Article

new media & society
1–29

DOI: 10.1177/1461444818791326
journals.sagepub.com/home/nms
SAGE

Imagining big data: Illustrations of "big data" in US news articles, 2010–2016

Christian Pentzold, Cornelia Brantner and Lena Fölsche
University of Bremen, Germany

Abstract

Imagining "big data" brings up a palette of concerns about their technological intricacies, political significance, commercial value, and cultural impact. We look at this emerging arena of public sense-making and consider the spectrum of press illustrations that are employed to show what big data are and what their consequences could be. We collected all images from big data-related articles published in the online editions of *The New York Times* and *The Washington Post*. As the first examination of the visual dimension of big data news reports to date, our study suggests that big data are predominantly illustrated with reference to their areas of application and the people and materials involved in data analytics. As such, they provide concrete physical form to abstract data. Rather than conceiving of potential ramifications that are more or less likely to materialize, the dominant mode of illustration draws on existing, though often trite, visual evidence.

Keywords

Big data, image type analysis, media discourse, public understanding of technology, visual communication

The increasing engagement with sets of aggregate data in organizations ranging from telecommunication providers, retailers, and banks to health institutions, state agencies, and scientific institutions is accompanied by debates about the character, conditions, and consequences of data-based operations.

Corresponding author:
Christian Pentzold, University of Bremen, Bremen 28359, Germany.
Email: christian.pentzold@uni-bremen.de

In these areas of contestation, the ambiguous term *big data* points to a palette of issues. For a start, we could say that the term is used to describe massive digital datasets that require innovations in analytical techniques in order to exploit them and create new forms of value. Big data's vastness is not about absolute size but about the required scale of analysis. Big data are seen as a next step in datafication, that is, Van Dijck (2014) explains, "the transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis" (p. 198). They inspire ambitions to make more accurate and reliable predictions in order to solve complex problems, from climate change to terrorism (Kitchin, 2014). Moreover, big data represent a regulatory challenge in terms of the extensive accumulation of information by state agencies and corporate ventures.

In sum, big data form a "cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and mythology that provokes extensive utopian and dystopian rhetoric" (Boyd and Crawford, 2012: 662). Yet the role that media discourse plays in this regard is largely overlooked. Studying the public understanding of big data is necessary because the notion is not a neutral term but originates from business contexts; its relevance is largely based on commercial considerations (Mayer-Schönberger and Cukier, 2013). Thus, the term can serve as a "powerful frame for discourse around knowledge" (Markham, 2013: §1), one that is less about the characteristics of data themselves but about the shifts in technologies and mindsets that embrace data first of all as a vital economic input.

Consequently, we focus on an emerging arena of public sense-making and consider the spectrum of images employed to illustrate press articles on big data. These visuals, we argue, are an object of analysis in their own right and with an inherent representational logic. Establishing a powerful imagery to visually capture big data-related topics is, we assume, both a key task and a challenge as journalists seek to give palpable form to a discrete phenomenon and thus turn big data into a public issue. Our examination contributes to a more profound investigation of the concrete representations of abstract information representing what big data actually are and what they mean for society. The analysis is rooted in cultural analysis and science and technology studies (Bowker, 2013). Scrutinizing their objective facticity, these approaches hold that data, while being abstract, presuppose interpretation and materialization.

Toward a public understanding of big data

There seems to be no lack of definitions of what big data are. Instead there are numerous attempts to provide more or less exhaustive designations, like the widely used and criticized "three Vs" of big data, referring to their alleged volume, variety, and velocity (Lupton, 2015). Yet it is difficult to provide a straightforward answer to the question "what are data?" Instead, we might better ask "when are data?" because data "exist in a context, taking on meaning from that context and from the perspective of the beholder" (Borgman, 2016: 18). As such, to be associated with any aspiration or concern, big data have to be made into a socially relevant phenomenon and problem as they have no value or meaning in isolation.

In this respect, Portmess and Tower (2015) claim that the term big data is in itself "a trove of suggested meanings for semantic exploration" (p. 3). They argue that the metaphorical contents of the discourse around the term "carry suggestive implications for exploring different ways of envisioning our relationship to emerging information technologies" (p. 3). Likewise, considering big data in historical and political dimensions, Beer (2016) invites us to see them as "an interweaving of a material phenomenon and circulating concept" (p. 4). The metaphors around big data do not only shape the collective mindsets on big data, but also their governance (Hwang and Levy, 2015). "How what counts as data (and data's referent) is a social process with political overtones," Boellstorff and Maurer (2015: 3) thus postulate.

Against this background, the scant attention to the imagery circulating around big data is especially conspicuous given the precept that data are abstract. While their abstract quality makes it difficult to think or write about data in general, "it follows from their abstraction that data ironically require material expression," Gitelman and Jackson (2013: 6) state. So we suppose that journalistic reports require visual displays in order to envision what big data actually are and what their implications would be for citizens, commerce, or society writ large (Coleman, 2010; Messaris and Abraham, 2001). Consequently, our study first asks: What types of images are employed in order to illustrate articles on big data (*RQ*1)?

Overall, this study resonates with analyses of the meaning work around science and emerging technologies (e.g. Cacciatore et al., 2012; Druckman and Bolsen, 2011). A chief communicative function of media frames lies in the contextualization of quite abstract issues, such as nanotechnology or molecular science, by offering patterns of interpretation. They select and highlight certain aspects and thus predispose understanding and stimulate public action. In this respect, big data might function as a prime motive for framing strategies too because, as Beer (2016) posits, the very concept of big data "shapes decisions, judgments and notions of value" (p. 5). Accordingly, the language around the concept is rich with metaphorical rhetoric that speaks of the "dataverse," "data deluge," "data explosion," and of big data as the "new oil" (Lupton, 2013, 2014; Puschmann and Burgess, 2014).

Compared to written texts, visuals are perceived in a holistic and associative manner and attain superior salience compared to verbal material and, thus, can be highly effective for articulating ideological messages (Brantner et al., 2013; Messaris and Abraham, 2001). Nevertheless, there are only a few studies on the visual framing of abstract themes in technology and science, let alone big data. Most of the work that examines the visual framing of less tangible issues deals with the depiction of climate change (e.g. Rebich-Hespanha et al., 2015; Wessler et al., 2016). It finds that climate images show identifiable people (most often politicians, but also scientists, citizens, business leaders, and celebrities), causes of climate change (such as through iconic images of smokestacks), and the impacts of climate change.

In case the visual dimension is discussed at all, the lack of creativity in depicting big data is criticized. For example, the Tumblr blog *bigdatapix* (2017) assembled a collection of visuals only to conclude that "Big Data is visualized in so many ways … all of them blue and with numbers and lens flare." As a case in point, searching for images of "big data" on Google brings up a visually coherent sample (see Figure 1). We see word clouds,

Figure 1. Screenshot of image search results for "Big Data" on Google, 2 August 2017. Search was done using an incognito browsing tab.

binary code, network structures, (watching) eyes, and a dominance of blue (for similar results on Bing, see Sylvia, 2017). The same kinds of visuals appear when searching for "big data" on the image platforms Fotolia, Flickr, and Pinterest as well as on Wikipedia.

Contrary to this monotony, we however assume that stereotypical and homogeneous illustrations in professional journalistic reports on the topic do not (or not yet) exist. Instead, we expect to find a variety of pictorial representations. We pose that journalists explore different ways of illustrating their reports on big data-related issues in order to provide meaningful visuals on the elusive nature of data and to reflect on the fundamental ambiguity of what big data actually are and how their implications can be imagined. We assume that experiments with the visual repertoire, especially in professional journalism, are also grounded in industry-wide standards aiming to provide poignant illustrations and photographs instead of indifferent off-the-shelf pictures (Hariman and Lucaites, 2007; Zelizer, 2005).

With a focus on linguistic metaphors, Puschmann and Burgess (2014) point to the interpretive flexibility of concepts associated with big data and the "ongoing contestation over their exact meanings and values" (p. 1691). Their analysis of news items found that the most prevalent tropes were either big data as a force of nature needing to be controlled or a nourishment/fuel to be consumed. When data are conceptualized as living organisms, Lupton (2013) argues, "they appear more benign, part of 'good nature', but also again as potentially wild and uncontained, growing out of our control" (para. 7). Against this backdrop, we secondly ask: What fields of reference are mobilized to illustrate big data and evoke certain attributes and characteristics (*RQ*2)?

We assume that the imagery will portray big data as a force of nature or nourishment/fuel. We expect these two kinds of metaphors are complemented by references to surveillance and control. Hence, we suppose to find images about "Big Brother watching" or the panopticon (Lohr, 2012). Notwithstanding the presumably diverse range of individual images on big data, we also believe that the collection can be classified into two broad groups: one group that underscores their beneficial, positive sides, and another group with illustrations that stress the negative aspects of big data. As such, they visually reproduce the oft-described dichotomy of utopian and dystopian scenarios surrounding technological innovations.

Furthermore, studies of photojournalistic routines have underscored the importance of portrait photographs and depictions of persons in newspaper illustrations (Grittmann and Ammann, 2011). We are, therefore, interested to what extent notable tech leaders, but also stereotypical images of the IT workforce, occur in big data illustrations. Consequently, our third question is: how are people featured in big data-related news imagery (*RQ*3)? We also ask if the visual coverage of big data people is biased as regards gender distribution (*RQ*4) and expect that the news outlets under consideration do not skew the actual gender distribution in the IT workforce. This reflects Kim et al.'s (2016) insight that news photographs of scientists in *The New York Times* science section mirrored the actual proportion of women scientists.

Data and method

Materials

We collected all images with captions published in the online editions of two influential US daily newspapers, *The New York Times* and *The Washington Post*, between January

2010 and December 2016. A string search for "big data" in the newspapers' online archives resulted in a total of 872 articles containing 956 pictures. We excluded those that did not address big data-related issues explicitly and/or that did not contain images. With 282 articles remaining, we used a corrected sample of 450 images as the basis of our analysis—219 from *The New York Times* and 231 from *The Washington Post.* Due to the calibration of the newspapers' websites, articles or photographs might still be missing from the sample because pages could have been de-published or pictures may no longer be available.

We refer to these elite news media because they not only mirror events but (still) constitute the public agenda and influence the framing of issues and thus affect how themes are understood by larger segments of the US-American public and abroad (McCombs, 2014; Vonbun et al., 2016). Despite the decline in newspaper circulation in affluent societies, including in the United States, and the fall of revenues, quality newspapers are managing to reach larger audiences across their online and offline outlets and also in part through syndication and aggregation of their content. They are a prime resource for journalists, politicians, and other opinion leaders. As such, they not only set frameworks for understanding big data-related topics but also explore, we assume, the spectrum of imaginaries. Starting our exploratory analysis with two flagship periodicals will therefore actually not capture a representative set of visuals circulating around big data in the press and social media but will help us to gauge the scope of possible images. In this sense, the material from these elite venues is, we believe, rich in variety and can inspire other efforts to extend prevalent big data illustrations.

Procedure

Image type analysis. We conducted an image type analysis on the total sample of 450 images. It combined interpretative and standardized steps and allowed for an in-depth classification and quantification of image types. The analysis centered on structural patterns, not individual pictures. The advantage of this approach was that a larger number of pictures could be reviewed than an investigation of each individual image, without losing sight of the visual peculiarities of each type (Brantner et al., 2017; Grittmann and Ammann, 2011). In general, an image type captures visuals with a similar meaning or content. The method builds on the idea that photojournalistic practices are routinized activities that iterate a limited and recurring repertoire of images, despite the changes the profession has undergone following digitization (Mäenpää, 2014). As will be shown later, the analysis was able to determine an emerging palette of images even though big data represent a new topic and there was no established signature class of iconic pictures.

We inductively developed image types from the material using a picture card sort technique (Fincher and Tenenberg, 2005). Screenshots of the images were collocated into groups of visually similar sets in an iterative procedure. The evolving types were constantly probed for internal homogeneity and consistency as well as for their differentiation from other types. Note that the polysemy of image types must be taken into account when assigning illustrations to them, as images can resonate with different types. We addressed this problem by discussing ambiguous cases before assigning them to the most appropriate image type.

Picture variables. In a concurrent step, we introduced and coded quantitative variables for the image types. The coding system reflected visual aspects of each type and was devised after the basic classification of types. Not all quantitative variables applied to all image types. For example, we employed person variables only for those images depicting people, coding their sex, social distance, gaze, and expression (for the list of person variables, see Table 2). We are aware of the problems associated with classifying the sex of depicted persons, especially regarding the reproduction of male and female gender norms (Brantner et al., 2017). The variables for gaze and social distance referred to the relationship between image and beholder. As such, social distance was evaluated using a three-point scale ranging from intimate or personal distance to social distance through to public distance. This was determined by the distance between the position of the camera and the represented person, which is commonly associated with different degrees of involvement and intimacy. As regards gaze, we expected that images connect viewers and depicted persons when persons are looking directly at the viewers, whereas they make the viewers invisible beholders of impersonal objects when persons are looking away from them (Bell and Milic, 2002).

We coded whether the news image represented a photograph or a nonphotographic image (drawing, diagram/infographic). As will be shown later, some image types only consisted of photographs while others included diagrams or drawings alone. We coded whether image sources were credited or not and which sources were used, that is, if the image was taken by photographers or compiled by (data-) journalists working for the news outlet itself, if it came from a press or photojournalism agency (e.g. Associated Press, Reuters), if it was made by artists (illustrators, design studios, cartoonists), if it originated from a stock photo agency (e.g. Getty Images), or if it was reproduced by courtesy of companies, depicted persons, or other agents.

To gauge the reliability of the coding procedure, two coders independently coded 10% of the material. We used Krippendorff's Alpha, which was between .86 (social distance) and 1.0 (gender, image source), indicating a reliable measurement.

Results: pictures about or of big data

As we expected, big data are a novel yet trending issue in public discourse (see Figure 2). We found no articles concerning big data and containing cognate images in 2010. After a slight increase in 2011 and 2012, coverage reached a first peak in 2013 with 74 articles and 114 images, respectively. Later, the two newspapers developed different dynamics. *The Washington Post* continued to cover the topic with roughly the same intensity in the following years but showed a steep decline in 2016. In comparison, *The New York Times* reported less intensively in the years 2014 and 2015 but exhibited a second peak in 2016, with 32 articles and 70 images. Presumably, the Edward Snowden revelations triggered the heightened media awareness in 2013.

Of the 450 images found in the news discourse on big data, more than two-thirds (69.3%) are photographs. The remaining images are either drawings, computer generated illustrations, and infographics (28.0%), or screenshots (2.7%).

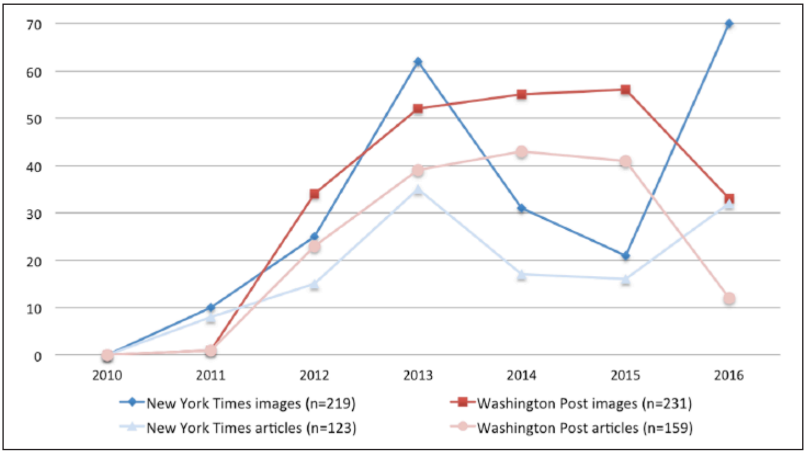

Figure 2. Article and image distribution timeline.

Regarding the sources of the illustrations, 44.4% (n=200) were produced by (photo-) journalists themselves and only 39 images (8.7%) seem to have been purchased from stock image agencies; 60 (13.3%) came from press agencies or photo agencies; 74 (16.4%) were images reprinted by courtesy of institutions, companies, or depicted persons; 64 (14.2%) originated from artists and illustrators; and 13 (2.9%) named no source or were in the creative commons.

We identified 13 image types that could be grouped into four clusters. We labeled them *big data visualizations (13.9%), big data technologies (19.8%), big data processes (10.4%)*, and *big data people (29.4%)*. A fifth cluster was added for images portraying a multiplicity of *big data application contexts*. This is the most frequent sort of cluster (26.4%).

Among the image types, the most common type depicts *protagonists* engaging with big data in one way or another (20.9%). Next, there are pictures of the *materiality* of big data's sites and infrastructures (9.1%) as well as images best described as *infographics* (10.0%). The description of the image types follows their order of appearance in Table 1.

Big data visualizations

The first cluster of *visualizations* comprises three image types that we named *infographics* (10.0% of illustrations), *large numbers* (1.3%), and *artistic renditions* (2.7%).

Infographics. The figures assembled in this type are rooted in established journalistic and scientific practices of using diagrams to give visual form to numerical information (Halpern, 2014). They are also closely tied to recent innovations in data journalism.

In this type, which was the third most popular sort of imaging big data, the challenge of displaying what big data looks like and of arriving at visually comprehensive tropes has spawned several enterprises. They aim at conditioning and formatting data, such as the most common words in online dating profiles or wine-drinking preferences. While some of these are simply entertaining or telling illustrations, others seek to scrutinize

Table 1. Frequency of image types.

Cluster	Image type	NYT (n=219)	WP (n=231)	Total (n=450)
Visualizations				*13.9%*
	Infographics	11.9%	8.2%	10.0%
	Large numbers	2.3%	0.4%	1.3%
	Artistic renditions	3.7%	1.7%	2.7%
Technologies				*19.8%*
	Apps	3.2%	6.1%	4.7%
	Materiality	6.4%	11.7%	9.1%
	IT logos	1.4%	6.5%	4.0%
	Human machines	2.7%	1.3%	2.0%
Processes				10.4%
	Datafication	6.4%	7.8%	7.1%
	Datafied individuals	5.0%	1.7%	3.3%
People				29.4%
	Protagonists	19.2%	22.5%	20.9%
	The IT workforce	7.3%	2.6%	4.9%
	Computer nerds and geeks	3.2%	0.9%	2.0%
	Big data people away from technology	2.3%	0.9%	1.6%
Application context		25.1%	27.7%	26.4%

complex affairs with large amounts of data, for instance, the voting decisions of US citizens or the spread of infectious diseases. The type contains diagrammatic illustrations like graphs, maps, and charts. They are usually accompanied with additional information about the underlying data and the lessons to be drawn from them. At times, the figures feature animated content or present simulations of data-driven processes (see Figure 3).

Large numbers. Big data cannot be displayed or pictured in total due to their volume and abstract quality. A compensation for this problem and a means to express the vast dimension of data is by way of numeric indication. Only about 1% of images in the sample used this form of ostentatious demonstration. However, we still can infer some level of popularity of this type because it featured prominently on book covers and was used to illustrate op-eds and commentaries. It contains images that work with capital letters, bold color schemes, and tall numbers. They act as a visual substitute for the magnitude of data and emphasize the assumed "bigness" of big data (see Figure 4).

Artistic renditions. The spectrum of visualizations used to represent big data also includes artistic approaches. They test different ways of simulating, displaying, and experiencing data without the ambition of providing statistically exact visual information like we find in infographics. Instead, while the illustrations contained in this comparatively small group often gesture toward the character of data visualizations, they do not refer to or use any sort of data in particular. The type encompasses images

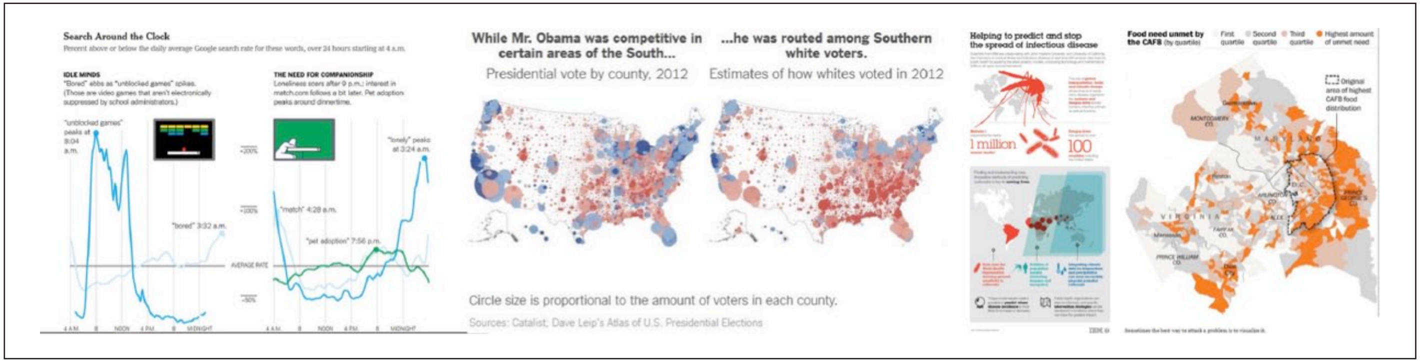

Figure 3. Examples of the image type *infographics*.
Left: The New York Times, July 4, 2015, Source: analysis of Google data by Seth Stephens-Davidowitz; 2nd from left: The New York Times, June 9, 2016, Sources: Catalist; Dave Leip's Atlas of U.S. Presidential Elections; 3rdfrom left: VentureBeat, September 29, 2013, Image Credit: IBM; right: The Washington Post, June 29, 2015, Image Credit: Capital Area Food Bank.

that are reminiscent of the colorful abstractions of modernist art. The appropriations mimic or caricature the appearance of presumably exact data visualizations. They moreover challenge the assumed objectivity of infographics in giving visual form to abstract data. Consequently, they are often included in articles in order to provide succinct illustrations that invite more than one reading and show different degrees of artistic alienation from visual stereotypes. The image on the left in Figure 5, showing a work by illustrator Chad Hagen that he terms a "nonsensical infographic," is a poignant example of this.

Big data technologies

This second cluster captures four image types: *apps* (4.7% of all illustrations), *materiality* (9.1%), IT logos (4.0%), and *human machines* (2.0%).

Apps. One way of making the technological dimension of big data seemingly clear is to show the domestic interfaces people commonly use when communicating via digital media. Hence, this mode of illustration picks up big data-based programs and customer applications familiar to users in order to provide images of how big data are involved in people's lives. The images in this type show screens of mobile phones or computers. Different services like search engines, e-commerce websites, or social networking sites are running on these graphical user interfaces. Besides screenshots and close-up shots, other images use more abstract depictions of generic operating screens with tiles, web forms, or drop-down menus (see Figure 6).

Materiality. This type also engages with the material side of big data, but takes a broader perspective on the tangible infrastructures, buildings, and programmable machines necessary for data production, usage and processing, storage, and transformation. They include both the immense complexes of data warehouses and headquarters buildings and small devices that have become intimate companions in daily life. One portion of the illustrations displays digital gadgets, mobile or handheld sensors, and wearables such as computers, smartphones, cameras, drones, or smart gear that both enable the production of big data and provide services based on big data analysis. Another portion indicates the hardware necessary for processing and storing data, for example, microchips and server farms. In addition, a few pictures show architectural sites like the US National Security Agency (NSA) premises in Fort Meade, MD, or the data facilities of IT businesses (see Figure 7).

IT logos. Another way of giving tangible form to big data is to use photographs of logos of organizations and firms involved in big data matters. In this type, which is mainly used by *The Washington Post*, the names of tech businesses and agencies symbolize the industrial and administrative complex of data analytics (see Figure 8). This includes data generators (like Twitter or Facebook), technology and software companies (like Yahoo or IBM), and state agencies such as the NSA.

Human machines. Some of the more provocative illustrations of the potential of big data to surface in the sample speculate about the intelligence, agency, and autonomy of

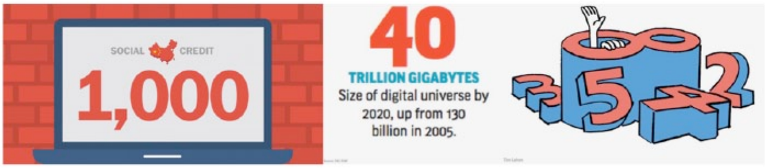

Figure 4. Examples of the image type *large numbers*.
Left: The Washington Post, October 22, 2016, Image Credit: Rachel Orr/The Washington Post; iStock; middle: The New York Times, April 11, 2013, Source: IDC/EMC; right: The New York Times, May 2, 2015, Image Credit: Tim Lahan.

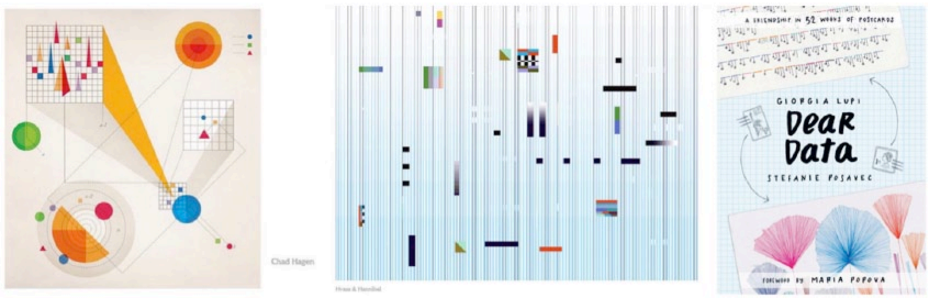

Figure 5. Examples of the image type *artistic renditions*.
Left: The New York Times, February 11, 2012, Image Credit: Chad Hagen; middle: The New York Times, March 27, 2015, Image Credit: Hvass & Hannibal; right: The Washington Post, August 27, 2016, Image Credit: Princeton Architectural Press.

computational machines. In a few cases, this also extends to questioning the man-made decisions and intentions behind seemingly smart and self-determined engines. Visually, this set consists of photographs as well as drawn images illustrating aspects of the envisioned subjectification of machines and some of their anthropomorphous traits. We find illustrations of computers becoming superior players in board games and of robots as the artificial teacher or better soldier. Some images of the *ego ex machina* connote the threatening prospect of machines assuming human capabilities or replacing humans (see Figure 9).

Big data processes

Two types of images fall into the third cluster: *datafication* (7.1% of the illustrations) and *datafied individuals* (3.3%).

Datafication. This group visualizes datafication processes that translate empirical circumstances into data formats and vice versa (Bowker, 2013). As such, the images seek to capture the intersection of presumably real and virtual realms of action and accountability taking place in the datafication of procedures and social relations. The type does not contain photographs of actual practices but instead graphic renditions of data-oriented

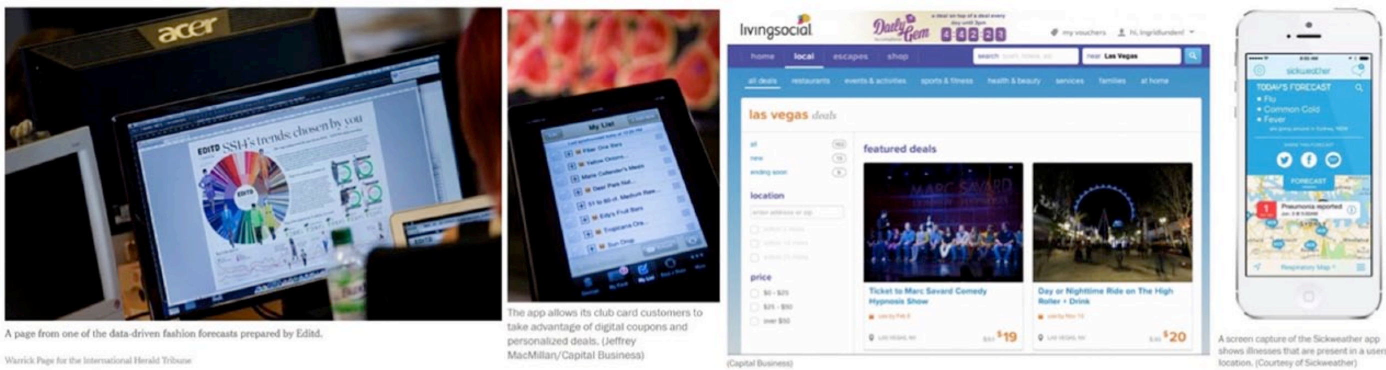

Figure 6. Examples of the image type *apps*.
Left: The New York Times, August 26, 2013, Image Credit: Warrick Page for the International Herald Tribune; 2nd from left: The Washington Post, June 15, 2012, Image Credit: Jeffrey MacMillan/Capital Business; 3rd from left: The Washington Post, July 13, 2014, Image Credit: Capital Business; right: The Washington Post, October 9, 2014, Image Credit: Courtesy of Sickweather.

Figure 7. Examples of the image type *materiality*.
Left: The New York Times, September 29, 2012, Image Credit: Hewlett-Packard; 2nd from left: The Washington Post, November 24, 2015, Image Credit: Sensoria Fitness; 3rd from left: The Washington Post, October 5, 2015, Image Credit: Alex Brandon/AP; right: The New York Times, June 8, 2013, Image Credit: Rick Bowmer/Associated Press.

Figure 8. Examples of the image type *IT logos*.
Left: The Washington Post, October 29, 2014, Image Credit: Justin Sullivan/Getty Images; middle: The Washington Post, March 17, 2015, Image Credit: Ben Margot/AP; right: The New York Times, June 8, 2013, Image Credit: David Burnett/Contact Images.

activities. They are mostly found in commercial contexts and in the operations of state agencies and intelligence services. Together, they form a sort of pioneering industrial and administrative complex where the technical innovations in datafication are developed and potential use scenarios explored. The drawings and cartoons imagine processes of dataveillance, monitoring, and the evaluation of people (Lupton and Williamson, 2017; see Figure 10, second and fourth image from left). Many illustrations employ the stereotype of the binary number system 0 and 1. Some use visual metaphors that present datafication in terms of aquatic or meteorological imagery (see Figure 10, first and third image from left). They thus connect to related verbal tropes like data deluge or data clouds (Lupton, 2014).

Datafied individuals. People also feature in data-based processes. The move from interaction based on observation or verbal communication to one based on data has commonly been associated with an increase in abstraction and decontextualization. As a result, data collection allows the profiling of people using only small amounts of personal information as a "series of discrete signifying flows" (Haggerty and Ericson, 2000: 612). Yet despite the virtual presence of such reassembled "dividuals" (Deleuze, 1992: 5), these likenesses still refer to physical bodies and personal identities. However, this referral does not occur via direct representation but through categorization, modeling, and projection. Reflecting these circumstances, the type includes drawings and cartoons that seek to portray the translation of personal traits or actions into abstract data and their presentation. They often point to the flaws and inaccuracies occurring in these forms of measuring and representation, and they indicate the legal, economic, or political ramifications these processes imply (see Figure 11).

Big data people

As we expected, the depiction of individuals in the IT workforce was a frequently used illustration strategy. We found four image types that focused on people dealing with big data, namely, *protagonists* (20.9% of the illustrations), the *IT workforce* (4.9%), *computer nerds and geeks* (2.0%), and *big data people away from technology* (1.6%). Table 2 shows the coded image variables for the four image types in this cluster.

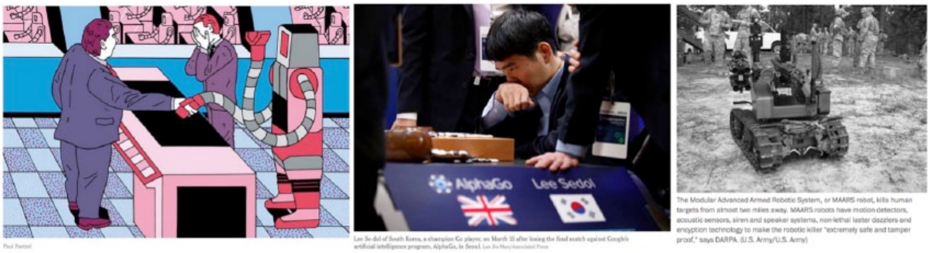

Figure 9. Examples of the image type *human machines*.
Left: The New York Times, April 18, 2015, Image Credit: Paul Paetzel; middle: The New York Times, March 25, 2016, Image Credit: Lee Jin-Man/Associated Press; right: The Washington Post, October 9, 2015, Image Credit: U.S. Army.

Table 2. Coded picture variables for the four image types depicting big data people.

Picture variables		Image types		Tests
		Protagonists ($n=94$)	Other three image types depicting people ($n=36$)[a]	
Sex	Female	9.6%	8.3%	Fisher's exact test = 12.329**
	Male	85.1%	63.9%	
	Male and female	5.3%	22.2%	
	Not assignable	0.0%	5.6%	
Persons depicted	One	83.0%	41.7%	Chi^2 = 24.241***
	Two	12.8%	30.6%	
	Three or more	4.3%	27.8%	
Distance	Intimate or personal	36.2%	5.6%	Chi^2 = 19.207***
	Social distance	44.7%	41.7%	
	Public distance	19.1%	52.8%	
Gaze	Direct at viewer	55.3%	2.8%	Chi^2 = 57.643***
	Full profile but away *f.* viewer	29.8%	13.9%	
	Face and gaze away *f.* viewer	10.6%	44.4%	
	Mixed or not discernable	4.3%	38.9%	
Expression	Laughing or merry	42.6%	16.7%	Chi^2 = 45.031***
	Friendly	26.6%	0.0%	
	Concentrated or serious	28.7%	44.4%	
	N.A.	2.1%	38.9%	

[a]As the three smaller image types depicting people—the IT workforce ($n=22$), Nerds and Geeks ($n=8$), Big data people outside big data ($n=6$)—did not significantly differ regarding the coded variables; the table only differentiates between the larger image type protagonists and the other three together. Two pictures did not show any people and were excluded from the analysis.
***$p<.001$.
**$p<.01$.

Before we go into the details, we will deal with the question of whether the visual news coverage of *big data people* is biased with regard to gender distribution (*RQ*4). Only 9.2% of the pictures depict only women, whereas almost four out of five (79.2%) feature only

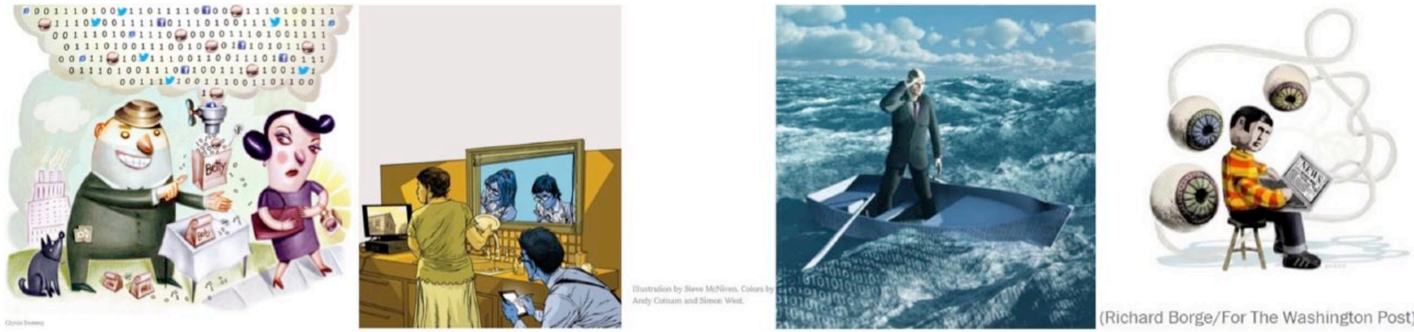

Figure 10. Examples of the image type *datafication*.
Left: The New York Times, June 23, 2013, Image Credit: Glynis Sweeney; 2nd from left: The New York Times, June 20, 2013, Image Credit: Illustration by Steve Mc Niven, colors by Andy Cotnam and Simon West; 3rd from left: VentureBeat, August 27, 2012, Image Credit: Bruce Rolff/Shutter Stock; right: The Washington Post, April 5, 2014, Image Credit: Richard Borge/For The Washington Post.

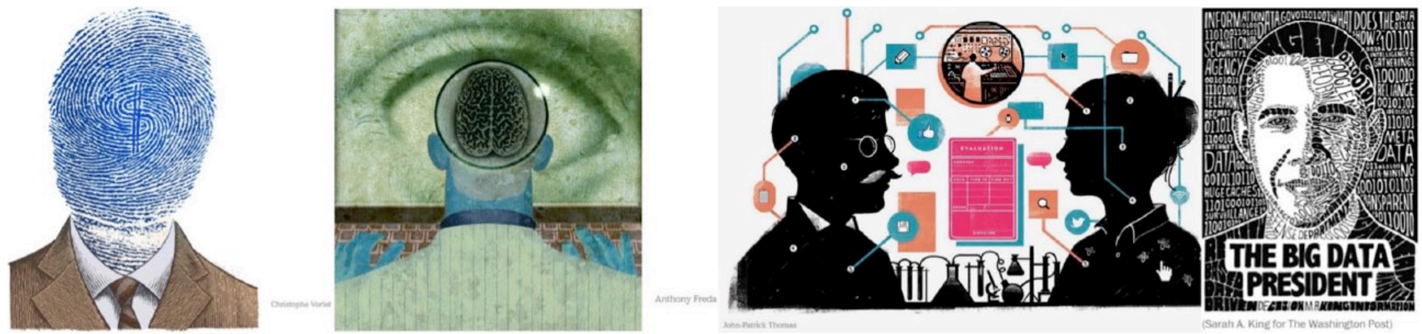

Figure 11. Examples of the image type *datafied individuals*.
Left: The New York Times, April 30, 2011, Image Credit: Christophe Vorlet; 2nd from left: The New York Times, March 23, 2013, Image Credit: Anthony Freda; 3rd from left: The New York Times, April 20, 2013, Image Credit: John-Patrick Thomas; right: The Washington Post, June 14, 2013, Image Credit: Sarah A. King.

men, and both sexes appeared in another 10% (in 1.5% of the images the gender could not be determined). In sum, of all people featured in the illustrations, 175 were male and 28 were female. Contrary to our expectation, the relative number of women in the US IT workforce—according to the *National Center for Women & Information Technology* (2017) 26%—is not reflected in the proportion of women shown in the images (13.8%). Instead, the rate of representation in quality news media is lower than that in the professional sector; the *Chi-square* test indicated that the distribution of the media is significantly different from the actual distribution (Chi^2 = 15.743; $df = 1$; $p < .001$).

Protagonists. One of the prime modes of illustrating articles on big data-related issues is to use portraits of key personnel. It involves business leaders, startup entrepreneurs, public advocacy leaders, scientists, and pundits. Presumably, that way big data shall be given a human face (Smolan and Erwitt, 2013).

Visually, this type composes of close-up or medium-shot photographs. In terms of the coded variables, the persons are mainly depicted at personal (36%) or social distance (45%) and less often at a public distance (19%). The focus in these illustrations is on the personal identity and appearance of prominent individuals, who are mostly shown alone (83% of the images show just one person). At times, these portraits also refer to the working environment or the métier of the people. Usually, it seems as though the protagonists are posing for the camera, which demonstrates that these photographs are not candid snapshots but staged portraits. People look directly at the camera in most of the photographs (55.3%) or are shown in profile (29.8%). Moreover, they are often depicted smiling (42.6%) or with a friendly expression (26.6%). These features should allow the viewer to establish a relationship with the depicted persons (see Figure 12).

The IT workforce. A related strategy places people into tech settings. Humans working with and in IT are thus visually linked to material technological infrastructures and devices. This approach toward depicting the intricate entanglement of people, data practices, and datafied environments is often employed in long-form journalism like reportages that seek to provide interpretation and contextual stories.

Visually, this group is based on long shots or figure shots, and the people are mainly depicted at a social or public distance. The images appear to be snapshots taken during routine work activities. Hence, they show people in occupational situations (meetings, personal conversations, monitor work) and professional spaces. The tech setting and tool use are clearly visible without being the focus of the image. People are not personally identifiable as in the protagonists' image type but are shown as representatives of collectives engaged with data analytics. This includes programmers, researchers, operators, teachers, and students. The impersonal nature of this type of depictions is underlined by the fact that the people never look directly at the viewer; they rarely smile but instead concentrate on their work. Often, they are shown as working in teams of two or in larger groups. Consequently, the images do not involve the viewer in the action because the depicted people are interacting with one another or with computers (see Figure 13).

Figure 12. Examples of image type *protagonists.*
Left: The New York Times, April 14, 2011, Image Credit: Noah Berger; 2nd from left: The Washington Post, October 6 2015, Image Credit: Courtesy of Carnegie Mellon; 3rd from left: The New York Times, February 8, 2014, Image Credit: Preston Gannaway for The Washington Post, via Getty Images; right: The New York Times, June 19, 2013, Image Credit: Jim Wilson.

Figure 13. Examples of image type *the IT workforce*.
Left: The New York Times, March 24, 2012, Image Credit: J. Emilio Flores; middle: The New York Times, July 18, 2012, Image Credit: Joshua Lott; right: The Washington Post, June 20, 2014, Image Credit: Image Credit: Jeffrey MacMillan.

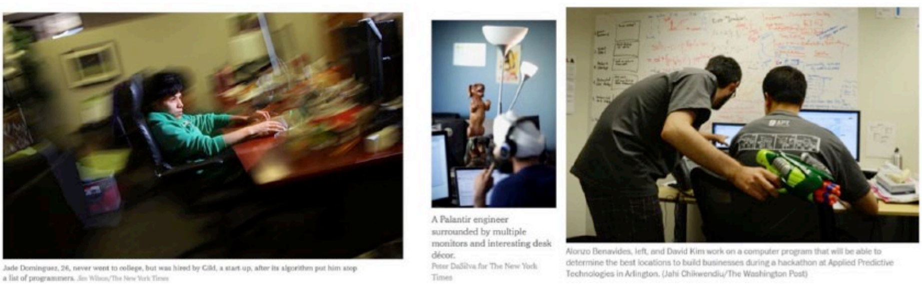

Figure 14. Examples of image type *computer nerds and geeks*.
Left: The New York Times, April 27, 2013, Image Credit: Jim Wilson; middle: The New York Times, Mai 31, 2014, Image Credit: Peter DaSilva; right: The Washington Post, October 18, 2013, Image Credit: Image Credit: Jahi Chikwendiu.

Computer nerds and geeks. Another type plays with stereotypical images of people strongly involved in tinkering with digital data and software. They embed them in a context filled with technology and more random things commonly associated with nerd culture.

As in the previous class, these photographs appear to be snapshots. The people focus on screens and are situated in work/leisure spaces filled with toys, tech devices, or board games. The protagonists are shown at a social or public distance and, again, do not face the viewers directly (see Figure 14). Of the 21 IT workers depicted in this image type, only two are women, and both are shown in the company of men. Thus, this image type reinforces the stereotype of the male computer nerd who is either Caucasian (White) or Asian (Kendall, 2011). However, it has to be added that in relation to the overall amount of imagery, this photojournalistic strategy is rare and makes up only 2% of all images.

Big data people away from technology. An even smaller set of images also centers on the people involved in developing and using big data. Yet instead of incorporating them into programmable machinery, these images depict the analytical work beyond the devices and computer interfaces. Most pictures show teams in front of whiteboards covered with handwriting in different colors. While digital media are absent, the

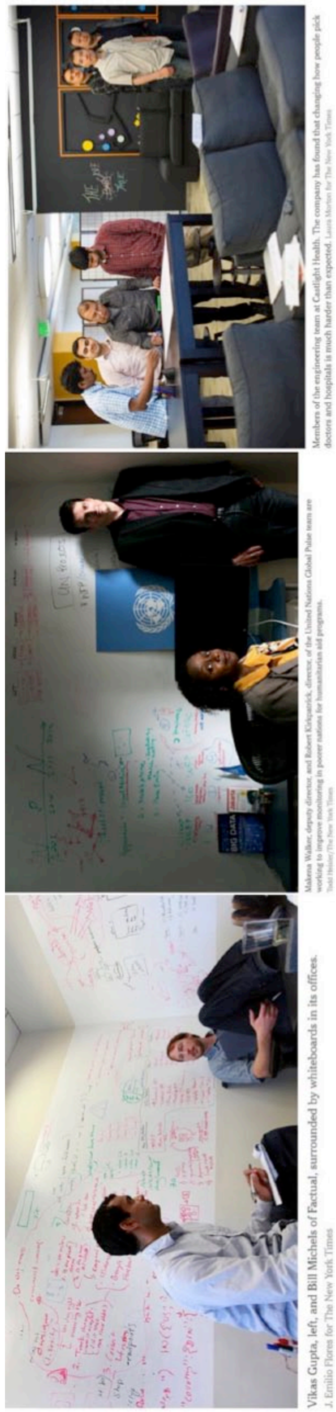

Figure 15. Examples of image type big data people off technology.
Left: The New York Times, March 24, 2012, Image Credit: J. Emilio Flores; middle: The New York Times, August 7, 2013, Image Credit: Todd Heisler; right: The New York Times, June 16, 2016, Image Credit: Image Credit: Laura Morton.

whiteboards testify to the complex formulae and analytical processes behind big data (see Figure 15).

Big data application contexts

During the sorting process, a group of images emerged that showed no visible cues to the material side of big data, big data processes, or people involved in big data-based operations. Thus, these images do not represent a type in the proper sense. This cluster is actually defined by the conspicuous absence of indexical images of big data as they visualize the plethora of application contexts in which big data do or potentially can play a role. They hence provide evidence of big data's versatility. As a consequence of this nonrepresentational pictorial commonality, the captions, or the surrounding text of the illustrations are essential for inferring the implicit link to big data. So where necessary, we used an article's title and context to identify the field of big data-usage.

Statistically, the cluster is the most prevalent image form ($n=119$). The prominent areas of application show executive (police, military, intelligence services, federal government), judicial (courts), and legislative (congress) contexts (29.3%). They are followed by diverse examples of scientific applications ranging from economics to literature studies (10.1%), industry (11.5%), and the service sector such as retailing, food catering, banking, hospitality as well as healthcare (14.3%). The rest of the illustrations fall into smaller groups, which represent between 6.8% (campaigning) and 0.7% (e.g. journalism, fitness) of the cases in this type. Two-thirds of the images depict one or more persons (67.8%). As it already became apparent with the *datafication* images, there seem to be two prevalent contexts of big data inquiries: state-run agencies and commercial firms feature as the prime arenas for the shift to computational tools and methods for analyzing big data. In these two interrelated fields a combination of technological, ideological, and procedural changes led to the expansion of data and the development of tools in order to harness data (Mayer-Schönberger and Cukier, 2013). Their pioneering role also becomes evident in the media focus on these two areas which dominate the reports on big data's societal impact.

Discussion

In face of the apparent novelty of big data analytics and the still nascent state of the public understanding and critique of their widespread and fundamental ramifications, the image type analysis proved a valuable method for identifying the visual repertoire used in US quality online news reporting on big data. We were able to formulate and describe 14 image types (*RQ*1 and *RQ*2). These types reflect an evolving segmentation and closure of the potential visual spectrum that, we argue, is rooted in the journalistic challenge of acquainting viewers with visual tropes. Hence, these common repertoires help to organize and simplify news work because journalists can resort to an available portfolio of generative types that may be selectively assigned to cover emerging events or stories. This corpus delineates the scope of intelligible imagery that the audience comes to associate with big data.

As we assumed, stock images are rarely used in the media coverage. According to their own professional standards, the two US elite news media have rather looked for adequate pictures to give visual form to big data issues. This practice stands in stark contrast to the images found through a Google image search, which contained hardly any photographs in the first 240 retrieved results (see Figure 1). Against this visually quite homogeneous tableau of imagery available online, the two newspapers exhibit a concern with how to provide more succinct and diverse visuals that not only resonate with the topics discussed but also play a role in actually problematizing the social and technical ramifications associated with big data.

However, in line with other studies that looked into the framing of science and technology and their common strategies of showing well-known people and the physical causes and apparent impacts of what are perceived as immaterial phenomena, the big data illustrations seek to provide concrete visual representations of less tangible entities or processes (Rebich-Hespanha et al., 2015; Wessler et al., 2016). We found that depictions of people, materialities, and application contexts serve as concrete visual surrogates for the virtuality and immateriality of big data. The arcane technical operations of agencies like the NSA but also of commercial players like IBM and Google are represented through visible manifestations of data production and their personal operators (Tufekci, 2014).

Therefore, visualizations and graphics of datafication processes that represent the abstract transformation of social action into big data are less significant. They only account for about one-quarter of the images. In contrast, the majority of illustrations demonstrate the most significant mode of visualization, which involves displaying concrete physical forms of abstract data: portraying the protagonists helps to give big data a human face (Smolan and Erwitt, 2013). Picturing data centers and server racks acknowledges the real-world nature and location of only seemingly untethered data clouds.

Arguably, these types of images fail to evoke more potential and evolving ramifications. These might not yet exist and are more or less likely to materialize. Notwithstanding the probability of future scenarios, such imaginings would open up the possibility to ponder the social and technological significance of datafication and related processes of automation and algorithmization in the present. In contrast, the illustrations around big data we found are much more bound to existing, but often trite, evidence, whereas more imaginative representations that might also dispense with photographs are underrepresented. In the process, they ignore the possibility of also challenging the apparent state of being by conceiving of visual prospects that would, in turn, help to illuminate current affairs and their contingencies.

In this regard, some inspiration might come from the way, news about government surveillance are visualized. In this area, Kilker (2016) recognized a shift of the visual evidence from the generic imagery of CCTVs and video screens to more compelling diagrams and data simulations. These did not show big data paraphernalia but actually made use of data analytics. In our study, 1 in 10 images was such an *infographic*. They belong to the cluster of *big data visualizations*, which together with its kindred cluster of *big data processes* forms a group of representations that seek to assimilate data-based insights or that try to emulate data-related practices such as categorization, modeling, and projection.

In these clusters, we were also able to identify two broad evaluative tendencies. One stressed the beneficial aspects of big data that can bring about improvements for certain fields of application. We too found images with negative connotations, like in the image types *datafication* or *human machines* that depicted the threatening implications of an assumed superiority of robots over humans. However, the more interesting finding seems to be the fact that the majority of the images took a neutral stance toward big data. Our analysis could not confirm neither the dominance of monotonous images of big data (see Figure 1), nor the primacy of metaphorical imagery of data as a natural force or nourishment/fuel that seem to dominate on the verbal level of news (Lupton, 2013, 2014; Puschmann and Burgess, 2014). *Datafication* was the only of the 13 image types that also uses a visual rhetoric of big data drawing on such kind of metaphors—but this only occurred in 10 cases, accounting for 2.2% of the total big data imagery. This reminds us of the fact that in a holistic analysis of frames, pictures and texts should be considered as complementing modes of communication and sense-making which do not transport evaluations in isolation but resonate with verbal arguments (Coleman, 2010). Subsequent analyses should take into account the multimodality of news discourses and the at times controversial negotiations entailed in societal sense-making processes. The task will be to grasp the constitutive interplay of verbal and visual modes within discursive contests around competing notions, ideologies, and imageries of what big data is (Messaris and Abraham, 2001). As such, it seems necessary to look more closely at media framing that not only takes place among newspaper articles but takes shape in the polyphony of platform and app communication that intersects with broadcast messages and visuals. An examination of the temporal dimension of the unfolding discourse could help us to understand how issue cycles resonate with events, political debates, and technological progress.

Furthermore, we asked how people working in big data are portrayed in the news imagery (*RQ*3). We found four different image types that can be grouped together in a larger cluster of big data-people. The largest represented a typical image repertoire frequently found in photojournalistic coverage. Here, the big data-industry is personified through prominent individuals. Contrary to our assumption that the actual gender distribution would not be skewed in the visual news discourse, photographs of people from the IT workforce showed the already male-dominated industry almost as a pure man's world (*RQ*4). In light of our findings, it seems that while there may be evidence for changing practices as regards a subsided gender bias in science journalism, journalists in other areas have not followed this trend (Kim et al., 2016). Yet the blame might not just lie with the media, as this bias may also be the result of the industries' protagonists' visibility and presentation. However, it is a reminder for journalists to be more sensitive to gender in future coverage of the tech industry.

Conclusion

Our study started from the idea that large data sets and the associated analytical operations have intricate relations to public debates in which their cultural significance, social value, and political relevance are negotiated. Media discourses form an integral part of the current reality of big data practices and configure potential future scenarios.

In these dynamics, publics are “simultaneously the informant, the informed and information of big data” (Michael and Lupton, 2016: 104). Their digital traces feed into data-based analytics, which will inform the services and content offered to them in turn. Moreover, public discourses are not only increasingly data-driven. They also structure the semantic repertoire of understanding big data by circulating frames of interpretation and evaluation that make people aware of the political dimension of datafication practices and mobilize them in struggles over relationships of data, power, and meaning. As such, public discourses can help to underscore the idea that data are not given facts; they are instead transformed in ways “that cut across notions of nature and culture” (Boellstorff and Maurer, 2015: 3f.).

By appreciating big data as a matter of contingent articulation that happens in discourse, the study can help to dismantle claims about a given and irrevocable facticity of data formats and data analytics in order to explore options for reimagining their status and implications. As of now, illustrations with reference to big data’s areas of application as well as the people and the materials involved in data analytics prevail in the visual spectrum. Relatively absent are more imaginative visuals that seek to go beyond the visible paraphernalia, utensils, and personnel. In a sense, these concrete objects stand in slightly peripheral relation to the abstract nature and escalating potentiality associated with big data (Gitelman and Jackson, 2013). Efforts to extend the figurative vocabulary can draw inspiration from artistic approaches probing new ways of simulating, displaying, and experiencing large data arrays. Rudimentary, they already featured in some of the images types, most notably *artistic renditions, datafied individuals*, and *datafication*.

With regard to the emerging metaphors, which sporadically occur in the visual material too, a critical ethos and epistemology are needed to analyze the problems of the intrinsic propositions associated with big data (Hwang and Levy, 2015; Lupton, 2013, 2014). The new scientific and societal paradigm of datafication, Van Dijck (2014) notes, is based on problematic ethical, analytical, and ontological requirements. “However compelling some examples of applied Big Data research, the ideology of dataism shows,” she argues, “characteristics of a widespread belief in the objective quantification and potential tracking of all kinds of human behavior and sociality through online media technologies” (p. 198). Societal discourses on surveillance and privacy issues, which at times surface in the material, especially in a portion of the images showing *datafied individuals* and *datafication*, not only challenge the popularization of datafication as a neutral paradigm but remind us that society and science have to deal critically with the mythologization of big data and their future exploration.

Authors’ Note

Cornelia Brantner is now affiliated to Institute for Knowledge Communication and Applied Research, Vienna, Austria.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Christian Pentzold https://orcid.org/0000-0002-6355-3150

References

Beer D (2016) How should we do the history of big data? *Big Data & Society* 3: 1–10.

Bell P and Milic M (2002) Goffman's gender advertisements revisited: combining content analysis with semiotic analysis. *Visual Communication* 1(2): 203–222.

Boellstorff T and Maurer B (2015) Introduction. In: Boellstorff T and Maurer B (eds) *Data, Now Bigger and Better!* Chicago, IL: Prickly Paradigm Press, pp. 1–7.

Borgman CL (2016) *Big Data, Little Data, No Data*. Cambridge, MA: The MIT Press.

Bowker G (2013) Data flakes. In: Gitelman L (ed.) *"Raw Data" is an Oxymoron*. Cambridge, MA: The MIT Press, pp. 167–172.

Boyd d and Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.

Brantner C, Geise S and Lobinger K (2013) Fractured paradigm? Theories, concepts and methodology of visual framing research: A systematic review. In: *Paper presented at the annual conference of the International Communication Association* (ICA), 17-21 June, London, UK.

Brantner, C, Lobinger K and Stehling M (2017) Memes against sexism? A multi-method analysis of the memes and selfies in the feminist protest hashtag #distractinglysexy and its resonance in mainstream news media. In: *Paper presented at the annual conference of the International Communication Association* (ICA), 25-29 May, San Diego, California.

Cacciatore MA, Anderson AA, Choi D-H, et al. (2012) Coverage of emerging technologies: a comparison between print and online media. *New Media & Society* 14(6): 1039–1059.

Coleman R (2010) Framing the pictures in our heads. In: D'Angelo P and Kuypers JA (eds) *Doing News Framing Analysis: Empirical and Theoretical Perspectives*. New York: Routledge, pp: 233–261.

Deleuze G (1992) Postscript on the societies of control. *October* 59: 3–7.

Druckman JN and Bolsen T (2011) Framing, motivated reasoning, and opinions about emergent technologies. *Journal of Communication* 61(4): 659–688.

Fincher S and Tenenberg J (2005) Making sense of card sorting data. *Expert Systems* 22(3): 89–93.

Gitelman L and Jackson V (2013) Introduction. In: Gitelman L (ed.) *"Raw Data" is an Oxymoron*. Cambridge, MA: The MIT Press, pp. 1–14.

Grittmann E and Ammann I (2011) Quantitative Bildtypenanalyse [*Quantitative image type analysis*]. In: Petersen T and Schwender C (eds) *Die Entschlüsselung der Bilder* [Deciphering Images]. Cologne: Halem, pp. 163–177.

Haggerty KD and Ericson RV (2000) The surveillant assemblage. *British Journal of Sociology* 51(4): 605–622.

Halpern O (2014) *Beautiful Data: A History of Vision and Reason since 1945*. Durham, NC: Duke University Press.

Hariman R and Lucaites JL (2007) *The Public Image: Photography and Civic Spectatorship*. Chicago, IL: The University of Chicago Press.

Hwang T and Levy K (2015) "The cloud" and other dangerous metaphors. *The Atlantic*, 15 January. Available at: https://www.theatlantic.com/technology/archive/2015/01/the-cloud-and-other-dangerous-metaphors/384518/ (accessed 11 December 2017).

Kendall L (2011) "White and nerdy": computers, race, and the nerd stereotype. *The Journal of Popular Culture* 44(3): 505–524.

Kilker J (2016) All about whom? Stock photos, interactive narratives and how news about governmental surveillance is visualized. *Visual Communication Quarterly* 23(2): 76–92.

Kim H, Kim S-H, Frear C, et al. (2016) News photos of scientists skew race but not gender. *Newspaper Research Journal* 37: 261–274.

Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and their consequences*. London: SAGE.

Lohr S (2012) The age of big data. *The New York Times*, 11 February. Available at: http://nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html (accessed 11 December 2017).

Lupton D (2013) Swimming or drowning in the data ocean? Available at: https://simplysociology.wordpress.com/2013/10/29/swimming-or-drowning-in-the-data-ocean-thoughts-on-the-metaphors-of-big-data/ (accessed 11 December 2017).

Lupton D (2014) Liquid metaphors for big data seek to familiarise technology. Available at: http://blogs.lse.ac.uk/impactofsocialsciences (accessed 11 December 2017).

Lupton D (2015) *Digital Sociology*. London: SAGE.

Lupton D & Williamson B (2017) The datafied child: The dataveillance of children and implications for their rights. *New Media & Society* 19(5): 780–794.

McCombs M (2014) *Setting the Agenda*. Cambridge: Polity Press.

Mäenpää J (2014) Rethinking photojournalism: the changing work practices and professionalism of photojournalists in the digital age. *Nordicom Review* 35(2): 91–104.

Markham A (2013) Undermining 'data: A critical examination of a core term in scientific inquiry. *First Monday* 18(10). Available at: http://firstmonday.org/article/view/4868/3749

Mayer-Schönberger V and Cukier K (2013) *Big Data*. New York: John Murray.

Messaris P and Abraham L (2001) The role of images in framing news stories. In: Reese S, Gandy OH Jr and Grant AE (eds) *Framing Public Life*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 215–226.

Michael M and Lupton D (2016) Toward a manifesto for the "public understanding of big data". *Public Understanding of Science* 25(1): 104–116.

National Center for Women & Information Technology (2017) By the numbers. Available at: http://ncwit.org/resources/numbers (accessed 11 December 2017).

Portmess L and Tower S (2015) Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of big data. *Ethics and Information Technology* 17: 1–9.

Puschmann C and Burgess J (2014) Metaphors of big data. *International Journal of Communication* 8: 1690–1709.

Rebich-Hespanha S, Rice RE, Montello DR, et al. (2015) Image themes and frames in US print news stories about climate change. *Environmental Communication* 9(4): 491–519.

Smolan R and Erwitt J (2013) *The Human Face of Big Data*. Sausalito, CA: Against all Odds Production.

Sylvia JJ (2017) *Posthuman media studies: an affirmative approach to informational ontology, big data and processes of subjectivation*. PhD Thesis, NC State University, Raleigh, NC. Available at: https://repository.lib.ncsu.edu/handle/1840.20/34458 (accessed 11 December 2017).

Tufekci Z (2014) Engineering the public: big data, surveillance, and computational politics. *First Monday* 19. Available at: http://firstmonday.org/article/view/4901/4097 (accessed 11 December 2017).

Van Dijck J (2014) Datafication, dataism and dataveillance. *Surveillance & Society* 12(2): 197–208. Available at: https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/datafication (accessed 11 December 2017).

Vonbun R, Kleinen-von Königslöw K and Schoenbach K (2016) Intermedia agenda-setting in a multimedia news environment. *Journalism* 17(8): 1054–1073.

Wessler H, Wozniak A, Hofer L, et al. (2016) Global multimodal news frames on climate change. *International Journal of Press/Politics* 21(4): 423–445.

Zelizer B (2005) Journalism through the camera's eye. In: Allan S (ed.) *Journalism: Critical Issues*. Maidenhead: Open University Press, pp. 167–177.

Author biographies

Christian Pentzold is an associate professor for Media and Communication Studies at the Center for Media, Communication and Information Research (ZeMKI), University of Bremen. More information about his work can be found here: christianpentzold.de.

Cornelia Brantner was an interim professor at the University of Bremen, Germany in 2017 and is currently Head of Department at IWAF (Institute for Knowledge Communication and Applied Research) in Vienna, Austria. Her research and teaching focuses on digital communication, public spheres, journalism, visual communication, and social movements.

Lena Fölsche is research associate at the Center for Media, Communication and Information Research (ZeMKI), University of Bremen. Her focus is on media society and digital practices.