Digital Journalism

ISSN: 2167-0811 (Print) 2167-082X (Online) Journal homepage: http://www.tandfonline.com/loi/rdij20

Algorithmic Transparency in the News Media

Nicholas Diakopoulos & Michael Koliska

To cite this article: Nicholas Diakopoulos & Michael Koliska (2016): Algorithmic Transparency in the News Media, Digital Journalism, DOI: 10.1080/21670811.2016.1208053

To link to this article: http://dx.doi.org/10.1080/21670811.2016.1208053

Published online: 27 Jul 2016.

ALGORITHMIC TRANSPARENCY IN THE NEWS MEDIA

Nicholas Diakopoulos and **Michael Koliska**

The growing use of difficult-to-parse algorithmic systems in the production of news, from algorithmic curation to automated writing and news bots, problematizes the normative turn toward transparency as a key tenet of journalism ethics. Pragmatic guidelines that facilitate algorithmic transparency are needed. This research presents a focus group study that engaged 50 participants across the news media and academia to discuss case studies of algorithms in news production and elucidate factors that are amenable to disclosure. Results indicate numerous opportunities to disclose information about an algorithmic system across layers such as the data, model, inference, and interface. Findings underscore the deeply entwined roles of human actors in such systems as well as challenges to adoption of algorithmic transparency including the dearth of incentives for organizations and the concern for overwhelming end-users with a surfeit of transparency information.

KEYWORDS algorithmic journalism; algorithmic transparency; computational journalism; ethics; media accountability; robot journalism

Introduction

The direct and indirect influences of digital technologies on the news media have been widely documented. Many newspapers have suffered or disappeared entirely as the former business model of classifieds, advertising, and subscriptions became less tenable in an online environment (Meyer 2006). News outlets overall had to adopt production methods that included more of an audience and online focus, which led to more interactive news features (Boczkowski 2005; Deuze 2005). Scholars have argued that the increased digitization and interactivity in online news, through hyperlinks, sourcing and document disclosure, and social media and audience comments, have contributed to greater openness or transparency in journalism (Bennett 2014; Deuze 2005; Karlsson 2010, 2011). Journalism ethics are evolving to incorporate transparency as a key value. Educational and professional organizations, such as the Society of Professional Journalists (SPJ), the Poynter Institute, and the Radio Television Digital News Association (RTDNA), have all adopted transparency as a new core value for professional journalists (McBride and Rosenstiel 2014; RTDNA 2015; SPJ 2014). The elevation of transparency to a new central standard in journalism has also been fueled by the belief that transparency could remedy the loss in credibility and trust (Hayes, Singer, and Ceppos 2007; Kovach and Rosenstiel 2014; Plaisance 2007) the news media has experienced in the recent past (Pew 2012).

Routledge Taylor & Francis Group

Digital Journalism, 2016
http://dx.doi.org/10.1080/21670811.2016.1208053

Deuze (2005, 455) defines transparency as the "ways in which people both inside and external to journalism are given a chance to monitor, check, criticize and even intervene in the journalistic process." Transparency allows audiences to see more of the news production process[1] and the journalists behind the news stories. However, such transparency is problematized by the adoption of computational journalism and the integration of algorithmic components, products, or platforms that help scale and optimize news production, curation, and dissemination (Diakopoulos, forthcoming). Such algorithmic systems challenge the norm of transparency due to the opacity in their automated decision-making capabilities (Diakopoulos 2015).

Computational Journalism—defined here as *finding, telling, and disseminating news stories with, by, or about algorithms* is being adopted in a number of ways by the news media. One aspect of this is automated journalism—also sometimes referred to as "algorithmic journalism" (Dörr 2016)—which employs software or algorithms with little human intervention to generate news stories for everything from crime reporting, to earthquake alerts and company earnings reports (Graefe 2016). The Associated Press employs such news writing software to automatically produce thousands of news stories on sports and financial topics (Associated Press 2015). Other platforms such as Yahoo or the Big Ten Network also increasingly rely on algorithms to produce game recaps for basketball and football games (Narrative Science n.d.) or fantasy-football stories (Yahoo n.d.). The *Los Angeles Times* was among the first to use an algorithm to report on an earthquake in 2014 (Oremus 2014), and used algorithmic reporting tools to produce short posts for their homicide blog as early as 2010 (Young and Hermida 2015). Besides automated journalism, the broader field of computational journalism suggests other ways in which algorithms are infusing the news media. For instance, simulations and data-driven predictions are now finding their way into news reports on topics like politics, finance, or public health. *The New York Times*, as well as BuzzFeed and Mashable all use algorithms in various ways to decide which news content will be distributed (and when) via social media (Albeanu 2015; Nguyen, Kelleher, and Kelleher 2015; Wang 2015). Journalistic tasks like finding and verifying sources and content on social media are increasingly taking advantage of algorithmically driven tools that can help cope with the issues of scale (Park et al. 2016; Thurman et al. 2016). News outlets also increasingly use algorithms in A/B testing of headlines and computational journalism is being deployed directly on social media and other platforms using news bots—automated agents that participate in news and information dissemination (Lokot and Diakopoulos 2016). These examples are just a sampling of the myriad ways that algorithmic approaches to content are infusing the news media.

As algorithms become more entrenched in the way the news media works, algorithmic transparency is in a nascent stage and has yet to develop an accepted standard of how and when it should be employed. It oftentimes manifests in computational and data journalism as an effort to help the public understand the methods used in the process of reporting a data-driven story. For instance, ProPublica has published white papers describing statistical methodologies with some of its investigations (Grochowski Jones and Ornstein 2016) and BuzzFeed, FiveThirtyEight, and other news organizations maintain repositories on Github where they open-source their data and code used in some (but not all) data-driven articles. *The New York Times* maintains a blog called "Open" that publishes technical articles about how some of its news technologies, such as its recommendation engine, operate. Other news organizations, like the BBC, have published academic papers to describe the details of some of their algorithmic tools (Shearer, Simon, and Geiger 2014).

The research presented in this paper is a first attempt to reconcile the increasing use of algorithms in the news media with a desire to enable transparency as a route to accountability (Ward 2014). Transparency is just one approach toward the ethics and accountability of algorithms (Dörr and Hollnbuchner 2016), however, it is gaining traction as a norm and practice within the industry, and is thus the focus of our study. In particular we ask: how exactly might the news media become transparent about the algorithms it uses in news production? We are explicitly interested in developing a palette of information disclosure options regarding algorithms so that professional journalists have a set of tangible considerations as they approach algorithmic transparency in different institutional and news production contexts. To do so we analyze the results of nine focus groups comprised of media practitioners and researchers from leading national news outlets and universities in the United States. Participants discussed three case studies that describe concrete uses of algorithms in the media including automated news writing, algorithmic curation, and simulation in storytelling. This research proposes a framework for *algorithmic transparency*—what we define as *the disclosure of information about algorithms to enable monitoring, checking, criticism, or intervention by interested parties*. Results indicate opportunities for information disclosure about algorithms across four key areas: data, model, inference, and interface. Moreover, the findings reinforce the complex interplay of human decisions that become embedded in such systems, and underline the potential implications of more transparency for news organizations. This article aims to inform both scholars and practitioners by raising awareness for algorithmic transparency and considering effective ways to inform news media audiences about algorithmic processes. Though focused on algorithms in the professional news media, the framework that we develop and present may also be applicable to uses of algorithms in other domains across government and industry.

Transparency

Transparency has become an important value in today's society from governance and businesses to journalism and computer sciences (Schudson 2015). Even though transparency is not a new concept for holding governments and institutions accountable, its recent renaissance has been accompanied by changes in communication technologies (Fung, Graham, and Weil 2007). The internet and the growing number of digital communication devices make storing, accessing, and analyzing information about organizations and individuals increasingly easy (Bennis 2013). Digital technologies have changed the access to and scrutiny of information by anyone with internet access, which Meijer (2009) broadly described as "computer-mediated transparency."

Transparency is generally considered a means to see the truth and motives behind people's actions (Balkin 1999) and to ensure social accountability and trust (Breton 2006). On a very basic level, transparency allows access to more information which can influence power relationships between governments and citizens, business and customers, and in our case between news outlets and audiences (Bennis 2013). The access to more information also reduces uncertainty in social relations and theoretically increases trust (Cotterrell 1999), which is crucial in the maintenance of a functional society (Fukuyama 1995). But while transparency can be seen as beneficial to engendering trust, seeing the inner workings of a government, business, or newsroom can result in negative implications such as undermining competitive advantages or creating costs without concomitant gains (Granados and Gupta 2013).

Transparency and the News Media

With a growing culture of transparency in society and a decline of trust in the news media, news organizations have turned to transparency as a way to open up news production processes to the public (Craft and Heim 2009). *The New York Times* public editor Margaret Sullivan (2013) declared that *The Times* embraces transparency through social media, Web-based chats with journalists, and through the public editor who advocates for readers. According to Allen, the implementation of such transparency mechanisms in the news media has two major functions:

> [I]t is an important part of the discovery of social truth, but it is also a way to gain access to the truth about the manufacturing of news. For journalists, transparency functions as a system of accountability and as a way of increasing legitimacy with citizens, both key institutional values. (Allen 2008, 324)

Transparency in the news media can thus be understood as an ethical principal for journalists (Plaisance 2007) and as a means to create positive perceptions of journalism (Karlsson 2010). Supporters of transparency see it as a solution to major problems journalism is facing today, such as the decline in journalistic authority (Singer 2007) and the loss of credibility in the news media (Hayes, Singer, and Ceppos 2007). Kovach and Rosenstiel (2014, 291) suggest transparency "gives the public a basis on which to judge whether a particular kind of journalism is the kind they wish to encourage and trust." Transparency has been described as "the new objectivity" (Weinberger 2009), a better form of truthtelling (Singer 2007), and as an opportunity to show that professional journalistic content is superior to non-professional content (Karlsson 2011).

Transparency includes the disclosure of errors and failures, which is thought to build trust (Silverman 2013). Yet, Broersma (2013) argues journalism would lose its authoritative voice by divulging shortcomings. It would ultimately invite doubt into its own discourse and break its paradigm. "That is why pleas for transparency about choices underlying reporting and admitting that reports, at best, temporarily resemble truth, subvert journalism" (33).

While most of the discussions around transparency in the news media are normative, research has shown that many journalists have not yet embraced transparency (Chadha and Koliska 2014) and prefer objectivity as a guiding journalistic principal (Hellmueller, Vos, and Poepsel 2013). Research has not yet found strong evidence that transparency affects news audiences' credibility perceptions (Karlsson, Clerwall, and Nord 2014; Roberts 2007). Yet, transparency has been elevated to a core normative ethic of journalism culture in the United States (RTDNA 2015; SPJ 2014). Moreover, outside norms from open-source culture are impacting computational and data journalism in particular, creating support for the shift toward transparency that includes open-source programming, open data, and public involvement in news production (Coddington 2015; Lewis and Usher 2013).

Beyond normative ethical considerations, news organizations may also consider transparency as a strategic business practice. Holtz and Havens (2009) argue that showing how companies do things has become more important than what they actually produce. Corporations increasingly open up by publishing Corporate Social Responsibility reports (Owen 2005) to improve their reputation by suggesting that they are "doing well by doing good" (Fombrun 2005). Tapscott and Ticoll (2003) argue that transparency has become a force that businesses cannot afford to ignore as responsible behavior and business integrity driven by transparency makes economic sense. On a practical level, this includes accessibility to leaders, disclosure of success and failures, and the promotion of ethical behavior. Granados and Gupta (2013) suggest that disclosure options such as distortion, bias, or concealment can be strategically employed. Strategic or managed transparency may not always be a genuine attempt to be truthful but can be used to create the appearance and perception of trustworthiness.

Algorithmic Transparency

News organizations are increasingly employing algorithms in the production of news to gather, organize, make sense of, create, and disseminate stories (Diakopoulos, forthcoming). Such technology enables scale and offers a competitive advantage for the business in producing content more efficiently. But what is good for business creates tension with the normative call for transparency since such "black boxes" often lack explainability in their automated decision making (Diakopoulos 2015). One of the main concerns regarding computational systems is the existence of embedded biases, which researchers have long criticized (Friedman and Nissenbaum 1996). These values, biases, or ideologies (Mager 2012) can create consequences for the formation of publics and the fair and unbiased provision of information (Gillespie 2014).

The notion of algorithmic transparency in the news media is an attempt to articulate the mechanisms by which information about algorithms may be made public. Disclosing information about how algorithms drive various computational systems would allow users to determine the values, biases or ideologies in operation in order to understand underlying points of view of a news product. McBride and Rosenstiel (2014) argue that transparency is ideal for a news environment that increasingly fosters journalism with a point of view. In other words, as long as biases are made public for audiences to evaluate, journalism would still fulfill its mission.

Individual studies have examined the role of information disclosure and explanation of specific algorithms in recommendation systems (Cramer et al. 2008; Schaffer et al. 2015; Tintarev and Masthoff 2007), personalization (El-Arini et al. 2012), ranking (Diakopoulos, Cass, and Romero 2014), or scoring (Kizilcec 2016). Results show that transparent explanations in recommender systems can serve to enhance acceptance of specific recommendations (Cramer et al. 2008) and improve a user's impression of recommendation quality, but also diminish the user experience (Schaffer et al. 2015). Moreover, expectation violation, which is thought to drive additional information seeking, was found to be a moderator in the effect of procedural transparency on increased trust (Kizilcec 2016). In this research we do not focus on specific technicalities around a particular algorithm, but rather explore information disclosure around a range of algorithms in the news media, including broader impacts on end-users and news organizations.

A Study of Algorithmic Transparency in the Media

In order to better understand the challenges and tensions surrounding algorithmic transparency, we convened 50 participants from the news media and academia to discuss three different case studies relating to the use of algorithms in the media. Nine focus groups considered the cases and generated a wide array of ideas relating to what might be disclosed about algorithms as well as the feasibility, potential impact, and other ramifications of disclosure. This exploratory study is a first attempt to examine the varied challenges surrounding algorithmic transparency in the news media. For this purpose, this research seeks the answer to three questions:

RQ1: What elements of algorithms could be made public?

RQ2: What are the limitations of algorithmic transparency?

RQ3: What are possible comprehensive disclosure mechanisms for algorithmic transparency?

Study Methodology

Participants. Fifty participants were invited to a one-day long set of focus groups on the topic of algorithmic transparency in the media. Participants were selected based on their public reputation and knowledge of computational journalism and use of algorithms in the news media. Participants were not compensated, though some were granted travel stipends. In an effort to commingle practitioner and scholarly perspectives, 28 participants were representatives of industry and 22 came from academia. Academic participants came from North American universities and included disciplinary experts from journalism studies or computational journalism, information science, and computer science (e.g. from Columbia, Stanford, New York University, Northeastern, Rutgers, City University of New York, University of Maryland, Harvard, Princeton, etc.). Industry participants primarily included reporters and editors who had worked on computational or data journalism projects (e.g. from *The New York Times*, *Washington Post*, *Texas Tribune*, NPR, *Boston Globe*, Associated Press, etc.), but also included data scientists or product managers working for news organizations or for technology companies creating information products (e.g. from CNN, Mashable, Vocativ, SmartNews, Chartbeat, Bloomberg, etc.). Individuals from several of the major technology and platform providers relevant to the field such as Facebook, Twitter, and Automated Insights were invited and either declined to participate or did not respond to the invitation.

Stimuli. Three case studies (CS) were developed to address diverse aspects of how algorithms are currently used in the professional news media to aid in content creation, organization/curation, and dissemination. These included automatically generated news content (CS1); algorithmically enhanced curation (CS2); and simulation, prediction, and modeling in storytelling (CS3). CS1 dealt largely with the use of automatic writing software such as that created by companies like Automated Insights or Narrative Science to produce news texts from structured data. CS2 focused on the use of algorithms to select, curate, recommend, and personalize content streams like the Facebook newsfeed, including algorithmically generated news homepages, and the use of algorithms to rank and curate social media comments on news sites. Finally, CS3 focused on the use of simulation and modeling in news stories, including examples such as political prediction models or models presented as interactive data visualizations. One page of background information about each of these case studies was provided to participants before and during the workshop to stimulate the elicitation of factors that might be made transparent about algorithms. The handout included a brief explanation of the technological processes involved in the various case studies to provide a common knowledge base for all participants.

Procedure and Methods

Focus groups were conducted for each case in order to gather data about knowledge, attitudes, and practices (Puchta and Potter 2004) from different perspectives and to stimulate interactivity, participation, and group insights (Morgan 1996). As algorithmic transparency is a highly specialized issue we sampled participants with a similar professional background to garner detailed expert opinions (Kemper, Stringfield, and Teddlie 2003).

Participants were broken into three facilitator/moderator-led subgroups of 14–18. Larger groups were chosen for logistical reasons and because fairly large groups elicit a wide range of responses (Morgan 1996). Moderators included the authors of this paper as well as another researcher at the institute where the focus groups were hosted. Groups spent about an hour considering each case study. After each session, groups were randomly mixed, ensuring that no individual participated in the same case study twice. Moderators first provided examples and questions to elicit an interactive discussion on the various aspects that could be disclosed about algorithms. Ideas for information disclosure about the algorithms presented were solicited, gathered, and displayed to collectively categorize and typologize the various viewpoints. Finally, the groups discussed ramifications related to manipulation, cost, and presentation of the potentially disclosed information. All participants re-convened at the end of the three focus group sessions and were debriefed as one larger group.

Each discussion was transcribed in real-time by stenographers that were instructed to leave any identifying information such as names or affiliations out of the transcripts. Participants were made aware that transcripts would be anonymized for analysis. We analyzed the nine transcriptions (three case studies three sessions each) through iterative qualitative code, affinity diagramming, typologizing, and memoing (Lofland and Lofland 1994). Open coding of participants' statements in the context of the overall conversations resulted in a codebook that was iteratively refined. All three CS1 sessions were coded independently by each of two researchers and discrepancies in the application of codes were discussed until consensus was reached. The first session of each of the remaining two case studies was coded by both researchers and discussed to resolve discrepancies. Then, one researcher coded the final two sessions of CS2 and the other of CS3.

Study Findings

Iterative analysis of the case study transcripts resulted in the emergence of an empirically grounded typology. By considering the three cases in conjunction, the typology covers a wider set of issues than if we had considered each case in isolation. Moreover, our findings suggest substantial overlaps of transparency issues across all three cases albeit with some cases exposing various issues more saliently than others. The typology recognizes the distinct roles but also deep entanglements of humans and algorithms that must be teased out in order to achieve transparency around real-world systems that so tightly weave humans and machines together. Participants pointed out that as human creations, algorithms involve human design decisions at many levels that may need to be disclosed. The typology organizes findings according to an input–output framework across four distinct phases. These phases reflect the pipeline of information as it ultimately moves toward an end-user and include: data (inputs), model (transformation), inference (output), and interface (output). Findings also show consequences of disclosure of various types of information. In the following subsections we detail the types of information participants deemed important to disclose about human–algorithm systems in the news media, and summarize these findings in Table 1. We do not attribute specific statements to individuals, but where appropriate indicate the case study, e.g. CS2.

Humans and Algorithms

Human involvement is essential to the design and ongoing functioning of algorithmic systems. At a high level transparency might involve explaining the goal, purpose, or intent of a given algorithm, including editorial goals that articulate rationale for the selection, inclusion, exclusion, or optimization of various inputs or outputs to the algorithm. Focus group participants frequently pointed out that the way different actors are integrated into the overall process should be made transparent in some cases. For automatically generated news articles (CS1), labeling whether a human editor had reviewed the output before publication was seen as informative for determining the credibility of a news product. Similarly, participants wanted to know the identity of the people that have control over an algorithmic system, oversee it, or are accountable for it. Authors or designers of such systems could be made public similarly to journalists' bylines. While it may be difficult to disambiguate and assign credit or blame in large collective efforts (Nissenbaum 1996), such disclosure could serve to reward individuals' reputations and exert social pressure toward normative behavior. In some cases an organization may buy software rather than create it, implying that the human involvement for that organization is not the algorithm design or coding but rather its configuration and parameterization—an aspect several participants suggested could also be disclosed. Other potential avenues for transparency that were discussed included describing the social or cultural context of algorithm development, the temporal evolution including how and why adjustments are made over time, as well as human-intuited "fudge factors" that are sometimes added by people to make algorithmic outputs line up better with expectations. A participant summed up several disclosable human-oriented facets that came up repeatedly in the discussion: "Things like who made it, what was the thinking behind it, what human oversight sits atop the algorithmic decisions, what are the assumptions underlying the algorithms, are there hard coded rules…" (CS2).

TABLE 1
Summary of transparency factors across four layers of algorithmic systems

Layer	Factors

Data	• Information quality.
	  ○ Accuracy.
	  ○ Uncertainty (e.g. error margins).
	  ○ Timeliness.
	  ○ Completeness.
	• Sampling method.
	• Definitions of variables.
	• Provenance (e.g. sources, public or private).
	• Volume of training data used in machine learning.
	• Assumptions of data collection.
	• Inclusion of personally identifiable information.

Model	• Input variables and features. • Target variable(s) for optimization. • Feature weightings. • Name or type of model. • Software modeling tools used. • Source code or pseudo-code. • Ongoing human influence and updates. • Explicitly embedded rules (e.g. thresholds).
Inference	• Existence and types of inferences made. • Benchmarks for accuracy. • Error analysis (including e.g. remediation standards). • Confidence values or other uncertainty information.
Interface	• Algorithmic presence signal. • On/off. • Tweakability of inputs, weights.

Given the difficulty in clearly delineating human influences in algorithmic systems, a topic for discussion by participants became the algorithmic presence—whether and when an algorithmic process was being employed at all. Participants wanted transparency when personalization algorithms were adapting a user's view of a content stream. "I kind of want to know if I'm always being watched or sometimes being watched or not being watched at all," remarked one participant in CS2 with regards to how personalization algorithms might be "watching" their activities. The idea of an "algorithm at work" badge or icon for content curated or personalized by an algorithm was suggested.

The automatic generation of news content (CS1) sparked a number of comments that touched on the different expectations for transparency depending on whether a human or an algorithm had generated the content. As one participant adroitly put it: "We have retractions and corrections all the time when they are human-ran stories. I think it is worth discussing what is the baseline here and what are we expecting of algorithms that we don't expect of human writers?" (CS1). On the other hand, "One bad journalist can put out four, five terrible misrepresented articles in a day but an algorithm can put out a thousand," noted another participant (CS1). Several participants pointed out if a human reporter makes a mistake you can fire them or chastise them to get them to improve. For algorithmic systems such feedback is more complicated as it must be channeled back to engineers who may make changes that in turn need to be re-validated.

Input–Output Pipeline

Algorithms work on various levels in producing, curating, or managing information. The input–output pipeline describes crucial stages and aspects of information that are fed into a system (data), transformed (model), output as a classification or predicted score (inference), and presented to the end-user (interface). Keeping in mind our underlying goal of enumerating the range of potential pieces of information to disclose around algorithms, in the subsections below we describe all four aspects and how information about them may be disclosed.

Data. Algorithmic systems have an undeniable dependency on data, which drives applications like machine learning, personalization, simulation, and automatically written stories. Focus group discussions highlighted some key dimensions that may be disclosed about an algorithmic system including the information quality and validity of the data feeding that system. Specifically, this includes variables such as data accuracy, uncertainty (e.g. error margins), timeliness (e.g. when was the data collected), completeness or missing elements, sampling method, provenance (e.g. sources), and volume (e.g. of training data used in machine learning), as well as assumptions or limitations inherent to the data collection. Participants also wanted to see how data has been transformed, vetted, cleaned, or otherwise edited (either automatically or by people). Discussants indicated that such transparency information should take the form of textual descriptions in addition to disclosure of the data itself. Metadata about the data was also deemed useful to inform decisions about raw data disclosure such as if the data stems from public or private sources, or includes personally identifying information. Participants were particularly interested to know about the datafication of their user activity and how it influences the end product, "You want to see the data profile that they're using on you that the algorithm is reacting to" (CS2). It was suggested that transparency of such information would have the added benefit of allowing users to correct information that was incorrect in their personal profile.

Many of the factors of potential disclosure noted above are essential to the functioning of the algorithmic system and are rooted in human processes such as how the data was chosen, collected, sourced, or sampled. Since algorithmic systems rely on quantification of the world in order to operate, human processes are needed to set the ground rules, definitions, and boundaries of that quantification in order to enable algorithmic operation at scale. As such, transparency information surrounding the data would necessarily need to consider the human influences on that data.

Model. The model itself as well as the modeling process and methodology are essential aspects of the algorithmic system that discussants suggested could be disclosed. Modeling involves building a simplified or optimized reality of the world, often using data and a process that predicts, ranks, associates, or classifies. A key aspect of any model that may be disclosed is the input scheme—the features or variables used by the algorithm. Participants wanted to know what exactly the model is optimizing, how different features are weighted, what type of the model was built (e.g. linear versus non-linear), and what software modeling tools were used since these embed different assumptions or limitations. For models trained on data, transparency around that data as enumerated above would provide a more complete picture of the model. While some participants talked about optimization functions and machine learning features, others in the groups saw such factors more as editorial criteria. For example, thresholds that are used by a model in determining the semantics of output for an automated writing algorithm might determine the interpretation of a value, i.e. whether a predicted value is a "moderate increase" or a "dramatic rise." The thresholds matter for the eventual interpretation of the output text by end-users.

Participants repeatedly mentioned source code as a potential avenue for transparency around the modeling process, "I want to see the source code for your data pipeline. I want to see the source code for your visual presentation," and "we could do pseudo-code as a way of interpreting for the public" (CS3). In effect, the code (and its revision history) is the artifactual output that reifies the modeling process, thus providing a vital trace of how that process works from inputs, to transformations, and eventual outputs. At the same time, code was discussed as a difficult technical description limiting the usability of such information for the broader public.

Participants recognized that much of the work in modeling ties back into human efforts, including the rationale for weightings or choice of model, the inclusion or exclusion of variables in the final model, the model validation procedure, and assumptions (statistical or otherwise) made in the process. This may also include how models evolve over time, "how did the models then learn from previous iterations, does the 2014 model differ from the 2012 and what do they learn from 2012 and incorporate into 2014?" (CS3). A challenge for algorithmic transparency comes back to the ambiguity of whether a decision is human-made or a result of an algorithmic optimization: "distinguishing between hand-coded rules and [machine] learned things" (CS2). There are two distinct avenues for transparency, one concerning the editorial human influences and decisions in modeling and methodology, and the other concerning how those editorial choices are *implemented* in code. There was a belief by some participants that such editorial decisions in the modeling process would lead to deeper insights about the news organization: "I personally [would] love to see what's been left out, rejected, because that would give a huge insight into the mindset of the news organization" (CS3).

Inference. Output inferences made via an algorithmic process, such as classifications, numerical predictions, or recommendations, often raised questions about error rates, uncertainty, and accuracy. Discussants acknowledged that much could be learned from understanding an algorithm's performance on outliers as well as typical cases. In terms of the Facebook newsfeed algorithm, one participant remarked, "the algorithm is making assumptions about who you're closest to or what content you want to see and those inferences may not be completely accurate" (CS2). It was suggested that transparency around inferences should include not only accuracy but also what types of inferences are being made at all.

In addition to disclosing summary statistics, study participants proposed that further error analysis around inferences could include detailing protocols for remediating errors, as well as characterizing error sources such as if they stem from human, data, or algorithm. Moreover, it was suggested that since classifiers oftentimes produce a confidence value, such information should also be disclosed in aggregate to indicate the average range of those confidence values and thus uncertainty in outcomes. Participants pointed out that for users to contextualize such transparency information it should also include a description of the significance of that uncertainty or how that uncertainty can be understood or taken into account by the overall process.

Interface. The final layer of output interacts directly with end-users: the interface that the algorithm portrays to the user. Any transparency information revealed about an algorithm must ultimately take some "tangible or visual" form in order to be presented to the end-user. Participants discussed options such as integrating information about an algorithm directly in the user interface by providing cues for clicking into FAQs to examine issues surrounding data, model, and inferences. Publishing periodic transparency reports or creating new roles for ombudspeople whose job it would be to communicate transparency information about algorithms were also proposed. At its simplest, the "algorithmic presence" could be signaled in the interface through an icon next to content that had been affected by an algorithmic process, such as personalization.

Participants discussed interactivity as a key mechanism for revealing extra layers of transparency information, for instance by allowing users to drill into content to see underlying data: "if you have an automatically written sports story maybe you could click through and see the box scores" (CS1); or if the transparency information needed to be a written description: "You could click on something that would give you a FAQ about that particular algorithm, who programmed it, why it's making choices that it's making … stuff that will be easily digestible for the average consumer" (CS2).

Interactivity can also allow users to perturb and tweak the input to the algorithm in order to better understand the outputs, or as one participant put it, "how can I interact with the algorithm to get it to treat me differently" (CS2). Users can use that interactivity to contribute to a social sensemaking endeavor around the algorithm: "people can try different things and they can say, look, I changed this parameter and used a different data-set and I notice that your analysis is biased" (CS3). In personalized systems, it was suggested this might take the form of allowing users to see how the system would appear if it were personalized for different archetypes of users.

Another conceptualization of interactivity involved being able to override the algorithm, to flip it on or off and to see the effect on the results. But participants pointed out that if the product itself is defined by that algorithm, such as a news recommendation app, an on/off switch would be unfeasible from a business standpoint. Yet, participants saw interactivity as a route to be able to see and compare the outputs of an algorithmic system based on whether that algorithm was on or off, or on the inputs and parameters that the user was changing. Overall, interactivity was seen as a way to open up some of the editorial decisions about an algorithm for scrutiny.

A tension emerged concerning transparency and information disclosure at the interface layer. Participants sparred with the idea that by disclosing more information they would be overloading the user, perhaps ruining the user experience: "The underlying problems of these algorithms we're trying to solve is overload and so the thing that I struggle with is coming up with a way to not add to the overload when we're helping people understand" (CS2). Many stressed that transparency information should be disclosed in a form that is translated for a general public audience, or made available on demand to those technical readers that really care about the nuts and bolts. But this is a delicate line to walk: "If it's a blurb, you risk it being too simplified and people … interpreting it wrong because you weren't precise enough in the hopes of being brief" (CS2).

Several focus group participants were skeptical that end-users would even care about algorithmic transparency, as one participant put it: "if you put yourself in the average Joe's shoes, what issue is it that they're seeing with the algorithms? I'm not sure…" (CS2). Though this sentiment was balanced by comments from others such as, "The audience has changed. The audience wants transparency now" (CS1). Another participant (CS2) remarked: "If it's detailed enough for somebody to understand the science of algorithms to really use critically, maybe 1 percent of the readership would be able to use it." In the case of algorithmic transparency indeed many people may not actually understand the disclosed information, which raises the challenge of trying to meet the information needs of diverse users while not "polluting" the user experience with a surfeit of information for the uninterested.

Consequences and Challenges to Algorithmic Transparency

After participants discussed aspects of algorithms that might be disclosed, they were prompted by the facilitators to start evaluating those factors along dimensions such as feasibility of disclosure (technical as well as general feasibility), openness to manipulation or gaming, cost or effort needed, and mechanism for presentation.

Chief among the moderating factors for more transparency were business issues: what would motivate a news or media organization to disclose details about their algorithms? Costs identified in producing transparency information included: data preparation, documentation writing, source code polishing, and benchmark testing. "How does being transparent offset loss of revenue," remarked one participant in CS2. Other disincentives to transparency pointed out were: privacy issues which might arise from disclosing data that is not properly anonymized, as well as the legal implications of admitting errors or uncertainties that could open the door for lawsuits. There was some concern that disclosing uncertainty information might lead readers to think the results were unreliable: "If I'm a reader and it says we're 80 percent certain of this, which might be very legitimate certainty to a scientist but to the reader they're like, then the article is useless" (CS3).

Participants also suggested that disclosing aspects of how a proprietary algorithmic system works may hurt an organization's technological competitive advantage, or open the system to manipulation by third parties. Yet it was recognized that people will game the system no matter what, and that by disclosing information publicly it would level the playing field or, as one participant (CS2) put it, "everybody has the same chance now because we all know the rules of the game."

While the business disincentives mentioned were plentiful, no clear and compelling value proposition for an organization to self-motivate and disclose transparency information about algorithms was proposed. Some suggested that if an organization started to lose income or users that might be an incentive to disclose information to help win those users back. Trust, credibility, reputation, and legitimacy were factors that participants clung to as positive outcomes of more transparency information:

> I don't know that it affects the public perception of the story in one way or another, but I think that it might shape the internal process and improve the legitimacy and the credibility of the projects in some way. (CS3)

Yet participants were also aware of transparency as a strategic act in which organizations might only disclose information that bolstered legitimacy but withhold information that somehow hurt the public's image of the organization. Overall, transparency was seen as something that can contribute to holding news organizations accountable, but it does not seem to have a compelling business case.

Discussion and Conclusion

Our results indicate that there are numerous aspects of algorithmic systems that could be disclosed in an effort to advance a journalistic truth-telling process that increasingly hinges on the norm of transparency. While our findings indicate two key dimensions as obstacles to algorithmic transparency: (1) a lack of business incentives for disclosure, and (2) the concern of overwhelming end-users with too much information, we were able to identify several opportunities for disclosure across four main layers of the input–output pipeline of algorithms: data, model, inference, and interface. Recent theoretical research on the ethics surrounding algorithmic journalism (Dörr and Hollnbuchner 2016) suggests aspects of data and code transparency that the current study empirically enumerates. Participants repeatedly emphasized the essential role humans play in all of these four areas. Recent work by Ananny (2015) to develop an ethical framework around algorithms cedes this point by considering the sociotechnical aggregation of code, practices, and norms as the unit of ethical analysis. Others have made inroads to studying the ethics of algorithmic systems through ethnographic techniques that allow for thick description of the dynamic human design processes around algorithmic system development (Neyland 2016). To the degree that they can be, a central challenge for the future study of algorithmic transparency will be to disentangle the roles and decisions of humans versus algorithms in specific algorithmic systems in order to highlight the hybrid nature of these systems.

The various criteria of disclosure suggested by this research proffer an opportunity to advance the deontological ethics of algorithms (Ananny 2015; Sandvig et al. 2015) by suggesting standards for what might be disclosed about algorithmic systems, including aspects of human involvement. The typology of factors enumerated in Table 1 forms a palette of options that may be disclosed about algorithms as part of an incipient algorithmic transparency standard. Such a standard can be informative not only for the disclosure of information about algorithms but perhaps also for the design of algorithms tuned to facilitate such disclosures. At the same time we must be careful to state that our framework may not be entirely comprehensive since it relies on the insights from participants around three case studies. While we chose these cases to be distinct and to cover diverse ways in which algorithms are used across the media production pipeline, they may not span the entire space of algorithms in use in news systems. Still, our results should spur future experimentation and evaluation of how such disclosures might be adopted by news organizations and subsequently received by end-users as a part of editorial guidelines or professional standards. This is a pragmatic path forward as one survey of journalists found that the largest impacts on journalist behavior were company editorial guidelines (59 percent thought there was some or high impact), and professional codes of ethics (50 percent) (Powell and Jempson 2014).

The value systems of new media platforms or apps do not necessarily share the same institutional origins as traditional media. Foundational norms and ethics around issues such as the public interest, diversity, or transparency may not be present (Napoli 2015). The liminal press described by Ananny and Crawford (2015) in their study of news app designers suggests that the value systems they observed were more oriented toward end-user information needs, experiences, and engagement. Our results indicated a similar focus, with conversations underscoring the end-user value proposition for information disclosure, and implications for harming the user experience by providing too much irrelevant information. Self-regulation in news media accountability and algorithmic governance is not well incentivized (Fengler and Russ-Mohl 2014; Saurwein et al. 2015), but a focus on user experience may provide a motivating path to implement transparency.

Rather than evoke normative arguments, news organizations might align information disclosures to enhance the user experience (Diakopoulos, Cass, and Romero 2014). This would have the added advantage of subverting cost disincentives because the overall product would benefit by having a more appealing and satisfying user experience. Such an approach would consider the tasks that users typically engage with in a news application, such as browsing for news, accessing recommendations, and sharing, and orient system transparency toward those tasks. This would require explicit task modeling to know what decisions and tasks a user is indeed engaged with in a specific context. This approach may be fruitful for transparency disclosures that are integrated directly into a user interface. A limitation of the current study is that we cannot say anything definitive about the real audience demand for algorithmic transparency. Future work should evaluate the algorithmic transparency factors reported in this paper (summarized in Table 1) in order to assess demand by end-users as well as how each factor affects the usability, learnability, user satisfaction, acceptance, and credibility of the system.

Yet, while such a focus on user experience might provide a business-oriented argument that is appealing to news organizations, the approach does not address a wider notion of media accountability. There are aspects of transparency information that are irrelevant to an immediate individual user context, but which are nonetheless of importance in media accountability for a broad public such as fair and uncensored access to information, bias in attention patterns, and other aggregate metrics of, for instance, error rates. In other words, some factors may have bearing on an individual whereas others have import for a larger public. Different disclosure mechanisms, such as periodic reports by ombudspeople may be more appropriate for factors like benchmarks, error analysis, or the methodology of data collection and processing, since they may not be of interest to or even comprehensible for many users yet demanded by those who value an expert account and explanation (Tilly 2006). Such reports hold organizations and institutions accountable but are not necessarily concerned with the user experience on a news-item level. A multi-layered "pyramid" model involving hints of transparency at the tip, the user-interface level, would then open up to progressively broader and denser descriptions of system transparency as the motivated end-user worked their way down to the base of the pyramid. This would support different types of end-users taking advantage of transparency information—not just a generic general end-user audience but an informed and interested range of actors such as other journalists, critics, activists, or policy makers.

An effective policy for algorithmic transparency in the news media will no doubt require additional research, development, and experimentation. Practical approaches will combine different disclosure formats that depend on the factor being disclosed and the likely audience for that information. In summary, this research makes clear that algorithmic transparency requires the consideration of multiple stakeholder perspectives including organizational incentives and costs, end-user tasks and utility, and ethical normative practices.

ACKNOWLEDGEMENTS

From the Tow Center for Digital Journalism at Columbia University we wish to thank Fergus Pitt for his deft help in moderating the focus groups and Claire Wardle for helpful feedback on an earlier version of this paper.

FUNDING

This research is supported by the Tow Center for Digital Journalism at Columbia University [grant number PG006417].

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

NOTE

1. As news production process we define the research, decisions, creation, and distribution of news stories using various sources, tools (from pencil to algorithms), and platforms.

REFERENCES

Albeanu, Catalina. 2015. "Inside Velocity: Mashable's Predictive Social Analytics Tool." *Journalism.co.uk*. https://www.journalism.co.uk/news/inside-velocity-how-mashable-s-predictive-social-analytics-tool-fits-in-the-newsroom/s2/a565861/.

Allen, David S. 2008. "The Trouble with Transparency." *Journalism Studies* 9 (3): 323–340.

Ananny, Mike. 2015. "Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness." *Science, Technology, & Human Values* 41 (1): 93–117.

Ananny, Mike, and Kate Crawford. 2015. "A Liminal Press: Situating News App Designers within a Field of Networked News Production." *Digital Journalism* 3 (2): 192–208.

Associated Press. 2015. AP, NCAA to Grow College Sports Coverage with Automated Game Stories. Associated Press. http://www.ap.org/content/press-release/2015/ap-ncaa-to-grow-college-sports-coverage-with-automated-game-stories.

Balkin, Jack M. 1999. "How Mass Media Simulate Political Transparency." *Cultural Values* 3 (4): 393–413.

Bennett, Philip. 2014. "Truth Vigilantes: On Journalism and Transparency." In *Transparency in Politics and the Media*, edited by N. Bowles, J. Hamilton, and D. A. L. Levy, 103–122. London: I.B. Tauris.

Bennis, Warren. 2013. "Leadership in a Digital World: Embracing Transparency and Adaptive Capacity." *MIS Quarterly* 37 (2): 635–636.

Boczkowski, Pablo J. 2005. *Digitizing the News: Innovation in Online Newspapers*. Cambridge, MA: MIT Press.

Breton, Albert. 2006. *The Economics of Transparency in Politics*. Aldershot, England: Ashgate.

Broersma, Marcel. 2013. "A Refractured Paradigm. Journalism, Hoaxes and the Challenge of Trust." In *Trust and Participation in a Transformed News Landscape*, edited by C. Peters and M. J. Broersma, 28–44. New York: Routledge.

Chadha, Kalyani, and Michael Koliska. 2014. "Newsrooms and Transparency in the Digital Age." *Journalism Practice* 9 (2): 215–229.

Coddington, Mark. 2015. "Clarifying Journalism's Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting." *Digital Journalism* 3 (3): 331–348.

Cotterrell, Roger. 1999. "Transparency, Mass Media, Ideology and Community." *Cultural Values* 3 (4): 414–426.

Craft, Stephanie, and Kyle Heim. 2009. "Transparency in Journalism: Meanings, Merits, and Risks." In *The Handbook of Mass Media Ethics*, edited by L. Wilkins and C. G. Christians, 217–228. New York: Routledge.

Cramer, Henriette, Vanessa Evers, Satyan Ramlal, Maarten van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. "The Effects of Transparency on Trust in and Acceptance of a Content-Based Art Recommender." *User Modeling and User-Adapted Interaction* 18 (5): 455–496.

Deuze, Mark. 2005. "What is Journalism?: Professional Identity and Ideology of Journalists Reconsidered." *Journalism* 6 (4): 422–464.

Diakopoulos, Nicholas. 2015. "Algorithmic Accountability: Journalistic Investigation of Computational Power Structures." *Digital Journalism* 3 (3): 398–415.

Diakopoulos, Nicholas. Forthcoming. "Computational Journalism and the Emergence of News Platforms." In *The Routledge Companion to Digital Journalism Studies*, edited by S. Eldridge and B. Franklin. https://www.routledge.com/The-Routledge-Companion-to-Digital-Journalism-Studies/Franklin-Eldridge-II/p/book/9781138887961.

Diakopoulos, Nicholas, Stephen Cass, and Josh Romero. 2014. "Data-Driven Rankings: The Design and Development of the IEEE Top Programming Languages News App." Proceedings Symposium on Computation + Journalism.

Dörr, Konstantin. 2016. "Mapping the Field of Algorithmic Journalism." *Digital Journalism* 4 (6): 700–722.

Dörr, Konstantin, and Katherina Hollnbuchner. 2016. "Ethical Challenges of Algorithmic Journalism." *Digital Journalism*. doi:10.1080/21670811.2016.1167612.

El-Arini, Khalid, Ulrich Paquet, Ralf Herbrich, Jurgen Van Gael, and Blaise Agüera y Arcas. 2012. "Transparent User Models for Personalization." Proceedings International Conference on Knowledge Discovery and Data Mining (KDD), 678–686.

Fengler, Susanne, and Stephan Russ-Mohl. 2014. "The (Behavioral) Economics of Media Accountability." In *Journalists and Media Accountability: An International Study of News People in the Digital Age*, edited by S. Fengler, T. Eberwein, G. Mazzoleni, and C. Porlezza, 213–230. New York: Peter Lang Publishing.

Fombrun, Charles J. 2005. "Building Corporate Reputation through CSR Initiatives: Evolving Standards." *Corporate Reputation Review* 8: 7–12.

Friedman, Batya, and Helen Nissenbaum. 1996. "Bias in Computer Systems." *ACM Trans. on Information Systems* 14 (3): 330–347.

Fukuyama, Francis. 1995. *Trust: The Social Virtues and the Creation of Prosperity*. New York: Free Press.

Fung, Archon, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. New York: Cambridge University Press.

Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by G. Tarleton, Pablo J. Boczkowski, and Kirsten A. Foot, 167–194. Cambridge, MA: MIT Press.

Graefe, Andreas. 2016. Guide to Automated Journalism. *Tow Center for Digital Journalism Report*. http://towcenter.org/research/guide-to-automated-journalism/.

Granados, Nelson, and Alok Gupta. 2013. "Transparency Strategy: Competing with Information in a Digital World." *MIS Quarterly: Management Information Systems* 37 (2): 637–641.

Grochowski Jones, Ryann, and Charles Ornstein. 2016. *Matching Industry Payments to Medicare Prescribing Patterns: An Analysis*. ProPublica Whitepaper. https://static.propublica.org/projects/d4d/20160317-matching-industry-payments.pdf?22.

Hayes, Arthur S., Jane B. Singer, and Jerry Ceppos. 2007. "Shifting Roles, Enduring Values: The Credible Journalist in a Digital Age." *Journal of Mass Media Ethics* 22 (4): 262–279.

Hellmueller, Lea, Tim P. Vos, and Mark A. Poepsel. 2013. "Shifting Journalistic Capital?" *Journalism Studies* 14 (3): 287–304.

Holtz, Shel, and John C. Havens. 2009. *Tactical Transparency: How Leaders Can Leverage Social Media to Maximize Value and Build Their Brand*. San Francisco, CA: Jossey-Bass.

Karlsson, Michael. 2010. "Rituals of Transparency." *Journalism Studies* 11 (4): 535–545.

Karlsson, Michael. 2011. "The Immediacy of Online News, the Visibility of Journalistic Processes and a Restructuring of Journalistic Authority." *Journalism* 12 (3): 279–295.

Karlsson, Michael, Christer Clerwall, and Lars Nord. 2014. "You Ain't Seen Nothing Yet." *Journalism Studies* 15 (5): 668–678.

Kemper, Elizabeth, Sam Stringfield, and Charles Teddlie. 2003. "Mixed Methods Sampling Strategies in Social Science Research." In *Handbook of Mixed Methods in Social & Behavioral Research*, edited by A. Tashakkori and C. Teddlie, 273–296. Thousand Oaks, CA: Sage.

Kizilcec, Rene F. 2016. "How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface." Proceedings Conference on Human Factors in Computing Systems (CHI).

Kovach, Bill, and Tom Rosenstiel. 2014. *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. 3rd ed. New York: Three Rivers Press.

Lewis, Seth C., and Nikki Usher. 2013. "Open Source and Journalism: Toward New Frameworks for Imagining News Innovation." *Media, Culture & Society* 35 (5): 602–619.

Lofland, John, and Lyn H. Lofland. 1994. *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. 3rd ed. Belmont, CA: Wadsworth.

Lokot, Tetyana, and Nicholas Diakopoulos. 2016. "News Bots: Automating News and Information Dissemination on Twitter." *Digital Journalism* 4 (6): 682–699.

Mager, Astrid. 2012. "Algorithmic Ideology: How Capitalist Society Shapes Search Engines." *Information, Communication & Society* 15 (5): 769–787.

McBride, Kelly, and Tom Rosenstiel. 2014. *The New Ethics of Journalism: Principles for the 21st Century*. Thousand Oaks, CA: CQ Press.

Meijer, Albert. 2009. "Understanding Modern Transparency." *International Review of Administrative Sciences* 75 (2): 255–269.

Meyer, Philip. 2006. *The Vanishing Newspaper: Saving Journalism in the Information Age*. Columbia, SC: University of Missouri Press.

Morgan, David L. 1996. "Focus Groups." *Annual Review of Sociology* 22: 129–152.

Napoli, Philip M. 2015. "Social Media and the Public Interest: Governance of News Platforms in the Realm of Individual and Algorithmic Gatekeepers." *Telecommunications Policy* 39 (9): 751–760.

Narrative Science. n.d. "Meet a Few of Our Fan Favorites." *NarrativeScience.com*. https://www.narrativescience.com/sports.

Neyland, Daniel. 2016. "Bearing Account-Able Witness to the Ethical Algorithmic System." *Science, Technology, and Human Values* 41 (1): 50–76.

Nguyen, Dao, Andrew Kelleher, and Adam Kelleher. 2015. "Introducing Pound: Process for Optimizing and Understanding Network Diffusion." *BuzzFeed Tech Blog*. http://www.buzzfeed.com/daozers/introducing-pound-process-for-optimizing-and-understanding-n.

Nissenbaum, Helen. 1996. "Accountability in a Computerized Society." *Science and Engineering Ethics* 2 (1): 25–42.

Oremus, Will. 2014. "The First News Report on the L.a. Earthquake Was Written by a Robot." *Slate*. http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html.

Owen, David. 2005. "CSR after Enron: A Role for the Academic Accounting Profession?" *European Accounting Review* 14 (2): 395–404.

Park, Deokgun, Simranjit Singh, Nicholas Diakopoulos, and Niklas Elmqvist. 2016. "Interactive Identification of High Quality Online News Comments." Proceedings Conference on Human Factors in Computing Systems (CHI).

Pew. 2012. *Further Decline in Credibility Ratings for Most News Organizations*. Accessed November 9, 2015. http://www.people-press.org/files/2012/08/8-16-2012-Media-Believability1.pdf.

Plaisance, Patrick Lee. 2007. "Transparency: An Assessment of the Kantian Roots of a Key Element in Media Ethics Practice." *Journal of Mass Media Ethics* 22 (2–3): 187–207.

Powell, Wayne, and Mike Jempson. 2014. "More Accountability in the Digital Age? The Influence of New Technologies." In *Journalists and Media Accountability: An International Study of News People in the Digital Age*, edited by S. Fengler, T. Eberwein, G. Mazzoleni, and C. Porlezza, 115–128. New York: Peter Lang Publishing.

Puchta, Claudia, and Jonathan Potter. 2004. *Focus Group Practice*. London: Sage.

Roberts, Chris. 2007. Measuring the Relationship between Journalistic Transparency and Credibility. Columbia, SC: University of South Carolina.

RTDNA. 2015. "New RTDNA Code of Ethics." *Radio Television Digital News Association*. Accessed November 3, 2015. http://rtdna.org/content/rtdna_code_of_ethics#.VZRTgO1VhBc

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langobort. 2015. "Can an Algorithm Be Unethical?" Paper presented to the 65th annual meeting of the International Communication Association.

Saurwein, Florian, Natascha Just, and Michael Latzer. 2015. "Governance of Algorithms: Options and Limitations." *Info* 17 (6): 35–49.

Schaffer, James, Prasanna Giridhar, Debra Jones, Tobias Höllerer, Tarek Abdelzaher, and John O'Donovan. 2015. "Getting the Message?: A Study of Explanation Interfaces for Microblog Data Analysis." Proceedings Conference on Intelligent User Interfaces (IUI).

Schudson, Michael. 2015. *The Rise of the Right to Know: Politics and the Culture of Transparency, 1945–1975*. Cambridge, MA: The BelkPress of Harvard University Press.

Shearer, Matt, Basile Simon, and Clement Geiger. 2014. "Datastringer: Easy Dataset Monitoring for Journalists." Proceedings Symposium on Computation + Journalism.

Silverman, Craig. 2013. "Corrections and Ethics." In *The New Ethics of Journalism—Principles for the 21st Century*, edited by K. McBride and T. Rosenstiel, 151–161. Thousand Oaks, CA: Sage.

Singer, Jane B. 2007. "Contested Autonomy." *Journalism Studies* 8 (1): 79–95.

SPJ. 2014. "Code of Ethics." *Society of Professional Journalists*. Accessed November 3, 2015. http://www.spj.org/ethicscode.asp

Sullivan, Margaret. 2013. "When Reporters get Personal." *The New York Times*. Accessed January 5. http://www.nytimes.com/2013/01/06/public-editor/when-reporters-get-personal.html

Tapscott, Don, and David Ticoll. 2003. *The Naked Corporation*. New York: The Free Press.

Thurman, Neil, Steve Schifferes, Richard Fletcher, Nic Newman, Stephen Hunt, and Aljosha Karim Schapals. 2016. "Giving Computers a Nose for News: Exploring the Limits of Story Detection and Verification." *Digital Journalism*. doi:10.1080/21670811.2016.1149436.

Tilly, Charles. 2006. *Why? What Happens When People Give Reasons … and Why*. Princeton, NJ: Princeton University Press.

Tintarev, Nava, and Judith Masthoff. 2007. "A Survey of Explanations in Recommender Systems." Proceedings Data Engineering Workshop.

Wang, Shan. 2015. "The New York times Built a Slack Bot to Help Decide Which Stories to Post to Social Media." *NiemanLab*. http://www.niemanlab.org/2015/08/the-new-york-times-built-a-slack-bot-to-help-decide-which-stories-to-post-to-social-media/.

Ward, Stephen J. A. 2014. "The Magical Concept of Transparency." In *Ethics for Digital Journalists: Emerging Best Practices*, edited by L. Zion and D. Craig, 45–58. New York: Routledge.

Weinberger, David. 2009. Transparency is the New Objectivity. http://www.hyperorg.com/blogger/2009/07/19/transparency-is-the-new-objectivity.

Yahoo. n.d. "Matchup Recaps." *Yahoo.com*. Accessed October 27, 2015. https://help.yahoo.com/kb/SLN8766.html

Young, Mary Lynn, and Alfred Hermida. 2015. "From Mr. and Mrs. Outlier to Central Tendencies: Computational Journalism and Crime Reporting at the Los Angeles Times." *Digital Journalism* 3 (3): 381–397.

Nicholas Diakopoulos (author to whom correspondence should be addressed), Philip Merrill College of Journalism, University of Maryland, USA; Corresponding author. E-mail: nad@umd.edu. ORCID http://orcid.org/0000-0001-5005-6123

Michael Koliska School of Communication and Journalism, Auburn University, USA. E-mail: mkoliska@auburn.edu