Ethical Dimensions of Visualization Research

Michael Correll
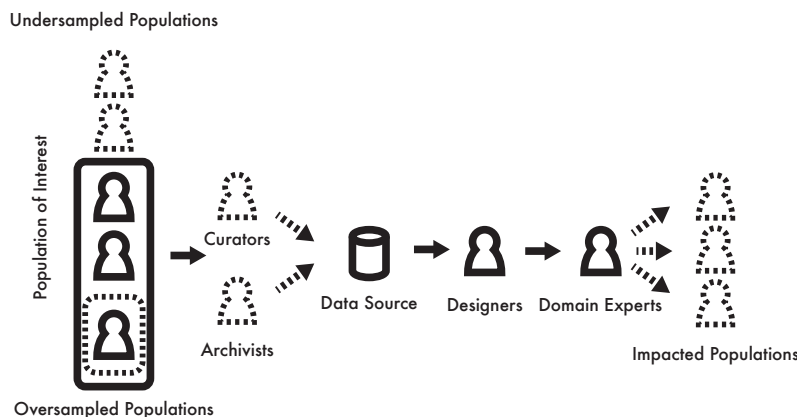Tableau Research
Seattle, WA
mcorrell@tableau.com

Figure 1: Visualizations projects are often described or evaluated as though they are straightforward paths from data collection to design to the intended user (solid outlines). This neglects or makes invisible critical populations, impacts, and labor (dashed outlines) that can contribute to the ethical character of a project. We have an obligation, where possible, to make these invisible facets and contributions visible.

ABSTRACT

Visualizations have a potentially enormous influence on how data are used to make decisions across all areas of human endeavor. However, it is not clear how this power connects to ethical duties: what obligations do we have when it comes to visualizations and visual analytics systems, beyond our duties as scientists and engineers? Drawing on historical and contemporary examples, I address the moral components of the design and use of visualizations, identify some ongoing areas of visualization research with ethical dilemmas, and propose a set of additional moral obligations that we have as designers, builders, and researchers of visualizations.

CCS CONCEPTS

• **Human-centered computing** → **Visualization theory, concepts and paradigms**; • **Security and privacy** → *Social aspects of security and privacy.*

KEYWORDS

Information Visualization; Visual Analytics; Ethics

1 INTRODUCTION

In the wake of leaked information about the NSA's spying program, Rogaway wrote "The Moral Character of Cryptographic Work" [86]. In that paper, he argues that the work of academics and engineers in cryptography has an inescapable moral character: it shifts power amongst social groups, and so has an inherent political impact on society, for good or ill. Critical movements in cartography [27, 28, 103] and data science [18, 29] have begun to analyze how the the use (and abuse) of data shifts structures of power. Visualization work has the same capacity, and so must also be analyzed with respect to its moral character. As per Rogaway:

I suspect that many of you see no real connection between social, political, and ethical values and what you work on. You don't build bombs, experiment on people, or destroy the environment. You don't spy on populations. You hack math and write papers. This doesn't sound ethically laden. I want to show you that it is.

In this paper I will draw on the history of visualization and analytics to illustrate that *all* visualization research, no matter how superficially apolitical or trivial, has a moral character. I will then illustrate how this moral character is reflected in conflicts between virtues that arise in current emerging areas of visualization research, and how they might be balanced. My goal with this work is both to promote caution and contemplation in visualization research (in that we should stop doing unethical work) but also to present new opportunities for research and growth (in that we should study the broader impact of our work and look for new areas to explore and problems to solve).

In the first two sections of this paper I will address the common feeling that data and data visualization, respectively, are apolitical or somehow ethically neutral, and that therefore we lack moral obligations with regards to how data are collected and visualized. It is this tendency to view our work as the mere reporting or structuring of objective fact that is most dangerous to me. Heidegger [51] specifically calls out the danger of this perspective:

> Everywhere we remain unfree and chained to technology, whether we passionately affirm or deny it. But we are delivered over to it in the worst possible way when we regard it as something neutral.

In the final two sections of the paper, I will address current trends in visualization research that appear to have ethical implications. I will then use these case studies as a basis to propose additional obligations that visualization researchers have in addition to their existing moral obligations as scientists, teachers, and citizens.

2 AGAINST THE NEUTRALITY OF DATA

It is tempting to claim that visualization is an ethically neutral activity because we are merely reporting the data, and data are just facts about the world. It's not our problem how these facts are collected, or who uses them. We're just the middle-man (or, more nefariously, the man in the middle [26]) between a stakeholder and their data. Provided that we did not introduce bias or intentionally deceive when presenting our data, we completed our duties. However, data are not naturally occurring phenomenon. The world does not spontaneously quantify, curate, or data-mine itself. Rather,

Figure 2: A table from Thomas Paine's 1775 pamphlet *Common Sense*[79]. These data were collected for an initial accounting purpose, but are used by Paine to argue for the relative weakness and fragility of the English Navy, and the potential strength of the American Navy, as part of an argument for independence and revolution.

the process of observing the world and quantifying it is a political act, and deserves ethical consideration [6].

Heidegger identifies quantification as the hallmark of modern technology: the turning of things (and people) into "standing reserves" of resources [51]. A river is not just a flowing thing to be admired, it's a certain amount of megawatts of power if connected to a hydroelectric dam. An acre of forest is not just a scenic location, but a reserve of charcoal and lumber and so on. Modern technological systems are not (just) alienating, but an entire reframing of how we relate to the world around us in terms of exploiting and utilizing resources. And people are, of course, no exception. The collection of mass data about people is a way of turning them into a standing reserve (of ad revenue, of content creators, of soldiers, of bodies).

This collection of data, and the distillation of people into data, has tremendous political power. Gottfried Achenwall coined the term "*Statistik*" to be the "science of the state" in his 1752 work *Constitution of the Present Leading European States*. The collection of vital statistics was initially intended to be undertaken by the state for such organizational purposes as determining the size of a tax base, or the amount of trees available for naval vessels. This initial data collection was by no means apolitical: the proper data set can help start wars (Fig. 2). Nor has this centralized and political use and meaning of statistics disappeared in the digital age: one of the first uses of computing machines to process statistical population data were the machines that IBM's

subsidiary Dehomag developed for the Nazi regime, which were used to expedite and support the Final Solution [11, 34]. The relative emotional distance of collecting and reporting on *data* (as opposed to managing and reporting on *people*) arguably contributed to the uniquely *bureaucratic* horrors of the Holocaust [99], and to what Hannah Arendt refers to as the "banality of evil" [3].

Conversely, *refraining* from collecting data likewise has political and ethical consequences. Within academia, the convenience sampling of so-called WEIRD populations (Western, Educated, Industrialized, Rich, and Democratic) constrains the broader applicability of findings [52], and excludes populations from consideration in later designs. This imbalance in data collection can result in unequal outcomes, as with the example of the over-representation of white faces in computer vision benchmarks resulting in commercial products that fail to accurately detect or model the faces of people with darker skin [20]. Absence of data can be engineered for political ends: the Trump administration's attempt to add a question about citizenship to the U.S. census is likely an attempt to dissuade non-citizens from answering the census in fear of retaliation [101], and so therefore to guide the distribution of state resources away from areas with larger immigrant populations.

There is no such thing as an objective view from nowhere: rather, knowledge is *situated* [49] within our perspectives and circumscribed by the limits of our experience. Therefore, data are not neutral and objective facts about the world— there is no such thing as "raw" data [46]. Data are always collected or processed by *someone*, for some aim. Often the work that goes into collecting and structuring data is made invisible [31, 33]. Often, too, are the purposes for which these data are collected and used given less importance than data as an abstract puzzle to be solved or a collection of insights to be gathered. The emerging field of "critical data science" [29] seeks to examine how data reinforce or challenge systems of power, and to "undo" [17] assumptions that collecting more data inevitably results in an increase in efficiency or a decrease in bias.

3   AGAINST THE NEUTRALITY OF VISUALIZATION

Well-designed visualizations are often conceived of as clear depictions of objective data. Drucker [38] views this framing as particularly dangerous:

> While it may seem like an extreme statement, I think the ideology of almost all current information visualization is anathema to humanistic thought, antipathetic to its aims and values. The persuasive and seductive rhetorical force of visualization performs such a powerful reification of information that graphics such as Google Maps
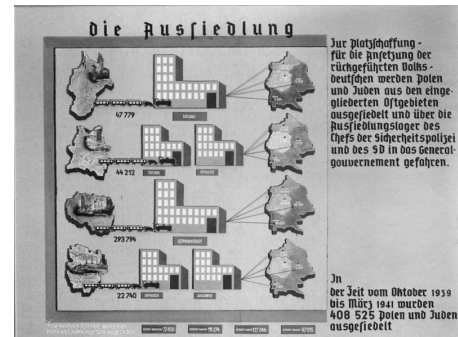
(a) Source-destination map [19]
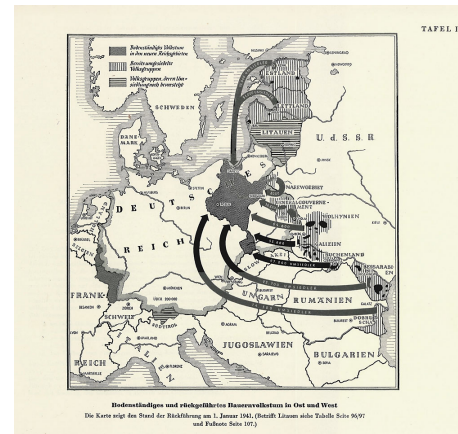
(b) Flow map [73]

Figure 3: Two visualizations from the Nazi regime's "Heim ins Reich" (Home to the Reich) campaign. This campaign was meant to promote the resettlement of ethnic Germans from other parts of Europe to newly conquered territories in Poland. The first map merely mentions that the existing Polish and Jewish population will be resettled; the second map does not mention them at all. Also invisible are the original borders of the annexed Polish state.

> are taken to be simply a presentation of "what is," as if all critical thought had been precipitously and completely jettisoned.

In other words, visualizations often depict data as a *given*, a collection of facts about the world that brook no argument or disagreement. Visualizations are often used as part of a rhetorical appeal to the authority and expertise of the people communicating the data [85], and can stifle critical or contradictory voices who do not have their own data sets to point to. Even the language we use to discuss and critique visualizations can echo implicit biases and inequalities present in society at large [53]. Designers often exclude representation of factors like the uncertainty of the data or the variability

of forecasts for reasons of complexity, scope, or anticipated innumeracy in the audience [15, 48], which can contribute to the perception that the data are immutable truths about the world, rather than designed artifacts representing one flawed, incomplete, and potentially idiosyncratic set of structured observations. The clean lines and structured layouts of traditional visualizations communicate authority and certainty in implicit but measurable ways [61]. While visualizations can be used to promote exploration and further questioning (as in the "martini glass" [88] structured narrative visualization), often designers must use unconventional designs to promote self-critique or skepticism [104].

Another concern is that data visualization, by presenting the data (rather than the people behind the data), can result in "cruel" and "inhuman" [37] charts. That is, by treating a chart of casualty figures as no different qualitatively than a chart of employment statistics, visualizations can hide the ethical and human suffering underlying the data. The infographics of the Nazi regime are a particularly heinous (but by no means unique) example of this erasure. For example Fig. 3 shows the colonization of conquered lands in full detail while relegating information about the forced resettlement and likely death of the original occupants to a caption.

Visualization creates an inherent separation between the people impacted by the data and the people consuming the data. The abstraction, quantification, and digital presentation creates what Baudrillard calls "virtualization" [8]: an air of unreality about the needs and suffering of people of concern. Likewise, Cairo [106] mentions that "I am just very skeptical to the idea that data visualization is a medium that can convey (or even care about conveying) or increase 'empathy'," and recent experiments by Boy et al. [16] suggest that even designs where the human component of data are made more prominent can fail to significantly impact our empathy with human suffering.

All visualizations are rhetorical, and have the potential to persuade [80]. Minor choices in how these charts are designed and presented can alter the message that people take away [56], occasionally without conscious knowledge: e.g., the biasing title of a visualization may not be recalled, but can still impact the remembered contents of a chart [64].

Visualization researchers may attempt to sidestep the rhetorical power of charts by separating visualizations into genres of infographics (that are meant for general audiences and can be used for persuasion) and statistical graphics (that are meant for experts and are actively discouraged from having adornments or embellishments [7]). However, relatively unadorned visualizations in the style of statistical graphics have a long history of use by politicians to bolster their arguments (as in Fig 5).

Likewise, visualizations do not have to be explicitly placed in a political or argumentative context in order to be intended
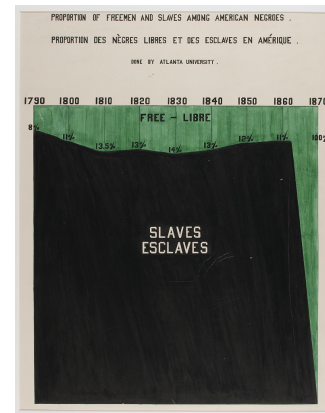
Figure 4: A visualization by W.E.B. Du Bois for the 1900 Paris Exhibition[39]. Despite being relatively straightforward charts without much commentary, Du Bois intended that these visualizations depict the progress and dangers of the African-American population[69] for a moral and political purpose.

as persuasive. For instance, while the chart in Fig. 4 may appear to be a statement of demographic fact, its author, civil rights activist W.E.B. Du Bois, intended it to implicitly function as part of an argument about the status and trajectory of African-Americans [40]:

> Thus all art is propaganda and ever must be, despite the wailing of the purists. I stand in utter shamelessness and say that whatever art I have for writing has been used always for propaganda for gaining the right of black folk to love and enjoy. I do not care a damn for any art that is not used for propaganda. But I do care when propaganda is confined to one side while the other is stripped and silent...

In our own work, the assumption that attempting to persuade with visualizations is only the goal of the propagandist, and that scientific visualization and statistical graphics are therefore above such considerations, cedes rhetorical ground to the groups that do not have such scruples.

In response to these issues, and drawing on similar concerns and methods from the "critical cartography" [28, 75] movement in GIS, Dörk et al. have called for a "critical infovis" [36] movement with the goal of making explicit the values and politics of visualizations. Likewise, D'Ignazio and Klein articulate the notion of "data feminism" [32] where the power imbalances in the process of designing and deploying data visualizations are centered.
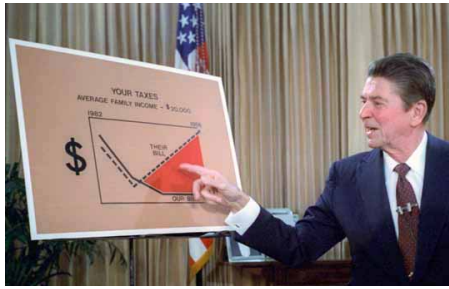
(a) Ronald Reagan uses a chart as part of a public address from the Oval Office in support of the Economic Recovery Tax Act of 1981, showing the difference between the Republican and Democratic tax cut plans.
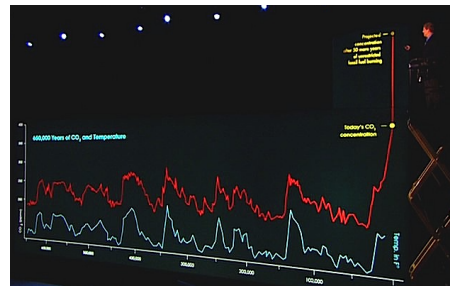
(b) Al Gore uses a chart (and a scissors lift) as part of his movie *An Inconvenient Truth* to show the connection between temperature and carbon dioxide, and the unprecedented scale of recent increases.

Figure 5: Charts lend authority and the perception of objectivity to arguments. Designers of all charts, not just infographics or educational graphs, must be mindful of how visualizations can be used to persuade.

4 CONCERNING TRENDS IN VISUALIZATION RESEARCH

There are several emerging areas of interest in the visualization community where work (or the lack of work) is a cause for concern. These areas are places where *power* and *responsibility* are being allocated in ways that could lead to unethical or irresponsible practice and outcomes. While a full review of all topics of visualization research, and their associated ethical considerations, is out of the scope of this paper, I selected these areas as representing ongoing areas of research where there are values and virtues in *conflict*. That is, emerging topics where there may not be a single clear path forward (as in rule-based deontological ethics), but where researchers will have to balance and cultivate opposing ethical principles (as in virtue ethics [58]). Virtue ethics does not generate prescriptive rules to follow or objective measures of success [71]. Rather, this framing suggests mutual (occasionally conflicting) values to cultivate.

I conclude each topic with a list of *design dilemmas*: open-ended expressions of ethical implications that might arise from visualization research in these areas.

Automated Analysis

One primary goal of visualization is the affordance of "insights": complex, deep, qualitative, unexpected, and relevant [77] revelations. In order to support insights, systems are beginning to explore the concept of automatic recommendations and analyses [93, 108]. The promise of these methods is that analysts can instantly discover important relationships in data, without having to spend many hours exploring trivial or uninteresting patterns.
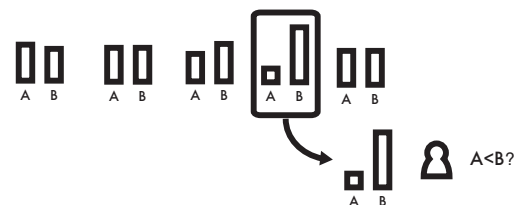
Figure 6: Systems that seek to automatically locate "insights" in datasets can save time for users, and assist users without strong backgrounds in statistics. However, they can promote noise over signal, and lead to unjustified conclusions. How do we empower users without supporting potentially dangerous decision-making?

However, analytics systems (or their consumers) may lack the statistical tools to *validate* insights. Therefore, analysts may come away with conclusions from visualizations that are empirically false or statistically unsupported [10, 107]. Automatic methods can exacerbate this problem [10], and create what Pu and Kay call "p-hacking machines" [83].

Unfortunately, "p-hacking machines" are alluring from an end-user perspective. Finding *something* is a better user experience than finding *nothing*. People may lack the statistical expertise to properly make use of factors that might contextualize the importance of patterns in visualizations, like confidence intervals [9] or probability information [74]. Very few visual analytics systems guide users not just to interesting data, but also through the process of analyzing such findings statistically [95]. Fewer still take into account decision-making biases and attempt to correct for analytical paths not taken [97].
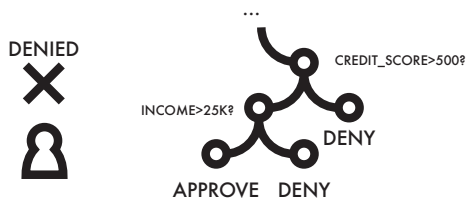
Figure 7: Visualizing Machine Learning models creates a conflict between *transparency* in decision-making, and managing the *complexity*. This trade-off also appears to come at the expense of *accuracy*. Is it more important to have a more understandable model, or a more accurate one?

An ethical concern with this research is therefore that we are enabling bad behavior (unjustified, incorrect, and potentially damaging conclusions from data) without adequate care for the people that can be harmed by decisions based on these conclusions, or adequate understanding about the literacy and capabilities of the people who we are empowering with these automated tools.

A further concern is that the our interactive systems only exacerbate the "garden of forking paths"[45] problem that has contributed to the replication crisis in the sciences. Visualization research itself has many of the same issues as problematic work in other fields [66]. By not creating robust ways of visualizing findings we therefore risk our own credibility as well.

However, automatic insights, by allowing people to quickly discover important facets of their data, can empower people without the time or expertise to discover these findings alone. Systems with excessive guidance or constraints also reduce the agency of the user. There is therefore a potential conflict between democratizing data analytics and promoting statistically sound decision-making (Fig. 6).

Design Dilemmas: How much *guidance* should analytics systems provide to users? How *prescriptive* should such systems be in forbidding or advising against actions that are likely to lead to statistically spurious conclusions?

Machine Learning

Machine learning methods are powerful tools for structuring and making predictions with data, and are present in many critical areas of our society, from finance to college admissions. The resulting models are often opaque, and fail to gracefully allow appeals from people who have be wrongfully or prejudicially categorized [78]. We have a moral duty (and, in some cases, a legal duty [96]) to communicate decision-making based on ML to the populations that are impacted by it. Communication in this way provides much needed context for decisions that seem misguided or callous, and allow

those impacted by the decisions to appeal their decisions or seek better outcomes [68]. Legal scholars such as Citron have argued for the right to "Technological Due Process" [24] in the face of opaque algorithmic decision-making.

Despite this obligation, and the historical positioning of visualization as a way of presenting statistical information to wider audiences, much of the prominent work on visualizing ML focuses on expert users [91, 102]. An ethical concern is that we are therefore empowering the creators of ML models, but are not empowering the people affected by these models. That is, we are not focused on transparently communicating why a particular model made a choice (about one's eligibility for a loan, or eligibility for parole) to audiences without deep statistical expertise.

Venues such as the 2018 Workshop on Visualization for AI Explainability (http://visxai.io/) are beginning to collect scholarship in this area, and online platforms such as Distill (https://distill.pub/) are beginning to collect public-facing "explainers" of ML concepts, but currently there are no standard methods and few success stories of visually communicating algorithmic decisions to the general audience. Even the very definition of what it means for an ML model to be "interpretable" is ill-defined and sometimes contradictory [70]. On the ML side, work on explainability is often considered in terms of numerically representing the contribution of particular features [72, 84], despite the fact that long lists of feature contributions may be difficult to interpret, or fail to speak to the domain expertise of the analyst. Methods in common use in the field, such as saliency maps [63] or model prototypes [62], are often poor conceptual models of ML behavior. Optimizing for human understanding of models often requires empirical testing and optimization independent of the modeling itself [23].

Simple models may be more *explainable* [47], but they are often less *accurate*. The values of transparency and utility may therefore be in conflict (Fig. 7). We want to give the people impacted by our models the opportunity to correct errors, identify points of unfairness, and in general have agency in the decisions that affect them. On the other hand, reducing complexity to afford explainability could result in performance losses that result in worse outcomes. Likewise, there are costs even for successful explanations of ML models and decision-making. Bad actors can game the system at the expense of those who are participating fairly, as with the "cabal" [59] of romance writers who engaged in a number of questionable activities (such as self-plagiarizing and misleading advertising) in order to consistently appear at the top of Amazon's ranking algorithms. Full transparency in models, especially in models built from demographic data, can also compromise the privacy of those who have had their data collected (as in social network data [98]).
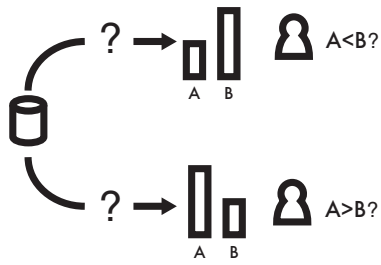
Figure 8: When we visualize the *end result* of a visualization design, but not the *process* by which it was created, we risk propagating false, misleading, or unreproducible findings. On the other hand, showing too many extraneous details may weaken the rhetorical impact, and increase the complexity, of visualizations. How do we use data to convince people, but without taking away agency?

Design Dilemmas: How much *abstraction* or *approximation* should we use when communicating complex ML models? What *standards* or *expectations* should we cultivate when choosing to visualize algorithmic decision-making?

Provenance

Visual analytics systems are increasing in both complexity and importance. Combined with the "garden of forking paths" problem mentioned above, this large number of potential actions means that it is becoming increasingly difficult to articulate exactly what steps an analyst took in order to produce a particular chart or to arrive at a particular conclusion. This need becomes even more important as analytical systems become more tightly integrated with machine learning, which can be non-deterministic in its output, or highly dependent on hyperparameters in its input.

An unmet ethical challenge in visualization is therefore to visualize the provenance of data and decision-making. Communicating the decisions that an analyst took, and affording different decisions, is a key component of both affording criticism and supporting transparency in data-driven decision-making. Much of visualization work is instead focused on affording exploration and analysis, rather than communication of how this exploration and analysis was performed [67]. There is initial work in increasing the transparency of visualizations: systems like Vistrails [21] and Hindsight [42] represent initial steps at visualizing scientific workflows and user viewing histories, respectively. Similarly, "literate visualization" [105] has the goal of making the design decisions that lead to a final visualization documented and transparent. However, very few visual analytics systems are built with the goal of analytical transparency in mind.

Notebook-style interfaces such as Jupyter and Observable and other literate programming environments such as R markdown represent important ecosystems for the transparent communication of analyses, but require coding or scripting expertise to construct. In contrast, popular visual analytics systems (such as Tableau, PowerBI, and Spotfire) heavily rely on GUIs and do not require coding expertise to use. There is therefore a gap between ease of analysis and ease of documentation, further muddying the waters between exploratory and confirmatory analytics.

A connected challenge is the rhetorical one of how to convince viewers not to be taken in by unreliable data or information (such as "fake news" [1]) or how to mitigate cognitive biases in decision-making [35]. Here there is an ethical balance between supporting the agency and desires of the viewer (who may not appreciate being intentionally guided away from the information they want to see) and the desire to communicate information that is both correct and useful. Visualizing provenance information and making explicit the analytical choices made by a system is one way of navigating between these two competing values (Fig 8).

Design Dilemmas: How, and how many, *alternate design or analytical decisions* should we surface to the user? Should we *audit* or *structure* the provenance of a visualization in order to surface irregularities?

5 WHAT ARE OUR OBLIGATIONS AS VISUALIZATION RESEARCHERS?

Visualization, especially visualization research, operates at the intersection of science, communication, and engineering. We have certain ethical obligations as scientists (for instance, to avoid breaches of consent and excesses of harm laid out in codes of research conduct such as the Declaration of Helsinki [5]). Likewise, we have ethical obligations as engineers (for instance, to avoid doing shoddy work or work dangerous to the public good) as laid out in professional codes of conduct such as the ACM's code of ethics [2]. Lastly, insofar as we are presenting data to the public, we have ethical obligations as journalists (for instance, to issue corrections and disclose conflicts of interest), as laid out in codes of conduct such as the SPJ's code of ethics [41].

Beyond the obligation of the component parts of visualization praxis, we also have obligations in that we have a great deal of power over how people ultimately make use of data, both in the patterns they see and the conclusions they draw [26]. We are often the first and only contact a person might have with an underlying store of data. This gives us special access to impacted populations, and special responsibilities as we control the curation, presentation, rhetorical content of the visualizations we create. Visualizations systems we create also embody design principles concerning democratization, transparency, clarity, and automation that lend their use a unique moral signature.

In the following subsections I will present three ethical challenges of visualization work, related to visibility, privacy, and power. I will briefly describe how visualization work impacts these realms, and suggest some virtues or related principles that can ameliorate the negative impact of visualization work in these spheres. In the spirit of virtue ethics, I do not view these principles as unimpeachable or absolute. Therefore, I end each section with a list of potential caveats, where adherence to these principles can have unwanted ethical impacts, or where there exist virtues whose cultivation may directly conflict with the principles I propose.

Make the Invisible Visible

There are many potentially invisible aspects of a visualization (Fig. 1). These non-visualized components such as the choices (and labor) that went into collecting and curating the data, or the populations that could be impacted by the decisions made by viewers of these visualizations, have a non-trivial impact on the good (or harm) that a visualization can do. I echo the view of Dörk et al. [36] that we have a responsibility to make the invisible visible:

We ought to **visualize hidden labor**. Properly acknowledging and rewarding people for their labor is a key component of fairness. Certain kinds of labor (especially those performed by marginalized groups) are under-represented or under-valued in our current schemes of commodification or valuation. For instance, the "emotional labor" [54] of people in service and nursing-related professions is often overlooked. Echoing D'Ignazio and Klein [33], I believe that the labor that goes into collecting, curating, and archiving data, and the further work of analyzing the data, is often invisible in visualizations. Visualizations are often presented as finished products, with the steps in their construction (and alternative steps not taken) hidden from view. Beyond fairness in attribution, making this labor visible is also of benefit to the progress of the field as a whole. Making the labor of analysis and data prep visible will contribute to reproducibility and openness, and will also facilitate in the creation of established standards. Making design and research visible (including surfacing intermediate or failed prototypes) will serve as points of inspiration for other designers, or warnings about potential unfruitful avenues of effort.

We ought to **visualize hidden uncertainty**. Uncertainty is an inescapable component of data collection and analysis, yet it is often hidden from the end user for reasons of complexity or anticipated literacy. Recent work in novel visualizations of uncertainty has specifically targeted general audiences [48, 60] and shown that the ability of the general public to make use of uncertainty information can be quite high given the right kind of visual presentation. Weather data, personal informatics, and electoral polling are all examples where the general public is presented with data with an inescapable component of uncertainty. The way that this uncertainty is presented can have measurable impacts on decision-making. For instance the perceived risk of hurricanes is impacted by how their paths are depicted [87], and voter turnout in elections can be impacted by the race's perceived closeness [4], which in turn can be impacted by how the polling data are presented [25]. To encourage better decision-making (or more uncertainty-aware decision-making), we must investigate the design space of uncertainty visualization and how to measure its impact on our audiences, just as we would also wish to measure task speed or accuracy [57].

We ought to **visualize hidden impacts**. The ACM Future of Computing Academy has called for writers of academic papers to make the potential impacts (especially the potential negative impacts) of academic work to be made explicit [50]. Visualization work often focuses on the positive aspects of a system (for instance, its ease of use, or the speed or accuracy with which analysts conduct their tasks), but rarely on the potential of these systems for harm or misuse. For instance, what harm could be done using the new classes of insights afforded by the visualization system? What groups or individuals may be impacted by the work but do not fit the data model in use? Are there immoral or predatory domains that could make use of the proposed techniques? Beyond these potentially extreme examples of negative impacts, there are potential negative impacts even for visualizations that have been judged to have succeeded in their goals. For instance, will people be harmed if and when the system ceases to be supported (a common occurrence for systems developed in academia)? What are the opportunity costs (in terms of training, collaboration time, and student or developer labor) associated with the system? Do all of the components of the ultimate value of the visualization [92] have a positive sign, from a consequentialist perspective?

Caveats: Visualizations are already complex and multifaceted artifacts, and designers frequently struggle with issues of comprehensibility and the literacy of their audience. Visualizing additional facets of the data, even for laudatory reasons of transparency and accountability, only exacerbate these issues. In addition to making visualizations more difficult to interpret (and so limiting their audience), depicting uncertainty and counter-narratives without proper care can weaken the rhetorical impact of visualizations. **Managing complexity** is therefore a virtue in design that can be in direct opposition with the desire to visualize the invisible.

Collect Data With Empathy

Ceglowski places many of the ills of the current internet at the hands of "investor storytime"[22]: the fable that existing data models are not quite up to the task of performing feats like microtargeting ads or predicting user behavior, but that

they will be *if only we collect more, and more personal* data. This goal results in an increasing pressure to collect as much data as possible to improve models or build bigger pictures or discover more context. The resulting ecosystem of omnipresent data collection means that it is becoming easier and easier to breach the privacy of users with relatively little effort [94]. Visualization is well-positioned to explicitly push against the pressure to collect more and more data by better communicating data's value and impact to analysts. We also have the option of contextualizing and curating data in a way that is respectful of our larger sets of shared values:

We ought to **encourage "small data."** boyd and Crawford claim that "bigger data are not always better data" [18]. The collection of additional data is not just an expenditure of time and resources, it can also intrusively erode the privacy and agency of the people who are subject to this data collection. Even if a particular dataset is used for a population's benefit, merely setting or eroding expectations of how much data one "needs" to perform analysis can have negative repercussions for vulnerable populations in the future. Designers of visualizations and analytics systems should be able to communicate how much data is "enough," and condition analysts to accept tradeoffs of accuracy or certainty in exchange for concision and protection.

We ought to **anthropomorphize data**. Much of visualization's power comes from the power of abstraction, but this creates a gap between populations and how they are represented in visualizations: quantization and virtualization of human beings stymies empathy. Designers must attempt to cross this gap between map and territory, especially for visualizations with high moral stakes such as those concerning human suffering. Proposed solutions to recenter human beings in data visualizations using person-shaped glyphs has not been shown to produce any additional empathic response [16], but including actual human beings in visualizations can help communicate complex phenomena [90] and contribute to human interest in, and memorability of, visualizations [13, 14]. Visualizations designers may have to borrow techniques from journalism and rhetoric, and propose novel designs or interventions, in order to foster empathy and spur action using visualizations.

We ought to **obfuscate data to protect privacy**. Visualization designers often have privileged access to sensitive datasets, and are then charged with communicating these datasets to wider public spheres. The privacy and consent of the people whose data we collect or measure is therefore of paramount ethical importance. People are often unfamiliar with how much data is really collected or collectible through public APIs or other services. These inadvertent exposures of data, combined with the ability of visualizations to highlight previously unseen patterns and trends, can result in severe breaches of confidentiality and privacy, as with the recent

use of fitness tracking company Strava's user heatmap and public API to identify the locations and internal layouts of U.S. military bases [55]. Preserving the privacy of the people in our dataset may involve novel designs [30] or interactive workflows [98]. Both classes of techniques involve aggregating, fuzzing, or otherwise restructuring data to preserve privacy. A related component of this obfuscation is then communicating the upper limit of accuracy or detail to analysts.

Caveats: Restricting the type and amount of data that we collect has a direct impact on the quality and scope of our analyses. It may be laudatory to avoid collecting unnecessary information, but seemingly irrelevant fields can result in serendipitous discoveries that might not have otherwise been possible. Limiting the scope of data collection also entails a form of selection bias that could result in biased or unjust conclusions arising from a lack of context. Aggregation to preserve privacy also can result in seemingly contradictory conclusions such as Simpson's Paradox. Our obligation to **provide context and analytical power** can therefore stand in direct opposition to the empathic collection of data.

Likewise, our empathic judgments can be biased or otherwise fraught. For instance, Nagel's concept of *moral luck* [100] notes the role that blind chance plays on our moral judgments. For instance, attempted murderers receive lighter sentences than murderers who succeed in killing their victims, even if their intent and actions were identical. Similarly, Bloom [12] argues that our visceral empathic reactions can be misused in service of violent or discriminatory ends. Cultivating our *sympathy* may therefore cause us to come in conflict with **institutional fairness**. Patil et al [81] argue for a checklist-based system of ethical reflection to circumvent these complexities.

Challenge Structures of Power

Satirist Peter Finley Dunne suggested that one of the jobs of newspapers is to "comfort the afflicted" and "afflict the comfortable." Data visualizations do not often achieve these goals. The process of collecting data still requires money, time, access, and storage, which inherently gives the advantage and priority to governments and corporations with access to those resources. Many of the resources for performing academic research likewise originate from powerful governments or corporations. The academic visualization focus on esoteric or otherwise complex datasets means that the intended audience of a visualization is often those with high scientific, mathematic, and visual literacies rather than wider and more general audiences. Visualization work should be concerned with imbalances in power, and focus on distributing power in more equitable ways, and to more ethical ends:

We ought to **support data "due process"**. Citron's [24] notion of a technological due process highlights the danger that automated and data-driven decision-making has

on our norms of decision-making and procedure. Datasets only imperfectly capture important information relevant to decsion-making, but may then become reified by algorithms or visualizations into unappealable assertions about the state of the world. People impacted by these systems have the same rights as they do with other forms of decision-making, and deserve some say in how they are (or are not) represented. Legal frameworks such as the EU's GDPR are beginning to codify rights such as "the right to explanation," "the right to privacy," and the "right to be forgotten." However, these existing laws capture our intuitive notions about these rights only imperfectly [96]. Our ethical obligations therefore may or may not reflect the letter of the law. Compounding this issue is that many of these algorithmic systems are of sufficient size or complexity that there are no clear procedures for visualizing them, especially for audiences that lack statistical expertise or extensive context about the domain. We therefore also have design research and pedagogical responsibilities to ensure that we are giving people agency and representation in ways that are useful and understandable.

We ought to **act as data advocates**. Visualizations have rhetorical strength and political power. Government agencies and corporations have explicit resources dedicated to the design and publication of annual reports and data reporting. Marginalized groups do not often have access to the same set of resources, and so are under-represented in data-based conversations. Conversely, groups may have financial or political interest in muddying the waters around debates of fact such as climate change or humanitarian crises. Just as it is considered laudatory to donate time or money to charitable causes, we should also donate a portion of our expertise in the presentation of information to advocate or amplify causes we believe in. This could be a relatively low-cost endeavor. For instance, many visualization papers use a similar set of standard datasets to illustrate or evaluate their designs. These datasets often have limited relevance or importance [65]. Alternative datasets about issues of current concern would also suffice to show that a system operates correctly, but could increase the visibility of ongoing injustices.

We ought to **pressure or slow unethical analytical behavior**. In response to abuses by U.S. Immigration and Customs Enforcement (ICE) agency, Amazon employees circulated an internal memo asking CEO Jeff Bezos to cut ties with the agency [89]. Google researcher Jack Poulson, *inter alia*, resigned from Google over ethical concerns about the design of a search engine that censors internet content in mainland China [44, 82]. Public resignations and dissent can surface perceived ethical lapses in companies, but many may lack the financial or political security to engage in such tactics. Likewise, organizational power may seek to circumvent or exclude those with ethical concerns from decision-making: Google's senior management reportedly kept its censorship

project secret from the internal teams that typically engage with the ethical implications of Google's work [43]. Even for those outside of such organizations, the very companies or governments engaged in ethical lapses are also frequently the source of funds and exposure for visualization research both in academia and industrial research. In such cases where one is unwilling or unable to risk retaliation to stop or speak out against unethical work, other options are the intentional sabotage or slowdown of labor. Within the context of visualization work, this could be overestimation of budgets of money and time, underestimation of result quality, or unnecessary delays (say, for additional user testing or ablation studies). Such sabotage may involve a conflict between *ethical*, *professional* and perhaps even *legal* duties, and should not be undertaken lightly.

Caveats: Conspiracy theorists, political extremists, and corporate interests (such as the tobacco and oil industries) make use of the margins of discourse, counter-narratives, and doubt to advance agendas that rely on the general public discarding the opinion of experts. These bad-faith epistemologies combined with the increasingly fractal nature of academic study has resulted in what Nichols calls "the death of expertise" [76]: a ongoing and increasing hostility to scientific and technocratic sources of knowledge. While people and organizations collecting data may be in positions of power compared to the general public, they might be at a further power disadvantage compared to these organizations that do not want the public to have access to or comprehension of particular information. The goal of **promoting truth and suppressing falsehood** may require amplifying existing structures of expertise and power, and suppressing conflicts for the sake of rhetorical impact.

6 CONCLUSION

This work presents some of the pressing ethical considerations of visualization work, but it functions as neither a complete survey of this space nor an exhaustive and prescriptive decision criteria to guarantee that a visualization was designed or deployed ethically (given the disagreements that rational people can have over what constitutes a moral course of action, no such criterion is likely to exist). Future work requires both developing a new pedagogy for instilling the right values in visualization designers and researchers as well as post-hoc studies of the ethical impact of visualization work in the existing moral landscapes.

It is my intention that this work functions as a both a synthesis of existing critical and ethical views of data and visualization as well as a call to action to be mindful of the ethical implications of our work, and to cultivate the right values and virtues in our work moving forward.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36.

[2] Ronald E Anderson. 1992. ACM code of ethics and professional conduct. *Commun. ACM* 35, 5 (1992), 94–99.

[3] Hannah Arendt. 2006. *Eichmann in jerusalem*. Penguin.

[4] John Ashworth, Benny Geys, and Bruno Heyndels. 2006. Everyone likes a winner: An empirical test of the effect of electoral closeness on turnout in a context of expressive voting. *Public Choice* 128, 3-4 (2006), 383–405.

[5] World Medical Association. 2013. World medical association declaration of helsinki: Ethical principles for medical research involving human subjects. *JAMA* 310, 20 (2013), 2191–2194. https://doi.org/10.1001/jama.2013.281053 arXiv:/data/journals/jama/929397/jsc130006.pdf

[6] Solon Barocas and danah boyd. 2017. Engaging the ethics of data science in practice. *Commun. ACM* 60, 11 (2017), 23–25.

[7] Scott Bateman, Regan L Mandryk, Carl Gutwin, Aaron Genest, David McDine, and Christopher Brooks. 2010. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2573–2582.

[8] Jean Baudrillard. 1995. *The Gulf War did not take place*. Indiana University Press.

[9] Sarah Belia, Fiona Fidler, Jennifer Williams, and Geoff Cumming. 2005. Researchers misunderstand confidence intervals and standard error bars. *Psychological methods* 10, 4 (2005), 389.

[10] Carsten Binnig, Lorenzo De Stefani, Tim Kraska, Eli Upfal, Emanuel Zgraggen, and Zheguang Zhao. 2017. Toward Sustainable Insights, or Why Polygamy is Bad for You.. In *Proceedings of the Conference on Innovative Data Systems Research (CIDR)*.

[11] Edwin Black. 2001. *IBM and the Holocaust: The strategic alliance between Nazi Germany and America's most powerful corporation*. Random House Inc.

[12] Paul Bloom. 2017. *Against empathy: The case for rational compassion*. Random House.

[13] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2016. Beyond memorability: Visualization recognition and recall. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 519–528.

[14] Michelle A Borkin, Azalea A Vo, Zoya Bylinskii, Phillip Isola, Shashank Sunkavalli, Aude Oliva, and Hanspeter Pfister. 2013. What makes a visualization memorable? *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2306–2315.

[15] Nadia Boukhelifa and David John Duke. 2009. Uncertainty visualization: why might it fail?. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. ACM, 4051–4056.

[16] Jeremy Boy, Anshul Vikram Pandey, John Emerson, Margaret Satterthwaite, Oded Nov, and Enrico Bertini. 2017. Showing People Behind Data: Does Anthropomorphizing Visualizations Elicit More Empathy for Human Rights Data?. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 5462–5474.

[17] danah boyd. 2016. Undoing the neutrality of big data. *Florida Law Review* 67 (2016), 226–232.

[18] danah boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679.

[19] Bundesarchiv. 1940. Planung und Aufbau im Osten. https://en.wikipedia.org/wiki/Heim_ins_Reich#/media/File: Bundesarchiv_R_49_Bild-0025,_Ausstellung_%22Planung_und_ Aufbau_im_Osten%22\protect\kern+.1667em\relax_Schautafel.jpg. Photograph by M. Krajewsky.

[20] Joy Buolamwini. 2018. When the Robot Doesn't See Dark Skin. *New York Times* (Jun 2018). https://www.nytimes.com/2018/06/21/ opinion/facial-analysis-technology-bias.html

[21] Steven P Callahan, Juliana Freire, Emanuele Santos, Carlos E Scheidegger, Cláudio T Silva, and Huy T Vo. 2006. VisTrails: visualization meets data management. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*. ACM, 745–747.

[22] Maciej Ceglowski. 2014. The Internet with a Human Face. *talk at Beyond Tellerrand in Düsseldorf, Germany, on* 20 (2014).

[23] Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L Boyd-Graber, and David M Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems*. 288–296.

[24] Danielle Keats Citron. 2007. Technological due process. *Wash. UL Rev.* 85 (2007), 1249.

[25] Michael Correll and Michael Gleicher. 2014. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2142–2151.

[26] Michael Correll and Jeffrey Heer. 2017. Black hat visualization. In *Workshop on Dealing with Cognitive Biases in Visualisations (DECISIVe), IEEE VIS*.

[27] Jeremy W Crampton. 2011. *Mapping: A critical introduction to cartography and GIS*. Vol. 11. John Wiley & Sons.

[28] Jeremy W Crampton and John Krygier. 2006. An introduction to critical cartography. *ACME: an International E-journal for Critical Geographies* 4, 1 (2006), 11–33.

[29] Craig Dalton and Jim Thatcher. 2014. What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data'. *Society and Space* 29 (2014).

[30] Aritra Dasgupta and Robert Kosara. 2011. Adaptive privacy-preserving visualization using parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2241–2248.

[31] Deanna Day. 2017. The History of Data is the History of Labor. *The New Inquiry* (Mar 2017). https://thenewinquiry.com/blog/ the-history-of-data-is-the-history-of-labor/

[32] Catherine D'Ignazio and Lauren Klein. 2019. *Data Feminism*. MIT Press. 2018 Draft.

[33] Catherine D'Ignazio and Lauren F Klein. 2016. Feminist data visualization. In *Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore. IEEE*.

[34] Jesse F Dillard. 2003. Professional services, IBM, and the Holocaust. *Journal of Information Systems* 17, 2 (2003), 1–16.

[35] E. Dimara, G. Bailly, A. Bezerianos, and S. Franconeri. 2018. Mitigating the Attraction Effect with Visualizations. *IEEE Transactions on Visualization and Computer Graphics* (2018), 1–1. https: //doi.org/10.1109/TVCG.2018.2865233

[36] Marian Dörk, Patrick Feng, Christopher Collins, and Sheelagh Carpendale. 2013. Critical InfoVis: exploring the politics of visualization. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2189–2198.

[37] Sam Dragga and Dan Voss. 2001. Cruel pies: The inhumanity of technical illustrations. *Technical communication* 48, 3 (2001), 265–274.

[38] Johanna Drucker. 2012. Humanistic theory and digital scholarship. *Debates in the digital humanities* (2012), 85–95.

[39] W.E.B. Du Bois. 1900. Proportion of freemen and slaves among American Negroes. http://hdl.loc.gov/loc.pnp/ppmsca.33913. A series of statistical charts illustrating the condition of the descendants of former African slaves now in residence in the United States of

America.

[40] W.E.B. Du Bois et al. 1926. Criteria of Negro art. *Crisis* 32, 6 (1926), 290–297.

[41] Elizabeth Farley, Fiona Grady, Dean S Miller, Rory O'Connor, Howard Schneider, Michael Spikes, Constantia Constantinou, et al. 2014. SPJ Code of Ethics. *The Power of Images* (2014).

[42] Mi Feng, Cheng Deng, Evan M Peck, and Lane Harrison. 2017. Hind-Sight: Encouraging exploration through direct encoding of personal interaction history. *IEEE transactions on visualization and computer graphics* 23, 1 (2017), 351–360.

[43] Ryan Gallagher. 2018. Google Shut Out Privacy and Security Teams From Secret China Project. *The Intercept* (Nov 2018). https://theintercept.com/2018/11/29/google-china-censored-search/

[44] Ryan Gallagher. 2018. Senior Google Scientist Resigns Over "Foreiture of our Values" in China. *The Intercept* (Sep 2018). https://theintercept.com/2018/09/13/google-china-search-engine-employee-resigns/

[45] Andrew Gelman and Eric Loken. 2013. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University* (2013).

[46] Lisa Gitelman. 2013. *Raw data is an oxymoron.* MIT Press.

[47] Michael Gleicher. 2013. Explainers: Expert explorations with crafted projections. *IEEE Transactions on Visualization & Computer Graphics* 12 (2013), 2042–2051.

[48] Miriam Greis, Jessica Hullman, Michael Correll, Matthew Kay, and Orit Shaer. 2017. Designing for Uncertainty in HCI: When Does Uncertainty Help?. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, 593–600.

[49] Donna Haraway. 1988. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist studies* 14, 3 (1988), 575–599.

[50] B. Hecht, L. Wilcox, J.P. Bigham, J. Schöning, E. Hoque, J. Ernst, Y. Bisk, L. De Russis, L. Yarosh, B. Anjum, D. Contractor, and C. Wu. 2018. It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. *ACM Future of Computing Blog* (Mar 2018). https://acm-fca.org/2018/03/29/negativeimpacts/

[51] Martin Heidegger. 1954. The question concerning technology. *Technology and values: Essential readings* 99 (1954), 113.

[52] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. Most people are not WEIRD. *Nature* 466, 7302 (2010), 29.

[53] Rosemary Lucy Hill, Helen Kennedy, and Ysabel Gerrard. 2016. Visualizing junk: Big data visualizations and the need for feminist data studies. *Journal of Communication Inquiry* 40, 4 (2016), 331–350.

[54] Arlie Russell Hochschild. 1983. *The Managed Heart: Commercialization of Human Feeling.* The University of California Press.

[55] Jeremy Hsu. 2018. The Strava Heat Map and the End of Secrets. *Wired* (Jan 2018). https://www.wired.com/story/strava-heat-map-military-bases-fitness-trackers-privacy/

[56] Jessica Hullman and Nick Diakopoulos. 2011. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2231–2240.

[57] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. 2018. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE transactions on visualization and computer graphics* (2018).

[58] Rosalind Hursthouse. 1999. *On virtue ethics.* OUP Oxford.

[59] Sarah Jeong. 2018. Bad Romance: How a cabal of authors profited by gaming Amazon's Kindle Unlimited algorithm. *The Verge* (Jul 2018). https://www.theverge.com/2018/7/16/17566276/

[60] Matthew Kay, Tara Kola, Jessica R Hullman, and Sean A Munson. 2016. When (ish) is my bus?: User-centered visualizations of uncertainty

in everyday, mobile predictive systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 5092–5103.

[61] Helen Kennedy, Rosemary Lucy Hill, Giorgia Aiello, and William Allen. 2016. The work that visualisation conventions do. *Information, Communication & Society* 19, 6 (2016), 715–735.

[62] Been Kim, Rajiv Khanna, and Sanmi Koyejo. 2016. Examples are not Enough, Learn to Criticize! Criticism for Interpretability. In *Advances in Neural Information Processing Systems.*

[63] P.-J. Kindermans, S. Hooker, J. Adebayo, M. Alber, K. T. Schütt, S. Dähne, D. Erhan, and B. Kim. 2017. The (Un)reliability of saliency methods. *NIPS workshop on Explaining and Visualizing Deep Learning* (2017). arXiv:stat.ML/1711.00867

[64] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM, 438.

[65] Robert Kosara. 2018. How to Get Excited About Standard Datasets. https://eagereyes.org/blog/2018/how-to-get-excited-about-standard-datasets.

[66] Robert Kosara and Steve Haroz. 2018. Skipping the Replication Crisis in Visualization: Threats to Study Validity and How to Address Them. In *Proceedings of BELIV 2018: Evaluation and Beyond – Methodological Approaches for Visualization.*

[67] Robert Kosara and Jock Mackinlay. 2013. Storytelling: The next step for visualization. *Computer* 46, 5 (2013), 44–50.

[68] Colin Lecher. 2018. What Happens When An Algorithm Cuts Your Health Care. *The Verge* (Mar 2018). https://www.theverge.com/2018/3/21/17144260/

[69] David Levering Lewis and Deborah Willis. 2010. *A small nation of people: W.E.B. Du Bois and African American portraits of progress.* Zondervan.

[70] Zachary C Lipton. 2016. The mythos of model interpretability. *arXiv preprint arXiv:1606.03490* (2016).

[71] Robert B Louden. 1984. On some vices of virtue ethics. *American Philosophical Quarterly* 21, 3 (1984), 227–236.

[72] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems.* 4765–4774.

[73] Konrad Meyer and Georg Blohm. 1942. *Landvolk im Werden.* Deutsche Landbuchhandlung.

[74] Luana Micallef, Pierre Dragicevic, and Jean-Daniel Fekete. 2012. Assessing the effect of visualizations on bayesian reasoning through crowdsourcing. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2536–2545.

[75] Ian Muehlenhaus. 2013. The design and composition of persuasive maps. *Cartography and Geographic Information Science* 40, 5 (2013), 401–414.

[76] Thomas M Nichols. 2017. *The death of expertise.* Tantor Media, Incorporated.

[77] Chris North. 2006. Toward measuring visualization insight. *IEEE computer graphics and applications* 26, 3 (2006), 6–9.

[78] Cathy O'Neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Broadway Books.

[79] Thomas Paine. 1776. *Common sense.*

[80] Anshul Vikram Pandey, Anjali Manivannan, Oded Nov, Margaret Satterthwaite, and Enrico Bertini. 2014. The persuasive power of data visualization. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2211–2220.

[81] DJ Patil, Hilary Mason, and Mike Loukides. 2018. *Ethics and Data Science.* O'Reilly.

[82] Jack Poulson. 2018. I Quit Google Over Its Censored Chinese Search Engine. The Company Needs to Clarify Its Position on Human Rights. *The Intercept* (Sep 2018). https://theintercept.com/2018/12/01/google-china-censorship-human-rights/

[83] Xiaoying Pu and Matthew Kay. 2018. The garden of forking paths in visualization: A design space for reliable exploratory visual analytics. In *Proceedings of the Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization (BELIV)*.

[84] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 1135–1144.

[85] Anne R Richards. 2003. Argument and authority in the visual representations of science. *Technical Communication Quarterly* 12, 2 (2003), 183–206.

[86] Phillip Rogaway. 2015. The Moral Character of Cryptographic Work. *IACR Cryptology ePrint Archive* 2015 (2015), 1162.

[87] Ian T Ruginski, Alexander P Boone, Lace M Padilla, Le Liu, Nahal Heydari, Heidi S Kramer, Mary Hegarty, William B Thompson, Donald H House, and Sarah H Creem-Regehr. 2016. Non-expert interpretations of hurricane forecast uncertainty visualizations. *Spatial Cognition & Computation* 16, 2 (2016), 154–172.

[88] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148.

[89] Hamza Shaban. 2018. Amazon employees demand company cut ties with ICE. *Washington Post* (June 2018). https://www.washingtonpost.com/news/the-switch/wp/2018/06/22/amazon-employees-demand-company-cut-ties-with-ice/

[90] Sarah Slobin. 2014. What If the Data Visualization is Actually People? *Source* (Apr 2014). https://source.opennews.org/articles/what-if-data-visualization-actually-people/

[91] Justin Talbot, Bongshin Lee, Ashish Kapoor, and Desney S Tan. 2009. EnsembleMatrix: interactive visualization to support machine learning with multiple classifiers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1283–1292.

[92] Jarke J Van Wijk. 2005. The value of visualization. In *Visualization, 2005. VIS 05. IEEE*. IEEE, 79–86.

[93] Manasi Vartak, Sajjadur Rahman, Samuel Madden, Aditya Parameswaran, and Neoklis Polyzotis. 2015. SeeDB: efficient data-driven visualization recommendations to support visual analytics. *Proceedings of the VLDB Endowment* 8, 13 (2015), 2182–2193.

[94] Paul Vines, Franziska Roesner, and Tadayoshi Kohno. 2017. Exploring ADINT: Using Ad Targeting for Surveillance on a Budget-or-How Alice Can Buy Ads to Track Bob. In *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*. ACM, 153–164.

[95] Chat Wacharamanotham, Krishna Subramanian, Sarah Theres Völkel, and Jan Borchers. 2015. Statsplorer: Guiding novices in statistical analysis. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2693–2702.

[96] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2017. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology* 32, 2 (2017).

[97] Emily Wall, Leslie M Blaha, Lyndsey Franklin, and Alex Endert. 2017. Warning, bias may occur: A proposed approach to detecting cognitive bias in interactive visual analytics. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*.

[98] Xumeng Wang, Wei Chen, Jia-Kai Chou, Chris Bryan, Huihua Guan, Wenlong Chen, Rusheng Pan, and Kwan-Liu Ma. 2018. GraphProtector: A Visual Interface for Employing and Assessing Multiple Privacy Preserving Graph Algorithms. *IEEE transactions on visualization and computer graphics* (2018).

[99] Mark Ward. 2016. *Deadly documents: Technical communication, organizational discourse, and the Holocaust: Lessons from the rhetorical work of everyday texts*. Routledge.

[100] Bernard AO Williams and Thomas Nagel. 1976. Moral luck. *Proceedings of the Aristotelian Society, Supplementary Volumes* 50 (1976), 115–151.

[101] Mchael Wines. 2018. Why Was a Citizenship Question Put on the Census? 'Bad Faith,' a Judge Suggests. *New York Times* (Jul 2018). https://www.nytimes.com/2018/07/10/us/citizenship-question-census.html

[102] Kanit Wongsuphasawat, Daniel Smilkov, James Wexler, Jimbo Wilson, Dandelion Mané, Doug Fritz, Dilip Krishnan, Fernanda B Viégas, and Martin Wattenberg. 2018. Visualizing dataflow graphs of deep learning models in TensorFlow. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 1–12.

[103] Denis Wood. 2010. *Rethinking the power of maps*. Guilford Press.

[104] Jo Wood, Petra Isenberg, Tobias Isenberg, Jason Dykes, Nadia Boukhelifa, and Aidan Slingsby. 2012. Sketchy rendering for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2749–2758.

[105] Jo Wood, Alexander Kachkaev, and Jason Dykes. 2018. Design Exposition with Literate Visualization. *IEEE transactions on visualization and computer graphics* (2018).

[106] Mushon Zer-Aviv. 2015. DataViz–The UnEmpathetic Art. https://responsibledata.io/2015/10/19/dataviz-the-unempathetic-art/.

[107] Emanuel Zgraggen, Zheguang Zhao, Robert Zeleznik, and Tim Kraska. 2018. Investigating the Effect of the Multiple Comparisons Problem in Visual Analysis. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 479.

[108] Donghua Zhu and Alan L Porter. 2002. Automated extraction and visualization of information for technological intelligence and forecasting. *Technological forecasting and social change* 69, 5 (2002), 495–506.