115 | feminist data studies: using digital methods for ethical, reflexive and situated socio-cultural research

Koen Leurs

abstract

What could a social-justice oriented, feminist data studies look like? The current datalogical turn foregrounds the digital datafication of everyday life, increasing algorithmic processing and data as an emergent regime of power/knowledge. Scholars celebrate the politics of big data knowledge production for its omnipotent objectivity or dismiss it outright as data fundamentalism that may lead to methodological genocide. In this feminist and postcolonial intervention into gender-, race- and geography-blind 'big data' ideologies, I call for ethical, anti-oppressive digital data-driven research in the social sciences and humanities. I argue that a reflexive data scholarship can emerge from the reintegration of feminist and postcolonial science studies and ethics of care ideals. Although it is not a panacea for all ails of data mining, I offer a road map for an alternative data-analysis practice that is more power-sensitive and accountable. By incorporating a people-centric and context-aware perspective that acknowledges relationships of dependency, reflects on temptations, and scrutinises benefits and harm, an 'asymmetrically reciprocal' (Young, 1997) research encounter may be achieved. I bring this perspective to bear on experiences of a two-year research project with eighty-four young Londoners on digital identities and living in a highly diverse city. I align awareness of uneven relations of power and knowledge with the messy relation of dependency between human and non-human actors in data analysis. This framework productively recognises that digital data cannot be expected to speak for itself, that data do not emerge from a vacuum, and that isolated data patterns cannot be the end-goal of a situated and reflexive research endeavor. Data-driven research, in turn, shows the urgency for renewed feminist ethical reflection on how digital mediation impacts upon responsibility, intersectional power relations, human subjectivity and the autonomy of research participants over their own data.

keywords

introduction

Digital data sets are never innocent, complete or self-explanatory. In this intervention into 'big data' ideologies, I call for ethical, anti-oppressive, data-driven research in the social sciences and humanities. In order to challenge the attitude of an omnipotent objectivity promulgated by some big data enthusiasts, I argue that social-justice oriented data scholars can respond by reintegrating feminist and postcolonial science studies and ethics of care into their analytical frames.

This reflective article on data-driven ethnographic research is based on a two-year qualitative study with eighty-four young Londoners about their digital identities. My study asked how diverse young Londoners identify with one another across their differences, using digital technologies. With this question in mind, I conducted fieldwork in Tottenham and other diverse London boroughs. I focussed on Tottenham because that is where, in 2011, a peaceful protest following the police shooting of Mark Duggan, a local mixed-race young man, escalated into the so-called 'BlackBerry Messenger riots'. Headlines such as 'London riots: BlackBerry Messenger and Twitter used as weapons' (Magee, 2011) and 'Over 150 arrested after London hit by huge race riots' (Taylor, 2011) are indicative of scholarly, media and governmental reports arguing that race relations and social media fuelled outbreaks of public order offences.

Much has been said about the lives and technology use of young people; however, their own voices are largely absent from these discussions. For this purpose I conducted face-to-face interviews and collaboratively made visualisations of Facebook friendship networks with young Londoners. Being confronted with homogeneous or heterogeneous networks on their screens, these visualisations triggered reflections about living in a diverse city. Through the methods used, young people involved in the study not only decided for themselves what data they wanted to share, but more importantly, they also directed the course of the study as they became co-researchers of their own digital practices.

I am certainly not claiming any novelty; like all research involving people, digital data-driven research involves complex power relations and mixed emotions in raising questions of ethics, voice and accountability. For example, on a dreary Wednesday evening early in 2014, I met with long-time buddies Danny[1] (14 years old) and Charlie (16 years old) in a local youth centre in Tottenham, North London. Danny described himself as 'loud', 'Christian' and 'born here'. Charlie is 'energetic' and described himself as 'Christian' and noted that his dad was 'born in Scotland, my mum was born in Wales'. We sat down in a small office otherwise used by youth workers. In the room next door, young people from the area were producing a radio show. Throughout the building, hip-hop, soul, Afrobeat and K-pop were blaring from large loudspeakers. The two friends wore sweaty tracksuits, as they had just returned from the basketball court, and were sharing a bag of chips drenched in ketchup. The atmosphere was relaxed. Both boys sang along to the familiar tunes they heard. The conversation was mostly about their families, their daily

---

[1]Research participants are introduced using the pseudonyms and self-identifications they shared.

school lives and their uses of digital media. In particular they expressed their love for athletics and playing video games. Suddenly, Charlie spoke in a more serious tone. Somehow triggered by the friendship visualisation showing on his screen, Charlie felt it was more important to share his highly ambiguous relationship with social media platforms like ASKfm and Facebook:

> For me the worst thing is Facebook, cos there is so many people dying over Facebook, because the social networks with their sexuality and gayness, if they are bi-sexual, so many people are dying over it but still so many people still use it. On ASKfm, a girl committed suicide, cos of cyberbullying, cos on her website people were writing comments anon yeah anonymously anonymous questions saying 'are you a Lesbian', 'Would you die' and anything. And she just got too much and she killed herself. I was back in, think about when I was in their situation, which I was in year 7, if I would be one of those people saying I'm gonna, I'm going to kill myself, I think I would do it, because I'm not the person to hold back. But luckily, if I didn't get help to control it, to ignore the people doing it, to this day now, without any support from my family, the teachers and that, I don't think I would be here now, to be honest. (Charlie, 16 years old)

Trying to fight back tears, Charlie, who had not yet explicitly talked about his sexuality or romantic relationships, elaborated how bullying on the basis of sexuality had impacted his life, as well as his sisters' lives. Danny tried to comfort Charlie by saying, 'don't worry man. I know you are an emotional person, man'. I felt responsible for bringing up this traumatic experience, and I started to worry about whether I was doing Charlie any harm or whether he experienced the exchange as something positive. Above all, I also thought of ways to comfort him. Then his phone rang, and after speaking to his mum he felt better and moved on as if nothing had happened.

This encounter illustrates how friendship network visualisations as standalone digital data, removed from their context, can tell a completely different story in comparison with how users make meaning of their digital data themselves. Data may be aggregated to find patterns; however, aggregate data is still connected to embodied experience, even though there is the claim that it is removed from identity and personal meaning-making (see, for example, Pink *et al.*, 2016; Zook *et al.*, 2017). Returning to my desk after fieldwork, I was eager to make sense of the rich data that young people had shared with me, but I also felt that I should consider this data in the context of the many painful and joyful experiences that this fieldwork had produced. I re-immersed myself in the digital humanities and computational social science literature on digital methods and data research; however, I quickly felt disoriented. Scholars either worry that 'big data' research lacks attention to history, culture, context, structure or agency, leading to 'methodological genocide' (Uprichard, 2015, n.p.), or they praise objective, disembodied and value-free insights that can be gleaned from easily accessible data as '"naturally" created by Internet users' (Mahrt and Scharkow, 2013, p. 21). This polarised debate offers little means to acknowledge the challenges, power relations and opportunities involved in doing a digital, data-driven socio-cultural study.

Developing a research participant-centred approach, I will argue, is one way to better attend to the epistemological, methodological and ethical pros and cons of data-driven research. Data-driven research revolves around three dynamics: 1) the automatic generation of data sets resulting from increased digitisation of everyday practices such as the logging of interactions and transactions including social media use, navigation, banking, shopping and transportation, biometrics, facial

recognition and surveillance; 2) the processing of this data through algorithms; and 3) operationalising data as a new regime of power/knowledge.

Dominant big data discourses reflect positivist, transcendental empiricism and disembodiment value-freeness. In data analysis, these ideologies tend to produce limited, masculine, able-bodied, heteronormative, middle-class and Western-centric representations of subjectivity and identity. Digital methods can be used to 'diagnose cultural change' (Rogers, 2013, p. 5), but they also yield large profits for transnational corporations. They are discriminatory and complicit with state and military surveillance. Consider, for example, 'digital deportability' (Tsianos and Kuster, 2012) enabled through 'augmented borders' (Ajana, 2015): in Australia the Border Risk Identification System (BRIS) identifies 'high-risk' travellers, while in thirty-one EU countries, the European Dactyloscopy (EURODAC) biometric database system is used to distinguish between desired travellers and undesired aliens through the algorithmic social sorting of digitised fingerprints. These are only some of the reasons why 'the matter of social media software [and datafication] is ... ripe for feminist analysis' (Bivens, 2015a, p. 715). Therefore, instead of seeing data-mining as neutral and self-explanatory, in this article I argue data-driven knowledge production practices are inherently subjective, power-ridden and context-specific and, above all, only produce partial truths. Recognising and reflecting upon the contingencies of digital method techniques and the epistemologies of digital methodologies will enable feminist data scholars across the humanities and social sciences to produce more robust and meaningful stories rather than universal truths or disembodied generalisations.

Below, I draw on feminist and postcolonial science studies to explore urgent ethical questions including, what is required of a self-reflexive digital data researcher? How can we draw on user-generated data to understand agency vis-à-vis structures of individuality and collectives across intersecting axes of difference? How can we strategically mobilise digital methods in a non-exploitative way to illuminate everyday power struggles, agency and meaning-making? My argument is structured as follows: First, on the level of epistemology, I suggest we can navigate between the dreams and nightmares projected on digital data-driven research by considering texts, users and materiality from a relational perspective. Like other branches of media and communication scholarship, feminist, postcolonial and critical race internet researchers commonly single out either texts, users or materiality in their analyses. Second, and subsequently, realising that no methodology is distinctively feminist or postcolonial and many methodologies can be engaged with in a reflexive way, I translate feminist and postcolonial ethics of care ideals to practically negotiate the contradictory and competing demands involved in the different phases of a data-driven research cycle. Third, I ground my road map for an alternative data-analysis practice in empirical data from researching digital identities with young Londoners.

aligning feminist internet studies with digital data-driven research

'Promise and peril: the Internet has brought both to women'; this was the apt diagnosis Morahan-Martin (2000, p. 683) reached nearly two decades ago. 'Big data' research, although clouded in both hyperbolic and polemic imaginaries of the inherent value of abundantly available data, needs to be similarly

assessed as simultaneously empowering and excluding, opening up new possibilities but also amplifying previously existing hierarchies for women and minorities. 'Big data' is a container term to signify large digital data sets, where definitions typically revolve around their distinctive 'volume', 'velocity', 'exhaustiveness', 'fine-grained resolution', 'relationality' and 'scaleability' (Kitchin, 2014, pp. 1–2). The total mass of digital data produced in 2013 and 2014 exceeded that previously recorded in human history (Andrejevic, Hearn and Kennedy, 2015). This includes data social media users volunteer through their free digital labour, directed data from surveillance systems, and data traces automatically produced in transportation, transactions and the use of digital devices (Fuchs, 2014; Kitchin, 2014).

Few scholars would contest that with the emergence of 'big data' and algorithmic computation, research practices have changed; it is no longer methods-as-usual. While sociologists argue data digitalisation demands a 're-configuration of expertise' (Savage, 2013, p. 8), anthropologists call for 'networked anthropology' and 'networked ethics' (Collins and Slover Durington, 2015), and researchers working in areas such as science, medicine, economics, media, communications and the digital humanities particularly display a celebratory 'forward-looking sentiment' (Svensson, 2012; Koh, 2015; Arora, 2016). Media researchers, for example, happily gather social media data as they provide unique 'archives of the everyday' (Beer and Burrows, 2013, p. 54). This data-rush mentality replicates dominant neo-liberal ideologies that equate technological advancements with social betterment.

Yet there is a growing critical perspective on the 'big data' bandwagon, especially coming from software studies and critical data scholars who are opening black-boxed processes of datafied research (Schäfer and van Es, 2017). While such scholars have critiqued big data 'to question how we think, see, and ultimately act on our knowledge' (Manovich, 2017), to think through the 'political consequences' of 'computational logics' (Berry, 2014, p. 40) and to demand 'de-reification' (Couldry, 2017), these criticisms are neither especially feminist nor postcolonial, as they only marginally address (if at all) the specific people to whom their claims apply. Thus, while researchers may agree that 'big data' ideologies erode situated understandings of 'research subjectivity' (Metcalf and Crawford, 2016), these general critiques do not address intersectional power relations such as gender, race, social class, age or location (Risam, 2015).

Feminist and postcolonial science studies are an obvious reference point to offer a much-needed corrective. Across disciplines, this literature has offered inspiration in demonstrating how scholarly inquiry is itself socio-culturally constructed and situated. The field also offers guidelines for more anti-oppressive ways to carry out digital data-driven research. Unfortunately, feminist and postcolonial scholarship and data studies remain as largely unconnected fields with little cross-fertilisation. The increasingly popular tool-, technique- and mining-driven approaches of the digital humanities and computational social science tend to replicate exclusionary modes of scientific inquiry. Feminist and postcolonial epistemologies can make us more aware of how knowledge is partial, situated and contextual. Conventional empiricist knowledge production is predicated on the 'the god-trick of seeing everything from nowhere' (Haraway, 1991, p. 189), meaning the researcher operates as disembodied, rational master subject with transcendental vision. He assumes detachment from a discrete, knowable world. Mirroring transcendental empiricist approaches, the 'data-rush' mentality (Mahrt and Scharkow, 2013, p. 20) in the digital humanities and computational social science tends to naturalise the politics of

knowledge production. This mentality—by uncritically replicating empiricist, apolitical stances—threatens to marginalise advancements that have been made by feminist and postcolonial researchers.

As a twenty-first-century invocation of the nineteenth-century *Methodenstreit* between theoretical-deductive and empirical-inductive research, this is an area of inquiry where views are particularly strongly divided. At one end of the continuum, scholars praise the positive potential of data analysis. Scholars who engage in the mathematical, quantitative study of big data claim that digitisation enables them to 'record the activities and interactions of the subjects directly', offering comprehensive (n = all), raw, pure and reliable data on 'real world network dynamics' as preferable to research participants' subjective responses that might 'contaminate' data. Big data are also praised for their capacity to bypass the 'laziness' and refusals of unwilling research subjects (Newman, Barabási and Watts, 2006, p. 5). Computational social scientists, for example, value Facebook as a research tool because, 'data can be recorded retrospectively in a convenient, accurate, and inexpensive way' (Kosinski *et al.*, 2015, p. 543). Those claiming that more objective, disembodied and value-free insights can be gleaned from data that are 'naturally' created by internet users (Mahrt and Scharkow, 2013, p. 21) replicate all-too-familiar historical Californian ideologies of utopian (and race-, gender-, class- and geography-blind) technological determinism.

At the other end of the continuum, scholars warn that governmental and corporate data-driven mass surveillance leads to a new totalitarianism of 'big data dictatorship' (Helbing in interview with *Centre for BOLD Cities*, 2016), and that 'big data' research that focusses solely on dominant patterns is neglectful of social divisions, oppression and marginalised voices, leading to 'methodological genocide':

> the methodological genocide that I am hinting at is epistemological, disciplinary, technological, urbanized, elitist, commercialized and, I dare say, often gendered. (Uprichard, 2015, n.p.)

Indeed there is a racist, classist and 'sexist data crisis' (McDonald, 2016). The digital humanities are responsible for replicating a white, masculine, middle-class and Western bias in knowledge production (McPherson, 2012), and computational social science methods similarly normalise 'gentlemanly' and 'technical' terms of expertise (Savage, 2013, p. 18). The difficult relationship between technologies and power differences grounded in gender, location and race can be traced back to the exclusionary, imperial and sexist designs of most of our technological systems. In parallel, data that may be gathered are never neutral or innocent. For example, in the context of South Asia, gender, location, race, caste and income, among other aspects, shape the 'data pyramid' and affect the production, circulation and storage of databased democracies, identities and geographies (Arora, 2016). Therefore, despite their seductive claims to tackle global problems including social divisions, inequality, racism and sexism, data studies, like any other technology, should not be considered as a panacea for social ills.

At the level of theory, critics rightfully lament how branches of contemporary critical theory—including new materialism, object-oriented ontology, posthumanism and new medium theory—'coincide' with the 'software of big business' (Galloway, 2013, p. 347). Mirroring the seductive bigness of data, such flat ontologies risk replicating a 'view from nowhere' when aiming for 'comprehensiveness'. In decentring the 'human and her attendant political and cultural concerns', power relations, experience and meaning-making may be lost in infinite assemblages (Andrejevic, Hearn and Kennedy, 2015, pp. 381–382). Arjun

Appadurai ([2016](#)) points to the particular epistemological Orientalising impact of Anglophone big data media studies that fuse Euro-American High Theory with military-industrial algorithmic High Technology. As a form of knowledge-based imperialism—through universalising automated aggregation of Western 'machine-based sociality'—less-privileged digital 'proxy social worlds' outside the West are marginalised and expected to 'catch up' with properly screenified singular modernity (*ibid.*, pp. 6–7). The radical, unlimited scope and distributed sensitivity promised by the illusion of flat ontologies is difficult to align with inductive and empirically grounded activist, anti-oppressive and/or emancipatory research ideals.

In the ever-accelerating quest for data mining and development of new tools, a moment of reflection is also needed to problematise the taken-for-granted ethics of data-driven research by asking, where are 'human subjects in Big Data research' hidden (Metcalf and Crawford, [2016](#))? However, in addition to the lack of feminist or postcolonial approaches within data science, data science itself receives scant attention within feminist and postcolonial research. Theorists of difference have not yet systematically scrutinised how big data, predictive analytics and digital methods operate; which knowledges they normalise; what they exclude; or how they can be appropriated (see, for example, Appadurai, [2016](#)). Nonetheless, outside of the field, feminist and postcolonial philosophy of science is readily recognised as a valuable resource to rethink data analysis epistemologies, methodologies and ethics. For example, cultural geographers draw heavily on feminist literature to argue that data analysis should develop a 'situated, reflexive and contextually nuanced epistemology' (Kitchin, [2014](#), p. 1).

But how can we mobilise feminist and postcolonial philosophies of science to develop a situated, reflexive and more ethically-sound epistemology and praxis of data-driven research? In the last three decades, feminist science critics have demonstrated how all scholarly inquiry is socio-culturally constructed, situated and partial. In particular, as 'counter cultures of science' (Harding, [1991](#), p. 10), they have exposed the ways in which dominant, positivist science standards are historically tied to imperialism and colonialism, the military-industrial complex, capitalism and cultural attributes of maleness (Haraway, [1991](#), 2011; Wajcman, [2004](#); Koh, [2015](#)). Most prominently for feminist data analysis, in the early 1990s Donna Haraway already recognised that technology-enhanced vision around surveillance, imaging, scanning and mapping contained a devouring totalitarian ideology of vision:

> Vision in this technological feast becomes unregulated gluttony; all perspective gives way to infinitely mobile vision, which no longer seems just mythically about the god-trick of seeing everything from nowhere, but to have put the myth into ordinary practice. And like the god-trick, this eye fucks the world to make techno-monsters. (Haraway, [1991](#), p. 189)

Haraway ([1991](#)) reframed objectivity as feminist by emphasising that the principles of 'accountability', 'positioning' and 'partiality' allow researchers to counteract hegemonic scientific notions of value-free truths. To evaluate anti-oppressive research, Bhavnani ([1993](#)) developed the criteria of resisting the 'reinscription' of power hierarchies, giving attention to the 'micropolitics' of research praxis and the integrality of 'difference'. These principles and criteria can be transposed to feminist and postcolonial data analytics. Indeed, achieving multiple partial views—combining and valuing equally big data and small data—can enable us to firmly stake out a nuanced position between uncritical celebration of the datalogical turn and outright dismissal of new methods as blind 'data fundamentalism' (Crawford, [2013](#)).

To achieve partial views on algorithmic data as situated knowledges, I will synthesise different strands of feminist scholarship on the internet.

Over the last twenty-five years, three strands of feminist internet studies have become especially prominent: feminist media studies, cyberfeminism and feminist technoscience; these subfields focus respectively on texts, audiences/users and infrastructures. Feminist media studies scholars productively use critical discourse analyses to conduct text-based analyses of expressions of power and contestation. Their focus is usually on self-representations including profile pictures, blogs, games, videos and fan fiction and issues of identity, stereotyping, voice, spectacle, exploitation, agency, hate, violence and systematic silencing (e.g. Shepherd *et al.*, 2015). Cyberfeminists focus largely on the myriad experiences of users, and employ cyber-ethnographic methods to destabilise and situate the universal user as variously located, embodied, gendered, racialised and classed (e.g. Gajjala, 2004). Feminist technoscience scholars largely focus on implications of materiality. For example, drawing on actor-network theory, they take up a relational perspective addressing humans and non-humans as assemblages (e.g. Bivens, 2015a). Most recently, sub-strands of each of these approaches have emerged, including feminist software studies and feminist studies of political economy. In addition, dialogue is slowly emerging between feminist internet scholars and critical race and postcolonial theorists, which is crucial for critiquing ideologies that ignore gender, race and geography and that equate technological advancements with social progress (e.g. Browne, 2015).

Although often conducted in isolation, these foci might be productively synthesised through feminist and postcolonial data analytics. Digital data are constituted at the interplay of material infrastructures, users and texts, and this tripartite focus can promote the production of multiple, situated knowledges. This is necessary because if we constrain our analysis either only to users or to textual representations but do not scrutinise power-ridden material interfaces or algorithms of social media platforms, 'we chip away at the problems feminism seeks to transform within and through social media, while the deeper manifestations continue to flourish' (Bivens 2015a, p. 2). In the next section, I draw on feminist and postcolonial ethics of care to reconsider data analytics for social-justice oriented research.

ethics of care and data analysis

This section proposes how we can mobilise data mining in practice, not in pursuit of universally valid truth claims or the discovery of law-like disembodied generalisations, but in an ethical, self-reflexive and situated attempt to achieve multiple partial views on everyday life practices and experiences. The thorough consideration of power, authority and ethical issues surrounding the research process is a strong feature of feminist and postcolonial scholarship, especially research of the qualitative, empirical variety (Ramazanoğlu and Holland, 2002; Edwards and Mauthner, 2012; Nagy and Hesse-Biber, 2012). This literature offers ideal-types for ethical, caring, committed and responsible researchers to establish trusting and anti-oppressive relationships between scholars and those studied.

A model of utilitarian research ethics—which is often reflected in disembodied, large-scale data-driven research projects—assumes impartiality and the neutrality of the researcher, and is focussed on outcomes based on distanced knowledge production. This model draws on a cost-benefit analysis and abides by expectations held by bio-medically/ethically oriented university institutional review boards (Kosinski *et al.*, 2015). These studies are based on abstract protocols in place to protect research participants against harm, as reasoned from the perspective of institutional audit culture, reputation management and safeguarding researchers from liability claims (Verhallen, 2016). Such a pragmatic approach has led to excesses, such as the infamous 'massive-scale emotional contagion' study of 689,003 Facebook users conducted by Cornell University scholars in collaboration with Facebook (Kramer, Guillory and Hancock, 2014). 'Consistent with Facebook's Data Use Policy', they argued that users agreeing to Facebook's Terms of Service (TOS) prior to setting up their account constituted 'informed consent' (*ibid.*, p. 8789). Unbeknownst to users, and without the choice to opt out, the scholars 'manipulated the extent to which nearly 7 hundred thousand people were exposed to emotional expressions in their News Feed' (*ibid.*, p. 8788). The ensuing debate drew attention to the semi-legal strategic bypassing of ethical obligations towards participants (i.e. users, consumers), which is typical not only for big-data marketing research and mass surveillance but also for some branches of academia.

In contrast, a feminist and postcolonial ethics of care is value-based and recognises the dependencies, partiality, political commitments and personal involvements of researchers. Care involves practices in which we engage 'to maintain, contain and repair our "world" so that we can live in it as well as possible', based on attentiveness, responsibility, competence and responsiveness (Tronto, 1994, pp. 126–136). Although often associated with women, the conflation of feminine traits with a feminist ethics of care in research needs to be questioned, not only because 'feminine' is a contested category but also because it champions a fully human, global morality which involves care for fellow humans and non-humans, such as animals, alike (Engster, 2006). Translated to the domain of research, feminist ethics confronts cultures of knowledge production, for example by focussing on concrete lived experiences, acknowledging dependencies, and assuming researchers are personally accountable and responsible for protecting research participants from risk or vulnerabilities (Edwards and Mauthner, 2012; Lomax, 2015). In addition, knowledge production is seen as a power-ridden, relational and context-specific dialogue rather than as neutral, impersonal and distanced data extraction (Hill Collins, 1991). Instead of having to be eliminated, compassion, emotionality and empathy are considered central to ethical judgements during the entire research cycle, which includes grant writing, conceptualisation and design, data gathering, analysis, publication and valorisation. This ethical perspective foregrounds striving towards 'egalitarian reciprocity' in researcher-researched relationships (Benhabib, 2002).

However, those critiquing an ethics of care approach as merely validating the perspective of Western, white, middle-class feminists or for perpetuating paternalistic, neo-imperialistic and hierarchical assumption of caring are right insofar as (usually privileged) researchers get to decide who receives this care. This is because caring commonly remains conceptually grounded in an ontology of the 'fundamentally relational character of being', a notion that glosses over histories of domination, gender and cultural differences (De Lissovoy, 2010, p. 289). Indeed, a common thread among decolonial, critical race, black feminist and activist researchers is their privileging of kindredness and a dialogic, situated, incomplete, plural and local research ethics practice to prioritise multiple community and indigenous

knowledges. These scholars rightly challenge 'epistemological assimilation' through particularising 'universal' ethics (including feminist ethics of care) as white, Eurocentric, middle-class and Western. Therefore, a more inclusive and progressive ethics champions solidarity through listening and cross-cultural dialogue at the local and global level, critical self-reflexivity of the researcher, and recognition of the situated relationalities of any research process (Mohanty, 1984; De Lissovoy, 2010; Tikly and Bond, 2013; Verhallen, 2016). Therefore, 'asymmetrical reciprocity' (Young, 1997) is a more realistic goal of critical research praxis. As scholars, we ultimately hold the power of definition and interpretation in our writing, and the relationships between us researchers and those whom we study are never truly balanced friendships, but always hierarchical, messy and subject to constraints of time and resources.

More specifically, Edwards and Mauthner (2012) offer practical guidance for an ethical feminist research practice. Sketching 'conditions of moral adequacy', they prompt researchers to be reflective and accountable while making decisions and grappling with dilemmas (*ibid.,* p. 21). Questions that need to be considered to ensure a more ethically sound research cycle include asking: who is involved and affected and in what contexts; how do people's social locations relate to one another; what needs exist and how do they interrelate; who is identified with and who is othered; how is one's study perceived; and what relationships are established with those involved. These questions are similarly relevant for feminist and postcolonial data analysis. Although it is not a panacea for all ills of data mining, the grid in the following table sums up how conditions of feminist moral adequacy can be translated into a road map for an alternative data-analysis practice, which is more situated, reflexive and accountable:

Road map for an alternative data-analysis practice

---

1) People are more than digital data

- People do not equal data traces; data traces are limited, often ahistorical and decontextualised representations of people.
- 'Big data' are only meaningful in interaction with in-depth 'small data' that value human subjectivities and meaning-making.
- Rather than only extracting data, collaborative data analysis values users as experts over their own practices and attempts to maintain connections between researchers and people under study.

2) Context-specificity

- Digital data are never 'raw' but are always already 'cooked' (Lupton, 2016, p. 4). The process of cleaning data to ensure software readability does not take place in a vacuum but reflects and shapes intersecting gender, race and classed power relations.
- Situations define moral problems (access, disclosure, privacy).
- User-generated content is generally published in informal spaces that users often perceive as private but may strictly speaking be publicly accessible. In any case, researchers are rarely the intended audience of user-generated content. Although often mentioned in Terms of Service agreements, users may be unaware about the possibility and reality of their data being mined for (commercial) research purposes. Researchers are responsible to inform users about how and why their data are gathered, coded, analysed, stored and possibly shared, and under which conditions.
- Digital data are performative, subjects are known in data-contexts through repetitive enactments and are slotted into known parameters (Day, 2014). These processual relationships between software/hardware/humans are bounded but not determined by socio-technical configurations (Wajcman, 2004; Barad, 2007) and can only be understood as part of wider visual/audio/textual/ hyperlinked/coded/algorithmic narratives.

(continued)

3) Dependencies and relationalities

- Digital data-environments, like social media sites, are characterised by distinctive 'platform values' (Leurs and Zimmer, 2017), presupposing inevitable dependencies.
- Dependencies include human and non-human actors including users, communities, networks, software, interfaces, algorithms, corporations, governments and fellow academics.
- Dependencies and relationalities are 'asymmetrically reciprocal' (Young, 1997). Encounters between researchers and research participants reflect uneven relations of power and knowledge, which also characterise complicated relations of dependency between human and non-human actors in data analysis.

4) Temptations

- Over-investment with digital personae under study may lead to misunderstandings of a multiplicity of selves.
- Researchers might fail to recognise that human subjects in research hold autonomy and authority over data, have the right to opt out, and can refuse to be coded.
- Over-investment in politics or outcomes might lead to over-identification with those studied.

5) Benefits and harm

- Researchers benefit in various ways from carrying out their studies (personally, career-wise, social standing). Researchers are accountable for creating exploitative relationships with research participants. Rather than 'do no harm', research should be beneficial to the people involved.
- Research participants may benefit in some way from collaborating or participating (personally, professionally or otherwise), causing harm when connections are broken at different points in a study. Those volunteering their data may feel betrayed when the researcher moves on to another project.

A feminist ethics of care demands attention to human meaning-making, context-specificity, inter/dependencies, temptations, as well as benefits and harm. A moral focus is on relationality, responsibility, inter-subjectivity and the autonomy of the research participants. These concerns offer new ways to theorise and empirically sustain the propositions that digital data cannot be expected to speak for themselves, that data do not emerge from a vacuum, and that isolated data on their own should not be the end goal of a critical and reflexive research endeavour. What a situated ethics of care for data analysis might look like will be further developed in the following substantive empirical section. This endeavour brings its own challenges, and I will discuss opportunities and constraints of striving for a feminist and postcolonial ethics in data analysis. In particular, I relate the guidelines outlined above to assess the 'micropolitics' (Bhavnani, 1993) of creating and analysing Facebook friendship visualisations with young people as part of my aim to develop anti-oppressive digital data research.

feminist data studies in practice

Reading most journal articles reporting on data-driven research, one gets the impression that gathering user-generated data is a straightforward process of using a software tool to scrape online texts. What often gets silenced in the methods sections of journal articles is how gathering digital data is a context-specific and power-ridden process similar to doing fieldwork offline. I can demonstrate this point using

my own, admittedly rather 'small-data' oriented study. The study asked how young Londoners perform digital identities and to what extent they network and identify with one another across racial, class and gender differences, using social media platforms. With this question in mind, I spent the autumn of 2013 to the winter of 2014 conducting fieldwork among eighty-four young people ranging from 12 to 21 years old. Although large-scale data mining on young urban dwellers can be done more efficiently in real time in a matter of hours, for example through fire-hose access to Tweets, recruiting eighty-four interviewees for a small-scale study on digital media use was a time-intensive eighteen-month-long process. The group consisted of forty-one young men and forty-three young women, living in the boroughs of Haringey, Hammersmith-Fulham and Kensington-Chelsea. These three boroughs are highly diverse and reflect working-class, middle-class and upper-middle-class environments, respectively. During face-to-face interviews in youth clubs, schools, churches and inside young peoples' homes, I made use of 'collaborative' paper, pencil and digital techniques to afford research subjects' 'greater narrative latitude' in knowledge production and a 'larger role in determining why and how research outcomes are produced' (Gubrium and Harper, 2013, p. 16).[2]

I gathered a varied, messy but also very rich data set, instead of a clean, readily-importable data set, by combining different techniques of data elicitation. First, the research participants were invited to draw a concept map by hand showing what their view of the internet looked like. These maps structured and informed the interview questions. Second, making use of the participants' personal computer, portable device or the laptop I brought, I created visualisations of personal Facebook friendship networks on the spot, together with those young people who had a Facebook account. This digital technique was used to grant research participants autonomy over the digital data they shared with me. The commercial but freely accessible Facebook application *TouchGraph Facebook Friend Graph* was chosen to generate visualisations of Facebook networks. After research participants logged into their Facebook account and opened TouchGraph, the algorithm first processed their Facebook network and subsequently grouped and colour-coded mutual Facebook friends into clusters. This digital-mapping exercise was more than a convenient technological solution. Aiming to generate situated knowledge about young Londoners and their digital identities, it was my intention to deploy digital methods that would cover texts, users and materiality in an anti-oppressive way. Below I demonstrate how an ethics of care for digital methods was relevant in this mapping process by reflecting on diverging fieldwork experiences.

people are more than data

Data-driven research often values aggregated, seemingly 'natural' volunteered data over the complexity of individual human subjectivity and meaning-making. In the case of data-driven research on Facebook practices, this means that humans are reduced to the available data traces they have left while engaging with the value-laden platform (Leurs and Zimmer, 2017). This incomplete but often perceived as total documentation operates as a powerful regime by interpellating people into known and limited parameters of gender, race, social class and location, thus subsuming subjects: data 'slowly becomes us, singularly and as a whole' (Day, 2014, p. 573). These datapoints leave little room for ambiguity; for example, data

---

[2] Although commonly used in internet studies, feminist fieldwork (Gubrium and Harper, 2013) and, especially, critical youth studies (Lomax, 2015), rather than empowering, the label 'participatory' (more so than terms like 'collaborative', 'activist', 'engaged' or 'emancipatory') can be said to reduce research subjects into the mainstream neo-liberal 'logic of productivity' (Tikly and Bond, 2013, p. 425; see also Cooke and Kothari, 2001).

researchers commonly consider user keyboard language settings as a proxy for geographic location and nationality. Or they aggregate common 'male' or 'female' Facebook practices but do not commonly reflect on the symbolic violence of the limited, binary options for self-profiling one's gender: either 'male' or 'female' (Bivens, 2015b). In-depth 'small-data' studies are important to chart how individual users personally value and engage with such settings and wider practices of algorithmic sorting. For example, there is no option to tick a box on Facebook to show one's ethnicity or geographical belonging on the profile, so users create separate pages for ethnic self-description and link to them from the profile. These specifics are difficult to aggregate into numbers, as Sarah (17 years old) for example explains:

> Sarah: They will all like pages which kind of relate back to their ethnic background and your hometown, the country where you come from, you can kind of write that on Facebook, and it will come up on their profile.
>
> Koen: And you yourself, are you interested in doing that or—
>
> Sarah: Yeah, I have liked the page and I have wrote down my ethnicity on Facebook.
>
> Koen: Why do you feel it's important to do that?
>
> Sarah: Because, I wouldn't say I'm British, like English British, because I'm coming from somewhere, I come from Pakistan originally, though I wasn't born there, I originally come from there, my nationality is British but my ethnicity is different.

Sarah's reflection was a response to my invitation to research participants to describe themselves. These self-representations illustrate the limited explanatory value of isolated and bounded digital data traces. For example, 18-year-old Tyreese explored his hybrid positionality by stating, 'I'm half, I'm mixed-other, but, I'm not really sure about my ethnicity, my mum's ancestry, she's fully Trinidadian, but she's very mixed race, and then she's also got Portuguese and parts of India in her. So I'm mixed from six-seven different countries'. In contrast, 16-year-old Jackson emphasised Britishness, indicating his desire to meet the norms and to pass as a white British boy: 'Well I've got a little bit of Caribbean in me, but other than that I'm white British with a bit of Northern, but I consider myself to be British, white British yeah, not Northern, not Caribbean, just that'. Thus, the level of autonomy human subjects involved in research should have over their representations may not be sufficiently safeguarded in large-scale data extraction approaches.

Furthermore, scholars have rightly lamented the problematic 'availability-bias' in data-driven research (Mahrt and Scharkow, 2013, p. 26). Focussing on recognising patterns, large-scale data mining assumes uniformity among users. The complexity of everyday life cannot always be fully captured in patterns or neat categories. For example, 16-year-old Jason noted he was 'born in the UK' and added that his 'father was born in Kenya in Africa and my mother was born in India, but they are both Indian'. Typical of his fellow peers, Jason stated that he was 'mainly on the computer whenever I'm at home'. However, as he was critical of platforms like Facebook and Twitter, he made use of them only strategically:

> Koen: And most of the other people that I interviewed mentioned kind of social media sites and things like that, but you haven't included them [in your paper and pencil Internet map], so are you using Facebook and Twitter and those?

Jason: I don't use them regularly, I just use them to talk just like on my phone, I don't really use it, cos I'm not a fan of what their terms are and stuff.

Koen: Yeah and can you explain?

Jason: They don't really delete your information, like they allow to store it permanently, even if you delete your profile, they are allowed to store it, and I just don't like the way how they are using our information and selling it, and then yeah. If you was getting a job interview and someone else finds your information, they see you posing with like something bad, then they will use that against you and you can potentially lose that job, so I just try and clear it as much as I could of it and not really use it.

Koen: Do you think young people should be better informed about that?

Jason: I believe they should, I mean the information being sold and being used against them, and stored for unlimited periods of time, and it's like the data belongs to Facebook, it doesn't belong to you anymore, you just give it up without knowing, and obviously it can be used against you in law as well.

Non-users like Jason confronted me with my own normative framework. I had wrongly assumed that all research participants fitted into the flat category of the 'digital native', and that all young people were heavy users who had all come to accept and embrace the new (lack of) privacy. Research participants like Jason were highly active on dedicated discussion boards and other atomised sites; however, they were non-users of mainstream platforms, resisting volunteering their data to profit-driven platforms and seeking to keep some control of their digital personae for specific intended audiences. They actively worked to avoid unintended consequences of endless future circulation of their digital footprints. Focussing on Facebook friendship visualisations, I was biased towards mapping user cultures on a mainstream proprietary platform—in this way I further marginalised possible 'countercultures of data' (Hoffman, 2016).

context-specificity

Digital data is performative and context-specific: the feelings shared by Charlie about suicidal tendencies in the context of social media peer cultures in response to the network visualisation can be made meaningful when understood as part of their wider asymmetrically reciprocal narratives, codes and algorithmic culture. Although often taken for granted as omnipresent in overdeveloped cities in the Global North, internet access was not a given in London. Social networking sites like Facebook and Twitter were blocked by a firewall in one of the youth centres where I held my interviews. It is ironic that youth clubs that aim to welcome youngsters to spend their leisure time and to do school work block a major channel of their communication. As an alternative internet access point, I created a Wi-Fi hotspot by sharing my tablet's internet connection. However, due to unstable 4G reception, the tool did not always run smoothly, unintentionally testing the patience of the research participants.

Slow internet connections and the lengthy loading and algorithmic processing of Facebook friendship data (usually less than two minutes, in some exceptional cases more than five minutes) resulted in rejection, reluctance or criticism, as I will discuss below in the section on temptations. It is worth mentioning that in the context of London and other large cities in the Global North and South, urban

ideals of citizenship are often projected on visionaries of smart/tech cities, social media and digital connectivity (see Prime Minister's Office, 2012). However, in their championing of digital civic participation for a more inclusive society, certain urban dwellers and specific user practices are privileged over others (Georgiou, 2016). During fieldwork I learned that, like public bike sharing, corporate- and city-sponsored wireless networks were most commonly available in the main tourist and higher income areas, while places like Tottenham lack these infrastructures.

Furthermore, data-driven research is susceptible to the changing whims of corporations. In April 2015, shortly after finalising fieldwork, Facebook changed its Application Programming Interface (API) and data retrieval policy and, as a result, TouchGraph was not able to access and visually process Facebook friendship data any longer. As such, it is now impossible to replicate the study exactly, although it can be approximated using other similar tools such as Netlytic.org. This development raises questions about whether it is ethical to work with the data I have gathered, and it also means I cannot hold the promise I made to the research participants: that they would be able to study their own networks using TouchGraph at a later time. This also again illustrates relationships of dependency in which we data scholars are necessarily involved with our platforms of study, which Marres (2012, p. 140) rightly critiques as an example of the increasing 'privatization of social research'. Conducting research on everyday internet use often implies abiding by the rules of large, transnational corporations. The tendency of privatisation raises the question of whether the development of non-extractive open-source software applications should be seen as an integral part of a feminist ethics of care in a data-sensitive research environment.

dependencies and relationalities

Although user-generated content on platforms like Facebook can be scraped and aggregated in seemingly anonymous large-scale databases, scholarly researchers are not commonly the intended audiences for this data. By agreeing to Facebook's obligatory TOS (Terms of Service) when setting up accounts, users agree to pay for a service by handing over their personal data; they passively grant permission to Facebook and third parties allowing them to commodify their data. Academic researchers are an example of such a third party. It cannot be expected that in their everyday, informal engagement with social media, users think about whether they mind that researchers might study their photos, videos or status updates. Although seductive big-data narratives praise the data mining of user-generated content for direct, reliable access to data that is 'real' and 'natural', it should be noted that data-scraping researchers are also complicit in exploiting the 'free labour' of human beings (Fuchs, 2014). In sum, people, hardware, software and data exist in a relationship of hierarchical dependency, and the presence of researchers in this relational model adds further complexity and messiness (Barad, 2007; Berry, 2014).

A feminist ethics of care highlights the accountability and responsibility of researchers towards research participants, and also acknowledges the autonomy subjects hold over their data. From this perspective, actively obtaining permission from research participants to use the data they generated is favoured. Therefore, I explicitly asked users for their permission, and although the majority of young people granted me access, this question usually triggered questions in response. Interviewees asked questions like 'How would you use it though?' ('Japanese-French' Skye, 16 years old), 'It won't post it on Facebook will it?' (Tyreese, 18 years old) as they wondered whether their anonymity was guaranteed and whether

TouchGraph would leave traces that could cause concern among friends. There is a need to actively gain enough trust to obtain permission, which shows how digital data researchers are inevitably dependent on the willingness and the varying knowledge of research participants and on their own capacities to develop research practices based on trust and rapport.

During the interviews, I described the workings of TouchGraph by showing a visualisation of my own personal Facebook friendship network. I discussed various clusters of friends including basketball teammates, family, research participants, as well as friends from various university networks across the world. I vividly remember mentioning that a Marie Curie Fellowship awarded by the EU allowed me to work at the London School of Economics and Political Science for two years to Sarah, a 17-year-old young woman, during our interview in February 2014. Sarah self-identified as follows: 'my nationality is British but my ethnicity is different. I'm from Kashmir'. She was doing her A-levels at the time; she was a youth counsellor and a deputy member of Youth Parliament. She was wearing a colourful headscarf and she spoke openly about her experiences. She reminded me of my privilege when she spoke about her relationship with LSE:

> I have applied to do accounting and finance at university, and I wanted to go to London School of Economics, but unfortunately they rejected me. (Sarah, 17 years old)

A painful, uneasy silence followed after she told me about her application. As a North-Western European, white, highly-educated, middle-class, heterosexual man, I was reminded of the distance between myself as a representative of this prestigious, elite university and Sarah. On top of that, I can rightfully be accused of replicating the uneven standard of extractive scholarship. The awarded Marie Curie grant is specifically oriented towards training young scholars to increase their chances at obtaining a tenure-track position in academia. After the postdoctorate, I was indeed hired as an assistant professor. Thus, in retrospect, I secured my personal upward mobility by studying young Londoners' experiences of urban cultural differences and digital media use. However, as I argue below, it is through being critically self-reflexive about one's positionality, locating research practices within wider power relations and structures, that we can begin to destabilise the normalised politics of knowledge production.

The interview took place at a youth centre in Tottenham, North London. The centre, where young people from the local area play sports, make online radio and use computers, is located a few hundred metres from where the London riots started in 2011 after police officers had shot the Afro-Caribbean neighbourhood resident Mark Duggan. In another research encounter, the 18-year-old 'British' and 'half-Iranian' Louise, who aspired to be a midwife, questioned me about reproducing stereotypically gendered caregiving roles in the family, after she learned that my partner and I had just had a baby: 'It's still most of the effort is on the mum, that's why you are interviewing here and not at home with the baby'. Uncomfortable research encounters such as this one are productive not only because it made me feel guilty and eager to go home, but also because they help to redirect our focus as researchers towards power relations and the politics of knowledge production.

Although young people did not shy away from emphasising difference and scrutinising my privilege, instances of meaningful connection were also common. Louise, for example, enthusiastically wanted to introduce her mum to me after she learned I had previously travelled to Iran: 'You've been to Iran!?

That's so cool. I've got family in the north. He's been to Iran!' Young people also expressed being appreciative of being able to voice their concerns to an interested outsider. For example, in our joint-interview Connor (15 years old), 'born here' but 'I'm quite a mixture', and Carmen (14 years old), 'India is my home', stated 'All the way from the Netherlands coming to London, coming to Tottenham. To speak to me and speak to Carmen. That's so cool!'

temptations

Facebook provides a trove of 'easy data' that can be gathered conveniently and inexpensively (Kosinski *et al.*, 2015, p. 543). Yet most data-driven, social media research is 'based on the assumption that users implicitly consent to the collection and analysis of their data by posting them online' (Mahrt and Scharkow, 2013, p. 26). Instead of starting from an aggregated data pool and thinking about consent only a posteriori, for example when trying to locate individual users in data sets, I started from the individual participant, and asked for consent before accessing research participants' data. Interviewees were individually asked for consent to collaboratively visualise and analyse their friendship networks. They were free to opt out from this exercise during the interview, and some did. But gaining informed consent was not always a self-evident component of the research ethics process.

Bringing in my intensely confusing research encounter with Liane and Kevin, two 12-year-olds, is helpful to draw out further the situated challenges around informed consent, autonomy, authority and expertise. The energetic cousins insisted on being interviewed together. Kevin and Liane were 'born in England', while their parents migrated to London from Romania. Kevin identified as 'different', noting that she feels 'I'm a boy like you', and Kevin agreed when described by Liane as a 'tomboy'. Liane described herself as 'fat' and 'wanna be skinny', but above all she mentioned feeling 'courage' and 'confident'. In their enthusiasm the cousins hardly allowed each other to finish their sentences. Eventually a sense of unevenness led to quarrelling and a generally hostile atmosphere. Re-listening to the interview audio recording and re-reading the transcript, I realise that the Facebook friendship visualisation exercise added fuel to the fire. After showing and discussing a visualisation of my personal network, which spans academic contacts, family and friends living across the world but is most densely concentrated in the Netherlands, the interviewees decided to visualise Liane's friendship network first, and Kevin's second. Having 1061 friends, it took a while for the TouchGraph algorithm to process Liane's data. Their patience was tested, with Liane asking questions such as 'Why do we have to wait, what are you doing!?' Kevin wondered when the interview would be over: 'Sir, how many minutes more?' Getting bored, they started to push each other around and Kevin told Liane to 'fuck off'; Liane shouted back 'don't tell me to fuck off'. After sensing the visualisation was almost done, I raised my voice and told them to 'calm down'. A bit startled, they started to wonder whether I was going to inform the youth centre staff about their behaviour: 'I'm sure he is going to tell, innit sir', but I tried to reassure them by saying, 'No, I would maintain your privacy and protect your anonymity'. Then a network structure of colour-coded friendship nodes appeared on the screen of my laptop (see Figure 1). A puzzled look came over Liane's face, as she stared at the brightly colour-coded clusters of grouped user names accompanied by their profile photos.

> Liane: It takes too long!

> Koen: Let's just let this load for a second, if we can have a look at both, ah there we go.
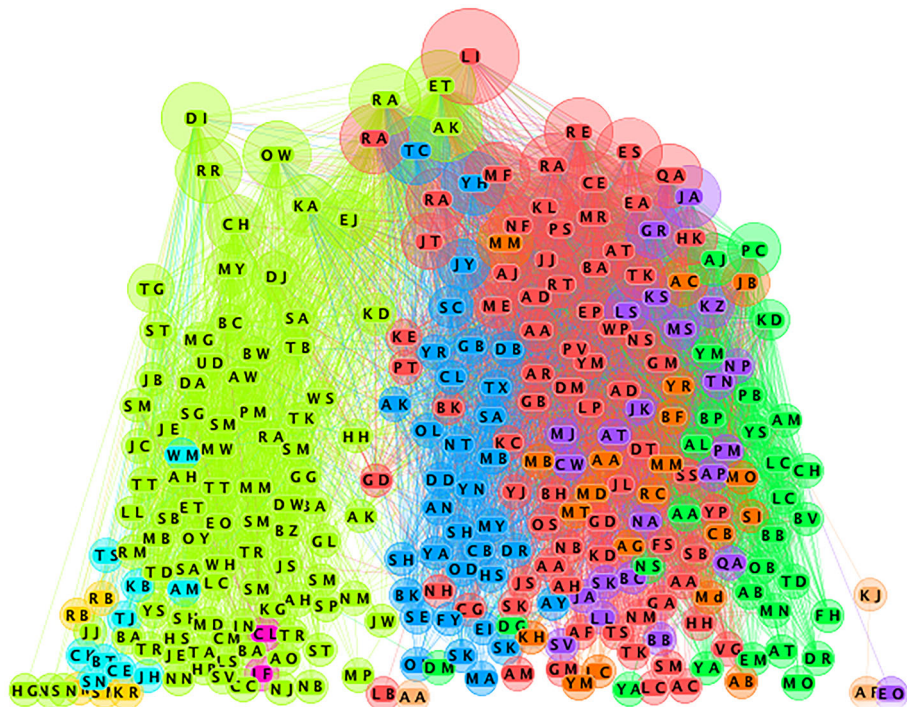
Figure 1   Anonymised visualisation of Liane's Facebook friendship network. *Source*: TouchGraph (2014).

Liane: Wooow! Shine bright like a diamond, beautiful see.

Koen: If you see these people and the different colours that they have.

Liane: Yeah?

Koen: You see that some people are blue, some are green.

Liane: What does that mean?

Koen: Are these people connected to each other, from school or?

Liane: They are cousins, they are friends, Romanian friends.

Koen: Who are the Romanian friends?

Liane: All of them!! See.

Koen: Is there a difference if you look at the colours?

Liane: I don't know what it is about.

Koen: No?

Liane: They are just friends, they talk to each other, I dunno.

This exchange illustrates my assumption about young people as experts over their own experiences, as being able to naturally 'tell it like it is', does not always hold. Feminist researchers with a psycho-social ontology of subjectivity have rightly recognised research participants (and interviewers alike) as 'defended subjects' with limited structured self-expertise (Hollway and Jefferson, 2009). In order to avoid assuming a rational unitary subject by default, research subjects should thus be given room to freely associate, structure and link their own narratives. In some cases, digitally rendered visualisations and algorithmic ordering principles as prompts might be perceived as at odds with people's processual subjectivities. Transposed to the digital context, we cannot always assume self-transparency, and we should critically reflect upon 'What happens when we ingest/absorb digital data about ourselves?' (Lupton, 2016, p. 4). I tried to explain to the cousins what we are shown through the data visualisation and why it is of interest to me, but to no avail. Kevin decided she did not want to proceed with visualising her network ('I don't want to'), nor discussing it ('I'm not gonna talk').

This refusal is generative for two purposes. First, it demonstrates that static understandings of consent are problematic, including implicit assumptions that users will permit external access to their data upon interacting with social media platforms. Although formal ethics requirements often require informing potential participants about one's study and that participants sign a form to indicate that they freely agree to take part prior to research, this initial permission should not be taken at face value. Kevin did sign a consent form, which included details on the friendship visualisation. However, informed consent is 'a process' that should be ensured throughout a research cycle (Plankey-Videla, 2012). Consent can be continued or withdrawn as a result of the research participants' better understanding of the nature of the study, increased or diminished rapport with a researcher, or loss of interest, as in the case of Kevin.

Second, Kevin's refusal redirects the focus towards power hierarchies in data research. As a research participant and data producer, Kevin claims autonomy and authority by opting out. Through this refusal, Kevin objects to academic objectification and subjection. Rather than silencing non-participants in our publications, writing about refusals is generative for interrogating the academic 'coding' practices we researchers sustain through our power of definition, representation and academic gaze (Gajjala, 2004, p. 25; Tuck and Yang, 2014, p. 812). 'Coding is something we [researchers] do to [human] objects' (Tuck and Yang, 2014, p. 812), and the process of often narrow, monolithic or typological coding of human subjectivity happens on unprecedented scales in some varieties of big data-driven research. The story of Kevin's refusal is particularly informative when juxtaposed against big data ideologies of comprehensiveness. Automated research that in principle includes all users (n = all) in any given group, platform or discussion is solely extractive and leaves no such room for subjectivity, meaning that 'inquiry as invasion is built into the normalised operations of the researcher' (*ibid.*, 2014, p. 811).

benefit and harm

Although I do not want to claim that data elicitation technologies can allow researchers to bridge the power gap with research subjects, young people involved in the study did appreciate that they were in charge of analysing their own Facebook friendship visualisation. For example, Anna (17 years old), who feels 'English', shared that 'I haven't done anything like this before, I really like it. I think it was good, it sort of put my mind in perspective'. For each research participant, the Facebook friendship visualisation showed various clusters of contacts, grouped according to certain shared dynamics including mutual

friends, activity and duration of the friendship. However, this is something the young people figured out when zooming in and out on the visualisation. TouchGraph does not explain how data is processed; visualisations are the result of corporate black-boxed algorithms. The implications of such forms of automated analysis can be considered in line with critiques of suppressive methodologies that were innovative in earlier eras.

Postcolonial technoscience scholars revealed how methodological devices such as biometric measurement were used to create hierarchies between population groups and applied as a rationale for imperial exploitation and domination (Harding, 2011). Anthropologists have been similarly critiqued for inscribing static cultural forms upon populations using methods including social networking analysis (Savage, 2013). Feminist- and Foucauldian-inspired analysis revealed that expertise and measurement of psychological tests and child development charts police normativity for example around sexual subjectivities (*ibid.*). Similarly, typologically-oriented data visualisations may lead to what Mohanty (1984, p. 333) in another context calls a 'suppression of the heterogeneity of the subject(s) in question', reaffirming hierarchical relationships of domination. Although data are never 'raw' but always 'cooked' and digested by humans and non-humans (Lupton, 2016), conventions imbue data visualisations with objectivity, validity and neutrality, based on an ideological sense of 'transparency, scientific-ness and facticity' (Kennedy *et al*., 2016, p. 716). In this process, in broadly comparable ways to colonial expansion, the observer holds the power to decide who has the 'right to look' at the populations in the graph and for what purposes. As it is primarily used for marketing and surveillance, mastering data through visualisation is exemplary for the 'military-industrial thirst for mapping and arranging populations' (Shepherd, 2015, p. 1) and resonates with Haraway's (1991, p. 189) observation of the devouring 'god-trick' ideology of 'techno-monster' vision. When engaging with data visualisations, we have to understand them as a contemporary invocation of a long historical trajectory of visually oriented, disciplinary modes of control and exploitation of the Other, dating back to colonial cartographic practices such as land and sea map-making.

I am sensitive to such discussions on spatial ordering, visual control and critiques of the roles of the overseer. By developing a collaborative approach to network visualisation, the interviewees engaged in mapping together with me. This approach, which can be called 'data elicitation' or 'data voice', served the purpose of prompting narrative reflection, in line with 'photo elicitation' and 'photo voice' activist research strategies (Gubrium and Harper, 2013; Lomax, 2015). I prompted interviewees to open up the black box of the software tool by asking them to reflect on possible ordering mechanisms behind the friendship clusters shown on the screen. Although beneficial as a thinking tool, standalone visualisations are not inherently meaningful, as discussed in Liane's experience, and they may be a source of epistemic violence when data classifications do not align with situated experiences. This thus corroborates Wilson's (2014) claim that data-gathering tools do not offer quick solutions for critical analysis.

Importantly, the visualisations also provided a substantial number of research subjects to engage in forms of counter-surveillance. For example, Chenise, a 21-year-old young woman who identifies as 'mixed race, so I'm not fully black', took it as an opportunity to check up on her friend list. At a certain point during our exchange she screamed:

> Chenise: That's my sister there, how did you get on my Facebook? I tried to keep my sister off my Facebook, I don't want you on my Facebook no more! Haha...
>
> Koen: And why is that?
>
> Chenise: We get along, and we don't get along. She reports everything back to my mum!

As I helped Chenise with her counterveillance, this exchange goes against the grain of big data ideologies that assume researchers are impartial in making transparent and disembodied data visualisations. It was my choice to bring in TouchGraph, and in the case of Chenise this decision resulted in a scuffle with her sister after our interview. An ethics of care can bring back into view power relations in data analysis and emphasises the importance of acknowledging the complex subjectivities of research participants.

conclusions

In times of increasing digital datafication, the question arises: how can data researchers better ensure that human subjects in research maintain their rights to make decisions and have a say over their own digital data? This article draws on feminist and postcolonial science studies to intervene in gender-, race- and geography-blind big-data ideologies and draws on feminist ethics of care to propose an alternative, research participant-centred data-analysis agenda. Often, number-crunching positivist 'discovery' is propagated as superior to small-scale, in-depth research, and this ideology commonly reflects cultural attributes of white, middle-class, masculine, able-bodied, heteronormative domination and Western-centric military-industrial imperialism. Feminist, critical race, de- and postcolonial and activist researchers have firmly put on the agenda the politics of knowledge production: knowledge is partial, situated and contextual, and power flows along intersectional axes of difference. Bringing this literature to bear on experiences of a two-year fieldwork project among eighty-four young Londoners on digital identities and multiculturalism, I have called for action and proposed guidelines for more humanising and less exploitative digital data analysis. This overview of much-needed warnings is not the perfect plan that cures all ills, but it offers a stronger commitment to reflecting on choices, dilemmas and obstacles in data-driven research. It does so by facilitating the apprehension of multiple partial truths through strategic, situated employments of digital methods.

Gathering data and identifying patterns using algorithms can never be the end goal of ethical, reflexive socio-cultural research. A feminist and postcolonial ethics of care draws our attention to digital data in relation to human meaning-making, context-specificity, dependencies and temptations, as well as benefits and harm. Data-driven research, in turn, shows the urgency for renewed feminist, critical race and postcolonial engagement and situated moral focus on how digital mediation impacts upon agency, relationality, responsibility, human subjectivity and the autonomy of research participants. Data does not somehow magically operate above and beyond issues of history and cultural difference. Notwithstanding the celebratory data-rush mentality of most digital humanities and computational social science research, knowledge is not out there to be scraped. There is no button for responsible data gathering or for critical analysis. However, when used in an anti-oppressive, agency-centred way, digital data can provide a fruitful entry point to bring out contradictions, paradoxes, meanings, beliefs, experiences and normative questions surrounding everyday life.

acknowledgements

author biography

Koen Leurs is Assistant Professor in Gender and Postcolonial Studies at the Graduate Gender Program, Department of Media and Culture Studies at Utrecht University, The Netherlands. He is a critical internet researcher interested in migration, diaspora, gender, race, urbanity, multiculturalism and youth culture using creative and participatory mixed methods. Leurs has published *Digital Passages: Migrant Youth 2.0: Diaspora, Gender & Youth Cultural Intersections* (Amsterdam: Amsterdam University Press, 2015) and *Everyday Feminist Research Praxis* (Cambridge: Cambridge Scholars Press, 2014). He is currently co-editing the Sage *Handbook on Media and Migration*, and journal special issues on 'Forced Migration and Digital Connectivity' for *Social Media + Society* and 'Connected Migrants: Encapsulation and Cosmopolitanism' for *Popular Communication*. He is the chair of the European Communication Research and Education Association (ECREA) Diaspora, Migration and the Media Section. From 2016 to 2019, he is the principal investigator for the Netherlands Organisation for Scientific Research (NWO) funded research project 'Young connected migrants. Comparing digital practices of young asylum seekers and expatriates in the Netherlands'. See www.koenleurs.net.

references

Ajana, B., 2015. Augmented borders. *Journal of Information, Communication and Ethics,* 13(1), pp. 58–78.

Andrejevic, M., Hearn, A. and Kennedy, H., 2015. Cultural studies of data mining. *European Journal of Cultural Studies*, 18(4–5), pp. 379–394.

Appadurai, A., 2016. The academic digital divide and uneven global development. *Annenberg School for Communication Project for Advanced Research in Global Communication Press*, 22 June. Available at: https://www.asc.upenn.edu/sites/default/files/documents/0106_15_PARGC_Paper4_FINAL.pdf [last accessed 9 September 2016].

Arora, P., 2016. The bottom of the data pyramid: big data and the Global South. *International Journal of Communication,* 10, pp. 1681–1699.

Barad, K., 2007. *Meeting the Universe Halfway*. Durham: Duke University Press.

Beer, D. and Burrows, R., 2013. Popular culture, digital archives and the new social life of data. *Theory, Culture and Society*, 30(4), pp. 47–71.

Benhabib, S., 2002. *The Claims of Culture: Equality and Diversity in the Global Era*. Princeton: Princeton University Press.

Berry, D., 2014. *Critical Theory and the Digital.* London: Bloomsbury.

Bhavnani, K.K., 1993. Tracing the contours: feminist research and feminist objectivity. *Women's Studies International Forum*, 16(2), pp. 95–104.

Bivens, R., 2015a. Under the hood: the software in your feminist approach. *Feminist Media Studies*, 15(4), pp. 714–717.

Bivens, R., 2015b. The gender binary will not be deprogrammed: ten years of coding gender on Facebook. *New Media & Society,* 20(10), pp. 1–19.

Browne, S., 2015. *Dark Matters: On the Surveillance of Blackness*. Durham: Duke University Press.

Centre for BOLD Cities, 2016. NRC interview with Dirk Helbing: 'We are building a data dictatorship'. *Centre for BOLD Cities,* 23 Aug. Available at: http://www.centre-for-bold-cities.nl/news/nrc-interview-with-dirk-helbing-we-are-building-a-data-dictatorship [last accessed 9 September 2016].

Collins, S.G. and Slover Durington, M., 2015. *Networked Anthropology.* London: Routledge.

Cooke, B. and Kothari, U., 2001. *Participation: The New Tyranny?* London: Zed Books.

Couldry, N., 2017. The myth of big data. In K. van Es and M. Schäfer, eds. *The Datafied Society: Studying Culture Through Data.* Amsterdam: Amsterdam University Press, pp. 55–68.

Crawford, K., 2013. The hidden biases in big data. *HBR.org*, 1 April. Available at: https://hbr.org/2013/04/the-hidden-biases-in-big-data [last accessed 25 March 2016].

Day, R.E., 2014. Indexing it all: the modern documentary subsuming of the subject and its mediation of the real. Conference paper presented at iConference, 4–7 March. Humboldt-Universität zu Berlin.

De Lissovoy, N., 2010. Decolonial pedagogy and the ethics of the global. *Discourse: Studies in the Cultural Politics of Education*, 31(3), pp. 279–293.

Edwards, R. and Mauthner, M., 2012. Ethics and feminist research: theory and practice. In T. Miller, M. Birch, M. Mauthner and J. Jessop, eds. *Ethics in Qualitative Research.* London: Sage Publications Ltd, pp. 14–28.

Engster, D., 2006. Care ethics and animal welfare. *Journal of Social Philosophy*, 37(4), pp. 521–536.

Fuchs, C., 2014. *Digital Labour and Karl Marx.* London: Routledge.

Gajjala, R., 2004. *Cyber Selves: Feminist Ethnographies of South-Asian Women.* Walnut Creek: AltaMira Press.

Galloway, P., 2013. The poverty of philosophy. *Critical Inquiry,* 39(2), pp. 347–366.

Georgiou, M., 2016. Conviviality is not enough: a communication perspective to the city of difference. *Communication, Culture & Critique,* 26 July. Available at: http://onlinelibrary.wiley.com/doi/10.1111/cccr.12154/full [last accessed 2 August 2016].

Gubrium, A. and Harper, K., 2013. *Participatory Visual and Digital Methods*. Walnut Creek: Left Coast Press.

Haraway, D., 1991. *Simians, Cyborgs, and Women: The Reinvention of Nature*. New York: Routledge.

Harding, S., 1991. *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Ithaca: Cornell University Press.

Harding, S., ed., 2011. *The Postcolonial Science and Technology Studies Reader*. Durham: Duke University Press.

Hesse-Biber, S.N., ed., 2012. *Handbook of Feminist Research: Theory and Praxis.* London: Sage Publications, Inc..

Hill Collins, P., 1991. *Black Feminist Thought*. London: Routledge.

Hoffman, A.L., 2016. CFP: "counter-cultures of data"—special issue of philosophy and technology. *Annaeveryday*, 28 June. Available at: http://annaeveryday.com/cfp-countercultures-of-data-special-issue-of-philosophy-technology/ [last accessed 5 July 2016].

Hollway, W. and Jefferson, T., 2009. Researching defended subjects with the free association narrative interviewing method. In H.J. Cook, S. Bhattacharya and A. Hardy, eds. *History of the Social Determinants of Health: Global Histories, Contemporary Debates*. Hyderabad: Orient Black Swan, pp. 296–315.

Kennedy, H., Hill, R.L., Aiello, G. and Allen, W., 2016. The work that visualisation conventions do. *Information, Communication & Society*, 19(6), pp. 715–735.

Kitchin, R., 2014. Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), pp. 1–12.

Koh, A., 2015. A letter to the humanities: DH will not save you. *Hybrid Pedagogy*, 19 April. Available at: http://www.digitalpedagogylab.com/hybridped/a-letter-to-the-humanities-dh-will-not-save-you/ [last accessed 7 July 2016].

Kosinski, M., Matz, S.C., Goslin, S.D., Popov, V. and Stilwell, D., 2015. Facebook as a research tool for the social sciences. *American Psychologist*, 70(6), pp. 543–556.

Kramer, A., Guillory, J. and Hancock, J., 2014. Experimental evidence of massive-scale emotional contagion through social networks. *PNAS*, 111(29), pp. 8788–8790.

Leurs, K. and Zimmer, M., eds., 2017. Platform values: an introduction to the #AoIR16 special issue. *Information, Communication & Society*, 20(6), pp. 803–946.

Lomax, H., 2015. Seen and heard? *Families, Relationships and Societies*, 4(3), pp. 493–502.

Lupton, D., 2016. Digital companion species and eating data: implications for theorising digital data-human assemblages. *Big Data & Society,* 3(1), pp. 1–5.

Magee, D., 2011. London riots: BlackBerry Messenger, Twitter used as weapons. *International Business Times*, 11 April. Available at: http://www.ibtimes.com/london-riots-blackberry-messenger-twitter-used-weapons-828319 [last accessed 11 April 2017].

Mahrt, M. and Scharkow, M., 2013. The value of big data in digital media research. *Journal of Broadcasting & Electronic Media,* 57(1), pp. 20–33.

Manovich, L., 2017. Cultural analytics, social computing and digital humanities. In M. Schäfer and K. van Es, eds. *The Datafied Society: Studying Culture Through Data.* Amsterdam: Amsterdam University Press, pp. 55–68.

Marres, N., 2012. The redistribution of methods. *The Sociological Review,* 60(S1), pp. 139–165.

McDonald, C., 2016. Is there a sexist data crisis? *BBC News*, 18 May. Available at: http://www.bbc.com/news/magazine-36314061 [last accessed 7 July 2016].

McPherson, T., 2012. Why are the digital humanities so white? In M.K. Gold, ed. *Debates in the Digital Humanities.* Minneapolis: University of Minnesota Press, pp. 139–160.

Metcalf, J. and Crawford, K., 2016. Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society,* 3(1), pp. 1–14.

Mohanty, C.T., 1984. Under Western eyes: feminist scholarship and colonial discourses. *Boundary*, 12(3), pp. 333–358.

Morahan-Martin, J., 2000. Women and the internet: promise and pitfalls. *CyberPsychology & Behavior*, 3(5), pp. 683–691.

Newman, M., Barabási, A.L. and Watts, D.J., 2006. *The Structure and Dynamics of Networks*. Princeton: Princeton University Press.

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T. and Tacchi, J., 2016. *Digital Ethnography: Principles and Practice*. London: Sage Publications Ltd.

Plankey-Videla, N., 2012. Informed consent as process. *Qualitative Sociology*, 35(1), pp. 1–21.

Prime Minister's Office, 2012. PM announces £50m funding to regenerate Old Street roundabout: major new investments from companies and government announced in Tech City. Press release. Prime Minister's Office, 10 Downing Street, 6 December. Available at: https://www.gov.uk/government/news/pm-announces-50m-funding-to-regenerate-old-street-roundabout [last accessed 15 July 2016].

Ramazanoğlu, C. and Holland, J., 2002. *Feminist Methodology.* London: Sage Publications Ltd.

Risam, R., 2015. Beyond the margins: intersectionality and the digital humanities. *DHQ: Digital Humanities Quarterly,* 9(2). Available at: http://www.digitalhumanities.org/dhq/vol/9/2/000208/000208.html [last accessed: 15 July 2016].

Rogers, R., 2013. *Digital Methods*. Cambridge: MIT Press.

Savage, M., 2013. The 'Social Life of Methods': a critical introduction. *Theory, Culture & Society,* 30(4), pp. 3–21.

Schäfer, M. and van Es, K., eds., 2017. *The Datafied Society: Studying Culture Through Data.* Amsterdam: Amsterdam University Press.

Shepherd, T., 2015. Mapped, measured, and mined: the social graph and colonial visuality. *Social Media + Society,* 1(1), pp. 1–2.

Shepherd, T., Harvey, A., Jordan, T., Srauy, S. and Miltner, K., 2015. Histories of hating. *Social Media + Society*, 1(2), pp. 1–10.

Svensson, P., 2012. Envisioning the digital humanities. *DHQ: Digital Humanities Quarterly,* 6(1). Available at: http://www.digitalhumanities.org/dhq/vol/6/1/000112/000112.html. [last accessed 15 July 2016].

Taylor, A., 2011. Over 150 arrested after London hit by huge race riots. *Business Insider*, 8 August. Available at: http://www.businessinsider.com/london-riots-mark-duggan-tottenham-2011-8?IR=T [last accessed 5 July 2016].

Tikly, L. and Bond, T., 2013. Towards a postcolonial research ethics in comparative and international education. *Compare: A Journal of Comparative and International Education*, 43(4), pp. 422–442.

TouchGraph, 2014. Mobile application software. Facebook version. New York: TouchGraph. Available at: https://apps.facebook.com/touchgraph [last accessed 11 April 2017].

Tronto, J., 1994. *Moral Boundaries: A Political Argument for an Ethic of Care*. New York: Routledge.

Tsianos, V. and Kuster, B., 2012. Thematic report 'Border Crossings'. *MIG@NET,* 30 October. Available at: http://www.mignetproject.eu/?p=577 [last accessed 25 March 2016].

Tuck, E. and Yang, K.W., 2014. Unbecoming claims: pedagogies of refusal in qualitative research. *Qualitative Inquiry*, 20(6), pp. 811–818.

Uprichard, E., 2015. Most big data is social data—the analytics need serious interrogation. *LSE Impact Blog*, 12 February. Available at: http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/12/philosophy-of-datascience-emma-uprichard/ [last accessed 11 April 2017].

Verhallen, T., 2016. 'Tuning to the dance of ethnography': ethics during situated fieldwork in single-mother protection families. *Current Anthropology*, 57(4), pp. 452–473.

Wajcman, J., 2004. *TechnoFeminism*. Cambridge: Polity Press.

Wilson, M., 2014. Morgan Freeman is dead and other big data stories. *Cultural Geographies,* 22(2), pp. 345–349.

Young, I.M., 1997. Asymmetrical reciprocity: on moral respect, wonder, and enlarged thought. *Constellations*, 3(3), pp. 340–363.

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S.P., Goodman, A., Hollander, R., Koenig, B.A., Metcalf, J., Narayanan, A., Nelson, A. and Pasquale, F., 2017. Ten simple rules for responsible big data research. *PLoS Computational Biology*, 13(3), pp. 1–10.