Interface/

Volume 1
Issue 1 *Theorizing the Web 2014*

Article 13

2015

How Evasion Matters: Implications from Surfacing Data Tracking Online

Janet Vertesi

Recommended Citation
Vertesi, Janet. (2015) How Evasion Matters: Implications from Surfacing Data Tracking Online. Jenny Davis & Nathan Jurgenson (eds.) Theorizing the Web 2014 [Special Issue]. *Interface* 1.1. 1 - 14. DOI: http://dx.doi.org/10.7710/2373- 4914.1013

Interface (2373-4914) is published annually by The Berglund Center (Pacific University) in collaboration with the Pacific University Libraries.

Interface/ Vertesi, Janet. (2015) How Evasion Matters: Implications from Surfacing Data Tracking Online. Jenny Davis & Nathan Jurgenson (eds.) Theorizing the Web 2014 [Special Issue]. *Interface* 1.1. 1 - 14. DOI: http://dx.doi.org/10.7710/2373-4914.1013

How Evasion Matters: Implications from Surfacing Data-Tracking Online

By Janet Vertesi
Princeton University

The past five years have seen the rise and expansion of a new infrastructural layer to the Internet. Inspired by the unprecedented success of Google Ad-words and Facebook's "social" data collection regime, this middle layer has expanded to include a plethora of bots, cookies, trackers, canvases, and other data sniffers intent on recording user clicks, likes, and purchases. Contemporary consumers do not even need to buy: browsing, searching, or clicking on a quiz can all be indicators that a user is looking for new shoes or a winter coat, triggering targeted ads. In the world of contemporary Internet companies, personal data reigns supreme.

This is not only an online phenomenon. Offline, too, loyalty card programs and credit card purchase data translate into targeted mail catalogs and direct-to-consumer marketing programs. Across social media companies and third-party data brokers, certain consumers are designated high-value targets. For example, according to a *Financial Times* report, information pertaining to prospective new mothers, who are likely to be making new and lasting brand choices, can sell for much higher prices than everyday users. [1] Attempts to coordinate online and offline datasets to better identify such high-value targets continue apace, leading personal data gleaned through commercial services to exchange hands among brokers for high prices.

I have spent the past three years attempting to evade various aspects of this middle-layer of the Internet; first, through studious avoidance of Google-related products and services, and most recently through concealing the impending arrival of a family member from data detection. I undertook this latter practice as an experiment in infrastructural inversion [2]: an attempt to make visible the embedded nature of this tracking infrastructure in daily life, as well as the values and assumptions that such technologies make about everyday users. While I have described the experiments elsewhere, [3] resisting these data dragnets [4] presented several key findings relevant to the study of Internet infrastructures and behaviors online that are important for our research community to address.

In this essay, I suggest three implications for our continuing studies of life online and offline that arose from practices of data-gathering evasion: first, considering the server as an actor in online interactions; second, reframing practices of resistance accordingly; and third, a resulting novel framework for personal data privacy policy and design.

Where is the server?

Evading data capture required concerted thinking about the web's underlying infrastructure. Because the post-Web-2.0 Internet is driven by the collection and sale of personal information, servers are incentivized to remember that input indefinitely, and to enable that data's transfer and correlation across various databases. Companies work hard to make such server architecture invisible to the user, to facilitate an otherwise "seamless" [5] interaction between site users: one that appears to be both unmediated and unrecorded. But this recording feature is no less important to Internet studies and user activity. Despite our willingness to study platforms, [6] algorithms, [7] and bots, [8] the server and its economically driven priorities remain arguably the most overlooked actors in studies of the web.

Surfacing the role of the server and its long-term memory is an important way to understand user interactions online, especially with respect to evasion techniques. Because the server is incentivized to track, trace, and pass data along for the purposes of advertising, social technologies take on a new valence as surveillance technologies. This means, on the one hand, that we must reach to a different literature to understand and explain interactions that occur in a tracked, traceable medium. Thus far, social media researchers have deployed interactional sociological principles -- such as Goffman's concept of “face work” [9]—to analyze interactions online. However, we might more productively reach for the vocabulary associated with regimes of visibility, tracking, and control, such as the literature on discipline, resistance, and system avoidance. [10]

On the other hand, it means that *no point-to-point interaction can be studied independently of its platform* as a contributing factor in users’ considerations of what to say—and not say—with respect to traceable media. We must therefore revisit the implications of studies that purport to analyze how people behave generally through datasets acquired online via third-party data sniffing. After all, these overheard “conversations” not only occur in a cultural context (i.e. of what it means to “do” social media interactions), but also in a context wherein individuals self-police their comments. [11] This policing is based not only on how they appear to their *friends* but how they hope to appear to the *server*.

This suggests productive and novel avenues for research. We might take a cue from danah boyd’s and boyd and Marwick’s study of teenagers’ Facebook use, where teens disguise their messages with cryptic prose to communicate only to people “in the know” but out in the open. [12] While boyd and Marwick use the example to argue for a networked notion of privacy, this technique can also reliably throw off culturally illiterate algorithms and stuff servers with irrelevant data. Most contemporary systems conceptualize privacy as person-to-person, with toggles enabling users to decide *with whom* to share their status or photos. But they do not yet designate *with what* we would like to share this data.

We might therefore ask: do users want their data to be visible *to the server*? Should this data be archived indefinitely, erased with a "sunset period" clause, or encrypted such not even the server and its algorithmic data detection bots can read, transfer, or analyze it? Currently some of these issues are addressed retrospectively as an issue of "the Right to Be Forgotten." [13] But this might be more productively construed as a way in which individuals can manage their relationships on the front end not only with each other, but with companies and their data servers as well. [14]

Resistance is Futile?

My attempts to purchase goods undetected purchase infant-related goods undetected by data tracking systems, for example by using cash or untraceable gift cards, revealed that the personal data dragnet is increasingly circumscribed by a network of actors who designate *traceable consumption* as "moral." This recalls work by Tarleton Gillespie on digital rights management. Gillespie describes how commercial technologies such as the iPod and iTunes, when first introduced, "make us moral" by inscribing users into a network of actors that includes Apple, the RIAA, and other agencies, each conspiring to make it so that every time the user deploys the iPod to purchase music, they do so in a way that is considered to be legitimate by those actors. [15] This also evokes early Latourian theory on how the seatbelt or the speed bump function to "make us moral" [16]: they inscribe us in a network of technologically-supported social assumptions of appropriate behavior, such that to use the technology is to participate in that regime, and therefore to be moral.

Similarly, the technologies of personal data capture interlace consumerism, technology, and citizenship. Under this framework, the implications of resistance or "opting out" become more severe. [17] Using Tor to evade consumption-based surveillance flags particular national security watch-lists; extensive withdrawals of cash from various ATMs can flag your debit card for consumer fraud; and signs at drugstores warn that excessive transactions involving gift cards can be reported to the authorities as potential instances of money laundering. Thus, avoiding data capture – like attempting to "de-script" [32] the seatbelt or the speed bump – can appear antisocial, immoral, or criminal. There is no evasion without repercussion.

This presents new questions for Internet researchers. First of all, it makes clear that "opting out" is not a true option, as it brings other social and even societal-level implications. After all, opting out of a system due to concerns about data privacy also implies opting out of any related systems, restricting individuals from civic, public, or commercial arenas. Like young urban youth caught once by the criminal justice system who attempt to avoid other linked public systems like hospitals and schools, [18] this "system avoidance" [19] can result in individuals resisting the use of other public infrastructures that bring other benefits. This produces a new implication for digital discrimination in addition to issues such as inclusion, exclusion, and disproportionate representation in datasets [20] or Internet literacy. [21]

Further, in many cases, it is vital to remain invisible. Much recent work has focused on the use of Facebook and Twitter in the context of protest, where many users in the field are already focusing their efforts on this data dragnet evasion as part of their resistance practices. In 2009 in Iran, Kevan Harris found that people tweeted dummy locations for protests in order to dupe law enforcement followers, opting for face-to-face whispers in crowds to spread the real locations for their assemblies. [22] In 2014 in Turkey, Zeynep Tufecki [23] found individuals circulating images of text instead of text itself on Twitter, under the assumption that the image could not be mined, textually analyzed, and aggregated quite so readily. Such resistance practices aim to circumvent the server's propensity for long-term memory or algorithmic detection. This brings to the fore the need to consider just how this reality of tracking via social/surveillance technologies is integrated into users' practices -- and to avoid assuming that tweets or messages recorded on a server represent ground truth. [24]

Data Balkanization and user privacy

Surfacing the server in my online interactions also produced a new way of thinking about where and how a user might strategically place their data. If complete evasion appears immoral and unnecessarily limiting (if not impossible in the long term), then how might these insights be applied to new data privacy opportunities?

One possibility is considering not simply the server in the singular, but servers in plural. After all, each company has their own hub that collects and stores user data. This data may travel from company to company, as a recent FTC report on data brokerage revealed, [25] but usually only does so along relational lines. [26] The problem for user privacy is not simply the fact that data may be transferred "down the line" outside of the trusting relationship between consumer and their company of choice. The problem lies in the potential for the aggregation and co-registration of multiple datasets through third party data acquisition. [27] If the power of big data is not in the single dataset, but in the ability to combine otherwise unrelated datasets to see new things, [28] then we might not fear the threat to privacy from the individual server so much as that from *connected* servers.

Taking networked servers and data transfer relationships seriously reveals a potentially new form of practical evasion: one that I call *data Balkanization*. [29] While data privacy scholars have attempted to evade collection in the first place or to confuse algorithms, [30] this approach to data privacy attempts to evade the repercussions of data combination by spreading traces across multiple servers that are disconnected from each other due to the status of the companies' relational ties. If the power of big data and ubiquitous computing lies in the combinability of datasets and the seamlessness of interactions, then obscurity lies in severing those ties and cherishing the seams. [31]

There are many such seams to be exploited. One involves the answer to the question: do the companies involved have a relationship with each other? In my own efforts to evade Google, I have resorted to using services such as Apple, Yahoo!, Bing, AIM, and Facebook. I know full well that these systems trace me on an individual basis. However, at present AOL and Microsoft, Apple and Google, Facebook and Yahoo! have limited, even hostile relationships with each other and are unlikely to share personal user data with their competition. Spreading traces across multiple servers *whose owners do not nurture relational ties* can increase user confidence at producing a kind of composite privacy through evading aggregation across platforms.

Data Balkanization flies in the face of contemporary systems' emphasis on convergence and ubiquitous computing. One-stop shops such as universal logins or the multiple services offer customer convenience, to be sure. And other points of convergence including the mobile phone or internet cookies provide cross-platform integration that make for a wondrous user experience. But they do so at the user's expense when they involve connecting otherwise disconnected datasets, threatening personal identification. Ultimately, a Balkanized approach to data privacy that spreads personal user data across multiple unrelated systems and relies upon the servers' disconnection to produce a modicum of user de-identification, may provide a way of de-scripting the network that does not imply overwhelming civil repercussions.

Future Thoughts

It is certainly true that for many people, evasion is the only option for protecting their privacy—and in some cases, their civil liberties as well. But evasion can also be a useful analytical tool in the service of infrastructural inversion. In this case, it revealed an otherwise invisible network of servers engaged in data storage and transfer that underlie the era of personalization algorithms. It exposed how those servers and the companies they represent entrap users in a web of technology and consumer activity that makes the collection of personal data through consumption "moral," while rendering attempts to evade detection criminally suspect. It also demonstrates how Internet researchers must take seriously how economically incentivized server activities are inserted into computer-mediated communication, and examine the repercussions for resistance. Finally, it presents strategies for user privacy not through evasion, but deploying knowledge of the network of servers to strategically place personal information in non-centralizeable or non-combinable locations. Each of these insights presents opportunities for further analysis, policy work, and design.

Notes

[1] http://www.ft.com/intl/cms/s/0/3cb056c6-d343-11e2-b3ff-00144feab7de.html

[2] Bowker, G. (1994). *Science on the run: Information management and industrial geophysics at Schlumberger, 1920–1940*. Cambridge, MA: MIT Press.

[3] For example, Vertesi, J. (2014). "How Not to Let the Internet Know You're Pregnant. Presentation at *Theorizing the Wrb* 2014, April 25, 2014. http://new.livestream.com/accounts/246408/events/2949381; http://time.com/83200/privacy-internet-big-data-opt-out/; and http://www.cnn.com/2012/10/07/opinion/vertesi-apple-maps/index.html respectively.

[4] Angwin, J. (2014). *Dragnet Nation*. A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance. New York: Times Books.

[5] Weiser, M. (1993). "Hot Topics: Ubiquitous Computing." *IEEE Computer* (October 1993): 71-72.

[6] See: Gillespie, T. (2010). "The Politics of Platforms." *New Media & Society* vol. 12 no. 3 347-364.

[7] For example: Barocas, S., Hood, S. & Malte Z. (2013). Governing algorithms: A provocation piece. *Governing Algorithms conference*, May 16-17, 2013; http://ssrn.com/abstract=2245322; Seaver, N. (2012). "Algorithmic Recommendations and Synaptic Functions." *Limn 2: Crowds and Clouds*. http://limn.it/algorithmic-recommendations-and-synaptic-functions/ ; Tufecki, Z. (2014a) "Engineering the Public: Big Data, Surveillance, and Computational Politics." *First Monday* 19(7).

[8] Geiger, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*. DOI: http://dx.doi.org/10.1080/1369118X.2013.873069

[9] See: Hancock, J., C. Toma, and N. Ellison. (2007). The truth about lying in online dating profiles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA: 449-452.

[10] Foucault, M. (1977). *Discipline and Punish: the Birth of the Prison*. New York: Random House.; Levy, K. (2014). "The Automation of Compliance: Techno-Legal Regulation in the U.S. Trucking Industry." Ph.D. thesis, Princeton University.; Brayne, S. (2014). "Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment." *American Sociological Review* June 2014 vol. 79 no. 3: 367-391.

[11] Many studies of online data sets treat this data as if it represented "naturally occurring" social interactions. While much can be gleaned from studying technologically-mediated interactions (Conversation Analysis was, for example, founded on analyzing recorded party line phone calls) there are limitations of technologically- and location-based conventions to consider. For other limitations on big data from social media, see Tufecki, Z. (2014b). "Big Questions for Social Media Big Data: Representativeness, Validity, and Other Methodological Pitfalls." In *ICWSM 14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*.

[12] boyd, d. (2013). *It's Complicated: The Social Lives of Networked Teens*. New Haven: Yale University Press.; boyd, d. and Crawford, K. (2011). "Six Provocations for Big Data." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431

[13] A European Commission law regulating post-hoc data retention and erasure upon request. See, for example, Mantelero 2013.

[14] Current work with Joseph Kaye, Samantha Jaroszewski, Vera Khovanskaya, and Jenna Song under a Yahoo! Faculty Research and Engagement Program grant is beginning to shed light on how users' conceptions of companies and their services play into their computer-mediated communications, alongside their individuals and networked management of interpersonal ties in data exchange. See Vertesi et al., "Data Narratives" (under review)

[15] Gillespie, T. (2007). *Wired Shut: Copyright and the Shape of Digital Culture.* Cambridge, MA: MIT Press.

[16] Latour, B. (1992) "Where are the Missing Masses? The Sociology of a Few Mundane Artifacts." In: Bijker and Law (Eds), *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp. 225-258). Cambridge, MA: MIT Press

[17] *Wyatt, S. (2003). "How Non-Users Matter." In: Pinch, T. and N. Oudshoorn, How Users Matter (pp. 67-80). Cambridge, MA: MIT Press.*; or, in ANT terms, network de-scription, see: Akrich, M. (1992). "The De-Scription of Technological Objects." In: Bijker and Law, Eds., *Shaping Technology/Building Society: Studies in Sociotechnical Change* (pp. 205-224). Cambridge, MA: MIT Press.

[18] Goffman, A. (2014). *On the Run: Fugitive Life in an American City*. Chicago: University of Chicago Press.

[19] Brayne, 2014.

[20] boyd, d. and Crawford, K. (2011). "Six Provocations for Big Data." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431

[21] Hargittai, E. (2003). The Digital Divide and What to Do About It. *New Economy Handbook*, Edited by D.C. Jones, San Diego, CA: Academic Press. 822-841. http://webuse.org/p/c04

[22] Harris, K. (2012). "The Brokered Exuberance Of The Middle Class: An Ethnographic Analysis Of Iran's 2009 Green Movement". *Mobilization: An International Quarterly* 17.4: 435-455.

[23] Tufecki, Z. (2014c) "Big Data to Ground Data." Presentation at *Theorizing the Web 2014*, Brooklyn, NY, April 25, 2014. http://new.livestream.com/accounts/246408/events/2949381

[24] Tufecki, 2014b

[25] http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

[26] I have discussed the relational approach to data transfer with respect to NASA missions (See: Vertesi, J. and Dourish, P. [2011]. "The Value of Data: Considering the Context of Production in Data Economies." In: *Proceedings of the ACM Conference on Computer-Supported Cooperative Work* [pp. 533-542]. New York: ACM Press). It applies to commercial transactions as well. If I have a relationship with a company (as a consumer, for example), then I give them my personal data in exchange for goods and services. Our data exchange is a part of our trusted relationship. They may, in turn, buy, sell, or exchange data with other companies with whom they have a relationship.

[27] "Third party clauses" in privacy agreements make it difficult to judge where user data is going. Sometimes companies include a clause about third party data transfer to indicate that their servers are hosted elsewhere, or that they contract out their phone service. In other cases, it means that they sell or share personal user data with companies that they have a relationship with, with whom the user has no relationship with at all.

[28] As we now know, data combinability presents devastating consequences for user privacy. Even given anonymized data, triangulating zip code, location, and call logs, for example, can result in a high accuracy of targeting unique users. With heightened personalization therefore comes heightened risks of identification -- or misidentification, which carries its own consequences.

[29] Also described in Vertesi et al, under review.

[30] In her study of data dragnets, Julia Angwin (2013) resorted to using creating fake identities and carrying her cell phone in a Faraday cage purse. Helen Nissenbaum and colleagues at NYU have recently built obfuscation technologies that aim to confuse the server with the collection of seemingly random associated user data. For example, a software script that clicks on *all*

displayed ads on a page can be a powerful way to cover one's tracks online. See http://obfuscationsymposium.org/

[31] Vertesi, J. (2014). "Seamful Spaces: Heterogeneous Infrastructures in Interaction." *Science, Technology, and Human Values*. March 2014 vol. 39 no. 2: 264-284.

[32] Akrich, M. (1992).