Article:

Tong, J. orcid.org/0000-0001-6843-3177 and Zuo, L. (2021) The inapplicability of objectivity: understanding the work of data journalism. Journalism Practice, 15 (2). pp. 153-169. ISSN 1751-2786

https://doi.org/10.1080/17512786.2019.1698974

Accepted version

The inapplicability of objectivity: understanding the work of data journalism[1]

Jingrong Tong

University of Sheffield

j.tong@sheffield.ac.uk

To cite: J. Tong and L. Zuo (2021) The inapplicability of objectivity: understanding the work of data journalism. Journalism Practice. 15(2), page 153-169. DOI: 10.1080/17512786.2019.1698974

Abstract
Data journalism is an emerging form of journalism, entailing the discovery of stories in data with the assistance of data algorithms. The burgeoning literature has largely interpreted the work of data journalism through the lens of objectivity. This paper, however, rejects the applicability of objectivity to data journalism. This inapplicability is the product of five factors: the extensive use of data and data algorithms in journalism; the unverifiability of data sources; the imbalance in data and data access; data journalists' insufficient knowledge of data contexts and algorithms; and their "design subjectivity" in the data processing process. Data reporting becomes a process of knowledge construction under the influence of these factors. The article argues that because of the "social constructionist" nature of data journalism, serving the public interest and democracy is a more appropriate principle than objectivity for data journalism. It suggests shifting academic attention from celebrating objectivity in data journalism to examining the epistemology of data journalists, the factors influencing data journalists' formation of knowledge in reporting, their defence of cultural authority, and the democratic meanings of data reports in future research. Such understanding also has implications for data journalism pedagogy and practice.

Keywords: data journalism, data, algorithms, objectivity, social construction of reality, design subjectivity

Data journalism is an emerging form of journalism which discovers stories in data with the assistance of data algorithms[2]. Since around 2008, data journalism has been increasingly integrated into newsrooms in and beyond the English-speaking world. For example, news outlets such as the *Guardian* and the BBC (both in the UK), the *Irish Times* (in Ireland), the *New York Times*, the *ProPublica* (both in the USA) and the *Süddeutsche Zeitung* (in Germany) have pioneered the practice of data journalism. Applying the principle of objectivity to data journalism, scholars (such as Parasie and Dagiral, 2012; Tandoc and Oh, 2017; Hammond, 2015; Cushion et al., 2017) argue that the extensive use of data and data algorithms helps improve objectivity in journalism.

By contrast, this article contends that objectivity is inapplicable to data journalism. The unprecedentedly extensive use of data and data algorithms in journalism; three problems surrounding data and data algorithms - the unverifiablity of data, the imbalances in data (representation) and data access; and data journalists' insufficient knowledge of data contexts and algorithms - as well as the "design subjectivity" on the part of data journalists in the data processing process challenge the appropriateness of the notion of objectivity for data journalism. Technically we may be able to find more "objective" evidence and less human involvement in data reports than in traditional reports, and attitudinally data journalists may want to remain objective. Nevertheless, the social construction of data and data algorithms, as well as data reporting being a process of constructing knowledge in and from data, make the notion of objectivity inapplicable to data journalism. This paper calls for academic attention to the "social constructionist" nature of data journalism, which has implications for data journalism research, pedagogy and practice.

Objectivity and journalism

Objectivity is one of the core journalistic norms and ideals, in particular in the American classic model of professional journalism and journalism practised in news agencies such as Reuters and public broadcasters such as the BBC (Hampton, 2008; Schudson, 2001). Objectivity stands opposite to the subjectivity of reporters (McQuail, 1992; McQuail, 1994). This term means the separation of personal opinions and values from facts and the removal of the subjective judgement of journalists about their reporting objects. Journalists are expected to balance reporting on different views like a "see-saw" and not to make judgements about opposing viewpoints cited in their reports (Mindich, 2000; Schudson, 2001; Schudson, 1978; Westerståhl, 1983). To summarise, "no subjectivity", "balance", "hard facts" and the absence of "value judgements" are the "four dimensions of objectivity" (Skovsgaard et al., 2013: 25-26). In addition, the "integrity" and willingness of journalists to commit to this norm is thought of as crucial in the success of objectivity (Ryan, 2001).

Objectivity is an institutionalised occupational norm. In the discussion about the application of objectivity in Reuters and the BBC, for example, Hampton argued that it was established as a "corporate" norm out of concerns both of "market niche" and "legislative mandate" in the 20th century (Hampton, 2008: 476-482). Further, objectivity is seen as "a cornerstone of the professional ideology of journalists in liberal democracies" (Lichtenberg, 1996: 225–42; Schudson, 2001; Carlson, 2007). At the heart of the identity of journalists, the norm of objectivity is accepted by individual journalists as a core part of their professional norms (Maras, 2013). In countries such as Germany, Britain, Italy, Sweden, Denmark and even China, for instance, objectivity is an important norm of journalists, although journalists may have different interpretations of this concept in ways that are pertinent to their specific context (Donsbach and Klett, 1993; Skovsgaard et al., 2013; Zhang, 2014; Tong, 2015).

In spite of its importance, the validity of the concept of objectivity is open to question for a number of reasons. Some scholars (such as Glasser and Ettema, 1989; Merrill, 1990) regard objectivity as only applicable to certain types of news reporting, such as "straight news reporting" (e.g. hard news reporting) but inapplicable to investigative journalism or advocacy journalism (Gauthier, 1993). For example, in the context of environmental reporting, which is often seen as advocacy journalism, objectivity is criticised for being inauthentic, as paying equal attention to both sides of arguments may lead to the inaccurate representation of reality and a loss of meaning in reports (Bavadam, 2010). Other scholars (such as Carlson, 2007; Lewis, 2012; Boudana, 2011; Tuchman, 1972) see the claim to objectivity as a strategy used by news organisations to control journalists or a tactic employed by journalists to avoid bearing responsibility for conveying opinions in their reports or to defend their journalistic authority. Even in the context of the United States, more recent research has revealed that journalists may not necessarily embrace objectivity as the most important journalistic principle, because factors, such as online journalism experience and gender, may shift their support of professional norms away from objectivity towards transparency (Hellmueller et al., 2013).

One major criticism of objectivity lies in the doubts of scholars about whether or not it is possible to achieve objectivity (Boykoff and Boykoff, 2007; Tuchman, 1972; Boudana, 2011; Bell, 1998). An important explanation of its unachievability is "the inevitability of 'social construction' of reality in any system" (Schudson, 1989: 274). Constructionist scholars disagree with the notion of objectivity in journalism. The theory of "social construction of reality" is the entirely opposite way of interpreting the work of journalism, seeing news production as a process where reality is constructed in and by news as a result of a series of decisions about the selection, inclusion and exclusion of information (Berger and Luckmann, 1966; Tuchman, 1972; Tuchman, 1978; Cohen and Young, 1973; Molotch and Lester, 1974; Gans, 1980). A wide range of factors from ideology, political interests, economic concerns, to journalists' personal interpretations may influence the way reality is constructed by news production (Schudson, 1989).

This type of criticism about objectivity is therefore very much concerned with the factors that are involved in actual journalistic practices and may influence the news which journalists produce. Merrill, for example, raises three points about the intractable problems of applying objectivity in journalism: the impossibility for journalists to report the complete context of a news story, the selections made by journalists, and the issues surrounding "the whole truth" (Merrill, 1990; Merrill, 1984: 104; also cited in Gauthier, 1993). The aspirations and efforts of journalists towards objectivity will not solve these problems (Lowenstein and Merrill, 1990).

Repudiating objectivity does not mean constructionist scholars are accusing journalists of reporting dishonestly. Instead, they argue that we need to evaluate journalism and its roles by using value systems other than objectivity. They consider serving democracy and the public interest and maintaining the health of the public sphere as a more important role for journalism. Public journalism, for example, is regarded as an alternative to the ideal of objective journalism (Ryan, 2001; Brewin, 2013). Journalism is expected to "help public life" and be “a primary force in the revitalization of public life” (Merritt, 1995). Adversarial journalism, in Ettema and Glasser’s characterisation of “custodians of conscience” is another type of good journalism (Ettema and Glasser, 1998). Transparency is deemed as important for interpretative reporting, which stands in opposition to objective reporting (Rupar, 2006).

Celebrating objectivity in data journalism

Barely giving attention to the “social constructionist” nature of data journalism, the burgeoning literature of data journalism has widely celebrated objectivity. A popular scholarly view (such as Parasie and Dagiral 2012; Tandoc and Oh, 2017; Carlson, 2019) sees journalistic norms such as objectivity as still evident in, and important for, data journalism. In spite of having recognised that incorporating data in journalism challenges the regime of objectivity, some scholars (such as Lesage and Hackett 2014: 42) believe objectivity continues to be relevant and the production of objectivity depends on “how journalists and other actors choose to work with data”. The heavy dependence on data, the involvement of “minimal personalisation”, the exclusion of the personal opinions of journalists and human news sources in news reports show that data journalism still adheres to objectivity and other traditional news values (Tandoc and Oh, 2017).

The use of data and algorithms is seen as making data journalism more objective and accurate (Meyer, 1991; Parasie and Dagiral 2012; Tandoc and Oh, 2017). This view, which can be traced back to precision journalism (Meyer, 1991), celebrates the promise of data and algorithms for achieving objectivity in journalism. Being free from “the conventional thinking and inherent biases implicit in the theories of a specific field” (Mayer-Schönberger and Cukier, 2013: 71), big data offers “renewed promise of objectivity” (Parasie, 2015: 365). The use of data and algorithms is thought of as being neutral, mitigating media bias, and thus improving the level of objectivity in news reports (Borges-Rey, 2016; Karlsen and Stavelin, 2014). For example, a content analysis of 260 data journalism stories published by the *Guardian* shows that data journalism reports rely on “the data to speak” rather than citing human news sources which results in “he-said/she-said reporting”(Parasie and Dagiral, 2012). This view also sees the use of algorithms as helping prevent human journalists from making interpretative errors and subjective judgements. The use of “figures and statistical analysis” is regarded as a sign of objectivity (Parasie and Dagiral, 2012). Therefore the practice of data journalism is largely thought of as being more objective and accurate than that of traditional journalism.

While it has its own merits, this popular view has not fully considered the problems associated with the substantial use of data and data algorithms, as well as the actual mechanics of data processing. These problems shake the foundations of the applicability of objectivity in data journalism but instead direct our attention to the “social constructionist” nature of data reporting. The remainder of the paper will discuss these relevant problems and their implications for data journalism.

The extensive use of data and data algorithms in data journalism

Journalism is always about handling data. However, the degree of data use in data journalism is unprecedented. By data, we mean any content or information that can be digitised. In most cases, data means structured, machine-readable data, or data that can be converted to structured, machine-readable data. But data can also include unstructured data, which requires particular skills and techniques to process. For example, a .csv table containing crime statistics is a form of structured data, which can be read by the machine, i.e. the computer; leaked documents about offshore company records can be subjected to optical character recognition (OCR) and made into pdfs, but in themselves they are unstructured data, which cannot be read by the machine. Therefore, advanced coding skills and techniques may be needed to process this kind of unstructured data. By data algorithms, we mean computational algorithms and other tools or applications that are used to process data. Excel, Access, Googlesheets, R, Python, OpenRefine, decision trees or z-test are all examples of data algorithms.

The extensive use of data and data algorithms in data reporting differentiates data journalism from traditional journalism in two ways. The first is about sources of information. Traditional journalists work on information from both primary and secondary sources. However, in theory, data journalists largely gain their knowledge about topics from data, which is obtained from secondary sources, as the first step leading to reporting (Sheridan Burns and Matthews, 2018). In addition, what data journalists handle for the purpose of reporting should ideally be mostly “raw” data rather than pivot tables or summary numbers provided by other professionals such as statisticians. “Raw” data refers to datasets that comprise detailed information collected for certain purposes, which has not yet been analysed. Most of the time (except when they cannot get hold of raw data), therefore, data journalists process and analyse “raw” data by themselves instead of relying on statisticians to analyse the data and provide findings for them, as exemplified in the series of riot reports published by the *Guardian* in 2011[3], the Teen Gun Violence[4] and NRA grants[5] reports covered by the Associated Press and the Colour of Debt commissioned by the *ProPublica* in 2015 (Kiel and Waldman, 2015). The need to get “raw” data is also one important reason why data journalists send Freedom of Information (FoI) requests to governments.

The fact that data journalists gain knowledge about the subjects of their reports from their own analysis of “raw” data means a new way of knowledge accumulation and production. Data journalists analysing data suggests they take on the role of data analysts. Finding stories in data through analysis and visualisation turns them into data interpreters. These new roles imply that their role in producing the knowledge about what happened is far more than their traditional journalistic role as reporters.

These new roles are conducted in tandem with a heavy use of data algorithms, which are essential in processing data, especially large-scale data. The use of data algorithms starts from the data collection stage through to the data visualisation stage. Data algorithms adopted by journalists “extend” their brains, borrowing the term used by McLuhan (McLuhan, 1964), and enable them to find stories in data. In this sense, data journalists practise journalism in a technological environment constructed by data and data algorithms of their choice, which is the second distinguishing feature of data journalism. As a result of such heavy use of data and data algorithms, there may be a considerable influence exerted by the problems posed by data and data algorithms as well as the subjectivity involved in the data processing process on the knowledge produced in their reports.

Three problems of data and data algorithms

There are three inter-connected problems related to the social constructionist nature and use of data and data algorithms: the unverifiablity of data; the imbalances in data (representation) and data access; and data journalists’ insufficient knowledge of the contexts of data and algorithms. These subtly but fundamentally influence the work of data journalism.

The unverifiability of data concerns the credibility and neutrality of the sources of data, in particular pre-existing datasets such as open data, public (record) data and leaked data. The reliability of data sources, certainly not being a problem exclusive to data journalism, is a pervasive issue in the current data-savvy environment. In spite of the unprecedented scale of data at our disposal (Anderson, 2008), data is sometimes full of errors (Bradshaw, 2014), and complete verification is often impossible. In the Opportunity Gap investigation into unequal access to educational opportunities in the USA, for example, the reporters "spent several weeks verifying the accuracy of the data". They found some problems in the data and even pushed the relevant office of the Department of Education to change "its process for gathering and verifying their data". However, they still could not be entirely sure about the credibility of the data, as they stated "we may not have accounted for every problem in the data".[6] When it comes to leaked data, the problem of unverifiablity is extraordinarily severe, as it would be 'mission impossible' for journalists to be completely certain of its credibility.

The huge difficulty in verifying data thus poses a fundamental challenge to the practice of data journalism, as journalists are expected to verify the information included in their reports. Observers and scholars have noticed the influence of the quality of data on the objectivity and accuracy of data journalism. For example, they (such as Parasie, 2015; Bradshaw, 2014) argue that the messiness and inaccuracy of data, which is beyond the control of data journalists, means journalists cannot guarantee the accuracy of their reports. These discussions contribute to our knowledge of the influence of data on data journalism. They however have not fully acknowledged that the unverifiability of data also results from the social constructionist nature of data.

Existing power and social relations shape what data exists and is used, how data is collected, curated and entered into databases, who collects and processes data, and who may access which data (Lupton, 2013; Manovich, 2012; Ruppert, 2012; Ruppert et al., 2017; Burns, 2015; Walter, 2010; Bolin and Schwarz, 2015). The social construction of the content and presentation of data impacts the credibility and meaning of data. For example, traffic data may be influenced by the distribution and operation of road sensors, the incident reports by the police, as well as how people create and curate the data before publishing it. It is uncertain if the number "3" on a spreadsheet under the heading of "traffic incidents" means there were only 3 incidents or only 3 incidents that had been reported by the police (but actually there may be five more incident reports that are still sitting on the desks of the police). Verifying the credibility of data would require data journalists to know this background information, which however is invisible in datasets. In this sense, the unverifiability of data originates from the "black box" of how data is produced, which is also related to the second problem.

The second problem is the imbalance in data (representation) and data access. The data (representation) imbalance refers to the problem of injustice in the production of data (Taylor, 2017; Johnson, 2014; Johnson, 2016; Heeks and Renken, 2018). Due to different levels of technological integration, different use of technologies as well as their existing social, economic and political positions and status, some individuals or social groups or areas may be included and represented more than others in data. In addition, data collection and production may be influenced by existing social dynamics and biases. A typical example of this is that when filling in a survey used to collect census data, by default, participants would only have two choices - male or female - to indicate their gender,[7] which means the data produced by the survey would actually reproduce the traditional social institution of gender. Its implication for data journalism is that data journalists may investigate the reality about what is included and visible in the data more than what is excluded from the data; and their reports may reproduce the systematic inequalities and biases embodied in data.

Another aspect of data imbalance refers to the imbalance in the access to data and in the provision of data by data holders. What data journalists can get hold of and then interrogate is limited or even decided by the accessibility of data. Different countries have different levels of data transparency, and the openness of societies also varies. Data journalists can get data in one country but may not be able to get the data about the same topic in another country. Even in a country which has freedom of information (FoI) laws or regulations, such as the UK or the US, not all public-record data is published as open data. In some cases, government bodies may not provide or may delay providing the data requested by journalists. Although public authorities are required to respond to FoI requests within 20 working days in the UK and US contexts[8], it is not guaranteed that data journalists can get the data in time or that they can get the data at all. For example, the EastHerts Council (in the UK) explains that there are some circumstances, under which they can refuse to respond, such as exceeding the 18-hour limit for time taken by staff to put the data together.[9] While talking about obstacles in the *Boston Globe* project investigating poor oversight and management by the Federal Aviation Authority (FAA) (in the USA), Jaimi Dowdell reckoned one main problem was caused by the delay or failure of governments to respond to their FoI requests.[10] Accessing some datasets but not others or accessing only part of the data may prevent journalists from understanding the full picture of a situation.

The third problem refers to insufficient knowledge obtained by data journalists of data contexts and algorithms. Knowledge insufficiency first of all means data journalists may not have the knowledge about the contexts of data collection and production. As discussed above, data is far from being objective and can be subjective (Lesage and Hackett, 2014). Particular data is collected and produced within a particular context for a particular purpose. Data is meaningless out of context. An appropriate understanding of data requires journalists to understand the context of data, without which, the interpretation of the chosen data may be distorted. One effective way of understanding the context of the data is to access the comprehensive metadata of the data, which is the detailed information about the process and context published by data holders. However, in reality, metadata may not be available. Data journalists may need to talk to the people who have close relationships with the data such as statisticians in the Office for National Statistics (ONS) or the US Census Bureau in order to understand the data and its context. However it is uncertain whether and to what extent they can be given the full context and gain a full and accurate understanding of the context and the data. This epistemological limitation may be an old problem, but in this context, the consequences of this limitation may be subtler in the face of “objective” evidence from data presented in reports.

Knowledge insufficiency also refers to an insufficient understanding of how data algorithms work. Processing data requires the use of computational algorithms that journalists choose to use, such as Google's search algorithms and social media companies' API algorithms. Like data, however, power relations also influence the creation and neutrality of algorithms (Walker, 2005). Some scholars (such as Diakopoulos, 2015; Zarsky, 2016) have noticed algorithmic bias created by algorithmic tools and decisions. Gillespie sees algorithms as "a new knowledge logic" which is "socially constructed and institutionally managed" (Gillespie, 2014: 192). The rules of these algorithms influence the findings of data journalists invisibly, which may not even be known by data journalists. The lack of sufficient knowledge of how algorithms work may hinder data journalists from making correct judgements about the findings and forming an appropriate understanding of the situation.

Because of the three problems surrounding data and data algorithms, even if data journalists want to remain neutral, the objectivity and accuracy of their work is undermined by issues with the data itself and with data algorithms, and this is beyond their control and cannot be entirely satisfactorily dealt with.

Choice-making in the data processing process

From choosing topics to choosing information or quotes, practising journalism is about making choices (Thomas, 2016; Hanusch, 2014; Perreault and Vos, 2018; Thomas, 2017). So is the practice of data journalism. Just like traditional journalists, data journalists need to make numerous choices about topics, data, data algorithms, approaches (such as hypothesis- or data-driven approaches) to data analysis (Parasie, 2015; Borges-Rey 2017), angles and so on. Making choices in this process is subjective and has the potential to influence the reports of data journalists, which reflect their interpretation of the data. This echoes other scholars' arguments about data processing, such as "in reality, working with Big Data is still subjective", and there is the "inherently subjective" process of data cleaning and the subjective interpretation of data (Boyd and Crawford, 2012: 667; Bollier, 2010; Lesage and Hackett, 2014).

The remainder of this section outlines the major types of choices that data journalists need to make in the five stages: data collection, cleaning, storage, analysis, and visualisation and how these choices are guided by reporting objectives but may in turn influence how they interpret data and what stories are found in the data.

Data collection

Data journalists start the data processing process by choosing which data to collect and how to collect it. On some occasions, data is made available, or, in some cases, leaked, to data journalists. The Panama Papers investigation is an excellent example of this, where a whistle-blower leaked a huge amount of data to journalists. However, in most cases, data collection is necessary. Data can be collected in different ways: downloading pre-existing and open-source spreadsheets from websites such as government bodies' websites, obtaining public data through FoI requests, manually collecting data from different sources and mashing them up onto one spreadsheet, automatically scraping data from the Internet, collecting data by using social research methods such as surveys and so on.

Different ways of collecting data imply different data coming to be at journalists' disposal. A simple example is the practice of Internet search. The selection of key words typed in to search on Google influences search results, which is the data that forms the basis of data journalists' analysis and influences the stories told in their reports. Likewise, using particular keywords or indicators to collect social data via APIs has the same effect.

Regardless of the means by which data journalists collect their data, their decisions about which data to collect and which ways to collect data are firstly decided by their reporting aims and objectives, or in other words, which questions they are asking. This makes data collection a subjective process. Take the Colour of Debt report commissioned by the *ProPublica* in 2015, which examined the disparity in the pursuit of lawsuits against debtors from white and non-white communities: this aim led the journalists to collect court case data from selected geographical areas and to decide what to include or exclude (Kiel and Waldman, 2015). In its Flatshare Bias investigation, the *Guardian* journalists did a survey to collect the data, in order to "test whether people with names associated with a specific religion were treated differently when applying for a house or flatshare" (Duncan, 2018). In data reports like this, the data was collected to serve a particular aim and it was influenced by that aim and by actual procedures, such as how the requests were worded and how the requests were sent in the survey, which are all the result of subjective choices. It is thus crucial to increase the transparency of the reporting process. Publishing the methodology as the reporters in these cases did is one plausible solution, although it does not mean the removal of the subjectivity involved in the choices made in the process.

Data cleaning and mashing-up (preparation)

Data journalists clean and mash up data to prepare it for analysis (Wickham, 2014). The choices about how to do this: to add, remove or edit data are driven by the reporting aim(s) of journalists, but they can influence the findings journalists will get. In the Colour of Debt investigative report, for example, reporters "grouped the suits by census tracts" and "added census data like race and median household income" to each tract, because they wanted to "explore whether these suits disproportionately impact black communities" (Kiel and Waldman, 2015; Waldman and Kiel, 2015). Reporters in this case chose to design their dataset in this particular way so that the dataset allowed them "to compare the per-capita judgment rates in mostly black neighbourhoods to that of mostly white neighbourhoods while holding income constant" (Waldman and Kiel, 2015: 2). The data was cleaned with an aim of meeting their reporting objectives. A different way of preparing the dataset therefore may have resulted in a different story about the topic. For example, if the reporters' interests concerned the potential differences made by gender or age, they might have brought in demographic data on gender and age. The reports might then have addressed the problems associated with gender or age inequality rather than racial inequality.

Data storage

More choices need to be made in the stage of data storage, guided by the objectives of data analysis and the reporting aim(s) of journalists as well as the scale, level of complexity and sources of data. Especially when dealing with large-scale datasets, journalists need to design particular structures or use particular data storage applications, which often have specific features, to create a database and store data in it. For example, data journalists who would like to explore the relationship between different types of entities in the database may want to store their dataset in relational databases such as MySQL, SQL Server or Oracle. The entities can be users, messages, geo-information, and payment figures and so on. In the Unfit for Duty investigation by the *Sarasota Herald-Tribune* in 2011, for instance, the data was stored in SQL Server to help build a flexible and searchable database which can be used to quickly aggregate tables and find information about police officers.[11] However if they were interested in text-based analysis, then it might be better for them to choose to store the same dataset in other tools, for example in Elasticsearch, which would allow them to explore large-scale text rather than the relationship between data entities. The specific way data is stored and the particular tools used will influence how data is analysed and the outcomes of data analysis. This is because different storage tools have different features and emphases.

Data analysis

In the data analysis stage, the reporting aim(s) of data journalists play a strong role here in influencing their choices of analytical aspects, analysis applications or statistical models. As a result of their choices, certain aspects of the data are prioritised and stressed over others. The design of the analysis shapes the results generated from it, which are the basis of data journalists' understanding of the topic.

Selected analytical tools have their own features and strengths. Things may be missed by one technology but can be picked up by another technology. MySQL, which is both a data storage and analysis application, for example, can provide journalists with an understanding of relationships but may be inefficient at offering an insight into the patterns in text in the data. On the other hand, using text mining applications like Kibana (with its matching data storage application Elasticsearch), or those like SPSS and Excel, which have strong statistical analytical functions, journalists could well find it difficult to grasp relationships between different types of concepts and entities. Therefore, while the choice of analytical tools is heavily influenced by the objectives of the analysis and the aim(s) of reporting, the strengths and weaknesses of tools may greatly shape how journalists understand the topic.

Apart from tools for data analysis, selecting particular features or aspects of the data to analyse, and statistical models, such as regression analysis, or correlation analysis, may also make a difference in generating results. An example is the Machine Bias investigation commissioned by the *ProPublica*. The explanation provided of their analysis clearly suggests the reporters decided to analyse particular aspects of the selected data but left out some other aspects: 'We analysed the COMPAS scores for "Risk of Recidivism" and "Risk of Violent Recidivism." We did not analyse the COMPAS score for "Risk of Failure to Appear."' They selected the data according to their criteria, including their definition of "recidivism". They tested racial disparities by running statistical models such as regression and logistic models and "a Cox proportional hazards model" (Larson et al., 2016). It is good to publicise the actual analysis process to increase transparency, although it remains uncertain how and to what extent their chosen focus and the aspects they selected to analyse have influenced their story. For example, the analysis of the COMPAS score for "Risk of Failure to Appear" and tests of disparities of other demographic factors such as gender or age may have led to the publication of different stories.

The use of "objective" statistical models or tools however does not mean data analysis is absolutely objective. Like data storage and analysis applications, each statistical model has its own strengths and weaknesses as well as priorities. They may be free from intellectual perspectives but they are not free from prioritising certain aspects of the data over others. In the science domain, Press and Tanur argued that even if different scientists were given the same data to analyse, their choices of "statistical models" or "assumptions about these models" might produce different conclusions (Press and Tanur, 2012). Likewise, Johnston argued choices made in classification turn it into "a subjective process", no matter what objective methods are used (Johnston, 1968).

Data visualisation

The stage of data visualisation involves journalists' choices of applications, types of visualisation, colours, scales, parameters, variables, value thresholds and so on. Not only the choice of bar charts, lines, 3-Ds, or maps but also that of different statistical functions or models and how to visualise them can make a difference in terms of producing the meanings of the data. Choosing to cluster a number of variables together produces a different result from deciding to visualise them separately. When mapping networks, visualising the networks by the degree of nodes can yield a network graph that is different to that produced by the weight (size) of nodes. A word cloud made up by the frequency of words will differ from that resulting from the significance of words. Therefore, choosing to visualise the data in a particular way may influence the presentation of the findings. "Facilitating understanding" (Kirk, 2016), data visualisation can be an effective way of communicating the meanings of the data to audiences, including data journalists themselves who are the first audiences of the data visualisations they produce. With appropriate visualisation, the meanings of the data can be suitably revealed and conveyed, while improper visualisation may distort the meaning of the data. The examples given by the *Economist*, for instance, show how the wrong choices of scale, colour, visualisation methods, and space may lead to misleading data visualisations. [12] In effect, the choices made at this stage will influence the data visualisations produced and thereafter the data journalists' own understanding of the data and the stories they want to present.

The real issue about objectivity

The discussion above shows that, comprising as it does numerous subjective human choices, the data processing process is an arena of subjectivity and choice-making, which leads to "design subjectivity" - the subjectivity of data journalists involved in the process in which data journalism projects are designed.

The process of data collection, cleaning, storage, analysis and visualisation in data reporting is a black box. Positive progress is being made toward breaking the black box open and being transparent. Although having not published their metadata, data journalists have started to publish their methodologies, and even datasets in some cases, to increase transparency. In addition, news organisations are also trying to revise their practice and ethical codes to deal with potential ethical problems surrounding data journalism. In 2017, for example, the Associated Press revised its stylebook and included data journalism as a new chapter[13]. It requires data analysis to be repeatable. While acknowledging the good intentions behind the positive move, however, we need to be aware of a pragmatic problem regarding whether or not the actual analysis can be easily repeated. Data analysis requires particular skills and specialisms, which many news editors may not have. The repeating of the whole process of data analysis may also be very time-consuming.

Nevertheless, the real issue in relation to objectivity is not the difficulty of repeating data analysis. Data-processing practices can be repeated in the face of the difficulty, if data journalists write data logs or record meta-data for their analysis in detail step by step. It will also be helpful if the level of data literacy within newsrooms increases greatly and if enough time is given for reproducing data analysis. However, being reproducible does not equate to achieving objectivity. To be repeatable by using the same set of data, algorithms and approaches only proves their practice to be sound and appropriate. It would not solve the real issue that originates from the problems surrounding data and data algorithms as well as "design subjectivity" in the whole process.

Together, the choice-making of data journalists and the "social constructionist" nature of intensely-used data and data algorithms subtly shapes the stories published and turns them into products of the social construction of reality. The aforementioned four essential elements of the concept of objectivity (Skovsgaard et al., 2013) collapse under such influences for two reasons. First, subjectivity and imbalance is inevitably involved in data, data algorithms and the data processing process; and second, due to the "social constructionist" nature of data and data algorithms, and the "design subjectivity" of data journalists, facts may not be 'hard'. Objectivity is thus not pertinent in the context of data journalism.

Discussion and conclusion: data reporting as social construction of reality

In this paper we argue that the notion of objectivity is inapplicable to data journalism and call for attention to the "social constructionist" nature of data reporting. Like traditional news reporting, data-driven stories present merely one version of reality revealed in and from the data. Differing from traditional news reporting, however, what data journalists need to choose have expanded from news sources, information and story angles to include data, analytical approaches, data algorithms and methods and so on. Apart from the conventional influences identified by Schudson (1989), the work of data journalists is also shaped by those that have had an impact on the data and the data algorithms used in the reporting process. This means that the "social constructionist" nature of data reporting starts from data and data algorithms and is thus preceding the actual reporting. Such precession suggests the potential influence of the work of other occupations such as data collectors, analysts and programmers on journalism through using data in reporting. This influence opens up space for debates on the questions of what influence data reporting, how data journalists defend and maintain cultural authority, as well as what data journalism is.

Understanding data reporting as social construction of reality has four implications. Firstly, it is unfair to judge the quality of data journalism within the regime of objectivity. In traditional journalism, in order to maintain objectivity, journalists may involve more news sources in their reports to balance the views presented by different social actors. However, in data journalism, the fact that the politics of data and data algorithms as well as the influencing factors are opaque and even preceding data reporting would make it more difficult, if not impossible, for data journalists to correct the subjectivity or even biases in data or data algorithms if they exist.

With the constant proliferation of data in our society, data journalism is no doubt an asset to both society and democracy, and deserves a welcoming embrace from journalism practitioners and observers. However, rather than placing an emphasis on objectivity, it would be more appropriate if we see data journalists as producers or interpreters of the meanings of data, which contribute to the maintenance of a healthy democracy and which inform readers about particularly important matters. The paper echoes the arguments of constructionist scholars about the role of journalism and journalistic norms, as discussed earlier. Like investigative journalism (Ettema and Glasser, 1987; Ettema and Glasser, 1998), the cultural authority of data journalism should come from its potential to better serve the public interest and democracy instead of from its claim to objectivity. In this sense, it would be more suitable to understand data journalism as public journalism that aims to serve democracy and to revitalise public life (Ryan, 2001). Rather than questioning whether or not data journalism is objective, it is argued that what is truly important for data journalism is to serve the public interest and democracy by engaging "people as citizens", enhancing "public discussion" or acting as "custodians of conscience" (Merritt, 1995; Ettema and Glasser, 1998).

The second implication is for data journalism research. In future research, it might be important to move away from celebrating objectivity in data journalism to looking into the actual choices made by data journalists, what factors may influence their practices and data reports, how data journalists know what they know about the data and the subjects of their reports (i.e. epistemology) as well as how data journalists defend and retain their cultural authority and what data journalism actually is. It might be also interesting to examine what (social) values are advocated, what reality is constructed in data reports and whether and to what extent these reports contribute to maintaining a healthy democracy.

The third is pedagogical implication with respect to the design of data journalism programmes to raise students' awareness of the above-discussed five influencing factors in addition to the traditional influences on news production. Such factors range from including or excluding certain data, focusing on certain aspects of the data collected but ignoring others, using particular algorithms or statistical models, to data journalists' particular understanding of the topic and interpretations of the patterns found in the data. Teaching students how to think about data journalism and making them aware of what may influence their data reporting is as equally important as teaching them practical skills.

The last but not least implication is for data journalism practice. The discussion in this paper offers an opportunity for us to reflect on the work of data journalism and what data journalists need to be careful about in order to have good practice and to produce high quality data reports. It would be important for data journalism practitioners to be aware of the five influencing factors. They need to pay extra attention to data, data algorithms as well as the subjectivity involved in the design of the data processing process, where certain aspects are prioritised but the importance of others is downgraded. In addition, the social construction of data reporting surely makes transparency particularly important for data journalism. To be transparent about the choices made in the whole process can improve the level of perceived credibility of data reports. Improved transparency around data analysis processes and the data would be helpful in enhancing the authority of data journalism in constructing knowledge in and from data.

In summary, this paper joins the debates about data journalism by arguing that objectivity is incompatible with data reporting and it is inappropriate if we evaluate the quality of data journalism by adhering to the tenet of objective reporting. Having said that, this paper is not arguing that data reporting is biased and dishonest nor does it reject the practice and importance of data journalism. It proposes to direct our attention to the "social constructionist" nature of the work of data journalism instead, by shifting our focus away from objectivity to understanding how data reports come into being, what meanings they have, how data journalists defend their authority and legitimacy, as well as whether and how data reports serve the public interest and contribute to democracy.

Bibliography

Anderson, C. 2008. The End of Theory. *Wired.*

Bavadam, L. 2010. Environment Stories, among the Most Challenging. *In:* ACHARYA, K. & NORONHA, F. (eds.) *The Green Pen: Environmental Journalism in India and South Asia.* New Delhi, Thousand Oaks, London and Singapore: SAGE.

Bell, M. 1998. The Truth is Our Currency. *The Harvard International Journal of Press/Politics,* 3**,** 102-109.

Berger, P. L. & Luckmann, T. 1966. *The Social Construction of Reality: A treatise in the sociology of knowledge,* New York, Penguin Books.

Bolin, G. R. & Schwarz, J. A. 2015. Heuristics of the algorithm: Big Data,user interpretation and institutionaltranslation. *Big Data & Society*.

Bollier, D. 2010. *The Promise and Peril of Big Data,* Washington, The ASPEN INSTITUTE.

Borges-Rey, E. 2016. Unravelling Data Journalism. *Journalism Practice,* 10**,** 833-843.

Boudana, S. 2011. A definition of journalistic objectivity as a performance. *Media Culture & Society,* 33**,** 385–398.

Boyd, D. & Crawford, K. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society,* 15**,** 662-679.

Boykoff, M. T. & Boykoff, J. M. 2007. Climate change and journalistic norms: A case-study of US mass-media coverage. *Geoforum,* 38**,** 1190 - 1204.

Bradshaw, P. 2014. Data Journalism. *In:* ZION, L. & CRAIG, D. (eds.) *Ethics for Digital Journalists : Emerging Best Practices.* London and New York: Routledge

Brewin, M. W. 2013. A Short History of the History of Objectivity. *The Communication Review,* 16**,** 211-229.

Burns, R. 2015. Rethinking big data in digital humanitarianism: practices, epistemologies, and social relations. *GeoJournal,* 80**,** 477-490.

Carlson, M. 2007. Blogs and Journalistic Authority: The role of blogs in US Election Day 2004 coverage. *Journailsm Studies,* 8**,** 264-279.

Carlson, M. 2019. News Algorithms, Photojournalism and the Assumption of Mechanical Objectivity in Journalism. *Digital Journalism*.

Cohen, S. & Young, J. (eds.) 1973. *The Manufacture of News: social problems, deviance and the mass media,* London: Constable.

Cushion, S., Lewis, J. & Callaghan, R. 2017. Data Journalism, Impartiality And Statistical Claims. *Journalism Practice,* 11**,** 1198-1215.

Diakopoulos, N. 2015. Algorithmic Accountability. *Digital Journalism,* 3**,** 398-415.

Donsbach, W. & Klett, B. 1993. Subjective objectivity. How journalists in four countries define a key term of their profession. *Gazette (Leiden, Netherlands)* 51**,** 53-83.

Duncan, P. 2018. How we carried out our survey of flatshare bias *The Guardian*.

Ettema, J. & Glasser, T. 1987. On the Epistemology of Investigative Journalism. *In:* GUREVITCH, M. & LEY, M. R. (eds.) *Mass Communication Review Yearbook.* London: SAGE.

Ettema, J. & Glasser, T. 1998. *Custodians of Conscience. Investigative Journalism and Public Virtue,* New York, Columbia University Press.

Gans, H. J. 1980. *Deciding What's News: a study of CBS evening news, NBC nightly news, Newsweek and Time,* London, CONSTABLE.

Gauthier, G. 1993. In Defence of a Supposedly Outdated Notion: The Range of Application of Journalistic Objectivity. *Canadian Journal of Communication,* 18.

Gillespie, T. 2014. The Relevance of Algorithms. *In:* GILLESPIE, T., BOCZKOWSKI, P. & FOOT, K. (eds.) *Media Technologies. Paths Forward in Social Research.* London: MIT Press.

Glasser, T. & Ettema, J. 1989. investigative Journalism and the Moral Order. *Critical Studies in Mass Communication,* 6**,** 1-20.

Hammond, P. 2015. From computer-assisted to data-driven: Journalism and Big Data. *Journalism*

Hampton, M. 2008. The “objectivity” ideal and its limitations in 20th-century British journalism. *Journalism Studies,* 9**,** 477-493.

Hanusch, F. 2014. *Lifestyle Journalism,* New York, Routledge.

Heeks, R. & Renken, J. 2018. Data justice for development:What would it mean? *Information Development,* 34**,** 90–102.

Hellmueller, L., Vos, T. P. & Poepsel, M. A. 2013. SHIFTING JOURNALISTIC CAPITAL? *Journalism Studies,* 14**,** 287-304.

Johnson, J. 2016. The question of information justice. *Communications of the ACM (Association for Computing Machinery),* 59**,** 27-29.

Johnson, J. A. 2014. From open data to information justice. *Ethics and Information Technology,* 16**,** 263–274.

Johnston, R. 1968. Choice in Classification: The Subjectivity of Objective Methods. *Annals of the Association of American Geographers,* 58**,** 575-589.

Karlsen, J. & Stavelin, E. 2014. Computational journalism in Norwegian newsrooms. *Journalism practice,* 8**,** 34-48.

Kiel, P. & Waldman, A. 2015. How We Analyzed Racial Disparity in Debt Collection Lawsuits: An explanation of how we analyzed whether debt collection lawsuits disproportionately impact black communities. *ProPublic.*

Kirk, A. 2016. *Data visualisation: a handbook for data driven design,* London, Sage.

Larson, J., Mattu, S., Kirchner, L. & Angwin, J. 2016. How We Analyzed the COMPAS Recidivism Algorithm. ProPublica.

Lesage, F. & Hackett, R. A. 2014. Between objectivity and openness-the mediality of data for journalism. *Media and Communication,* 2**,** 42-54.

Lewis, S. C. 2012. The Tension between Professional Control and Open Participation: Journalism and its Boundaries. *Information, Communication & Society,* 15**,** 836 - 866.

Lichtenberg, J. 1996. In Defence of Objectivity Revisited. *In:* CURRAN, J. & GUREVITCH, M. (eds.) *Mass Media and Society.* London, New York. Sydney, Auckland: Arnold,.

Lowenstein, R. L. & Merrill, J. C. 1990. *Macromedia: Mission, Message and Morality,* New York, Longman.

Lupton, D. 2013. Digital sociology: beyond the digital to the sociological. *In:* OSBALDISTON, N., STRONG, C. & FORBES-MEWETT, H. (eds.) *The Australian Sociological Association 2013 Conference Proceedings: Reflections, Intersections and Aspirations, 50 Years of Australian Sociology.* Melbourne: TASA.

Manovich, L. 2012. Trending: The Promises and the Challenges of Big Social Data. *In:* GOLD, M. (ed.) *Debates in the Digital Humanities.* Minneapolis, MN: University of Minnesota Press.

Maras, S. 2013. *Objectivity in Journalism,* Cambridge and Malden, Polity.

Mayer-Schönberger, V. & Cukier, K. 2013. *Big Data: A Revolution that Will Transform How we Live, Work and Think,* Boston and New York, Houghton Mifflin Harcourt.

Mcluhan, M. 1964. *Understanding Media: The Extensions of Man,* New York, McGraw-Hill.

Mcquail, D. 1992. *Media Performance: Mass Communication and the Public Interest,* London, Sage.

Mcquail, D. 1994. *Mass Communication Theory. An Introduction,* London, SAGE.

Merrill, J. C. 1984. Journalistic objectivity is not possible. *In:* MERRILL, J. C. & DENNIS, E. (eds.) *Basic issues in mass communication* New York: Macmillan.

Merrill, J. C. 1990. Semantics and objectivity. *In:* LOWENSTEIN, R. L. & MERRILL, J. C. (eds.) *Macromedia: Mission, message, and morality.* New York: Longman.

Merritt, D. 1995. *Public journalism and public life: Why telling the news is not enough,* Hillsdale, NJ, Erlbaum.

Meyer, P. 1991. *Precision Journalism: A Reporter’s Introduction to Social Science Methods.,* Lanham, MD, Rowman & Littlefield.

Mindich, D. 2000. *Just the Facts: How Objectivity Came to Define American Journalism,* New York, NYU Press.

Molotch, H. & Lester, M. 1974. News as Purposive Behavior: On the Strategic Use of Routine Events, Accidents, andScandals. *American Sociological Review,* 39**,** 101-112.

Parasie, S. 2015. Data-Driven Revelation? *Digital Journalism,* 3**,** 364-380.

Parasie, S. & Dagiral, E. 2012. Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago. *New Media & Society,* 15**,** 853–871.

Perreault, G. P. & Vos, T. P. 2018. The GamerGate controversy and journalistic paradigm maintenance. *Journalism,* 19**,** 553–569.

Press, S. J. & Tanur, J. M. 2012. *The Subjectivity of Scientists and the Bayesian Approach,* New York, Chichester, Winheim, Brisbane, Singapore, Toronto, John Wiley & Sons.

Rupar, V. 2006. How did you find that out? transparency of thenewsgathering process and the meaning of news. *Journalism Studies,* 7**,** 127-143.

Ruppert, E. 2012. The Governmental Topologies of Database Devices. *Theory, Culture & Society,* 29**,** 116–136.

Ruppert, E., Isin, E. & Bigo, D. 2017. Data politics. *Big Data & Society*.

Ryan, M. 2001. Journalistic Ethics, Objectivity, Existential Journalism, Standpoint Epistemology, and Public Journalism,. *Journal of Mass Media Ethics: Exploring Questions of Media Morality,* 16**,** 3-22.

Schudson, M. 1978. *Discovering the News: A Social History of American Newspapers,* New York, Basic Books.

Schudson, M. 1989. The sociology of news production. *Media, culture & society,* 11**,** 263-282.

Schudson, M. 2001. The objectivity norm in American journalism. *Journalism,* 2**,** 149–170.

Sheridan Burns, L. & Matthews, B. J. 2018. First Things First: Teaching Data Journalism as a Core Skill. *Asia Pacific Media Educator,* 28**,** 91–105.

Skovsgaard, M., Albæk, E., Bro, P. & Vreese, C. D. 2013. A reality check: How journalists' role perceptions impact their implementation of the objectivity norm. *Journalism,* 14**,** 22 - 42.

Tandoc, E. C. & Oh, S.-K. 2017. Small Departures, Big Continuities? *Journalism Studies,* 18**,** 997-1015.

Taylor, L. 2017. What is data justice? The casefor connecting digital rightsand freedoms globally. *Big Data & Society,* 1-14.

Thomas, R. J. 2016. In defense of journalistic paternalism. *Journal of Media Ethics,* 31**,** 86–99.

Thomas, R. J. 2017. Helpfulness as Journalism's Normative Anchor. *Journalism Studies*.

Tong, J. 2015. Being Objective With a Personal Perspective How Environmental Journalists at Two Chinese Newspapers Articulate and Practice Objectivity. *Science Communication,* 37**,** 747-768.

Tuchman, G. 1972. Objectivity as strategic ritual: An examination of newsmen's notions of objectivity. *American Journal of Sociology,* 77**,** 660-670.

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

Waldman, A. & Kiel, P. 2015. Racial Disparity in Debt Collection Lawsuits: A Study of Three Metro Areas. October 8, 2015 ed.: ProPublica.

Walker, J. 2005. Links and Power: The Political Economy of Linking on the Web. *Library Trends,* 53**,** 524-529.

Walter, M. M. 2010. The Politics of the Data. *International Journal of Critical Indigenous Studies* 3**,** 45-56.

Westerståhl, J. 1983. Objective news reporting: General premises. *Communication research,* 10**,** 403-424.

Wickham, H. 2014. Tidy Data. *Journal of Statistical Software,* 59**,** 1-23.

Zarsky, T. 2016. The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making. *Science, Technology, & Human Values,* 41**,** 118-132.

Zhang, S. I. 2014. Chinese-style pragmatic objectivity in war reporting. *Asian Journal of Communication,* online first.

---

[1] Although this paper does not involve the detailed analysis of interview data or other empirical data collected, the authors would like to acknowledge that the understanding of data journalism presented in this paper was influenced by their twelve interviews with data journalists in the UK and the US in 2018 and 2019 as well as research about available resources such as videos and audios of the speeches of data journalists, their self-reflexive articles and data reports published by the US and UK news outlets. The authors would like to thank the journalists for sharing experience and viewpoints with us and the two anonymous reviewers for their valuable and constructive comments.

[2] Data algorithms refer to computational algorithms and other tools/applications that are used to process data.

[3] https://www.theguardian.com/news/datablog/2011/aug/09/uk-riots-incident-listed-mapped#data

[4] https://www.apnews.com/Teengunviolence

[5] https://www.apnews.com/ce39136dad7c49d6977ba851018f5d92

[6] https://www.propublica.org/article/opportunity-gap-methodology

[7] There have been some changes. For example the Office for National Statistics (ONS), in the UK, will potentially include "a gender or transgender status question" in its 2021 census questions and topics. Accessed at June 3 2019, at https://www.ons.gov.uk/census/censustransformationprogramme/questiondevelopment/2021censustopicresearchupdatedecember2018#annex-3-summary-of-research-undertaken-for-gender-identity-topic-december-2017-to-november-2018 and https://www.ons.gov.uk/census/censustransformationprogramme/questiondevelopment/2021censustopicresearchdecember2017

[8] https://ico.org.uk/media/1165/time-for-compliance-foia-guidance.pdf

[9] https://www.eastherts.gov.uk/article/35311/Tips-for-successful-FOI-Requests

[10] https://dynamicsofwriting.com/2017/10/11/you-are-the-only-thing-stopping-you-from-doing-great-work-spotlight-fellow-jaimi-dowdell-talks-about-her-two-year-project-investigative-journalism-and-how-students-can-succeed-in-publishing-toug/

[11] http://cops.heraldtribune.com/Home/About

[12] https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368 and https://medium.economist.com/the-challenges-of-charting-regional-inequality-a9376718348

[13] https://www.ap.org/press-releases/2017/data-journalism-chapter-debuts-in-2017-ap-stylebook