Big Data, Social Norms and Discrimination:
Lessons from The eGirls Project

Jane Bailey and Valerie Steeves

Introduction

We have prepared this report on "Big Data, Social Norms and Discrimination: Lessons from The eGirls Project" for The Open Society Foundation. The findings reported upon relate to data collected in January and February of 2013 when researchers with The eGirls Project held a series of interviews and focus groups with girls and young women between the ages of 15 and 22. All participants used interactive online media (such as social networking, blogging and/or user generated video sites) as a regular part of their social lives. Half of our sample resided in an urban Ontario setting and half resided in a rural Ontario setting.

We interviewed six girls aged 15-17 and six young women aged 18-22, for 60-90 minutes each. An additional 22 participated in four focus group discussions, as follows: (1) seven girls aged 15-17 living in the urban setting; (2) five girls aged 15-17 living in the rural setting; (3) six young women aged 18-22 living in the urban setting and (4) four young women aged 18-22 living in the rural setting. Focus group discussions were approximately 90 minutes in length. A professional research house recruited our participants on the basis of sex, age (either 15-17 or 18-22) and location of residence (urban or rural). While participants were not recruited on the basis of self-identification with regard to other aspects of their identities, such as race, ethnicity, gender identity or sexual orientation, our participant group included members of racialized, linguistic, and various religious groups.

In the interviews and the focus groups, we explored, among other things, the types of visual and textual representations the participants used online to express their identity as young women, and the benefits and pitfalls they experience on social media. We also asked for their views on the issues and policy responses focused upon by policymakers and explored their understandings of networked privacy and equality. With participant permission, the interviews and focus groups were audiotaped and transcribed by our research assistants for analysis. The transcripts were then subjected to a thematic qualitative analysis. All identifying information was removed from the transcripts, and pseudonyms are used below to identify participants (Bailey, 2015: 26).

Harms caused by Big Data profiling as experienced by girls and young women on social media

Big Data and the shaping of girls' identity

Young people have embedded networked technologies seamlessly into their social lives, using social media to explore their identities, deepen their connection with friends and family and explore their interests (Steeves, 2005). Girls in particular are likely to use social media for communication and identity play: although Canadian girls and boys aged 13-17 are equally likely to have a Facebook account, girls are significantly more likely than boys to have an account on other social media sites, including Twitter (53% compared to 41% of boys), Instagram (55 percent compared to 32% of boys),

Tumblr (41% compared to 16% of boys) and Pinterest (22% compared to 4% of boys) (cite). Younger girls aged 11-12 are also more likely to have accounts on branded play sites that incorporate elements of social media (and therefore a higher level of informational disclosure) into their play (e.g. 48% to 22% on Webkinz, 44% to 22% on Moshimonsters, and 31% to 27% on Poptropica). Boys also tend to provide less information about themselves on social media and many lie about what they do post; girls, on the other hand, tend to post more and tell the truth (Steeves, 2014). These factors combine to make girls the ideal target for online marketers because they are more likely to reveal information about themselves on networked media as they go about their daily lives.

Moreover, almost all of the online spaces young people prefer to inhabit (49 of the top 50 favourites identified in 2013) are structured by seamless commercial surveillance that collects information explicitly and in the background for the purposes of creating detailed individualized profiles (Steeves, forthcoming). The idea is not just to advertise to young people, but to shape their identities to make them more susceptible to marketing messages (Steeves, forthcoming; Montgomery, 2015).

eGirls participants indicated that these commercial messages are rampant on social media and unanimously felt that the ubiquitous presence of diet ads, weight loss tips, and other "beauty aids," combined with content posted by models and clothing companies, narrowed the kinds of girls they could be online.

> Barbie, that's pretty, that's the perfect example that everyone uses. So like Barbie, top models, and everything, we all see — we always see those kind of [people], they're all amazing, . . . it's mostly really, really awesome people and, like, they're really pretty and really like skinny and everything, they're perfect (Emily, age 15).

These messages replicate a narrow representation of girls as sexual objects: "[They] have a man, who is perhaps fully clothed or maybe has his shirt off, he's rapping and then next to him are women in bikinis. OK. So the women are just objects, they're just complementary, he's the centre focal point and the women are just ornaments around him… [The message is] … That your whole, that being a woman is about how well you can please guys. … You know, so I'm thinking that, OK, to be a good woman I need to know how to do all these disgusting acts, I need to know how to be beautiful, I need to know how to lose weight, that's a big important one, if you're not skinny then no one is going to love you, that's what every magazine is about, 'oh she gained ten pounds'" (Allessandra, age 21). Monica (age 16) talked about the narrow range of the kinds of femininity she saw online: "Well, magazines and stuff, it's like weight loss is the whole idea of 'get into your bikini bod by the summer'. That's all they support. They don't support anything else." Cassandra (age 19) argued that "everything in ads is more directed towards girls," to encourage them to buy products to look like "all those beautiful women who have all these professional people doing their hair."

> 'If I get this, I'll look like Halle Berry.' And you get this, you're like, 'Oh my God, I'm not looking like Halle Berry.' So you're trying everything… So I don't know, girls are just … I don't know … just have to look good… It's just the way we work, I guess (Cassandra, age 19).

Moreover, the inability to replicate these images in their own lives left them feeling badly about themselves and reduced their confidence in their ability to enjoy networked publicity. As Cindy said, "you're like, oh man, I don't look like that. Um, but I could someday, you know, but

you just, you don't right now. So you might get down on yourself because of that."

Big Data and the magnification of discrimination

Although the pressure they described is clearly linked to persistent and pre-existing media stereotyping, our participants told us the impact of these stereotypes is magnified by social media, in three ways. First, the presence on social media of so many girls who sought to emulate this image increased the pressure to conform to the stereotype. Second, photoshopping and other technical affordances meant that girls could present a better-than-real picture of themselves. Even when our participants would talk about photos of other girls they knew were photoshopped, they were still mesmerized by the "perfection" of the performance itself. " . . . [There are] girls on Facebook . . . they'll have like five hundred likes on some of their pictures and . . . I'll sit there and like notice it at first and be, like, this person has to be, like, oh my God, they are so flawless." Third, the constant publicity encouraged on social media opens girls up to harsh judgment from peers. Girls were criticized both for failing to perform to the norm — Kiera (age 21) recounted a story of a girl in high school who was "just bashed" by a boy on Facebook because "She had a very authentic look, and she was never really scared to say what she wants or act in any way that she wants" — or for performing too well and crossing "the slut line".

> The presence of "more girls everywhere . . . trying to put, like, the prettiest girls on magazines and stuff" (Lynda) "I don't know, sometimes, it'll make you feel like crap. It's like, just again setting in, why can't I look like that? Why can't I be like that? Why don't I have these friends? Why am I not popular? And just drains everybody else" (Monica, age 16).

The commercial model behind social media plays into each of these effects. The environment is shaped to elicit ongoing disclosure of personal and non-personal information to create what Mosco (2009) calls the "immanent commodification" of young people's social interactions. The resulting data stream is used to craft commercial messages to make them more powerful and compelling, and to insert them into the social interaction itself. This creates an "ongoing interactive relationship — cognitive, emotional, and behavioural/physical — with brands" which is then constantly monitored so commercial messages can be "increasingly directed at individual consumers instead of demographic groups" (Montgomery, 2015).

> I think social media is great at giving girls this fantasy world but at the same time I think it's also really easy to sort of make them feel really bad about themselves (Cindy, age 20).

This significantly constrains girls' and young women's equal participation in networked society because it privileges those performances that can successfully emulate the stereotypes contained in mainstream marketing. Moreover, the success of those performances is immediately quantifiable through the number of "likes" they do (or do not) attract. Because the ongoing data stream captures these interactions, the social world becomes not just an environment for advertising, but a disciplinary mechanism that rewards those who reproduce the messages that have artificially been inserted into the environment itself.

Accordingly, the Big Data model amplifies discrimination by "[intensifying] girls' interactions with media representations and [restructuring] the environment in ways that privilege heteronormative performances of girl… [Moreover], the visual nature of social media alienates the feminine body through the hyper-visibility of the *image* of the body; this makes the body an object of judgment that is subject to scrutiny by others and the self, and exacerbates the negative effects of failed performances" (Steeves, and Bailey, in press).

Moreover, the need to pursue external validation, particularly from male peers, set girls up for conflict. Jill explained it this way: "A girl, let's say she's, I don't know, with a bunch of guys in a sexual pose, or . . . has tons of booze around her, or something. Someone will write a comment that will be, like, kind of subtle but showing that it's inappropriate, and a lot of people will join in, and you can get, like, up to seventy-five comments and everyone's joining in and fighting." The competitive nature of the networked environment problematized the category "girl" for our participants, who described "other girls" as mean, cruel, bitchy, complaining, sluts, and show-offs, and made the creation and inhabiting of strong, independent feminine identities in online spaces incredibly difficult to achieve.

A rights-based framing of the issues discussed in The eGirls project

The issues raised by the big data environment should be analyzed through a rights-based lens because big data practices interact with other environmental factors (such as social norms and practices) to affect various fundamental human rights and democratic values (Bailey, 2015; Steeves, 2007). Rights affected include not only privacy, access to information, the child's right to development of the personality and the right to participate fully in cultural and artistic life (Steeves, 2007), dignity, and personhood (Schneier, 2015), but also equality. Equality is affected because members of vulnerable communities appear to be disproportionately negatively affected by a lack of privacy and by the stereotypes and discriminatory prejudices (Steeves and Bailey, in press) that themselves can be coded into and/or produced and reproduced through algorithms in a big data environment (Gandy, 1997; Sweeney, 2014). Moreover, because these fundamental rights are at stake, it is essential to insist that the diverse array of persons affected (including youth) have a right to participate in the process of developing and implementing policy relating to big data practices (Bailey, 2015).

Policy recommendations for actors working on human rights issues related to big data, as uncovered by The eGirls project

Do not assume uniformity in terms of the experiences and effects of big data across users. For example, eGirls Project participants understood that the impacts of a big data environment affected girls differently from boys and that the seriousness of the impact varied with age (Steeves, 2015). Expecting differences across users based on their situated life experiences makes it especially important to respect international obligations to create policy processes that enable participation by diverse groups of people from a range of genders, races, ages, ethnicities, sexual identities and so forth (Bailey, 2015). Understanding differences in experiences and needs enables the development of policy approaches grounded in the situated knowledges of diverse community members (locally, nationally and globally).

Do not focus on a single "problem" or a single "fix". Instead, use the knowledge gained from diverse community participation to understand what the problems and benefits associated with living in a big data environment are, how community members prioritize associated problems and benefits, and community members' own strategies and proposed approaches for addressing them. Consider how the interaction of other social processes work to shape those diverse perspectives (Choo, 2010).

Do not assume that adults' understandings of the experiences of youth reflect youths' own experiences. For example, Canadian federal public policy dialogue around children and technology has placed significant emphasis on the risk of unknown sexual predators online (Bailey and Steeves, 2013). eGirls Project participants indicated some concern about unknown sexual predators online (especially with respect to their younger siblings and relations). However, they demonstrated far more concern about the impact of the widespread availability and scrutiny of data relating to them and the ways in which the online environment exposed them to the risk of reputational ruin (Bailey, 2015).

Approach questions relating to big data from a broader environmental perspective. Recognize both the range of factors and players involved, as well as how the interactions between them affect the impact of big data. For example, eGirls Project participants' experiences suggested a complex interplay between a market model driven to compel data disclosure, social norms that made it easier to attain success through repetition of mediatized stereotypes (up to a difficult-to-define point where the repetition went too far) and technical architectures that made control over data difficult and coded social success quantitatively through "likes", thereby incenting further disclosure (Bailey, 2015; Steeves, 2015; Bailey, Steeves, Burkell and Regan, 2013).

Publicly reveal how big data and its analytics work to allow community members to better understand how their data is being collected, stored, distributed and mined. The objectives of making this information accessible may lead community members to modify their behaviour accordingly, but more importantly can empower them to pressure for change in, and potentially regulation of, corporate and government behaviour (Bailey and Steeves, 2015).

Proactively address the underlying roots of discriminatory behaviours produced in the big data environment. Too often policy approaches focus on reactive responses aimed at penalizing individual users, blaming those targeted for attack for having disclosed too much and (in the case of youth) subjecting targets to further monitoring and surveillance by parents and other adults. For example, eGirls Project participants suggested that girls and young women were more likely to be negatively judged for their self-representations online, especially if they were sexualized (Steeves, 2015). They suggested a need to intervene on heterosexist stereotyping that privileged thin, white mediatized representations of femininity that were a prominent part of advertising they were targeted with in online social spaces (Bailey, 2015).

Focus on the role that corporations play in the way they structure online interactions to compel data disclosure and make privacy protection difficult. For example, eGirls Project participants noted that certain online social networking platforms demanded unnecessary information (such as birth dates), that apps sometimes demanded information only after allowing the user to get hooked on them, that user agreements and the technicalities of privacy settings often made it difficult for them to proactively

protect their data even where they wanted to, and that some platforms automatically integrated postings from other platforms, making it difficult for them to play different roles for different audiences (Bailey, 2015; Steeves, 2015).

Suggested additional research on the human rights implications of big data

Young people represent an important demographic for understanding the implications of big data, both because they are early adopters of online media, and also because they often unknowingly disclose considerable quantities of data about themselves in the course of their fully integrated online/offline existences. In this way, their perspectives and experiences offer an ideal window onto behavioural targeting practices and impacts. As such, research should be carried out to further explore these practices and impacts from the perspectives and experiences of a diverse array of young people, as well as to explore young people's understandings of privacy and equality in online environments.

Research should aim to go to the root of behavioural targeting practices, and to allow for better understanding of their logistics and analytics, as well as their impacts. To the degree possible, the findings from that research should be disseminated in a variety of publicly accessible formats. Moreover, researchers should explore new models for protecting privacy beyond the data protection model premised on individual consent. That model is rapidly becoming outdated and ineffective because the analytics of the algorithmic sort make it impossible to consent in an informed way to ongoing and complex systems of collection and use of data on a transaction-by-transaction basis.

Once the behavioural targeting of youth and its impacts on their lived experiences are better understood, researchers should investigate the intersection between big data practices and currently identified social problems such as "cyberbullying" and online harassment. Researchers should seek to better understand how the big data model interacts with other factors to shape an online environment ripe for attack and abuse. Within this stream of analysis, researchers should explore further the relationship between privacy and equality in a big data environment by investigating whether certain communities are disproportionately targeted and/or experience disproportionately harmful impacts. Achieving a better understanding of these interactions would allow for development of meaningful policy and educational responses, including responses aimed at industry and its practices, as well as responses aimed at discriminatory root causes of targeting.

Recommended readings about the rights implications of data profiling (top 5 highlighted)

Angwin, J. (2014). It's complicated: Facebook's history of tracking you. ProPublica. Available at: http://www.propublica.org/article/its-complicated-facebooks-history-of-tracking-you.
Austin, J. and Reed, M. (1999). Targeting children online: Internet advertising ethics. Journal of Consumer Marketing, 16(6), 590-602.
Bailey, J. (2015). A Perfect Storm: How the Online Environment, Social Norms and Law Constrain Girls' Online Lives. In Jane Bailey and Valerie Steeves, eds. eGirls, eCitizens, Ottawa: University of Ottawa Press, 21-54.

Bailey, J. and Valerie Steeves. (2013). Will the Real Digital Girl Please Stand Up? In Hille Koskela and Macgregor Wise eds. New visualities, new technologies: The new ecstasy of communication. Ashgate Publishing.

Bailey, J. and Steeves, V. (2015) (Eds.), eGirls, eCitizens. Ottawa: University of Ottawa Press.

Bailey, J., Steeves, V., Burkell, J. and Regan, P. (2013). Negotiating with gender stereotypes on social networking sites: From 'bicycle face' to Facebook. Journal of Communication Inquiry, 37(2), 91-112.

Bamford, J. (2013). They know much more than you think. The New York Review of Books. Available at: http://www.nybooks.com/articles/archives/2013/aug/15/nsa-they-know-much-more-you-think/?pagination=false.

Boellstorff, T. (2013). Making big data, in theory. First Monday, 18(10).

boyd, D. and Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological and scholarly phenomenon. Information, Communication and Society, 15(12), 662-679.

Bryant, A. and Raja, U. (2014). In the realm of big data… First Monday, 19(2).

Cai, X. and Zhao, X. (2013). Online advertising on popular children's websites: Structural features and privacy issues. Computers in Human Behaviour, 29(4), 1510-1518.

Capella, M. and Terlutter, R. (2013). The gamification of advertising: Analysis and research directions of in-game advertising, advergames, and advertising in social network games. Journal of Advertising, 42(2/3), 95-112.

Chester, J. (2007). Digital destiny: New media and the future of democracy. New York: The New Press.

Choo, H.Y. and Myra Marx Ferree. (2010) Practicing Intersectionality in Sociological Research: A Critical Analysis of Inclusions, Interactions, and Institutions in the Study of Inequalities Sociological Theory 28:2, 129-149.

Craig, T. and Ludloff, M. (2011). Privacy and big data. Sebastopol, CA: O'Reilly Media.

Crawford, K., Miltner, K. and Gray, M. (2014). Critiquing big data: Politics, ethics, epistemology. International Journal of Communication, 8, Special Section Introduction, 1663-1672.

Crawford, K. and Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. Boston College Law Review, 55(1).

Deibert, R. (2013). Big data: They reap what we sow. In Deibert, R. (Ed.), Black code: Inside the battle for cyberspace (50-68). Toronto: McClelland & Stewart.

Gandy Jr., O. (1993). The panoptic sort: A political economy of personal information. Boulder, CO: Westview Press.

Gandy Jr., O. et al. (1997). Race and risk: Factors affecting the framing of stories about inequality, discrimination, and just plain bad luck. Public Opinion Quarterly, 61(1), 158-182.

Gandy Jr., O. (2009). Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage. Burlington, VT: Ashgate Press.

Gillespie, T. (2014). The relevance of algorithms. In Gillespie, T., Boczkowski, P. and Foot, K. (Eds.), Media technologies: Essays on communication, materiality and society (167-194). Cambridge, MA: MIT Press.

Gillespie, T. (2010). The politics of 'platforms'. New Media & Society, 12(3), 347-364.

Kervin, L., Jones, S. and Mantei, J. (2012). Online advertising: Examining the content and messages within websites targeted at children. E-Learning and Digital Media, 9(1), 23-37.

Lesk, M. (2013). Big data, big brother, big money. IEEE Security and Privacy, 11(4), 85-89.

Loshin, D. (2013). Big data analytics: From strategic planning to enterprise integration. Saint Louis, MO: Morgan Kaufmann.

Lyon, D. (2014 forthcoming). Surveillance in the era of big data: Capacities, consequences, critique. Big Data and Society. 1(1).

Lyon, D. (2003). Surveillance as social sorting. London: Routledge.

Manovich, L. (2011). Trending: The promises and challenges of big social data. In Gold, M. (Ed.), Debates in the digital humanities (460-475). Minneapolis: University of Minnesota Press.

Marwick, A. (2013). Big data, data-mining and the social web. Governments, corporations and hackers: the Internet and threats to the privacy and dignity of the citizen. Power, privacy and the Internet, New York Review of Books Conference, Scandinavia House, New York: NY. 30-31 October, 2013.

Montgomery, Kathryn. (2015). Youth and Surveillance in the Facebook Era: Policy Interventions and Social Implications. Telecom Policy.

Mosco, V. (2014). Big data and cloud culture. In Mosco, V. (Ed.), To the cloud: Big data in a turbulent world (175-226). Boulder, CO: Paradigm Publishers.

Mosco, V. (2009). The Political Economy of Communication, 2nd ed. Thousand Oaks, CA: Sage.

Oboler, A., Welsh, K. and Cruz, L. (2012). The danger of big data: Social media as computational social science. First Monday, 17(7).

Pasquale, F. The Black Box Society: The Secret Algorithms That Control Money and Information. Boston: Harvard University Press.

Schneier, B. (2015). Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World. New York: W.W. Norton Company Inc.

Shade, L. and Shepherd, T. (2014). Tracing and tracking Canadian privacy discourses: The audience as commodity. In Kozolanka, K. (Ed.), Publicity and the Canadian state: Critical communications perspectives (215-234). Toronto: University of Toronto Press.

Steeves, V. (Forthcoming.) Terra Cognita: Surveillance on Young People's Favourite Websites. In Emmeline Taylor and Tonya Rooney, eds., Surveillance and Childhood: Understanding the ethical and social implications of new and emerging technologies. London: Ashgate.

Steeves, V. (2015). "Pretty and Just a Little Bit Sexy I Guess": Publicity, Privacy and the Pressure to Perform "Appropriate" Femininity on Social Media. In Jane Bailey and Valerie Steeves, eds. eGirls, eCitizens, Ottawa: University of Ottawa Press, 153-174.

Steeves, V. (2014). Young Canadians in a Wired World, Phase III: Life Online. Ottawa: MediaSmarts.

Steeves, V. (2014). Young Canadians in a Wired World, Phase III: Online Privacy, Online Publicity. Ottawa: Media Awareness Network.

Steeves, V. (2012). Hide and seek: Surveillance of young people on the Internet. In Lyon, D., Haggerty, K. and Ball, K. (Eds.), The international handbook of surveillance studies (352-392). London: Routledge.

Steeves, V. (2009). Data protection versus privacy: Lessons from Facebook's beacon. In Matheson, D. (Ed.), The contours of privacy (183-196). Newcastle Upon Tyne, UK: Cambridge Scholars Press.

Steeves, V. (2007). The Watched Chilf: Surveillance in Three Online Playgrounds. Proceedings of the International Conference on the Rights of the Child. Montreal: Wilson Lafleur. El niño observado: vigilancia en tres sitios de juegos de niños en Internet. Actas de la Conferencia Internacional sobre los Derechos del Niño. Buenos Aires: Universidad de Buenos Aires.

Steeves, V. (2005). Young Canadians in a Wired World — Phase II Trends and Recommendations. Ottawa: Media Awareness Network.

Steeves, V. and Jane Bailey. (In press) Living in the Mirror: Understanding Young Women's Experiences with Online Social Networking. In Emily Van De Muelen, ed., Expanding the Gaze: Gender, Public Space and Surveillance, Toronto: University of Toronto Press.

Sweeney, L. (2013). Discrimination in online ad delivery. Communications of the ACM, 56(5), 44-54.

Taylor, C. and Peter, T. (2013). Every class in every school: Final report on the first national climate survey on homophobia, biphobia, and transphobia in Canadian schools. Toronto: Egale CHRT.

Tene, O. and Polonetsky, J. (2012). Privacy in the age of big data: A time for big decisions. Stanford Law Review, 64, 63-69.

Tien, J. (2013). Big data: Unleashing information. Journal of Systems Science and Systems Engineering, 22(2), 127-151.

Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. First Monday, 19(7).

Turow, J. (2011). The daily you: How the new advertising industry is defining your identity and your net worth. New Haven, CT: Yale University Press.

Vis, F. (2014). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. First Monday, 18(10).