Collins and Evans' (2007) typology of expertise that presumes expertise is a 'real' feature of socialization and experience, Reich (2012) has argued that journalists can be understood as 'interactional' experts. Journalists' expertise lies in their ability to work with and among *other* types of experts, ultimately synthesizing and translating others' specialized knowledge for non-experts. Through their interactions and engagement with lay audiences, Reich argues, journalists also develop a *bipolar* form of interactional expertise. Finally, within the sociology of knowledge, there is a greater recognition for a competence-based approach focusing on what people can do, rather than what they can learn (Collins and Evans 2007; cf. Reich 2012). Skills are thus recognized as integral to expert distinction. Next, we revisit these constructivist (Eyal 2013) and normative (Collins and Evans 2007) frameworks for a brief discussion of three concepts important for understanding journalistic expertise amid big data (Lewis and Westlund 2015a).

1. ***Social interactions.*** As noted by Reich (2012), some journalists have primarily source-interactional expertise (e.g. long-time beat reporters) while others primarily have audience-interactional expertise (e.g. editors who hear from readers), and others some combination (e.g. columnists and commentators).

The social, cultural and technological nature of big data may affect the character of these interactions with sources and audiences. Audience metrics can facilitate decisions for news reporting and news distribution (e.g. Tandoc 2014; see also Nguyen and Vu, this volume). News companies are increasingly hiring data scientists to make sense of data – both *data as source material* for journalistic storytelling as well as *data on audiences* for business purposes.

2. *Networked interactions.* Networked interactions involve the more socio-technical interactions of expertise that might occur between journalists (as actors) and machines (as actants – see Lewis and Westlund 2015b). From Eyal's (2013) approach, the growing deployment of algorithms and automation in journalism might entail new arrangements of 'networked expertise', altering how we imagine what it is that journalists know and how they represent that knowledge to the world. In this vein, Anderson (2013b) has shown how the dividing lines of expertise between 'original' reporting and 'parasitic' news aggregation are hardly clear-cut. In fact, networks of social actors and technological actants, when viewed holistically, yield rather complicated renderings of journalistic expertise under different conditions of digitization. Ultimately, as human expertise is increasingly inscribed into technical systems used for news production and news distribution, it challenges what is 'human' and what is 'machine' about such expertise (Lewis and Westlund 2016).

3. *Skill sets.* Expertise is manifest in actual, practical skills (Collins and Evans 2007). Professional approaches to big data prioritize certain skills, such as data analysis, computer programming and visualization, drawn from disciplinary origins such as computer science, mathematics and statistics (Mayer-Schönberger and Cukier 2013). Consequently, news organizations striving to succeed with big data increasingly need expertise in areas such as computer programming and sophisticated back-end databases to comprehend data, and to publish it in ways that allow users to explore the data for themselves (Parasie and Dagiral 2013). *Data* and *code* thus constitute skills-based forms of expertise that news organizations are working to cultivate. Yet bridging the skills gap between journalists and technologists, or helping journalists develop such data-and-code skills, is neither easy nor broadly institutionalized as yet (Howard 2014; Lewis and Usher 2014, Westlund 2011). Journalistic expertise may evolve by having journalists learn to write basic software, and having external 'algorithmists' – expert reviewers of big-data analysis and predictions (Mayer-Schönberger and Cukier 2013) – critique computational journalism in a similar fashion as ombudsmen (cf. O'Neil 2016). Incorporation of such skills into the news production and distribution process may alter the notion of what truly *counts* as expertise in journalism.

Economics

Economics, with its etymology in the 'management of household', is a discipline that studies the behaviours of agents in households and organizations, focusing on how resources are managed to achieve certain ends. Typically, these agents are assumed

to act rationally, making choices about how to use limited resources towards desired outcomes and strategic goals. Applied to the context of communication, *media economics* is defined as 'the study of how media industries use scarce resources to produce content that is distributed among consumers in a society to satisfy various wants and needs' (Albarran 2002, 5), and thus includes questions of media management and media innovations.

Many legacy news media companies around the world – especially local newspapers – have historically benefited financially from their standing as oligopolies or monopolies in a distinct geographical market (Picard 2010). In the twenty-first century, such firms have been facing a shrinking advertising base, fragmenting audiences, and rising competition from mobile, social and digital media (Anderson, Bell and Shirky 2012). Amid a general call for media companies to innovate (Storsul and Krumsvik 2013; cf. Westlund and Lewis 2014), big data represents an opportunity for value creation through revised business processes as well as new products and services. Big data has obvious relevance for business-side revenue opportunities, allowing media companies to better understand and serve particular audiences and advertisers. Nevertheless, these developments come with concerns about the ultimate social and political outcomes, including how targeted advertising based on data mining leads to pressures on (news) media companies to personalize content in response (Couldry and Turow 2014). The march towards big-data personalization, in this view, threatens the very ecology of common knowledge upon which representative democracy depends. To date, and after the 2016 American presidential election, academics and pundits alike have expressed alarm about the pronounced individualization of media behaviour – whether in the form of selective exposure or in unintentional filter bubbles created by algorithm-controlled social platforms (Lewis and Carlson 2016).

Importantly, can big data afford new value creation without undermining the church–state divide that, for many journalists, is central to professional autonomy? Big data promises economic efficiency by enabling 'more observation at less cost' (Crawford, Miltner and Gray 2014, 1666) – as in the case of labour-saving forms of automated journalism (Carlson 2015; Westlund 2013). Moreover, it may be associated with augmenting, rather than displacing, human labour by catalysing new types of technologically enabled forms of news work (Powers 2012), or by allowing journalists to function more like 'knowledge managers' who better gather, organize and analyse disparate information flows in a community (Lewis and Usher 2013).

We hereby suggest two types of applications for envisioning the value creation opportunities for a journalism leveraging big data. The first involves social actors – especially journalists but also technologists – manually drawing upon large datasets to report and present news in ways that differentiate their work from the traditional storytelling paradigm, thereby creating value for distinct publics interested in new types of news and creating a competitive advantage vis-à-vis commodity news in the marketplace. This shift has been called 'method journalism' (Madrigal 2014), moving from an *area of coverage* (a topic, beat or location of interest) to focus instead on the *method of coverage*. Several news startups and initiatives are emblematic of this

change, built around method-oriented objectives, such as FiveThirtyEight and Vox (Madrigal 2014).

The second approach involves journalists and technologists employing technological capacities – in the form of technological actants – to algorithmically gather, link, compare and act upon big data of interest to audiences (cf. Lewis and Westlund 2015b). These algorithms can be tailored to fit with the personalized preferences and behaviours of individuals, promoting specific news articles and news categories to specific individuals (building upon processes described by Thurman 2011). Importantly, social actors within news media firms need to actively assess what technological actants they are to use for such purposes (if acquired from external providers), or how they are to be developed and configured (if developed internally). Diverse social actors need to collaborate in inscribing the technological actants with logics and news values for their operation, ensuring that journalistic values and ambitions are built into automation in a way that suits audiences (cf. Westlund 2012). Such actant-focused automation, as Anderson, Bell and Shirky (2012) have argued, offers an important yet underexplored avenue for news media to cut expenses (e.g. by no longer wasting resources on stories that a robot could write just as well) and simultaneously create value (e.g. by redeploying humans towards projects where they uniquely can contribute).

Ethics

The term 'ethics' derives from the Greek *ethos*, meaning 'character' or 'personal disposition' on the part of the individual, and relates to the Greek *mores* (or 'morals') and its emphasis on the customs of a group. As such, ethics is internally concerned with personal decision-making and externally situated in relation to the rules of society (Ward 2010). Ethics is thus concerned with appropriate practice within a framework of moral principles.

Ethical standards serve an essential function in orienting journalists, especially, to work in ways that promote honesty, accuracy, transparency and public service (Ward 2010). For journalists, ethical codes and conduct serve not only to guide their choices but also to define who they are as professionals. The big-data phenomenon is freighted with its own set of ethical quandaries – about issues such as user privacy, information security and data manipulation (Crawford et al. 2014) – that deserve scrutiny and reflection as journalists determine how to adapt to and along with the series of innovations associated with it. Next we briefly discuss three important areas of concern: (1) data transparency and quality, (2) social science research ethics, and (3) inscription of values into technological systems.

1. ***Data transparency and quality.*** Publishing data and making large datasets publicly available online is linked to transparency. An ethos of openness is shared among many data journalists (Howard 2014), even as journalists broadly have struggled to embrace such openness as a professional norm

(Lewis 2012). Data journalists often seek to make complete datasets and programming code open to public examination and collaboration. Lewis and Usher (2013) have suggested this may lead to journalism reinventing itself, integrating norms like iteration, tinkering, transparency and participation that are connected with the social, cultural and technological framework of digital technologies (cf. Lewis and Usher 2016). There are, however, often underlying problems with public data provided by diverse stakeholders, problems that may go unnoticed either because of the size of the data involved or because of the attractiveness of making it freely available. Crawford (2013), for instance, mentions deep structural signal problems when collecting big data from Twitter, as such data has little or no representation of less-connected communities. Moreover, big-data analysis may come with ethical challenges regarding disclosing private or sensitive information (Howard 2014). Altogether, even well-intentioned efforts to revise journalistic norms through big-data run up against ethical questions embedded in the organization, analysis and dissemination of such data.

2. ***Social science research ethics.*** Journalists have long drawn on social science methods, but they are doing so to a greater extent in this data-rich environment (Howard 2014). Meanwhile, at the same time that journalists are embracing such techniques, 'social scientists are undergoing a fundamental shift in the ethical structure that has defined the moral use of these techniques', rethinking what it means to protect individuals from harm and to allow for informed consent in a world of big-data research methods involving millions of human subjects (Fairfield and Shtein 2014, 38). Journalists, of course, are not subject to institutional review boards (IRB), and yet they should be cautious: just because certain content is publicly accessible does not mean that it was intended to be *made public* to everyone (boyd and Crawford 2012).

3. ***Inscription of values into technological systems.*** Finally, journalists should consider the ethics of technological systems design, or the inscription of values into technological systems (Nissenbaum 2001). To the extent there is a turn towards technology-led practices in journalism (see Westlund 2013), what happens as humans embed technological actants like algorithms with *some* assumptions, norms and values, and not others? To what extent (and how) should technological actants be designed to manage thorny issues such as fake news and manipulation, or otherwise be 'taught' to act ethically? Is there an ethics of algorithms (Kraemer et al. 2011)? Such an ethics will need to unpack various factors of selection, interpretation and anticipation, revealing 'how algorithms structure how we can *see* a concern, why we think it *probably matters*, and *when* we might act on it' (Ananny 2013, 6, original emphasis), all issues of deep relevance for journalistic decision-making. Altogether, this attention to code and structure behind technology encourages a study of the 'black boxes' of big data, in order to uncover matters such as biases, influences and power structures (Diakopoulos 2015).

The epistemology of big data and journalism: Current research[2]

Finally, we address the fourth 'E' of big data: epistemology. Epistemology, as a theory of knowledge, differs from ontology. In the philosophy of science, ontology refers to fundamental inquiries into the nature of existence, that is, the 'science of what is' (Smith 2001, 79). While the world undoubtedly exists in the form of nature, people or events, any attempt to represent the world will, in fact, turn into some sort of re-presentation, with inherent limitations. Epistemology thus points to the nature and boundaries of human knowledge about the world and the determination of truth in that process of re-presentation. The term derives from the Greek *episteme*, which means knowledge, and was developed on the basis of *epistanai*, which means to understand. A fundamental issue in epistemology concerns the work of legitimizing certain types of information as knowledge relative to others. The academy, like other knowledge-producing fields of practice, long has developed epistemologies that shape what counts in this regard (Schon 1995).

Journalism has become, and remains to be, one of the most influential knowledge-producing institutions in society. The concept of epistemology in journalism refers to the norms, standards and methods that determine how journalists know what they know, as well as how they develop and display that knowledge through their production practices and news products (Ekström 2002; Ettema and Glasser 1998). Ekström (2002) has shown how the epistemology of journalism involves the rules, routines and institutionalized procedures that are used in a specific social setting, and which impact how knowledge is produced and how claims about knowledge are explicitly or implicitly expressed. Indeed, how journalists justify their truth claims is just as important as the claims themselves (Ettema and Glasser 1998). Research into news production has shown how journalists develop methods for adjudicating knowledge claims in a routinized fashion (Tuchman 1978), such as by adhering to ideals such as objectivity and practices such as multiple-sourcing (Wiik 2010). Ultimately, epistemology concerns how journalists – and other agents of news production, such as news bots – make assessments of diverse information and sources, determining that the information gathered is sufficiently correct and reliable in order to be published as news (Ekström 2002; Ettema and Glasser 1998; cf. Godler and Reich 2012).

While there are many important studies of digital journalism, few focus on epistemology. Exceptions include Matheson's (2004) early work on the hosting of blogging by the *Guardian*. His approach to epistemology characterized a kind of

---

[2] This section builds on our journal article (Lewis and Westlund 2015a) as well as a grant proposal ('The epistemologies of digital news production') that was successfully funded by the Swedish Foundation for Humanities and Social Sciences in October 2016. The grant proposal was developed together with Professor Mats Ekström at the University of Gothenburg; his important reflections and contributions to our joint grant proposal are reflected in this discussion of epistemology and journalism.

journalism built around online connections more than 'fact' involving interpersonal relations and a different kind of authority, as the public was invited to participate in the construction of knowledge. More recently, Wahl-Jorgensen has analysed matters of epistemology in the study of journalism and emotion (see, e.g., 2015; 2016). A key point in her works is that in relation to the rise of citizen journalism, social media and user-generated content, journalism has developed new conventions of storytelling. These challenge the objectivity ideal and established ways of knowing, and involve new forms of truth claims, giving room for journalism in which emotion and personal experience are key ingredients in the narratives. Exemplifying this, a citizen sharing a personal experience through amateur footage with a mobile device may be seen as truly authentic, perhaps more so than the truth claims of a story produced by a professional journalist (Wahl-Jorgensen 2016). And, other recent research has focused on the intersection of data journalism and epistemology. Borges-Rey (2017) employed three conceptual lenses – materiality, performativity and reflexivity – for an interview study, concluding that while data journalism reinforces norms and rituals, it also displays a distinct character.

Aside from these and a few other studies, relatively little is known about epistemology in contemporary journalism, even while there is great concern among scholars, pundits and practitioners about the direction of journalism in the digital era. There have indeed been substantial shifts in the structural and technological conditions of journalism (see, e.g., Anderson, Bell and Shirky 2012; Meikle and Redden 2010; Westlund 2011, 2012), shifts that relate closely to weakening the financial base of legacy news media and their capacity to provide quality news reporting (Picard 2010; cf. McChesney 2012). In much of the Western world, journalism is under stress: legacy providers are shrinking, some failing altogether, leaving a range of stakeholders – media managers, policymakers and citizens – to wonder about the future of news. In countries such as Denmark, Norway and Sweden, where press subsidy and publicly financed media have long played an important part in the media system, politicians have begun redesigning media policies, launching media inquiries that investigate the situation and propose directions for future media policy.

Against that backdrop, it is fair to ask how journalists are coping with such pressures: is news quality worsening, maintaining status quo or improving? This is an important question because the democratic significance and future legitimacy of journalism is inexorably connected to the extent that journalists succeed in publishing relevant and 'accurate' news (Ekström 2002; Karlsson 2011).

While this chapter does not address the concept of quality per se, it focuses on how journalistic quality relates to epistemology through news production processes. To date, the literature on journalism and epistemology mostly consists of studies of news production processes at newspapers or news broadcasters before the rise of digital journalism (see, e.g., Ekström and Nohrstedt 1996; Ettema and Glasser 1987, 1998). Throughout the 2000s and 2010s there has most likely been not only continuity but also change in journalistic epistemology – that is, how journalists think about and act towards knowledge production and knowledge expression – particularly amid increasingly technologically oriented forms of news production.

How journalists justify the stories and facts they publish appears to vary with different forms of news, as seen in the case of investigative journalism (Ekström 2002; Ettema and Glasser 1987) and more specific sub-genres of reporting (Ekström and Nohrstedt 1996). It matters, therefore, to understand if and how the fundamental principles and work processes of what might be called 'traditional news journalism' have changed – and, in turn, how such changes may be manifest in emerging forms of journalism as well. The introduction of diverse technologies into news work has indeed raised questions about the relative knowledge value that is associated with specific forms of news work such as photojournalism and programmer-journalism (Powers 2012). Here, we will focus on the literature that takes up the intersection of epistemology and (big) data journalism.

The development of computer-assisted reporting (CAR), nearly a half-century ago, pointed to a hope for 'precision journalism' – the potential for achieving greater accuracy through a combination of computer and social science (Meyer 1973). While computers have improved significantly since then, journalists have maintained their established epistemology, which assumes that 'data have no journalistic value on their own' and therefore journalists must work to find the story 'hidden' in the data (Parasie and Dagiral 2013, 859). Such a view fits the normative paradigm of journalists as essential knowledge-producers for society (Lewis 2012). However, questions have emerged about the role that technology might play in developing the capacities for and practices of knowledge production in journalism – for example, in the form of augmented reality for digital storytelling (Pavlik and Bridges 2013) or technological systems for customizing diverse types of news for diverse types of audiences (Westlund 2013). Through it all, the point remains: technology has uncertain consequences for journalistic epistemology.

The extraction, combination and analysis of big data offer improved possibilities for specific forms of journalism, such as investigative journalism (cf. Parasie 2015), and with it a stronger link between the production of social facts in journalism and matters of 'science' and 'precision'. Importantly though, data should never be taken as a proxy for the 'science of what is', in the ontological sense, but rather as one form of epistemological knowledge in which numbers carry great significance. There is no such thing as 'raw data' (Gitelman 2013), and thus (big) data does not represent an objective truth. Figures yielded by big data – even if enormous, robust and highly correlated – require interpretation.

There is a need to study the implications of big data for the epistemologies of news production, on the one hand, and news distribution, on the other hand. In our previous research (Lewis and Westlund 2015b, 2016; Westlund and Lewis 2017), we have discussed the role of diverse agents – social actors and technological actants – in connection to the five stages of news work described by Domingo et al. (2008): access/observation, selection/filtering, processing/editing, distribution and interpretation. The first three stages, especially, are closely connected with journalistic epistemology, even when humans are not tightly involved. For example, a programmer-journalist at *The Los Angeles Times* developed an algorithm to record earthquake notifications,

process such alerts into epistemological facts, and facilitate easy editing and rapid publication.

Altogether, technology and datasets may be implicated in each of the five stages, in each case helping to determine if something is worthy and credible enough for publishing – a key element of epistemology. The access/observation phase may involve 'watchdogging in code' (Stavelin 2014), through which journalists use algorithms to continuously and automatically monitor what politicians are doing. Relevant both to access/observation and selection/filtering, journalists use different data-driven means to check 'facts'. For example, the Truth Teller prototype at the *Washington Post* is used by journalists to fact-check political speech in real-time by applying a speech-to-text algorithm to fact-oriented databases.

Turning to processing/editing, big data and related processes – for example algorithms and automation – complicate traditional notions of journalistic judgement, insofar as the technology adjudicates findings in the data (Carlson 2015). For news distribution, meanwhile, big data is connected with emerging representations of digital journalism such as infographics, interactive data visualizations and customizable probability models (Howard 2014; Usher 2016). These news products, in turn, carry certain epistemological assumptions about how audiences might acquire knowledge, as users are encouraged to 'play' with the data to comprehend a particular and personalized version of the news narrative. For instance, some news organizations have sought to make datasets more accessible, transparent and exploratory for users, in line with the ethos of open-source software and open-government advocacy (Lewis and Usher 2013; Parasie and Dagiral 2013). Processes of news delivery and audience engagement in this big-data context thus present new questions about the legitimation of knowledge in and through such data-driven participation.

The epistemology of big data and journalism: Directions for future research

This chapter has reprised our discussion of epistemology, expertise, economics and ethics as four key conceptual lenses for the intersection of journalism and big data. While representing distinct concepts, these lenses are interrelated when it comes to journalism practice and future research. For instance, research into the epistemology of data journalism is linked to research questions about the skill sets and authority of journalists vis-à-vis other social actors in the organization (i.e. expertise). This, in turn, may connect both to ethics (such as the norms and values that guide what actors and technological systems do) as well as to economics (such as innovation, efficiency and value creation). Having said this, we now turn our focus more exclusively to the nexus of epistemology, data and journalism. Ultimately, a more holistic approach to this area is missing in contemporary research, with a number of key questions waiting to be addressed.

In his analysis of television news, Ekström (2002) conceptualized journalistic epistemology in three parts: form of knowledge (i.e. medium-specific concerns – in his case those associated with television as a media form); production of knowledge (i.e. professional norms and routines); and public acceptance of knowledge claims (i.e. the conditions for social legitimacy). Each of these perspectives, brought into conversation with our data-oriented emphasis, leads to various questions for future research. For instance, how might data journalism be associated with particular types of knowledge claims? What are the institutionalized routines and procedures that social actors adopt to guide the production of data-backed knowledge claims? And, how are such routines conditioned to ensure that claims are legitimate and justified? Ultimately, what is the audience's role in the legitimation of knowledge in and through datasets?

More specifically, as an example, we can outline here directions for a study of epistemological similarities and differences between data journalism and 'traditional news journalism'. Such study can provide insights into changing as well as constant approaches to epistemology, yielding insights about determinations of journalistic quality and thus indications about the future legitimacy of journalism. This means examining, from a sociological perspective, how journalists know what they know; how various knowledge claims are articulated; and how facts and news stories are justified.

The epistemology of what one may call 'traditional journalism' corresponds to long-standing professional practice, yet one that most likely has transformed over time. The emergence of data journalism quite naturally will relate to such established professional practice, remediating old practices and values when approaching the new, while at the same time developing new approaches oriented to the affordances of computer programming, software development, algorithms and datasets – the stuff of big data, in a sense. Importantly, there likely are processes of cross-fertilization between data journalism and traditional journalism, though with the former being seen as 'the new' phenomenon being institutionalized into established practices within 'the old'. For instance, in what ways does data journalism – and the journalists making knowledge claims through datasets, large and small – apply epistemological values and practices from traditional journalism, and vice versa?

Future research should adopt a systematic approach that ensures clarity and coherence –for instance, by analysing journalistic epistemology in the context of news production frameworks like those described above. There is a need for theoretical concepts and questions that may guide the empirical data collection in relation to the stages outlined by Domingo et al. (2008). Following the lead of Ekström and Nohrstedt (1996), we suggest addressing three sets of questions for each stage and form of news journalism. First, what norms, standards and methods are applied when processing information, guiding decisions on whether the information is credible enough for publication or not? Second, what network of sources do journalists turn to and rely on, and how are these assessed? Third, what characterizes the knowledge claims articulated?

Answering such questions generally involves research methods that are qualitative and ethnographic, in line with classic and contemporary newsroom studies (Willig 2013). There are, of course, growing challenges to conducting such studies, particularly

as news work becomes harder to pin down methodologically: It is more embedded in machine systems that may be black-boxed to researchers; it is increasingly technologically dependent, obscuring the converging roles of human and machine; and, as news organizations increasingly rely on freelance labour, news work is less connected to a single site for ethnographic study (see discussion in Deuze 2007; Lewis and Westlund 2016). Altogether, a combination of observations at key sites, interviews with key informants on-site/off-site, and studies of the systems that mediate journalistic judgements is critical for capturing the nature of epistemology in various forms of journalism.

Ultimately, journalism matters to democratic society as a form of knowledge production about public affairs. As such, research into the epistemology of news production matters not simply for what it may contribute to the academic study of journalism and communication. It also speaks to this 'post-truth' moment for media, politics and society: an opportunity to understand how information is legitimated as 'real' or 'fake', to make sense of journalism's co-production of knowledge with publics, to identify dynamics of information production and circulation that lead to various interpretations on the part of audiences. In the end, to study the epistemology of journalism in an era of big data is to assess how technologies and their traces – from the algorithms of social media platforms to open data acquired online – are implicated in *how news is transformed into knowledge*, and what that means for the social authority, trust and legitimacy that may (or may not) be invested in journalism going forward.

References

Abbott, A. D. (1988), *The System of Professions: An Essay on the Division of Expert Labor.* Chicago: University of Chicago Press.

Albarran, A. B. (2002), *Media Economics: Understanding Markets, Industries and Concepts.* Ames, IA: Blackwell.

Ananny, M. (2013), 'Toward an Ethics of Algorithms: Observation, Probability, and Time.' Paper presented at the Governing Algorithms Conference, New York, May 16–17.

Anderson, C. W. (2013a), 'Towards a sociology of computational and algorithmic journalism'. *New Media & Society*, 15(7), pp. 1005–21.

Anderson, C. W. (2013b), 'What aggregators do: Towards a networked concept of journalistic expertise in the digital age'. *Journalism*, 14(8), 1008–23.

Anderson, C. W. (2015), 'Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism'. *Digital Journalism*, 3(3), 349–63.

Anderson, C. W., Bell, E. and Shirky, C. (2012), *Post-industrial Journalism: Adapting to the Present*. New York: Tow Center for Digital Journalism, Columbia University.

Bell, E. (2016), Facebook is eating the world. *Columbia Journalism Review*. http://www.cjr.org/analysis/facebook_and_media.php

boyd, d. and Crawford, K. (2012), 'Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon'. *Information, Communication & Society*, 15(5), 662–9.

Borges Rey, E. (2017), Towards an epistemology of data journalism in the devolved nations of the UK: Changes and continuities in materiality, performativity and reflexivity, *Journalism*. Online-first version, February 2017. Retrieved from https://doi.org/10.1177/1464884917693864.

Braun, J. A. (2015), *This Program is brought to You By...: Distributing Television News Online.* New Haven, CT: Yale University Press.

Bruns, A. (2012), 'Reconciling community and commerce?' *Information, Communication & Society*, 15(6), 815–35.

Carlson, M. (2015), 'The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority'. *Digital Journalism.* 3(3), 416–31.

Carlson, M. (2017), *Journalistic Authority: Legitimating News in the Digital Era.* New York: Columbia University Press.

Collins, H. and Evans. R. (2007), *Rethinking Expertise*. Chicago, IL: University of Chicago Press.

Couldry, N. and Turow, J. (2014), 'Advertising, big data and the clearance of the public realm: Marketers' new approaches to the content subsidy'. *International Journal of Communication*, 8, 1710–26

Crawford, Kate. (2013), 'The Hidden Biases in Big Data'. *Harvard Business Review*, 1 April, http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/

Crawford, K., Miltner, K. and Gray, M. L. (2014), 'Critiquing big data: Politics, ethics, epistemology'. *International Journal of Communication*, 8, 1663–72.

Deuze, M. (2007), *Media Work*. Boston: Polity Press.

Diakopoulos, N. (2015), 'Algorithmic accountability: Journalistic investigation of computational power structures'. *Digital Journalism*, 3(3), 398–415.

Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J. B. and Vujnovic, M. (2008), 'Participatory journalism practices in the media and beyond: An international comparative study of initiatives in online newspapers'. *Journalism Practice*, 2(3), 326–42.

Ekström, M. (2002), 'Epistemologies of TV-Journalism. A theoretical framework'. *Journalism: Theory, Practice and Criticism*, 3(3), 259–82.

Ekström, M. and Nohrstedt, S-A. (1996), *Journalistikens etiska problem* Rabén Prisma.

Ettema, J. and Glasser, T. (1987), 'On the epistemology of investigative journalism'. In: M. Gurevtich and M. Levy (eds), *Mass Communication Yearbook*, 6. London: Sage.

Ettema, J. S. and Glasser, T. L. (1998), *Custodians of Conscience: Investigative Journalism and Public Virtue.* New York: Columbia University Press.

Eyal, G. (2013), 'For a sociology of expertise: The social origins of the Autism epidemic.' *American Journal of Sociology*, 118(4),863–907.

Fairfield, J. and Shtein, H. (2014), 'Big data, big problems: Emerging issues in the ethics of data science and journalism'. *Journal of Mass Media Ethics*, 29(1), 38–51.

Gitelman, Lisa, (ed.) (2013), *'Raw Data' is an Oxymoron*. Cambridge, MA: MIT Press.

Godler, Y. and Reich, Z. (2013), 'How journalists think about facts: Theorizing the social conditions behind epistemological beliefs'. *Journalism Studies*, 14(1), 94–112

Howard, A. B. (2014), *The Art and Science of Data-Driven Journalism*. New York: Tow Center for Digital Journalism, Columbia University.

Howard, P. N. (2015), *Pax Technica: How the Internet of Things may Set Us Free or Lock Us Up.* New Haven: Yale University Press.

Karlsson, M. (2011), 'The immediacy of online news, the visibility of journalistic processes and a restructuring of journalistic authority'. *Journalism*, 12(3), 279–95.

Kitchin, R. and McArdle, G. (2016), 'What makes big data, big data? Exploring the ontological characteristics of 26 datasets'. *Big Data & Society*, 3(1), doi:10.1177/2053951716631130

Kraemer, F., K. van Overveld, and M. Peterson. (2011). 'Is There an Ethics of Algorithms?' *Ethics and Information Technology* 13(3), 251–260. doi:10.1007/s10676-010- 9233-7.

Lewis, S. C. (2012), 'The tension between professional control and open participation: Journalism and its boundaries'. *Information, Communication & Society*, 15(6), 836–66

Lewis, S. C. and Usher, N. (2013), 'Open source and journalism: Toward new frameworks for imagining news innovation'. *Media, Culture & Society*, 35(5), 602–19

Lewis, S. C and Usher, N. (2014), 'Code, collaboration, and the future of journalism: A case study of the Hacks/Hackers global network'. *Digital Journalism*, 2(3), 383–93.

Lewis, S. C. and Usher, N. (2016), 'Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers'. *Convergence: The International Journal of Research Into New Media Technologies*, 22(5), 543–60.

Lewis, S. C. and Westlund, O. (2015a), 'Big data and journalism: Epistemology, expertise, economics, and ethics'. *Digital Journalism*, 3(3), 447–66.

Lewis, S. C. and Westlund, O. (2015b), 'Actors, actants, audiences, and activities in cross-media news work: A matrix and a research agenda'. *Digital Journalism*, 3(1), 19–37

Lewis, S. C. and Carlson, M. (2016), 'The dissolution of news: Selective exposure, filter bubbles, and the boundaries of journalism'. In D. Lilleker, D. Jackson, E. Thorsen and A. Veneti (eds), *US Election Analysis 2016: Media, Voters and the Campaign,* 78. Poole, England: Center for the Study of Journalism, Culture and Community, Bournemouth University.

Lewis, S. C. and Westlund, O. (2016), 'Mapping the human–machine divide in journalism'. In T. Witschge, C. W. Anderson, D. Domingo and A. Hermida (eds), *The SAGE Handbook of Digital Journalism*, 341–53. London: SAGE.

Madrigal, A. C. (2014, June 10), Method journalism. *The Atlantic*. Retrieved from http://www.theatlantic.com/technology/archive/2014/06/method-journalism/372526/.

Manovich, L. (2012), 'Trending: The promises and the challenges of big social data'. In M. K. Gold (ed.), *Debates in the Digital Humanities*, 460–75. Minneapolis, MN: The University of Minnesota Press.

Matheson, D. (2004), 'Weblogs and the epistemology of the news: Some trends in online journalism'. *New Media & Society*, 6(4),443–468.

Mayer-Schönberger, V. and Cukier, K. (2013), *Big Data: A Revolution that will Transform how we Live, Work, and Think.* Boston: Houghton Mifflin Harcourt.

McChesney, R. (2012), 'Farewell to journalism?.' *Journalism Studies*, 13(5–6), 682–94.

Meikle, G. and Redden, G. (eds.) (2010), *News Online – Transformations and Continuities*. London: Palgrave Macmillan.

Meyer, P. (1973), *Precision Journalism: A Reporter's Introduction to Social Science Methods,* 1st ed. Bloomington: Indiana University Press.

Meyer, P. (1973), *Precision Journalism: A Reporter's Introduction to Social Science Method*s. Bloomington: Indiana University Press.

Nissenbaum, H. (2001), 'How computer systems embody values'. *Computer*, 34(3), 120–19.

O'Neil, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Random House.

Parasie, S. and Dagiral, E. (2013), 'Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago'. *New Media & Society*, 15(6), 853–71.

Parasie, S. (2015), 'Data-driven revelation: Epistemological tensions in investigative journalism in the age of "Big data"'. *Digital Journalism*, 3(3),364–380. doi:10.1080/2167 0811.2014.976408.

Pavlik, J. V. and Bridges, F. (2013), 'The emergence of augmented reality (AR) as a storytelling medium in journalism'. *Journalism & Communication Monographs*, 15(1), 4–59.

Picard, R. G. (2010), 'A business perspective on challenges facing journalism'. In David. A. L. Levy and Rasmus. K. Nielsen (eds), *The Changing Business of Journalism and its Implications for Democracy*. Reuters Institute for the Study of Journalism, University of Oxford.

Powers, M. (2012), '"In forms that are familiar and yet-to-be invented": American journalism and the discourse of technologically specific work'. *Journal of Communication Inquiry*, 36(1), 24–43.

Reich, Z. (2012), 'Journalism as bipolar interactional expertise'. *Communication Theory*, 22(4). 339–58

Schon, D. A. (1995), 'The new scholarship requires a new epistemology'. *Change* 27(6), 27–34.

Shah, D. V., Cappella, J. N. and Neuman, W. R. (2015), 'Big data, digital media, and computational social science: Possibilities and perils'. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 6–13.

Smith, B. (2001), 'Objects and their environments: From aristotle to ecological ontology'. In *The Life and Motion of Socio-Economic Units*. In F. J. Raper and J.-P. Cheylan (eds), 79–97. London: Taylor and Francis.

Stavelin, E. (2014), *Computational Journalism: When Journalism Meets Programming*. Unpublished dissertation. Bergen: University of Bergen.

Storsul, T. and Krumsvik, A. H. (2013), *Media Innovations. A Multidisciplinary Study of Change*. Gothenburg: Nordicom.

Tandoc, E. C. (2014), 'Journalism is twerking? How web analytics is changing the process of gatekeeping'. *New Media & Society*, 16(4), 559–75.

Thurman, N. (2011), 'Making "the daily me": Technology, economics and habit in the mainstream assimilation of personalized news'. *Journalism*, 12(4), 395–415.

Tuchman, G. (1978), *Making News: A Study in the Construction of Reality*. New York: Free Press.

Tufekci, Z. (2014), 'Engineering the public: Big data, surveillance and computational politics'. *First Monday*, 19(7). doi:10.5210/fm.v19i7.4901

Usher, N. (2016), *Interactive Journalism: Hackers, Data, and Code.* Champaign, IL: University of Illinois Press.

Wahl-Jorgensen, K.. (2016), 'Emotion and journalism'. In T. Witschge, C. W. Anderson, D. Domingo and A. Hermida (eds), *The SAGE Handbook of Digital Journalism*, 128–43. London: SAGE.

Wahl-Jorgensen, K. (2015), 'Resisting epistemologies of user-generated content? Cooptation, segregation and the boundaries of journalism'. In M. Carlson and S. C. Lewis (eds), *Boundaries of Journalism: Professionalism, Practices, and Participation.* New York: Routledge.

Ward, S. J. A. (2010), *Global Journalism Ethics.* Montreal: McGill-Queen's University Press.

Westlund, O. (2011), *Cross-Media News Work: Sensemaking of the Mobile Media (r)*
     evolution. Gothenburg: University of Gothenburg.
Westlund, O. (2012), 'Producer-centric versus participation-centric: On the shaping of
     mobile media'. *Northern Lights*, 10(1), 107–21.
Westlund, O. (2013), 'Mobile news: A review and model of journalism in an age of mobile
     media'. *Digital Journalism*, 1(1), 6–26.
Westlund, O. and Lewis, S. C. (2014), 'The agents of media innovation: Actors, actants,
     and audiences'. *The Journal of Media Innovations*, 1(2), 10–35.
Westlund, O and Lewis, S. C. (2017, forthcoming), 'Reconsidering news production: How
     understanding the interplay of actors, actants, and audiences can improve journalism
     education.' In R. Goodman and E. Steyn (ed.), *Global Journalism Education: Challenges*
     and Innovations in the 21st Century. Austin, TX: The Knight Center for Journalism in
     the Americas at The University of Texas at Austin.
Wiik, J. (2010), *Journalism in Transition. The Professional Identity of Swedish Journalists*,
     JMG Book Series no. 64. Gothenburg: University of Gothenburg.
Willig, Ida (2013), 'Newsroom ethnography in a field perspective'. *Journalism,* 14(3),
     372–87.