Complementary Social Science?

Quali-Quantitative Experiments in a Big Data World

Blok, Anders; Pedersen, Morten Axel

Complementary social science? Quali-quantitative experiments in a Big Data world

Anders Blok[1] and Morten Axel Pedersen[2]

Abstract

The rise of Big Data in the social realm poses significant questions at the intersection of science, technology, and society, including in terms of how new large-scale social databases are currently changing the methods, epistemologies, and politics of social science. In this commentary, we address such epochal ("large-scale") questions by way of a (situated) experiment: at the Danish Technical University in Copenhagen, an interdisciplinary group of computer scientists, physicists, economists, sociologists, and anthropologists (including the authors) is setting up a large-scale data infrastructure, meant to continually record the digital traces of social relations among an entire freshman class of students ($N > 1000$). At the same time, fieldwork is carried out on friendship (and other) relations amongst the same group of students. On this basis, the question we pose is the following: what kind of knowledge is obtained on this social micro-cosmos via the Big (computational, quantitative) and Small (embodied, qualitative) Data, respectively? How do the two relate? Invoking Bohr's principle of complementarity as analogy, we hypothesize that social relations, as objects of knowledge, depend crucially on the type of measurement device deployed. At the same time, however, we also expect new interferences and polyphonies to arise at the intersection of Big and Small Data, provided that these are, so to speak, mixed with care. These questions, we stress, are important not only for the future of social science methods but also for the type of societal (self-)knowledge that may be expected from new large-scale social databases.

Introduction

The information regarding the behaviour of one and the same object under mutually exclusive experimental settings may . . ., in an often-used expression within atomic physics, suitably be characterized as complementary, in that – although their description in everyday language cannot be subsumed into one whole – they nevertheless each express equally important aspects of the total sum of thinkable experiences regarding the object. (Bohr, 1957 [1938]: 38; authors' translation)

Such were the words of the renowned physicist Niels Bohr (1885–1962) in a talk he delivered in 1938 to an audience of Danish anthropologists. We cite this passage because we want to ask: how can Bohr's "principle of complementarity" be brought to bear on current interdisciplinary research fields in the social sciences, which comprise both quantitative and qualitative methods and data-sets? Indeed, may Bohr's far-reaching thinking inspire us to conjure a new epistemology as well as a new politics of "quali-quantitative" (Latour et al., 2012) methods for the social sciences?

[1]Department of Sociology, University of Copenhagen, Copenhagen, Denmark
[2]Department of Anthropology, University of Copenhagen, Copenhagen, Denmark

Corresponding author:
Anders Blok, Department of Sociology, Øster Farimagsgade 5, Copenhagen K, 1014 Denmark.
Email: abl@soc.ku.dk

It is important to note that, while Bohr was and still today is hailed for his contribution to modern physics, "[t]he full radical nature of Bohr's views has not always been recognized" by physicists (Pinch, 2011: 6–7). Indeed, within both the natural and the social sciences, there is widespread skepticism if not downright hostility towards bringing together radically different research methods and traditions in the hope of generating new and better insights into complex phenomena. The cross-disciplinary field of "computational social science" is a case in point. Consider the following two citations, taken from the agenda-setting 2009 *Science* paper co-authored by leading researchers from this nascent field, and from a recent column in *Huffington Post* by US anthropologist Paul Stoller (2013):

> In short, a computational social science is emerging that leverages the capacity to collect and analyze data with an unprecedented breadth and depth and scale. (Lazer et al., 2009: 722)

> The problem of Big Data is here to stay, which means that in the coming months and years we'll need a legion of ethnographically trained analysts to produce 'Thick Data' – to save us from ourselves. (Stoller, 2013)

Lazer et al. (2009), of course, echo broader visions (in)famously made on behalf of the scientific "Big Data revolution" such as Chris Anderson's controversial claim about a supposed "end of theory" (2008). On his part, Stoller draws heavily on a blog posting by Tricia Wang (2013), who argues that "ethnographers must engage with Big Data," for fear of being "minimized as a small line item on a budget, and relegated to the small data corner."

Our ambition in this piece is to sketch an alternative to these two mutually exclusive (and mutually hostile) understandings of the role and value of (seemingly) "big" quantitative data and (seemingly) "small" qualitative data methods and approaches in contemporary research and society. To do so, we pose the following questions: how do the *kinds* of knowledge that can be obtained from "Big" (computational) and "Small" (ethnographic) data, respectively, relate or not relate? Indeed, beyond the suggestion of Stoller and "like-minded" social scientists voicing *tout court* concern about Big Data, might the notion of *thick data* (c.f. op cit; see also Boellstorff, 2013) be put to better use to denote what *relates*, rather than what *separates*, innovative approaches to future social science experiments? Invoking Bohr's principle of complementarity, we contend that social relations, as objects of both scientific and everyday knowledge, indeed depend on the type of measurement device deployed for their observation. At the same time, we propose that new productive interferences may be expected to arise at the intersection of Big and Small Data, provided that these are, so to speak, mixed with care.

Below, we explore such questions by way of discussing a specific social science experiment, which attempts to mix computational and ethnographic methods into what might be dubbed "Complementary Social Science," namely the Copenhagen Social Networks Study. In this research project, an interdisciplinary group of computer scientists, physicists, economists, sociologists and anthropologists (including ourselves) has recently embarked on a novel attempt to combine quantitative width and qualitative depth in mapping the spatio-temporal details of a concrete social network.

Social fabric: Remixing mixed methods

The Copenhagen Social Networks Study is a large-scale cross-disciplinary research program jointly hosted by the Faculty of Social Sciences, University of Copenhagen, and the Department of Informatics and Mathematical Modelling at the Danish Technical University (DTU). Our 25-plus member big team collaborates to make continuous recordings of social interactions at all communication channels among an entire freshman class ($N > 1000$) of DTU students, using smart phones distributed to students (as well as to members of the project group) as measurement devices ("socio-meters"). This allows us to digitally map out the "complete" social network of an entire freshman class, including face-to-face encounters via Bluetooth, geo-location proximities via GPS, social network data via apps, and telecommunication data via call logs. The set-up also includes, by way of "embedding" an anthropologist within the freshmen group for an entire year, "thick" ethnographic fieldwork data on friendship and other social relations amongst the same group of students. Simultaneously, researchers track different components of the social fabric via the application of established survey methods (for an overview, see Stopczynski et al., 2014). This vast and heterogeneous collection of data on a large-scale social network infrastructure is being used to study:

(i) How information and influence are transmitted and transformed in the DTU "social fabric"
(ii) How friendships, networks and behaviors form, offline and online
(iii) How the researchers themselves study "Big Data" and handle issues of ethics and privacy

Within this framework, a team of sociologists and anthropologists (senior scholars, doctoral students, and undergraduates) from Copenhagen University, led by the two of us, conduct a joint "Ant-Soc" work package,

which aims to push current boundaries for mixing and cross-fertilizing quantitatively and qualitatively based social network research. We do this by exploring a number of closely interrelated research questions, themes and methods at the core of current concerns in sociology, anthropology, and science and technology studies (STS) – as well as in the cross-disciplinary fields of computational, digital, and experimental social science – including:

- Ethnographic fieldwork – how can ethnographic studies of friendship and other social network relations amongst students enrich or challenge computational approaches – and vice versa?
- Quali-quantitative methods – does the rise of computational social science lead to a reconfiguration of traditional splits between quantitative and qualitative research methods?
- Big data experiments – what kinds of anthropological and sociological experiments does our set-up enable, and how might such innovations enrich existing social-scientific designs?
- Social life of Big Data – what new ethical, political and organizational challenges and opportunities does the rise of large-scale social databases pose to the social sciences and society at large?
- Research collaboration – taken as an object of STS, what may be learned about cross-disciplinary collaboration from the research program itself?

In terms of social science methods, the unique cross-disciplinary and cross-institutional set-up of the Copenhagen Social Networks Study allows for drawing together two hitherto distinct literatures, on computational social science (e.g. Lazer et al., 2009) and mixed methods (e.g. Creswell, 2011), respectively. To the best of our knowledge, researchers in computational social science are yet to acknowledge the potential significance to their work of grounded participant-observation in the ethnographic tradition. Conversely, the mixed methods literature so far has not picked up the challenges and opportunities of large-scale digital data, and is therefore at risk of remaining trapped within an obsolete distinction between qualitative and quantitative (i.e. survey-based) approaches. This methodological remixing – which allows for new ways of stitching together computational and ethnographic data – is what we propose to dub "complementary social science."

Bohr revisited: A complementary social science?

What can we say about the relationship between "big" or "deep" (i.e. computational) and "small" or "thick"

(i.e. ethnographic) data approaches in light of our experiences from the Copenhagen Social Networks Study project so far? To begin with, the collaborative research vision we advocate here differs from the hostile stance towards Big Data by increasing numbers of qualitative social scientists. Valuable as the emerging Critical Big Data Studies Paradigm (as one might call it) undoubtedly will prove to be, it tacitly relies upon and thus reproduces a problematic bifurcation between "hard" quantitative evidence in need of further interpretation and "soft" qualitative data imbued with "the meaning" needed to close this hermeneutic gap. As Boyd and Crawford (2012: 670) write, echoing Stoller's and Wang's positions:

> [I]t is increasingly important to recognize the value of 'small data'. [...] Take, for example, the work of Veinot (2007), who followed one worker – a vault inspector at a hydroelectric utility company – in order to understand the information practices of a blue-collar worker. [...] Her work tells a story that could not be discovered by farming millions of Facebook or Twitter accounts.

While we very much sympathize with Boyd and Crawford's ambition to affirm the lasting value of ethnography and its critical potential, arguments such as the above also tend to reproduce prevailing assumptions that "big" and "small" data and methods are simply mutually exclusive. This runs counter to what we would like to think the Copenhagen Social Networks Study project allows for, which is to make "big" and "small" data mutually dependent and enhancing, and thus potentially recalibrate this and other unhelpful bifurcations between so-called quantitative and so-called qualitative data worlds. That, after all, is what complementarity in Bohr's definition was all about: an epistemological and ontological predicament, whereby two phenomena or processes are at one and the same time totally disparate and totally interdependent. As Karen Barad (2011: 444) puts the point: "Complementarity entails *two* important features: mutual exclusivity *and* mutual necessity. For two variables to be complementary they have to be *both* simultaneously necessary and mutually exclusive. Otherwise, what is the paradox?". Critical Big Data studies, it seems to us, risk losing the paradox of complementarity rather than benefitting from it.

Another problem with such an emerging "critical consensus" amongst qualitative social scientists is its overly narrow understanding of what critique is. Put bluntly, there is something unsatisfactory about reducing the intervention that sociologists and anthropologists like us can make on Big Data realities and discourses to a one-dimensional question of

deconstruction and debunking. Again, we are of course not claiming that allegedly "neutral" incipient hegemonic discourses on, e.g., so-called non-theoretical computational social science can and should not be questioned. All we are suggesting is that too much default Big Data bashing runs the risk of tacitly relying on an assumed vantage point from *outside* these discourses and practices – a quite conservative observational "fly on the wall" position from which the anthropologist and sociologist are taken to enjoy unique access to both her object of study and a broader "context" of which it is supposedly a "part" (cf. Riles, 1998). Take, for instance, Mike Savage and Roger Burrows's (2007) otherwise pertinent points about what they (following Nigel Thrift) dub "knowing capitalism"; surely, we might counter, there is more to the Big Data challenge than merely diagnosing a new form of capitalism?

In the Copenhagen Social Networks Study project, adopting such a conservative notion of critique would entail that physicists like Sune Lehman (the PI from DTU) were located fully *within* the large-scale social network database investigated by our interdisciplinary team, whereas sociologists and anthropologists, such as the two of us, would be positioned *on the outside* of this Big Data reality and discourse. All we would do (or pretend to be doing) as so-called critical anthropologists and sociologists, then, would be to observe purportedly "native" computational social scientists by *looking at them from the outside in*, seeking to identify, monitor, trace, describe and ultimately decode the more or less exotic ideas and practices of this scientific "tribe." Yet, exciting and self-satisfying as this would undoubtedly be, we would thereby end up replicating the dubious assumptions about mutual exclusivity between quantitative Big Data and qualitative small data social science that we set out to transcend in the first place.

Our vision of a complementary social science is meant to offer a way of avoiding such pitfalls. By attempting to mix and even merge our own research agendas, perspectives, and methodologies with our "native" computational social scientific "informants," we have from the onset of the project attempted to position ourselves *within* the large-scale social network database under investigation, not on the *outside* of it – as would be the aspiration of many conventional science studies approaches. For only by in this way strategically striving to partly *collapse* our own research interests with those of the other researchers involved in the Copenhagen Social Networks Study can we hope to move from mutual exclusivity to interdependency in the complementarity between computational and ethnographic approaches. We like to think of this as a germane strategy for gradually

reconfiguring the very epistemological, methodological, and political playing field onto which the recently much-hyped discourses of "Big Data" social science make claim. As an attempt to work around and thus distort what is on the inside and what is on the outside of various conventionalized bifurcations and contrasts (researcher vs. research object; natural science vs. social science/humanities; "hard" quantitative vs. "soft" qualitative data and approaches), our vision of a complementary social science contains the promise of an "immanent," non-skeptical critique (cf. Holbraad et al., 2013) of narrowly technical and neo-positivist celebrations of Big Data in research as well as non-research contexts.

Quali-quantitative methods

In a recent article aptly entitled "The whole is always smaller than its parts," Latour et al. ask:

> […] [I]s there a way to define what is a longer lasting social order *without* making the assumption that there exist two levels [of individuals and structures]. […] Instead of having to choose and thus to jump from individuals to wholes, from micro to macro, [we want to] occupy all sorts of other positions, constantly rearranging the way profiles are interconnected and overlapping. (2012: 591)

To a large extent, this passage captures what the two of us hope to achieve from our quest to laterally assemble into a single research reality (or "research relationality") the two hitherto bifurcated social scientific arenas of Big and Small data: by stubbornly resisting the temptation to perceive the cross-disciplinary Copenhagen Social Networks Study as comprised by two mutually exclusive methodological and epistemological domains – a quantitative and a qualitative one, respectively – we wish to insist on the potential for new and progressive forms of what Latour et al. (2012) dub "quali-quantitative methods." This, however, raises a new order of urgent questions.

It is becoming increasingly clear that, within algorithmically generated Big Data worlds such as the digital database generated in our research program, the "part" is indeed often *bigger* than the "whole," as Latour et al. suggest. This is illustrated, for instance, by the way in which complexity in quantitative social network mappings tends to *increase*, as opposed to decrease, the moment one looks not for "aggregate" static structures but for the replication, say, of ever more fine-grained "temporal motifs" in dynamic interaction patterns across smaller groups (Kovanen et al., 2011). But are such granularities necessarily *the same* in Big and Small data worlds?

By investing so heavily in the promises of new large-scale "digital trace" databases, Latour et al. may risk losing sight of the ways in which disparate data worlds rub off against and emerge from each other, rather than producing new seamless "wholes." Within computational social science, for instance, the focus on granularity "drives forward a concern with the microscopic, the way that amalgamations of databases can allow ever more granular, unique, specification" (Ruppert et al., 2013: 38). Accordingly, in the Copenhagen Social Networks Study project, we hope to *extend* the method of quali-quantitative research by making the very study of social relations – by means of ethnographic fieldwork, surveys, digital traces, and so on – part and parcel of the experimental set-up itself. By doing so, we put ourselves in a position to experiment on social science complementarity in practice: how, we ask, may our experimental settings of observation *themselves* be experimentally varied so as to create different kinds of mutual necessity between Big (computational) and Small (ethnographic) data? What forms of "granularity," "thickness," and "depth" arise from engaging in, rather than simply tracing, concrete work of computational data design, compilation and assemblage, *alongside* ethnographic descriptions (see also Kockelman, 2013)?

In pursuing this kind of experimental approach, new and difficult questions of research ethics emerge, in part because established conventions on how to deal responsibly with issues of privacy, confidentiality, etc., differ widely between (and to some degree also within) computational and ethnographic approaches. Here as well, however, much-needed dialogues and imports should go both ways, as computational researchers stand to learn from routine experiences, on the part of field-workers, of gaining access to "private" layers of people's lives far beyond what should ever be conveyed in "public" research. We need to ask: what could ethically appropriate, and indeed ethically desirable, forms of complementary social science data be?

At this relatively early stage of our collective endeavors, it is still premature to give definite answers to such epistemological and ethical questions. What we can provide, however, is a flavor of what we have in mind. For example: how might the combination of ethnographic embedding and computational data help us understand the way various affective moods, impulses, and rumors spread throughout a collectivity? As Ruppert et al. (2013: 35) suggest, the transactional data favored by most computational social science lends itself to entirely "non-individualist" accounts of social life, where the play of fluid and dynamic trans-actions is the focus of attention. In our setting, one question might be: what role does parties and partying play in the formation of social connections, and how

can we *quantify* the importance of "ambience," "atmosphere," "togetherness," and so on in this respect – "quali-quantities" which, from a standard sociological and anthropological point of view, would be purely qualitative? What, in turn, might such new data assemblages that "stitch together" data worlds produced through computational and ethnographic methods do to established concepts of "personhood," "sociality," and "politics"?

Conclusion

We are well aware that our vision of a complementary social science for the 21st century raises many new questions and numerous potential objections. Apart from the sheer "technical" challenges of remixing different methods, devices, infrastructures, and data forms (ranging from ethnographic field notes to database algorithms) into "thick" data, questions arise at the heart of the philosophy of science. For example: if it is indeed the case that, in computational social science, we leave behind the search for causal statistical modeling and enter a new world of visual "pattern recognition" (Ruppert et al., 2013: 36), then what happens to time-honored distinctions between numbers and narratives; description and explanation; and indeed, simulation and the real world? Such questions, it seems to us, are not just becoming ever more pertinent with ongoing digital social database developments (including recent scares and scandals). They can also only be addressed in the same "messy" interface between basic research, social and political engagement, and collaborative experimentation, which Niels Bohr pioneered in his time.

Furthermore, as noted, our vision of a complementary social science must respond to legitimate and urgent ethical and political concerns – raised by both critics and supporters of computational data – with regard to issues of surveillance, privacy, and future misuse of data. Once again, without claiming to have in any way "solved" these issues, the Copenhagen Social Networks Study team has already taken some important steps in trying to make data available to the research subjects themselves via apps and websites (including the official project website: https://www.sensible.dtu.dk/), as well as more interactive on- and offline forums, and by including members of the project team themselves in the experiment. In this sense, our notion of complementarity extends to the mutual dependencies of researchers and their subjects; to the greatest extent possible, research subjects should be cast as co-producers of knowledge about themselves, just as researchers should strive whenever feasible to render themselves subject to their own research questions and methods. Our collaborative research program,

{}7971707744187089987789788777777777777777777777

---

then, is not just about methods and results; it is also, more broadly or "thickly," an experiment in "data democracy."

Declaration of conflicting interests

The author declares that there is no conflict of interest.

Funding

then, is not just about methods and results; it is also, more broadly or "thickly," an experiment in "data democracy."

Declaration of conflicting interests

The author declares that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements

A first version of this text was presented at the University of Copenhagen's *An Open World: Bohr Conference 2013*. We thank organizer Ole Wæver and discussants during this event. The SensibleDTU project, initiating the Copenhagen Social Networks Study, was made possible by a Young Investigator Grant from the Villum Foundation (awarded to Sune Lehmann). Scaling the project up to 1000 individuals in 2013 was made possible by an interdisciplinary UCPH 2016 grant, Social Fabric (PI David Dreyer Lassen). We thank everyone partaking in the SensibleDTU and Social Fabric projects for valuable input to this text and the research underlying it, including research group leaders: Assoc. Prof. Sune Lehmann (DTU Compute), Prof. David Dreyer Lassen (Economics), Assist. Prof. Jesper Dammeyer (Psychology), Assoc. Prof. Joachim Mathiesen (Physics), Assist. Prof. Julie Zahle (Philosophy), and Assoc. Prof. Rikke Lund (Public Health). Finally, we thank two anonymous reviewers for apposite comments helping us strengthen our argument on complementarity vis-à-vis the mixed methods literature.

References

Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, 16.07, 23 June.

Barad K (2011) Erasers and erasures: Pinch's unfortunate 'uncertainty principle'. *Social Studies of Science* 41(3): 443–454.

Boellstorff T (2013) Making big data, in theory. *First Monday* 18(10). Available at: http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750 (accessed 31 July 2014).

Bohr N (1957[1938]) *Atomfysik og menneskelig erkendelse*. Copenhagen: Schultz Forlag.

Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication & Society* 15(5): 662–679.

Creswell JW (2011) Controversies in mixed methods research. In: Denzin NK and Lincoln YS (eds) *The SAGE Handbook of Qualitative Research*. 4th ed. London: SAGE Publications, pp. 269–283.

Holbraad M, Pedersen MA and Viveiros de Castro E (2013) The politics of ontology: Anthropological positions. (Theorizing the contemporary). *Cultural Anthropology*. Available at: www.culanth.org/fieldsights/462-the-politics-of-ontology-anthropological-positions (accessed 7 July 2014).

Kockelman P (2013) The anthropology of an equation. Sieves, spam filters, agentive algorithms, and ontologies of transformation. *HAU – Journal of Ethnographic Theory* 3(3): 33–61.

Kovanen L, Karsai M, Kaski K, et al. (2011) Temporal motifs in time-dependent networks. Available at: http://arxiv.org/abs/1107.5646 (accessed 7 July 2014).

Latour B, Jensen P, Venturini T, et al. (2012) 'The whole is always smaller than its parts' – A digital test of Gabriel Tarde's monads. *British Journal of Sociology* 63(4): 590–615.

Lazer D, Pentland A, Adamic L, et al. (2009) Computational social science. *Science* 323: 721–723.

Pinch T (2011) Karen Barad, quantum mechanics, and the paradox of mutual exclusivity. *Social Studies of Science* 41(3): 431–441.

Riles A (1998) Infinity within the brackets. *American Ethnologist* 25(3): 378–398.

Ruppert E, Law J and Savage M (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30(4): 22–46.

Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.

Stoller P (2013) Big data, thick description and political expediency. *Huffington Post*, posted 16 June 2013. Available at: www.huffingtonpost.com/paul-stoller/big-data-thick-descrption_b_3450623.html (accessed 7 July 2014).

Stopczynski A, Sekara V, Sapiezynski P, et al. (2014) Measuring large-scale social networks with high resolution. *PLoS ONE* 9(4): e95978.

Wang T (2013) Big data needs thick data. *Ethnography Matters*, posted 13 May 2013. Available at: http://ethnographymatters.net/2013/05/13/big-data-needs-thick-data/ (accessed 7 July 2014).