A Path to Understanding the Effects of Algorithm Awareness

Kevin Hamilton

Center for People and Infrastructures

University of Illinois, Urbana-Champaign

1308 W Main Street

Urbana, IL 61801-2307

kham@uiuc.edu

Karrie Karahalios

Center for People and Infrastructures

University of Illinois, Urbana-Champaign

1308 W Main Street

Urbana, IL 61801-2307

kkarahal@uiuc.edu

Christian Sandvig

University of Michigan

Room 5385 North Quad

105 South State Street

Ann Arbor, MI 48109-1285

csandvig@umich.edu

Motahhare Eslami

Center for People and Infrastructures

University of Illinois, Urbana-Champaign

1308 W Main Street

Urbana, IL 61801-2307

eslamim2@uiuc.edu

Abstract

The rise in prevalence of algorithmically curated feeds in online news and social media sites raises a new question for designers, critics, and scholars of media: how aware are users of the role of algorithms and filters in their news sources? This paper situates this problem within the history of design for interaction, with an emphasis on the contemporary challenges of studying, and designing for, the algorithmic "curation" of feeds. Such a problem presents particular challenges when, as is common, neither the user nor the researcher has access to the actual proprietary algorithms at work.

Author Keywords

Visualization; perception; social media; reverse engineering; design; algorithms; privacy; infrastructure; software studies.

ACM Classification Keywords

H.5.m. Information interfaces and presentation: Miscellaneous.

Introduction

At a recent party attended largely by professors and other Ph.D-bearing individuals who are uninvolved in

computer science, an informal survey of Facebook users revealed near total ignorance about the fact that the social media site filters one's news feed. "Why in the world would they do that?" asked one highly intelligent, prize-winning professor of medicine. Casually and in user studies, even some Graduate Students in Computer Science who routinely used Facebook were unaware that their feeds were filtered or curated.

For some, such lack of awareness is indicative of successful design - for shouldn't good interaction design be invisible? For others this invisibility indicates that something potentially controversial has been settled, decided, made static. How aware do users need to be of the algorithms at work in their daily internet use? How does understanding of algorithms affect use? And what are the best paths to answering these questions when so many algorithms are buried not only outside of human perception, but behind walls of intellectual property?

This paper will (1) review the precedent and context for studying "invisible" processes in support of design, (2) identify the differing arguments in support of revealing or obscuring computational processes to users, and (3) outline some paths forward to studying this subject in the context of the algorithms at work in internet search and Social Network Sites.

Studying Interaction with the Invisible

Multiple areas of design research require representation of the "hidden" patterns at work in human interaction with everyday infrastructures. From earlier efforts influenced by cognitive science to more recent work informed by actor-network or feminist theories,

designers of interactive experience for digital and physical spaces often look to base their production on knowledge of what "really happens" at sites of experience and consumption.

Though Human Factors research set an influential precedent for the HCI community through revealing the cognitive structures at work in interaction with machines, parallel efforts exist in Architecture and Urban Planning, where designers have historically looked to cognitive science and wayfinding studies as a source for designing memorable, navigable cities. Likewise, scholars of media have looked to demonstrate how viewers understand editorial processes as a way to facilitate the design of prosocial media content.

As accounts of human behavior change, designers and scholars have shifted what they expect to find in these studies. Where an influential urban planner such as Kevin Lynch imagined a shared cognitive map beneath individual subjective experiences of the city, more recent approaches expect a less unifed account of reception and action. Whether through "Third Paradigm HCI" [21,24], wherein researchers depend on epistemologies grounded in feminist, postcolonial, or critical race theories, or through design informed by computational neuroscience [25], the study of patterns in human interaction have turned from a search for universals to a concern for "situatedness." Some have even left behind the hope of basing design on the understanding, prediction or modeling of human behavior at all, opting instead for generative design approaches, or real-time user-driven iteration.

The study and representation of "hidden" patterns in human interaction as a basis for design has itself thus

become harder to see, through a proliferation and dispersion of methodologies, and through a turn from the quantitative to the qualitative. In older models, and especially those that emerged from time and motion studies, a photographic image might easily afford discovery of patterns and structure. One might even say that the mathematical functions emerging from a cognitive scientific understanding of minds as "processors" are a form of making "visible" previously hidden structures. But how does one visualize the structure of "situated actions" that emerge in ways that reflect their unique and often deeply embedded social contexts, where matters of race, class or gender are foundationally in play?

If the design of not only efficient but humane and just interfaces still requires some form of "making visible" the processes of human plans, desires, and actions – a requirement only a few would argue - the task of the HCI researcher has only grown more complicated. The need for new modes of visualization is urgent.

Seams, Transparency, and Blackboxing

While the visualization of human interaction with machines has grown more elusive, opinions differ on the necessity of visualizing the other side of these relationships – the work of machines. How visible or legible should computational processes be to users of a final product? How much do users need to know about the ordering principles and bases of the products and interfaces they utilize or inhabit?

Some approach this question through the metaphor of the "seam," arguing over the merits of seamless or "seamful" designs for a particular product or experience. Seamless designs, some argue, wherein

the black box is sealed or even invisible, will feel "effortless" or invisible when based on the right model of human behavior. The user, some argue, never needs to know the model of processing at work, and the role of their actions within it.

Others argue that between the intuitive, action-situated nature of human intention and the inadequacy of many cognitive models, the "seams" around an interface's construction should remain highly visible, so as to facilitate experimentation and innovative use. "Seamful" interfaces, they argue, are less universalizing, and allow for a broader range of human behavior, adaptive and evolving use [31].

If such arguments often hinge on a more phenomenological approach, others debate the merits of visible design logics from a political or social perspective. Drawing from the work of historians and sociologists of science such as Latour, Bowker or Starr, many see opacity in technologies as a call for inquiry into what processes of debate and concern have been arrested or settled behind the opaque surface of "black boxes." Julian Oliver and others have called for a new "critical engineering" that reveals such processes as a matter of serving the common good, returning a settled matter, such as which information a mobile app might share with a third party, into a space of debate.

Some designers of secure and critical systems even argue for such transparency as fundamental to the success of certain computational processes or interfaces. E-voting technologies, for example, depend for their authority on some form of explanation of the machine's processes in order to create trust [20].

Across such debates, opacity in computational processes can result from a myriad of factors. The "black box" protects intellectual property in some cases or removes a state process from public view in the interests of guarding authority. In many cases, opacity is simply a function of complexity, and difficult to avoid.

Machine processes elide visibility as easily as the processes by which humans interact with them, while the value of making such processes visible in the first place remains a matter of debate.

The design of algorithmic interfaces – as in search or Social Network Sites (SNSs) - begs for address of these questions, and brings a host of new ones. At stake in these designs is both *adaptability*, as often the same interface must serve so many different uses and users, and *security*, as the user needs to believe that the system is designed dependably, with her interests in mind. A study of the value of transparency in algorithmic interfaces will need to draw from a diverse range of methods and critiques.

Attention to Algorithms

A growing group of researchers has turned to the study of the implications of algorithmic interfaces [1]. Pioneering work on search engine results emphasized that it is impossible to separate normative social questions from the technical implementation of search algorithms [8]. Search engines intentionally and unintentionally shape what knowledge is easiest to find, sometimes with unexpected results, as some sources of knowledge are de-emphasized or even hidden.

While work on search continues [6], the normative investigation of algorithms has spread to a variety of other domains, including automated stock trading [16], health care, credit scores [12], scholarly journal rankings [2], and in fact most online activities [5]. New disciplinary perspectives are now joining scientists and engineers in grappling with the definition of an "algorithm," its efficacy, and whether one might be described as "wrong" or even "unethical" [14].

In a recent paper Kerr and Earle [23] call special attention to the ethics of algorithms as predictive processes, comparing *consequential* prediction – in which an algorithm displays possible outcomes for use in a person's decision-making – to *preferential* or *preemptive* prediction. In *preferential* prediction, a process anticipates a user's desires and offers options likely to please, where in *preemptive* prediction, an algorithm delimits a person's access without providing a choice, and often without the person's knowledge. Kerr and Earle call special attention to the dangers of opacity in preemptive algorithms, pointing to their potential for unjust application. Looking to history, one might recall here the Federal Housing Authority's reliance in the 1930s on algorithmic processes of rating potential loan recipients based on the racial composition of their home neighborhoods [32].

Even seemingly user-focused preferential processes might enact or reveal embedded social bias, as discussed in Mike Ananny's recent discovery that the Android app store recommendations implied a connection between gay social behavior and pedophilia [22]. As many debate the origins and implications of such connections, Nicholas Diakopoulos, calls for more "algorithmic accountability reporting" from journalists, and differentiates between the intentional and incidental application of "algorithmic power." [26]

The profit motive has also been essential in mobilizing research about algorithms for some time. Advertisers long to know the surest way to make a status update "sticky" in a space such as Facebook, keeping it near the top of feeds for a longer time. Experts now promise to help advertisers understand the phenomenon of news-propagation in social media (e.g., SocialFlow [18]). Articles in the trade press have speculated for years about Facebook's EdgeRank algorithm, revealed by Facebook engineers at a development conference in 2010. A slew of news stories in the summer of 2013 introduced yet new processes to succeed EdgeRank, accompanied even by photographs of the laborers at work on the next rollout, with an effect for some not unlike getting a peek inside the factory of fictional chocolatier Willy Wonka.

On Reverse Engineering
So how are we to bring the diverse methods, debates, and design rationales surrounding visibility of human and machine processes to bear on algorithms? What distinctive challenges do algorithmic newsfeeds, search results, and interfaces present to understanding human interaction, and establishing norms for design or even ethics?

The challenges facing HCI researchers overlap somewhat with those faced by journalists at the moment. The journalist's primary method in such cases – that of unveiling – will in some cases serve the traditional HCI need for visibility as well. To this end, the previously mentioned Diakopoulos study calls for more reverse engineering of algorithms by journalists, and provides case studies and techniques to this end. Indeed, one recent journalistic essay on Netflix in *The*

Atlantic was the result of a collaboration between a journalist and an HCI researcher [27].

Yet, as Nicholas Seaver pointed out in a response to the Netflix article, reverse engineering as a method of understanding algorithms has its limits [28]. "While reverse engineering might be a useful strategy for figuring out how an existing technology works," Seaver writes, "it is less useful for telling us how it came to work that way." As a research strategy, reverse engineering, Seaver continues, "misses the things engineers do that do not fit into conventional ideas about engineering." Reverse engineering may promise the sort of visibility that researchers crave, but certain matters within this process will remain in the dark.

HCI researchers who look to reverse engineering as a technique for studying algorithmic interfaces will also face some additional challenges that are less likely to occur in journalism. These include the technical challenges of producing an accurate enough approximation of an opaque algorithm, and the unique ethical challenges of working either "with" the proprietors of algorithms or "against" such proprietors through reverse engineering. As an "insider," a researcher looking to understand the construction of "algorithmic power" might well run into conflicts of interest with her partner or employer. As an "outsider" working through certain reverse engineering techniques, a researcher might well end up violating terms of service agreements, raising ethical questions about which at least some professional organizations, such as the ACM, have established clear expectations. (Though we don't take such policies as a last word, we are eager for more conversations about the ethics of this practice.)

In the case of algorithmically curated information sources, the task of study is complicated by the fact that such news sources are different for each user, composed based on her account settings, the composition of her social network, and records of past use. Recent studies in configuration settings for SNSs reveal that users are often unaware of available site features and are confused by unintuitive and frequently changing interfaces [18]. Yet even if the user knows the feed can be adjusted, it isn't clear what behavior a change in settings would produce. Where other systems facilitate real-time input and output, an algorithmically curated news feed relies primarily on changes from outside the user's control. A setting change may not take effect for some time. That these algorithms are also regularly updated by providers introduces another wrinkle, as the user can't even verify that the same process is in place over serial visits or uses.

Such complex factors contribute to the opacity of algorithms in these settings, and also present a challenge to the study of opacity's effects. In addition, the algorithms themselves are often trade secrets within a competitive marketplace. To isolate them for study requires not only the pursuit of a moving target, but potentially a breach of terms of service.

Research Questions

Certainly reverse engineering will have to play some role in the study of, and design for, algorithmic interfaces. But the questions are larger than such a technique can contain or accomplish.

Much as Diakopoulos called for more journalistic reporting on algorithms, we would like to call for renewed scholarship and design in this area, with attention to the following substantive and methodological questions:

1. How is a research project on algorithmic interfaces to proceed when access to the actual algorithm is limited?

2. Where, when and how are users made aware of algorithms?

3. How does perception translate into cognition and knowledge of the process at hand?

4. How important is accurate cognition to use?

Three ready approaches appear toward the goal of answering such questions, thus furthering understanding the influence of algorithm visibility and awareness on use. These include: (1) surveying users to determine their awareness of processes at work in their everyday consumption; (2) exposing hidden algorithmic processes to users and then studying the effects of knowledge on use; and (3) working with users to try and deduce the algorithmic processes at hand, as well as the design rationales behind them.

We have begun to try out a combination of these approaches, with some initial results already reached through the survey method. (Our ongoing work to this end suggests, for example that less than 25% of regular Facebook users are aware that their feeds are curated or filtered, and even less know how to affect that process.)

To conclude this paper, we'll offer some thoughts on the benefits and challenges of the other two approaches.

Exposing Algorithms

In the second approach outlined above, researchers might reveal the algorithms at work in a particular interface to users, and then study the effects of new knowledge on use. This would either require reverse engineering a process, or working with an algorithm's proprietor. Previously mentioned technical and ethical challenges will apply here, as do the concerns raised by Seaver's response to the Atlantic article.

If, taking a cue from designers of critical systems such as e-voting machines, a user requires a full explanation of a process in order to establish trustworthiness [20], then revealing the algorithm's processes alone would not suffice. In order to determine, and design for, the influence on knowledge of a process on use, the researcher would need to experiment with revealing to users not only the process at hand, but the design goals and context that led to the process at hand.

In other words, to understand the effects of knowledge of a process on use in the case of algorithms, one would need to plumb and augment the user's understanding not only of the process at hand, but of the entire design context and motivations out of which an algorithmic process emerged. Taking into consideration the work of Suchman, Schön, and other scholars of the design process, this might prove a challenging task indeed, given the relative inseparability of goals, plans and actions.

It's also worth noting that the exposure of a black box's contents often occurs during points of crisis or breakdown in use. So information gleaned through exposing algorithms to users may require understanding some additional contextualization in light of the anxieties surrounding infrastructural failure.

Perceiving Algorithms

Rather than revealing an algorithmic process at work to users and studying the effects of such knowledge, a third approach might enlist users in a process of inference and deduction about opaque processes. To use an analogy, consider how city planner Kevin Lynch developed design principles for urban design by asking city dwellers to sketch maps of their environments from memory [29]. In this way, Lynch learned something about what features of a city are more or less memorable in support of a "cognitive image." Based on an assumption that easily "imaged" cities make for better cities, he then moved to develop design recommendations for urban planners.

In this analogy a researcher into algorithmic interfaces might be able to develop understanding about how users become aware of algorithms *whether or not* she has a copy of the "true map." The researcher might even be able to make conclusions about the effects of knowledge on use in light of incorrect knowledge. In this way, researchers into algorithm awareness might also share some perspectives in common with Lynch's contemporary critics, who sometimes question whether a unified "image" of a city is possible or even desirable [30].

Likewise, HCI as a field has fundamental divisions over the role of mental models in the use of computers, and

HCI debates have echoed the trajectory of modern cognitive science. Work in usability once stressed the value of the user holding an *accurate* mental model of the system being used [10]. Researchers later conceptualized HCI's task as reciprocal: to intervene both in how the user models the computer and how the computer models the user [7]. Yet distinct from these approaches, a significant group saw their goal as eliminating the need for any mental model at all, causing the computer to be "intuitive," "self-explanatory," or to fade into the background (for a review of this debate, see [17]).

More recent work has largely abandoned the idea of the *accurate* mental model as unobtainable (or even because it is metaphysically troubling—there are potentially infinite ways to model a single system). Instead, work has shifted toward the signifiers that the system provides that allow a user to construct their understanding of it, which is always partial [11]. Mental models are now often thought inevitable and are evaluated in terms of utility for the user, rather than their verisimilitude [17].

In light of the questionable value or even possibility of establishing, or designing for, a clear "image" of the algorithmic process at work, a study based on "designing with" a user through collaborative deduction of an algorithmic process might yield helpful answers to the question of how users are aware of these processes at all, and whether awareness affects use.

An example from history might serve well here – that of a pedagogical device created by cyberneticist Ross Ashby, a machine that long-served the classroom at University of Illinois, and reached the world through the writing of Ashby's colleague, Gordon Pask. In the 1950s Ashby created a small device containing two switches and two lamps. As described by Jan Mueggenburg[9], "These inputs and outputs gave the box four variables with two states each, yielding sixteen different combinations of lever positions and lamps turned on or off." Ashby would challenge his students to deduce the relationship of inputs to outputs – which was impossible without opening the device. Instead, the device enabled students to study the process of deduction itself.

Likewise, we imagine a path forward that invites research participants to probe and query the nature of the algorithmic process at hand in, for example, a sorted and filtered Facebook feed. We've actually begun such a study in a way that also involves some reverse engineering, revealing aspects of the algorithmic process at hand. In our study, about which we'll say more in a future paper, we're synthesizing all three approaches outlined here.

A Path Forward
To determine the role of algorithm visibility or even comprehension in user experience may thus require careful attention to acts of seeing and sense-making by users, often in ignorance about the accuracy of a user's picture of an algorithmic process.

In aspects of both the second and third approaches outlined above, we imagine a path forward in which, rather than presenting a participant with an algorithm whose processes we already know, we present her with an opaque algorithm at work – such as that which populates a Facebook news feed – and ask them to work with us to form a picture of the processes at hand. In this way we hope to note and record acts of

user perception, and then to study how particular percepts lead to unique ways of sense-making about opaque processes.

Then we can move to evaluate how a feeling of literacy or legibility about algorithms on the part of a user influences interaction.

This path places the study of perception and literacy above the close study of the algorithms themselves as mathematical processes. Though we are not naïve to the sometimes negative political and social effects of opaque technological processes, our proposed path refrains from assuming malfeasance where in instances of opacity. We remain open about this latter question, in part because the history of technology shows that there are many explanations for why a technology remains "opaque" [3][15]. We also, with researchers such as Bill Gaver in his use of probes in the face of uncertainty or Pierce [13] in his discussion of "undesign," want to let human acts of seeing and sensation lead our efforts.

In this spirit, we imagine an approach in which researchers work with users to create a rough "map" of an algorithm together – even if the map itself is neither the primary goal, nor even accurate.

Acknowledgements

References

[1] S. Barocas, S. Hood, and M. Ziewitz. Governing Algorithms: A Provocation Piece. Paper presented to the conference, "Governing Algorithms: A Conference on Computation, Automation, and Control." New York University, New York. March 29. 2013.Anderson, R.E. Social impacts of computing: Codes of professional ethics. *Social Science Computing Review 10*, 2 (1992), 453-469.

[2] B. Brembs, K. Button, & M. Munafò. Deep impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience 7*: 291. 2013.

[3] Paul N. Edwards. Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems. In Modernity and Technology. eds. T. Misa, P. Brey, and A. Feenberg. MIT Press. 2003.

[4] W. Gaver, A. Boucher, S. Pennington, B. Walker. Cultural probes and the value of uncertainty. Interactions. (11)5. September 2004.

[5] T. Gillespie. Can an Algorithm Be Wrong? *Limn 2*. 2012. http://limn.it/can-an-algorithm-be-wrong/

[6] L. A. Granka. The Politics of Search: A Decade Retrospective. The Information Society 26(5): 364-374. 2010.

[7] B. Laurel. *Computers as Theatre*. Addison-Wesley: New York. 1991.

[8] L. Introna and H. Nissenbaum, Shaping the Web: Why the Politics of Search Engines Matters, *The Information Society*, 16(3):1-17, 2000.

[9] Jan Mueggenburg. Lively Artifacts | Feedback. Accessed September 17,2013. http://openhumanitiespress.org/feedback/science-technology/lively-artifacts

[10] D. Norman. *The Psychology of Everyday Things*. Basic Books: New York. 1988.

[11] D. Norman. *Living with Complexity*. MIT Press: Cambridge, MA.2010.

[12] F. A. Pasquale III. Restoring Transparency to Automated Authority. *Journal on Telecommunications and High Technology Law*, 9 (235). 2011.

[13] James Pierce. Undesigning Technology: Considering the Negation of Design by Design. *CHI2012*.

[14] C. Sandvig, K. Hamilton, K. Karahalios, & C. Langbort.. Re-Centering Algorithms. Paper presented to the conference, "Governing Algorithms: A Conference on Computation, Automation, and Control." New York University, New York. March 2013.

[15] Susan Leigh Star. The Ethnography of Infrastructure. American Behavioral Scientist. (43)377. 1999.

[16] C. Steiner. *Automate This: How Algorithms Came to Rule Our World.* Portfolio: New York. 2012.

[17] L. Suchman. Human-Machine Reconfigurations: *Plans and Situated Actions.* Cambridge University Press: New York. 2006.

[18] J.Watson, A. Besmer, and H. R. Lipford, H. +your circles: sharing behavior on google+. In Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS '12, ACM (New York, NY, USA, 2012), 12:1–12:9.

[19] Social Flow. Accessed September 17, 2013. http://www.socialflow.com

[20] Wolter Pieters. Explanation and Trust: What to Tell the User in Security and AI? *Ethics and Inf. Technol.* 10, 1 (2011), 53-64.

[21] S. Harrison, P. Sengers, and D. Tatar. Making Epistemological Trouble: Third-paradigm HCI As Successor Science. *Interacting with Computers 23*, 5 (2011), 385-392.

[22] Mike Ananny. The Curious Connection Between Apps for Gay Men and Sex Offenders. *The Atlantic*, April 14, 2011.

[23] I. Kerr and J. Earle. Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy. *Stanford Law Review Online 66* (2013), 65.

[24] Phoebe Sengers. What I Learned on Change Islands: Reflections on IT and Pace of Life. *Interactions 18*, 2 (2011), 40-48.

[25] B. Minnery and M. Fine. Neuroscience and the Future of Human-computer Interaction. *Interactions 16*, 2 (2009), 70-75.

[26] N.Diakopoulos. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. Tow Center for Digital Journalism Brief, Columbia University (2014). http://towcenter.org/blog/computational-journalism-and-the-reporting-of-algorithms/

[27] Alexis Madrigal. How Netflix Reverse Engineered Hollywood. *The Atlantic*, January 12, 2014.

[28] Nick Seaver. On Reverse Engineering: Looking for the Cultural Work of Engineers (2014). https://medium.com/anthropology-and-algorithms/d9f5bae87812

[29] Kevin Lynch. *The Image of the City*. MIT Press: Cambridge. 1960.

[30] Offenhuber, Dietmar, and Ratti C. Reading the City — Reconsidering Kevin Lynch's Notion of Legibility in the Digital Age, in The Digital Turn, in *Design in the Era of Interactive Technologies*, ed. Zane Berzina, Barbara Junge, Wim Westerveld, and Carola Zwick. Zürich: Park Books (2012).

[31] Matthew Chalmers and A. Galani. Seamful interweaving: heterogeneity in the theory and design of interactive systems. *DIS '04 Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (2004)*, 243-252.

[32] Jennifer Light. Discriminating Appraisals: Cartography, Computation, and Access to Federal Mortgage Insurance in the 1930s. *Technology and Culture 52*, 3 (2011), 485-52