Article

new media & society
2018, Vol. 20(3) 973–989

DOI: 10.1177/1461444816676645
journals.sagepub.com/home/nms
SAGE

Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability

Mike Ananny
University of Southern California, USA

Kate Crawford
Microsoft Research New York City, USA; New York University, USA; MIT, USA

Abstract
Models for understanding and holding systems accountable have long rested upon ideals and logics of transparency. Being able to see a system is sometimes equated with being able to know how it works and govern it—a pattern that recurs in recent work about transparency and computational systems. But can "black boxes' ever be opened, and if so, would that ever be sufficient? In this article, we critically interrogate the ideal of transparency, trace some of its roots in scientific and sociotechnical epistemological cultures, and present 10 limitations to its application. We specifically focus on the inadequacy of transparency for understanding and governing algorithmic systems and sketch an alternative typology of algorithmic accountability grounded in constructive engagements with the limitations of transparency ideals.

Keywords
Accountability, algorithms, critical infrastructure studies, platform governance, transparency

> The observer must be included within the focus of observation, and what can be studied is always a relationship or an infinite regress of relationships. Never a "thing."
>
> Bateson (2000: 246)

Corresponding author:
Mike Ananny, University of Southern California, Los Angeles, CA 90089, USA.
Email: ananny@usc.edu

Introduction

Algorithmic decision-making is being embedded in more public systems—from transport to healthcare to policing—and with that has come greater demands for algorithmic transparency (Diakopoulos, 2016; Pasquale, 2015). But what kind of transparency is being demanded? Given the recent attention on transparency as a type of "accountability" in algorithmic systems, it is an important moment to consider what calls for transparency invoke: how has transparency as an ideal worked historically and technically within broader debates about information and accountability? How can approaches from Science and Technology Studies (STS) contribute to the transparency debate and help to avoid the historical pitfalls? And are we demanding too little when we ask to "look inside the black box"? In some contexts, transparency arguments come at the cost of a deeper engagement with the material and ideological realities of contemporary computation. Rather than focusing narrowly on technical issues associated with algorithmic transparency, we begin by reviewing a long history of the transparency ideal, where it has been found wanting, and how we might address those limitations.

Transparency, as an ideal, can be traced through many histories of practice. From philosophers concerned with the epistemological production of truth, through activists striving for government accountability, transparency has offered a way to see inside the truth of a system. The implicit assumption behind calls for transparency is that *seeing* a phenomenon creates opportunities and obligations to make it accountable and thus to *change* it. We suggest here that rather than privileging a type of accountability that needs to look *inside* systems, that we instead hold systems accountable by looking *across* them—seeing them as sociotechnical systems that do not *contain* complexity but *enact* complexity by connecting to and intertwining with assemblages of humans and non-humans.

To understand how transparency works as an ideal, and where it fails, we trace its history and present 10 significant limitations. We then discuss why transparency is an inadequate way to understand—much less govern—algorithms. We finish by suggesting an alternative typology of algorithmic governance: one grounded in recognizing and ameliorating the limits of transparency, with an eye toward developing an ethics of algorithmic accountability.

Transparency as an ideal

Transparency concerns are commonly driven by a certain chain of logic: observation produces insights which create the knowledge required to govern and hold systems accountable. This logic rests on an epistemological assumption that "truth is correspondence to, or with, a fact" (David, 2015: n.p.). The more facts revealed, the more truth that can be known through a logic of accumulation. Observation is understood as a diagnostic for ethical action, as observers with more access to the facts describing a system will be better able to judge whether a system is working as intended and what changes are required. The more that is known about a system's inner workings, the more defensibly it can be governed and held accountable.

This chain of logic entails "a rejection of established representations" in order to realize "a dream of moving outside representation understood as bias and distortion" toward

"representations [that] are more intrinsically true than others." It lets observers "*uncover* the *true essence*" of a system (Christensen and Cheney, 2015: 77). The hope to "uncover" a singular truth was a hallmark of The Enlightenment, part of what Daston (1992: 607) calls the attempt to escape the idiosyncrasies of perspective: a "transcendence of individual viewpoints in deliberation and action [that] seemed a precondition for a just and harmonious society."

Several historians point to early Enlightenment practices around scientific evidence and social engineering as sites where transparency first emerged in a modern form (Crary, 1990; Daston, 1992; Daston and Galison, 2007; Hood, 2006). Hood (2006) first sees transparency as an empirical ideal appearing in the epistemological foundations of "many eighteenth-century ideas about social science, that the social world should be made knowable by methods analogous to those used in the natural sciences" (p. 8). These early methods of transparency took different forms. In Sweden, press freedom first appeared with the country's 1766 *Tryckfrihetsfororordningen* law ("Freedom of the Press Act") which gave publishers "rights of statutory access to government records." In France, Nicholas de La Mare's *Traite de la Police* volumes mapped Parisian street crime patterns to demonstrate a new kind of "police science"—a way to engineer social transparency so that "street lighting, open spaces with maximum exposure to public view, surveillance, records, and publication of information" would become "key tools of crime prevention" (Hood, 2006: 8). In the 1790s, Bentham saw "inspective architectures" as manifestations of the era's new science of politics that would marry epistemology and ethics to show the "indisputable truth" that "the more strictly we are watched, the better we behave" (Cited in Hood (2006: 9–10). Such architectures of viewing and control were unevenly applied as transparency became a rationale for racial discrimination. A significant early example can be found in New York City's 18th century "Lantern Laws," which required "black, mixed-race, and indigenous slaves to carry small lamps, if in the streets after dark and unescorted by a white person." Technologies of seeing and surveillance were inseparable from material architectures of domination as everything "from a candle flame to the white gaze" were used to identify who was in their rightful place and who required censure (Browne, 2015: 25).

Transparency is thus not simply "a precise end state in which everything is clear and apparent," but a system of observing and knowing that promises a form of control. It includes an affective dimension, tied up with a fear of secrets, the *feeling* that seeing something may lead to control over it, and liberal democracy's promise that openness ultimately creates security (Phillips, 2011). This autonomy-through-openness assumes that "information is easily discernible and legible; that audiences are competent, involved, and able to comprehend" the information made visible (Christensen and Cheney, 2015: 74)—and that they *act* to create the potential futures openness suggests are possible. In this way, transparency becomes *performative*: it does work, casts systems as knowable and, by articulating inner workings, it *produces* understanding (Mackenzie, 2008).

Indeed, the intertwined promises of openness, accountability and autonomy drove the creation of 20thcentury "domains of institutionalized disclosure" (Schudson, 2015: 7). These institutionalizations and cultural interventions appeared in the US Administrative Procedures Act of 1946, the 1966 Freedom of Information Act, the Sunshine Act, truth in packaging and lending legislation, nutritional labels, environmental impact reports and

chemical disclosure rules, published hospital mortality rates, fiscal reporting regulations, and the Belmont Report on research ethics. And professions like medicine, law, advertising, and journalism all made self-regulatory moves to shore up their ethical codes, creating policies on transparency and accountability before the government did it for them (Birchall, 2011; Fung et al., 2008; Schudson, 2015).

The institutionalization of openness led to an explosion of academic research, policy interventions, and cultural commentary on what kind of transparency this produces—largely seeing openness as a verification tool to "regulate behavior and improve organizational and societal affairs" or as a performance of communication and interpretation that is far less certain about whether "more information generates better conduct" (Albu and Flyverbom, 2016: 14).

Policy and management scholars have identified three broad metaphors underlying organizational transparency: as a "public value embraced by society to counter corruption," as a synonym for "open decision-making by governments and nonprofits," and as a "complex tool of good governance"—all designed to create systems for "accountability, efficiency, and effectiveness" (Ball, 2009: 293). Typologies of transparency have emerged:

- "*Fuzzy*" transparency (offering "information that does not reveal how institutions actually behave in practice … that is divulged only nominally, or which is revealed but turns out to be unreliable") versus "*clear*" transparency ("programmes that reveal reliable information about institutional performance, specifying officials' responsibilities as well as where public funds go") (Fox, 2007: 667);
- Transparency that creates "*soft*" accountability (in which organizations must answer for their actions) versus "*hard*" accountability (in which transparency brings the power to sanction and demand compensation for harms) (Fox, 2007);
- Transparency "*upwards*" ("the hierarchical superior/principal can observe the conduct, behavior, and/or 'results' of the hierarchical subordinate/agent") versus "*downwards*" ("the 'ruled' can observe the conduct, behavior, and/or 'results' of their 'rulers'") versus "*outwards*" ("the hierarchical subordinate or agent can observe what is happening 'outside' the organization") versus "*inwards*" ("when those outside can observe what is going on inside the organization") (Heald, 2006: 27–28);
- Transparency as *event* (the "inputs, outputs, or outcomes" that define the "objects of transparency") versus as *process* (the organizational "rules, regulations, and procedures" that define the conditions of visibility) (Heald, 2006: 29–32);
- Transparency in *retrospect* (an organization's "periodic" or "ex post account of stewardship and management") versus *real-time* (in which "the accountability window is always open and surveillance is continuous") (Heald, 2006: 32–33).

More recent scholarship has connected these typologies of transparency to pre-histories of the Internet, contemporary digital cultures, and the promises of open government. Finding World War II-era evidence of Heald's model of multi-dimensional transparency, Turner (2013) describes how counter-fascist artists, technologists, and political theorists invented immersive media installations designed to make new types of democratic

viewing possible. This story of enacting democratic ideals through the performance of "open" technological cultures has many chapters, including the growth of early virtual communities from counter-culture roots (Turner, 2006), the pursuit of radical transparency by open source programming collectives (Kelty, 2008), and Wikipedia editing communities (Tkacz, 2015), the activism of "hacker" movements premised on enacting libertarian self-realization through open coding ethics (Coleman, 2013), and the emerging intersection of software hacking movements with open government initiatives (Schrock, 2016). Much research on online participatory culture is infused with unexamined assumptions about the benefits of transparency, equating the ability to see inside a system with the power to govern it.

Currently, there is a strong strand of research that emphasizes "algorithmic transparency," illustrated best by Pasquale's (2015) *The Black Box Society*, but also underpinning proposals by Diakopoulos (2016), Brill (2015), and Soltani (Zara, 2015). Transparency can be at the level of platform design and algorithmic mechanisms or more deeply at the level of a software system's logic. Early computer scientists attempted to communicate the power of code and algorithms through visualizations designed to give audiences an appreciation of programming decisions and consequences. For example, in 1966, Knowlton created an "early computer animation explaining the instruction set of a low-level list processing language" with Baecker (1998) following up in 1971 with a series of animations comparing the speed and efficacy of different sorting algorithms. Such algorithm animations formed the early basis of later "taxonomy of algorithm animation displays" that showed the diversity of approaches computer scientists used to communicate using animation the inner workings of algorithms to students and non-specialists (Brown, 1998).

Some contemporary researchers in computer science consider algorithmic transparency a way to prevent forms of discrimination. For example, a recent study showed that men are more likely to see ads for highly paid jobs than women when searching on Google (Datta et al., 2015). The researchers conclude that even Google's own transparency tools are problematically opaque and unhelpful. Datta et al. (2015) conclude with a call for software engineers to "produce machine learning algorithms that automatically avoid discriminating against users in unacceptable ways and automatically provide transparency to users" (p. 106). Within the discourse of computer science, transparency is often seen as desirable because it brings insight and governance. It assumes that knowing is possible by seeing, and that seemingly objective computational technologies like algorithms enact and can be held accountable to a correspondence theory of truth.

Limits of the transparency ideal

But this ideal of transparency—as a method to see, understand and govern complex systems in timely fashions—is limited in significant ways.

Most fundamentally, the epistemological premise of transparency is challenged by the "pragmatic maxim" that sees truth as meanings achieved through *relations*, not representations. Pragmatists "focus on the contingencies of human practice, denying the availability of a transcendental standpoint from which we might judge the worth of those practices" (Bacon, 2012: 5).

Although pragmatism is open to charges of relativism, experientialism, and utilitarianism—seeing all ideas as potentially valid, privileging sense-makers' experiences, and too strictly equating an idea's truth with its workable deployment in social environments (Misak, 2007)—and seeing a system's inner workings can indeed provide insight and spur further investigation, proponents of experiential learning claim that a system's significance and power is most revealed by understanding both its viewable, external connections to its environments and its internal, self-regulating workings. Formal descriptions of observable phenomena are valuable, but if ideas "become true just in so far as they help us to get into satisfactory relation with other parts of our experience," (James, 1997: 100) then accountability-through-visibility is only possible by seeing truths through their "becoming or destruction, not by their intrinsic nature or correspondence" (Ward, 1996: 111) to a revealed representation presumed to be a truthful account of a how systems *really* work when free of their relations.

Pursuing views into a system's inner workings is not only inadequate, it creates a "transparency illusion," (Heald, 2006: 34) promising consequential accountability that transparency cannot deliver. Below, we isolate 10 limitations of the transparency ideal: not as a complete list, but what we see as the most common and entrenched shortcomings.

Transparency can be disconnected from power

If transparency has no meaningful effects, then the idea of transparency can lose its purpose. If corrupt practices continue after they have been made transparent, "public knowledge arising from greater transparency may lead to more cynicism, indeed perhaps to wider corruption." Visibility carries risks for the goal of accountability if there is no system ready and "capable of processing, digesting, and using the information" to create change (Heald, 2006: 35–37). Transparency can reveal corruption and power asymmetries in ways intended to shame those responsible and compel them to action, but this assumes that those being shamed are *vulnerable* to public exposure. Power that comes from special, private interests driven by commodification of people's behaviors may ultimately be immune to transparency—for example, the data broker industry has so far proven more powerful than calls for transparency because, in part, once "information has been swept into the data broker marketplace, it becomes challenging and in many cases impossible to trace any given datum to its original source" (Crain, 2018: 6).

Transparency can be harmful

Full transparency can do great harm. If implemented without a notion of *why* some part of a system should be revealed, transparency can threaten privacy and inhibit honest conversation. "It may expose vulnerable individuals or groups to intimidation by powerful and potentially malevolent authorities" (Schudson, 2015: 4–5). Indeed, "radical transparency" (Birchall, 2014) can harm the formation of enclave public spheres in which marginalized groups hide "counterhegemonic ideas and strategies in order to survive or avoid sanctions, while internally producing lively debate and planning" (Squires, 2002: 448). Hidden transcripts—like those that organized the underground railroad during US slavery (Bratich, 2016: 180)—"guard against unwanted publicity of the group's true opinions,

ideas, and tactics for survival" (Squires, 2002: 458). Companies often invoke a commercial version of this when they resist transparency in order to protect their proprietary investments and trade secrets (Crain, 2018) and prevent people from knowing enough to "game" their systems and unfairly receive goods or services (Diakopoulos, 2016: 58–59). Secrecy and visibility here "are not treated here as values in and of themselves, but as instruments in struggles" (Bratich, 2016: 180–181) and evidence of how transparency both reflects and shapes social power.

Transparency can intentionally occlude

Stohl et al. (2016) distinguish between *inadvertent opacity*—in which "visibility produces such great quantities of information that important pieces of information become inadvertently hidden in the detritus of the information made visible"—and *strategic opacity*—in which actors "bound by transparency regulations" purposefully make so much information "visible that unimportant pieces of information will take so much time and effort to sift through that receivers will be distracted from the central information the actor wishes to conceal" (pp. 133–134). We can think of the emblematic case where an organization is asked to share its records, so it prints out all records into hundreds of reams of paper that must be manually waded through in order to find any incriminating evidence. It is a form of transparency without being usable: a *resistant* transparency.

Transparency can create false binaries

Several of the key limitations of transparency as an instrument of accountability are linked to the seemingly dualistic nature of seeing. Fox (2007: 663) asks it this way: "what kinds of transparency lead to what kinds of accountability, and under what conditions?" A contemporary example can be found in the comparison of Wikileaks' revelation of the entire Afghan War Logs database (Leigh and Harding, 2011)—acquiescing to some redaction only at the demand of journalistic partners (Coddington, 2012)—to Edward Snowden's refusal to publicly reveal the entire database of National Security Agency (NSA) and Government Communications Headquarters (GCHQ) surveillance information, preferring to cultivate relationships with trusted journalists who determined what information needed to be released (Greenwald, 2014). Without nuanced understandings of the kind of accountability that visibility is designed to create, calls for transparency can be read as false choices between complete secrecy and total openness.

Transparency can invoke neoliberal models of agency

The ideal of transparency places a tremendous burden on individuals to seek out information about a system, to interpret that information, and determine its significance. Its premise is a marketplace model of enlightenment—a "belief that putting information in the hands of the public will enable people to make informed choices that will lead to improved social outcomes" (Schudson, 2015: 22). It also presumes that information symmetry exists *among* the systems individuals may be considering—that systems someone may want to compare are equally visible and understandable. Especially in

neoliberal states designed to maximize *individual* power and minimize government interference, the devolution of oversight to individuals assumes not only that people watch and understand visible systems, but that people also have ways of discussing and debating the significance of what they are seeing. The imagined marketplaces of total transparency have what economists would call perfect information, rational decision-making capabilities, and fully consenting participants. This is a persistent fiction.

Transparency does not necessarily build trust

Although transparency is often thought to engender trust of organizations and systems, there is little conceptually rich empirical work confirming this (Albu and Flyverbom, 2016). Specifically, different stakeholders trust systems differently, with their confidence depending upon what and when information is disclosed, and how clear, accurate, and relevant they perceive information to be (Schnackenberg and Tomlinson, 2014). Trust can also go the other direction. Some designers may not release detailed information about their systems, not due to trade secrets or competitive advantage, but because they lack trust in the ethics and intentions of those who might see them. Leonardo da Vinci refused to publish the exact details of his early submarine designs: "I do not publish nor divulge these, by reason of the evil nature of men, who would use them for assassinations at the bottom of the sea" (da Vinci, 2015: n.p.).

Transparency entails professional boundary work

Transparency is often limited by professionals protecting the exclusivity of their expertise. Such transparency can be performative—as Hilgartner (2000: 6) describes the "dramatic techniques" of science policy advisors practicing openness to establish authority as experts—or co-opted by special interests using open data (Levy and Johns, 2016) and scientific norms of data sharing to advance their own aims—as drug and tobacco companies do when they "institutionalize uncertainty" (Michaels, 2008: 176) by using public information to continuously manufacture alternate explanations of their product's effects, going so far as to admit that "doubt is our product" (p. x).

Professions have a history of policing their boundaries: controlling who has access to expertise, who is allowed to perform work, who will hold professionals accountable, and how conflicts among professionals will be resolved (Abbott, 1988). Along with these definitions and controls comes secrecy and a reluctance to make all parts of a profession visible. Doing so would reveal how much of professional work actually involves not the formal application of approved knowledge but, rather, an interplay between tacit and explicit knowledge that may raise questions about whether relevant regulations actually achieve the desired oversight. It may be impossible to really *see* professional practices without understanding that they are situated within contexts, contested by people with different kinds of expertise, and inseparable from the developmental histories of practitioners (Goodwin, 1994).

Transparency can privilege seeing *over* understanding

Seeing inside a system does not necessarily mean understanding its behavior or origins. A long-standing concern in educational reform movements is creating materials that help

learners not only see systems in action but also help them experiment with a system's components. In Froebel's "kindergarten method" of learning, he gave children physical objects designed to teach children how complex patterns emerged from their constituents (Stiny, 1980). Children saw not only the *existence* of patterns but also the components and combinations that *made* them. This tradition continued through Piaget's and Vygotsky's child development studies (Crain, 2000), Dewey's (1997) experiential learning, and Papert's (1980) LOGO programming language designed for children to learn about complex systems—and represent their *own* theories of complex systems—by building computational models that simulate and interact within their environments. Resnick et al. (2000) continue to invent computational "construction kits" but they explicitly hide certain processes and mechanisms while making others visible and manipulable (pp. 16–17).

Learning about complex systems means not simply being able to look inside systems or take them apart. Rather, it means dynamically interacting with them in order to understand how they behave in relation to their environments (Resnick et al., 2000). This kind of complex learning intertwines epistemological claim-making with material design, social contexts, and self-reflexivity—making sure that a system is not only *visible* but also debated and changeable by observers who are able to consider how they know what they know about it.

Transparency has technical limitations

Sometimes, the details of a system will be not only protected by corporate secrecy or indecipherable to those without technical skills, but inscrutable even to its creators because of the scale and speed of its design (Burrell, 2016; Crain, 2018). For example, engineers have developed deep learning systems that "work"—in that they can automatically detect the faces of cats or dogs, for example—without necessarily knowing *why* they work or being able to *show* the logic behind a system's decision. This can cause serious difficulties, such as when Google's Photo app unexpectedly tagged Black people "gorillas," (Dougherty, 2015) HP's facial recognition software repeatedly failed to recognize faces with dark skin tones, and Nikon's camera algorithms misperceived that Asian people were blinking (Wade, 2010). While the presumption was that these image recognition systems hadn't been trained on enough Black faces—and some companies recommended solving the problem with brighter foreground lighting—similar problems with Google's system had emerged earlier when people of all races were being misidentified as dogs. Engineers of these systems could not precisely say where in the problems were occurring—even though they had total access to the systems' designs and implementations. Diakopoulos (2016) draws attention to the fact that some aspects of algorithmic systems may never be disclosed because they never take durable, observable forms, for example, an "algorithm could compute a variable in memory that corresponds to some protected class such as race," but if the memory exists only temporarily "FOIA would be unable to compel its disclosure" (p. 59).

It may be necessary to access code to hold a system accountable, but seeing code is insufficient. System builders themselves are often unable to explain how a complex system works, which parts are essential for its operation, or how the ephemeral nature of computational representations are compatible with transparency laws.

Transparency has temporal limitations

Historians of transparency find that decisions about *when* to make systems visible reveals a great deal about what people think "the system" is and what kind of knowledge is required at different moments. For example, Schudson (2015) describes how the National Environmental Policy Act (NEPA) required the public to "enter the process not after the fact but in the very formation of proposed" action that *might* have environmental impacts (p. 226). Transparency can mean future relevance, anticipated revelation, ongoing disclosure, or *post hoc* visibility—different moments in time may require or produce different kinds of system accountability. The temporal dimension of transparency is further complicated by the fact that objects and systems change over time—especially rapidly in the context of networked computational systems. Even if an algorithm's source code, its full training data set, and its testing data were made transparent, it would still only give a particular snapshot of its functionality. This is particularly true for adaptive systems that "learn" as the amount and types of data they draw on increase—and for platforms with shifting interfaces, settings, capabilities, and number of users. There is no "single" system to see inside when the system itself is distributed among and embedded within environments that define its operation. Any notion of transparency or auditing without temporal dimensions misses seeing previous iterations, understanding how they worked, why they changed, and how their interacting components actually constituted different systems.

All of these limitations revolve around a central concern long studied by scholars of science and technology: how to understand seeing and insight as inseparably intertwined aspects of epistemology and knowledge production.[1] What systems *are* or *mean* depend upon the tools and perspectives people employ while looking—and they can be many things at once. To suppose that instruments and representations of knowledge are somehow separate from the practices and cultures of study is to misunderstand watchers as people without stakes in what they see.

To ask to "look inside the black box" is perhaps too limited a demand and ultimately an ill-fitting metaphor for the complexities of contemporary algorithmic systems. It sidesteps the material and ideological complexities and effects of seeing and suggests a kind of easy certainty that *knowing* comes from looking. And if accountability requires seeing a system well enough to understand it—and if Minsky (2007: 6) is right that if "you 'understand' something in only one way, then you scarcely understand it at all"—then using transparency for accountability begs the question of what, exactly, is being held to account. A system needs to be understood to be governed—and in as many different ways as possible.

Why transparency is an inadequate way to govern algorithmic systems

It is difficult to make a complex object of study transparent, especially in ways that create the authoritative knowledge that defensible accountability requires. Tensions and limitations in the ideal of transparency are long-standing and rooted in historical contexts, but they are also changing amid a new set of online cultures, digital infrastructures,

and networked institutions defining the conditions under which systems can be seen and understood—conditions of visibility governed in part through algorithmic systems (Gillespie, 2014). The power of these systems lies in the "careful plaiting of relatively unstable associations of people, things, processes, documents and resources" (Neyland and Möllers, 2016: 1)—but also as relationships without social articulation (Eslami et al., 2015), elite structuring masked as emergent sociality (Mackenzie, 2015; Striphas, 2015), automated media circulation (Hallinan and Striphas, 2014; Napoli, 2014), optimized labor (Kushner, 2013) and financial markets (Arnoldi, 2015), bracketed research methods (Driscoll and Walker, 2014), and rationalized expertise (Anderson, 2013; Levy, 2015).

In digital contexts, transparency is not simply about revealing information or keeping secrets but continually deploying, configuring, and resisting platforms, algorithms, and machine learning protocols that manage visibility. Such "disclosure devices" (Hansen and Flyverbom, 2015: 872) are neither exclusively human nor entirely computational, but networks of human and non-human agents that create "particular visibilities and possibilities for observation." They make "some parts of organizational and social life knowable and governable" and keep others out of sight (Flyverbom, 2016: 112). As Heemsbergen (2016) puts it,

> radical transparency's promise to end secrecy has not materialized. Instead, the social-material relations underpinning digital disclosures suggest they function to reconfigure visibilities of control and recognition rather than reveal extant objects. (p. 138)

An algorithmic system is not just code and data but an *assemblage* of human and non-human actors—of "institutionally situated code, practices, and norms with the power to create, sustain, and signify relationships among people and data through minimally observable, semiautonomous action" (Ananny, 2016: 93). This requires going beyond "algorithms as fetishized objects" to take better account of the human scenes where algorithms, code, and platforms intersect (Crawford, 2016: 14).

Holding an assemblage accountable requires not just seeing inside any one component of an assemblage but understanding how it works as a system. Connecting to our earlier discussion of the pragmatic approach to epistemology, we might reframe the question as: what kind of claims can be made to *understand* an actor-network, and how is this understanding related to but distinct from simply *seeing* an actor-network? As Ward (1996) describes, any understanding of "society" is the "outcome of the settlement of controversies," not a disconnected entity that gives truth to issues (p. 103). Knowledge is produced in domains of conflicting interpretations and complex social interactions and disputes (Crawford, 2016; DiSalvo, 2012). Because an actor-network and its understandings are partial achievements amid ongoing contestations, those who study science and knowledge must always "remain undecided as to what science, knowledge, and technology are actually made of" Ward (1996: 103–104). They are not objects of study that can be understood by peering inside to see what they "really" are.

If the truth is not a positivist discovery but a relational achievement among networked human and non-human agents, then the target of transparency must shift. That is, if a system must be seen to be understood and held accountable, the kind of "seeing" that an

actor-network theory of truth requires does not entail looking *inside* anything—but *across* a system. Not only is transparency a limited way of knowing systems, but it cannot be used to explain—much less govern—a distributed set of human and non-human actors whose significance lies not internally but relationally.

Conclusion

In this article, we have sketched an ideal of transparency through different political theories, regulatory regimes, epistemological models, and material systems, in order to develop a 10-part typology of its limitations. By highlighting these limitations, we can better understand the normative work that the ideal of transparency is expected to perform and how optimism around its potential might be tempered. In the particular context of platforms and data systems, we argue that making one part of an algorithmic system visible—such as the algorithm, or even the underlying data—is not the same as holding the assemblage accountable. If we recognize that transparency alone cannot create accountable systems and engaging with the reasons behind this limitation, we may be able to *use* the limits of transparency as conceptual tools for understanding how algorithmic assemblages might be held accountable.

The limitations identified here represent not just constraints on the ideal of transparency—reasons to distrust the idea that seeing equals knowing—but also openings: ways to make a model of algorithmic accountability from the transparency's limitations. Specifically, if transparency is ineffective because the power to shame is ineffective against those with the power to endure visibility, then the system's model of accountability should focus on the power imbalance transparency's impotence reveals. If transparency has the potential to do harm, then the system's model of accountability should focus on how the risk of suffering or causing harm is distributed among the actors susceptible to transparency. If large-scale, persistent transparency effectively occludes information and distracts observers from significant sources of power, then accountability models should focus on how systems direct attention during information overload. If transparency assumes false binaries between complete openness versus total closure, then the system's accountability should focus on which *types* or *degrees* of visibility lead to the desired normative outcome. If transparency assumes the active participation of individuals interested in and able to provide oversight, then the model of accountability might ask whether they have the motivations, skills, and associations required to achieve the collective oversight transparency promises.

If transparency empowers actors who are not trusted with the power their insights bring, then the accountability model might focus on how confidence and suspicion circulate within the system and which parts of this circulation require transparency. If transparency requires professional expertise to be understood and acted upon, then models of accountability might examine how the system develops and deploys different types of authority and specialized ways of seeing. If transparency equates seeing a system with understanding its dynamism, then accountability models might critique how systems change, who has power to define significant change, and how the power to see a system change differs from the power to experiment with changing a system. If a system is so complex that even those with total views into it are unable to describe its failures and

successes, then accountability models might focus on the whether the system is sufficiently understood—or understandable—to allow its deployment in different environments, whether more development time is needed, or if the system should be built at all. Finally, if a system's future behavior cannot be accurately predicted, if the conditions driving its past behavior cannot be reconstructed, or if it never exists in a stable form that can be held constant, looked into, or archived, then accountability models may need to develop ways of describing system responsibility at different time periods and with different rhythms.

In the spirit of those who argue for transparency as a way to hold media systems accountable, our aim is to *use* its limitations as a path toward accountability: to go beyond explaining why it is sometimes unnecessary and always insufficient to simply look inside structures and show instead how these limitations can be starting points for reconstructing accountability for systems that cannot be seen into, held still, or fully traced. This is an increasingly urgent concern for those aiming to hold accountable algorithmic assemblages.

Implicit in this project of reconstructing accountability through an explication of transparency's limitations is a set of normative ideals: a desired role for automation, an ideal actor-network configuration, and an ethics of algorithmic assemblages (Ananny, 2016). This is to be expected since no model of accountability can avoid the questions "accountable for what?" and "accountable to whom?" (Glasser, 1989: 179). The next step in this project is to deploy this reconstructed accountability and draw on transparency's limitations to ask to what ends, exactly, transparency is in service. For any sociotechnical system, ask, "what is being looked at, what good comes from seeing it, and what are we *not* able to see?"

Funding

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

Note

1. There are many productive lines of scholarship here, from the "visual and graphic documents of laboratory data, textual figures, [and] biographical accounts of scientific 'problems'," (Lynch and Woolgar, 1988: 99) to the intertwined representational materials and strategies of contemporary computer scientists (Coopmans et al., 2014: 1), mathematicians (Barany and MacKenzie, 2014), neuroscientists (De Rijcke and Beaulieu, 2014), computational biologists (Carusi and Hoel, 2014), and planetary scientists (Vertesi, 2014).

References

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

Albu OB and Flyverbom M (2016) Organizational transparency: conceptualizations, conditions, and consequences. *Business & Society*. Epub ahead of print 13 July. DOI: 10.1177/0007650316659851.

Ananny M (2016) Toward an ethics of algorithms: convening, observation, probability, and timeliness. *Science, Technology & Human Values* 41(1): 93–117. DOI: 10.1177/0162243915606523

Anderson CW (2013) Towards a sociology of computational and algorithmic journalism. *New Media & Society* 15(7): 1005–1021. DOI: 10.1177/1461444812465137

Arnoldi J (2015) Computer algorithms, market manipulation and the institutionalization of high frequency trading. *Theory, Culture & Society* 33: 29–52.

Bacon M (2012) *Pragmatism*. Cambridge: Polity Press.

Baecker R (1998) Sorting out sorting: a case study of software visualization for teaching computer science. In: Stasko J (ed.) *Software Visualization Programming as a Multimedia Experience*. Cambridge, MA: The MIT Press, pp. 369–381.

Ball C (2009) What is transparency? *Public Integrity* 11(4): 293–308.

Barany MJ and MacKenzie D (2014) Chalk: materials and concepts in mathematics research. In: Coopmans C, Vertesi J, Lynch M, et al. (eds) *Representation in Scientific Practice Revisited*. Cambridge, MA: The MIT Press, pp. 107–130.

Bateson G (2000) *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. Chicago, IL: University of Chicago Press.

Birchall C (2011) Introduction to "secrecy and transparency": the politics of opacity and openness. *Theory, Culture & Society* 28(7–8): 7–25.

Birchall C (2014) Radical transparency? *Cultural Studies ↔ Critical Methodologies* 14(1): 77–88.

Bratich J (2016) Occult(ing) transparency: an epilogue. *International Journal of Communication Systems* 10: 178–181.

Brill J (2015) Scalable approaches to transparency and accountability in decisionmaking algorithms: remarks at the NYU conference on algorithms and accountability. *Federal Trade Commission*, 28 February. Available at: https://www.ftc.gov/system/files/documents/public_statements/629681/150228nyualgorithms.pdf

Brown MH (1998) A taxonomy of algorithm animation displays. In: Stasko J (ed.) *Software Visualization Programming as a Multimedia Experience*. Cambridge, MA: The MIT Press, pp. 35–42.

Browne S (2015) *Dark Matters: On the Surveillance of Blackness*. Durham, NC: Duke University Press.

Burrell J (2016) How the machine "thinks": understanding opacity in machine learning algorithms. *Big Data & Society* 3(1). DOI: 10.1177/2053951715622512

Carusi A and Hoel AS (2014) Toward a new ontology of scientific vision. In: Coopmans C, Vertesi J, Lynch M, et al. (eds) *Representation in Scientific Practice Revisited*. Cambridge, MA: The MIT Press, pp. 201–221.

Christensen LT and Cheney G (2015) Peering into transparency: challenging ideals, proxies, and organizational practices. *Communication Theory* 25(1): 70–90.

Coddington M (2012) Defending a paradigm by patrolling a boundary: two global newspapers' approach to WikiLeaks. *Journalism & Mass Communication Quarterly* 89(3): 377–396.

Coleman G (2013) *Coding Freedom: The Ethics and Aesthetics of Hacking*. Princeton, NJ: Princeton University Press.

Coopmans C, Vertesi J, Lynch M, et al. (2014) Introduction. In: Coopmans C, Vertesi J, Lynch M, et al. (eds) *Representation in Scientific Practice Revisited*. Cambridge, MA: The MIT Press, pp. 1–12.

Crain M (2018) The limits of transparency: data brokers and commodification. *New Media & Society* 20(1): 88–104.

Crain W (2000) *Theories and Development: Concepts and Applications*. Upper Saddle River, NJ: Prentice Hall.

Crary J (1990) *Techniques of the Observer*. Cambridge, MA: The MIT Press.

Crawford K (2016) Can an algorithm be agonistic? Ten scenes from life in calculated publics. *Science, Technology & Human Values* 41(1): 77–92.

da Vinci L (2015) *Notebooks of Leonardo Da Vinci* (trans. JP Richter). Istanbul: eKitap Projesi.

Daston L (1992) Objectivity and the escape from perspective. *Social Studies of Science* 22(4): 597–618.

Daston L and Galison P (2007) *Objectivity*. Brooklyn, NY: Zone Books.

Datta A, Tschantz MC and Datta A (2015) Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies* 1: 92–112.

David M (2015) The correspondence theory of truth. *Stanford Encyclopedia of Philosophy*, 28 May Available at: http://plato.stanford.edu/entries/truth-correspondence/

De Rijcke S and Beaulieu A (2014) Networked neuroscience: brain scans and visual knowing at the intersection of atlases and databases. In: Coopmans C, Vertesi J, Lynch M, et al. (eds) *Representation in Scientific Practice Revisited*. Cambridge, MA: The MIT Press, pp. 131–152.

Dewey J (1997) *Experience and Education*. New York: Free Press.

Diakopoulos N (2016) Accountability in algorithmic decision making. *Communications of the ACM* 59(2): 56–62.

DiSalvo C (2012) *Adversarial Design*. Cambridge, MA: The MIT Press.

Dougherty C (2015) Google photos mistakenly labels black people "gorillas." *The New York Times*, 1 July. Available at: http://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/

Driscoll K and Walker S (2014) Working within a black box: transparency in the collection and production of big Twitter data. *International Journal of Communication Systems* 8: 1745–1764.

Eslami M, Rickman A, Vaccaro K, et al. (2015) "I always assumed that I wasn't really that close to [her]": reasoning about invisible algorithms in the news feed. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, Seoul, Republic of Korea, 18–23 April.

Flyverbom M (2016) Transparency: mediation and the management of visibilities. *International Journal of Communication Systems* 10: 110–122.

Fox J (2007) The uncertain relationship between transparency and accountability. *Development in Practice* 17(4–5): 663–671.

Fung A, Graham M and Weil D (2008) *Full Disclosure: The Perils and Promise of Transparency*. Cambridge: Cambridge University Press.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski P and Foot KA (eds) Media Technologies: Essays on Communication, Materiality, and Society. Cambridge, MA: The MIT Press, pp. 167–194.

Glasser TL (1989) Three views on accountability. In: Dennis EE, Gillmor DM and Glasser TL (eds) *Media Freedom and Accountability*. New York: Praeger.

Goodwin C (1994) Professional vision. *American Anthropologist* 96(3): 606–633.

Greenwald G (2014) *No Place to Hide*. New York: Metropolitan Books.

Hallinan B and Striphas T (2014) Recommended for you: the Netflix Prize and the production of algorithmic culture. *New Media & Society* 18: 117–137.

Hansen HK and Flyverbom M (2015) The politics of transparency and the calibration of knowledge in the digital age. *Organization* 22(6): 872–889.

Heald D (2006) Varieties of transparency. *Proceedings of the British Academy* 135: 25–43.

Heemsbergen L (2016) From radical transparency to radical disclosure: reconfiguring (in)voluntary transparency through the management of visibilities. *International Journal of Communication Systems* 10: 138–151.

Hilgartner S (2000) *Science on Stage: Expert Advice as Public Drama*. Stanford, CA: Stanford University Press.

Hood C (2006) Transparency in historical perspective. In: Hood C and Heald D (eds) *Transparency: The Key to Better Governance?* Oxford: Oxford University Press, pp. 1–23.

James W (1997) What pragmatism means. In: Menand L (ed.) *Pragmatism: A Reader*. New York: Random House, pp. 93–111.

Kelty C (2008) *Two Bits: The Cultural Significance of Free Software*. Durham, NC: Duke University Press.

Kushner S (2013) The freelance translation machine: algorithmic culture and the invisible industry. *New Media & Society* 15: 1241–1258.

Leigh D and Harding L (2011) *WikiLeaks: Inside Julian Assange's War on Secrecy*. New York: Public Affairs.

Levy KEC (2015) The contexts of control: information, power, and truck-driving work. *Information Society* 31(2): 160–174.

Levy KEC and Johns DM (2016) When open data is a Trojan Horse: the weaponization of transparency in science and governance. *Big Data & Society* 3(1). http://bds.sagepub.com/content/3/1/2053951715621568

Lynch M and Woolgar S (1988) Introduction: sociological orientations to representational practice in science. *Human Studies* 11(2): 99–116.

Mackenzie A (2015) The production of prediction: what does machine learning want? *European Journal of Cultural Studies* 18(4–5): 429–445.

Mackenzie D (2008) *An Engine, Not a Camera: How Financial Models Shape Markets*. Cambridge, MA: The MIT Press.

Michaels D (2008) *Doubt Is Their Product: How Industry's Assault on Science Threatens Your Health*. Oxford: Oxford University Press.

Minsky M (2007) *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster.

Misak C (ed.) (2007) *New Pragmatists*. Oxford: Oxford University Press.

Napoli PM (2014) Automated media: an institutional theory perspective on algorithmic media production and consumption. *Communication Theory* 24(3): 340–360.

Neyland D and Möllers N (2016) Algorithmic IF … THEN rules and the conditions and consequences of power. *Information, Communication & Society*, 20 (1): 45-62.

Papert S (1980) *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books.

Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Phillips JWP (2011) Secrecy and transparency: an interview with Samuel Weber. *Theory, Culture & Society* 28(7–8): 158–172.

Resnick M, Berg R and Eisenberg M (2000) Beyond black boxes: bringing transparency and aesthetics back to scientific investigation. *Journal of the Learning Sciences* 9(1): 7–30.

Schnackenberg AK and Tomlinson EC (2014) Organizational transparency: a new perspective on managing trust in organization-stakeholder relationships. *Journal of Management* 42: 1784–1810.

Schrock AR (2016) Civic hacking as data activism and advocacy: a history from publicity to open government data. *New Media & Society* 18: 581–599.

Schudson M (2015) *The Rise of the Right to Know*. Cambridge, MA: Belknap Publishing.

Squires CR (2002) Rethinking the black public sphere: an alternative vocabulary for multiple public spheres. *Communication Theory* 12(4): 446–468.

Stiny G (1980) Kindergarten grammars: designing with Froebel's building gifts. *Environment and Planning B* 7(4): 409–462.

Stohl C, Stohl M and Leonardi PM (2016) Managing opacity: information visibility and the paradox of transparency in the digital age. *International Journal of Communication Systems* 10: 123–137.

Striphas T (2015) Algorithmic culture. *European Journal of Cultural Studies* 18(4–5): 395–412.

Tkacz N (2015) *Wikipedia and the Politics of Openness*. Chicago, IL: University of Chicago Press.

Turner F (2006) *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago, IL: University of Chicago Press.

Turner F (2013) *The Democratic Surround*. Chicago, IL: University of Chicago Press.

Vertesi J (2014) Drawing as: distinctions and disambiguation in digital images of Mars. In: Coopmans C, Vertesi J, Lynch M, et al. (eds) *Representation in Scientific Practice Revisited*. Cambridge, MA: The MIT Press, pp. 15–36.

Wade L (2010) HP software doesn't see black people. *Sociological Images*, 5 January. Available at: https://thesocietypages.org/socimages/2010/01/05/hp-software-doesnt-see-black-people/

Ward SC (1996) *Reconfiguring Truth*. New York: Rowman & Littlefield.

Zara C (2015) FTC chief technologist Ashkan Soltani on algorithmic transparency and the fight against biased bots. *International Business Times*, 9 April. Available at: http://www.ibtimes.com/ftc-chief-technologist-ashkan-soltani-algorithmic-transparency-fight-against-biased-1876177

Author biographies

Mike Ananny is an assistant professor at the Annenberg School for Communication and Journalism and a Faculty Affiliate of Science, Technology, and Society, at the University of Southern California.

Kate Crawford is a principal researcher at Microsoft Research New York, a senior research fellow at NYU's Information Law Institute, and a visiting professor at MIT's Center for Civic Media.