

**BEKI GRINTER**

Georgia Institute of Technology | beki@cc.gatech.edu
 Rebecca "Beki" Grinter is a professor in the School of Interactive Computing and, by courtesy, the Scheller College of Business at the Georgia Institute of Technology.

A Big Data Confession

Big data is being positioned as a research agenda within HCI. But I confess that I'm troubled by developments in the relationship between people (and their data, and as data) and the algorithms of big data analysis.

A recent *Communications* article repurposed cellular telephony call detail records (CDRs) to determine hundreds of thousands of people's average commutes, carbon footprints, and locations [1]. The authors cited the prohibitive costs of collecting this data using traditional methods as justification for repurposing CDRs. While not conducted by HCI researchers, this report focuses on topics of interest to us: mobility and sustainability. It raised three concerns for me.

First, the repurposing of records separates the original rationale for data collection from the intent of the current algorithmic analysis. The customers, whose data was originally collected for billing and network-traffic-management purposes, participated in a study in which their data was transformed into information about their location and movements. They were not required to give their consent to participate, but should they have been? The authors describe efforts to anonymize the data, but this is difficult to do. Should participants have any ability to control the algorithms that generated this data (e.g., to make them not run on the human-identifiable data)? Further, should participants have the ability to prevent their data from being repurposed without their consent, a

protection against future, but currently unimagined, uses?

Second, this repurposing of data about people by corporations leads to new resources. For example, could this new location information be combined with other sources (e.g., census data about average income by zip code) to customize service offerings? How much are our movements and locations worth to other corporations? This repurposed data creates new resources and opportunities for the corporation, but what does it mean for the customer?

Third, what does it mean when only corporations can analyze big data? This research was made possible only through access to corporate CDR data. We might argue that this data is the privilege of corporate researchers and those who collaborate with them or could purchase the assets. But like others, I see a potential research divide between those with access, who then control the agenda and produce its knowledge (including what gets analyzed and does not), and those without [2]. As a research community, should we discuss strategies for mitigating the potential divide?

More generally, this study raises new questions about the relationship between the researcher and the researched and highlights the role of the corporation in that process. What should HCI do? Our commitment to the human experience has long required us to be mindful of participants' rights. More recently, we have focused



our attention on methods such as action research that promote research-participant partnerships and theories such as feminism that encourage attention to values such as empowerment, agency, and equity in our technological designs. Being at the intersection of the human and technological experience gives us unique and important



perspectives to bring to bear on the challenges presented by the opportunities of big data research.

Other activities have begun to transform people into big data. For example, massively online open courses (MOOCs) allow thousands of people to enroll in a single class, where they are algorithmically managed as elements of a big data

set (e.g., via automated grading). But previous research finds that the human experience of learning is what encourages women and minorities to stay in disciplines like computing. An open question remains as to whether MOOCs could erode those gains as currently conceived. I believe that HCI could promote a different agenda. Some commentators on MOOCs cite their potential global reach. Today that reach is treated as an opportunity for knowledge transfer from the professor to the masses of students. Could we leverage our knowledge about collaboration and build technologies that support learning within and across local and remote groups, making the process one of knowledge co-creation and relationship building, including mentorship, rather than a one-way transfer?

HCI is also engaged in research to program humans as a social computer. But what does it mean to treat people as software? Recently, colleagues asked what it would mean for their children to perform crowd work [3]. But the aspirations for paid crowd labor often focus on a different demographic than academics' offspring. However, HCI's recent turn toward feminism suggests alternatives, including that offered by Lilly Irani and Six Silberman [4], who developed technology that gave voice to crowd workers, highlighting how traditional approaches often make their work invisible.

Big data is growing in importance as a research theme, but there is an increasing number of people concerned about the challenges such an agenda creates. HCI is uniquely positioned to act. We can build on and broaden our collaborations with scholars in the social sciences, humani-

ties, and computing to promote human experiences of big data. For example, we can draw on our many years of ethical practices around the human experience of technology and work with scholars of ethics to ask what needs to change in order to adequately protect the rights of people in big data. We can continue to draw on theories that are well positioned to ask questions about who is being represented in a system and how, and offer alternatives like those proposed by Irani and Silberman [4]. More generally, I believe the HCI community is well situated to focus on questions of control, ethics, morality, equality, representation, and so forth that are embedded in big data and the algorithms that process it.

In conclusion, I am troubled by big data, but I also confess to being excited about the possibilities it holds for HCI because of the opportunity it creates for us to be advocates for a truly human experience.

Acknowledgments

I'd like to thank Shaowen Bardzell, danah boyd, Eric Gilbert, Lilly Irani, Phoebe Sengers, and Susan Wyche for their encouragement, feedback, and insight.

ENDNOTES:

1. Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J.M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A., and Volinsky, C. Human mobility characterization from cellular network data. *Commun. ACM* 56, 1 (2013), 74-82.
2. boyd, d. and Crawford, K. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society* 15, 5 (2012), 662-679.
3. Kittur, A., Nickerson, J.V., Bernstein, M.S., Gerber, E.M., Shaw, A., Zimmerman, J., Lease, M., and Horton, J.J. The future of crowd work. *Proc. ACM Conference on Computer Supported Collaborative Work* (San Antonio, TX). ACM Press, New York, 2013.
4. Irani, L. and Silberman, M.S. Turkopticon: Interrupting work invisibility in Amazon Mechanical Turk. *Proc. of ACM Conference on Human Factors in Computing Systems* (Paris, France). ACM Press, New York, 2013.