WHAT DOES A COSMOLOGY OF BIG DATA TELL US?

A. BIG DATA

"Big data describes datasets that are so large, complex, or rapidly changing that they push the very limits of our analytical capability. It's a subjective term: What seems "big" today may seem modest in a few years when our analytic capacity has improved. While big data can be about anything, the most important kinds of big data — and perhaps the only ones worth the effort — are those that can have a big impact through what they tell us about society, public health, the economy, scientific research, or any number of other large-scale subjects."

Joel Gurin, author of *Open Data*

"Big data is a broad term for *data sets* so large or complex that traditional *data processing* applications are inadequate. Analysis of data sets can find new correlations, to "spot business trends, prevent diseases, combat crime and so on."[1] *http://en.wikipedia.org/wiki/Big_data*

B. CLAIMS, TRENDS AND QUESTIONS

"Big data is well on its way to enormous. In 2015, companies will continue to see the evolution of this environment as many have already started down the path of embracing big data to improve revenues, control costs, and find new visions that drive business. The wave of data innovation is far from over."

"Real-time big data isn't just a process for storing petabytes or exabytes of data in a data warehouse. Real-time big data enables you to combine and analyze data from multiple sources so you can take the right action at the right time and right place." http://events.pentaho.com

Big data, Big claims. Quants lick their lips. Research councils rejoice. With so much data on file, cross-cutting so many groups and categories, computed in so many different directions at such high speed, yielding so much pattern and so many correlations, 'win-win' tie-ups are now being envisaged between social science research and the digital interface and, on its basis, between academia and the policy State.

Two questions: (i) does the amplification of data by way of VERY big numbers computed at EXTREMELY high speed mean a fresh role for number and a new ontology of data, or does this development index an extension and acceleration of existing quantitative methodologies; and then (ii) whether ontologically radical or not, do claims for Big Data deepen the epistemic primacy of number in the intellectual account of the social? Indeed, does the rise of Big Data spell the beginning of the end of ethnography and other numerically indifferent methodologies?

On these questions: most advocates of Big Data seem to laud its ability to find correlations on a larger scale than before and in quick time: in which case, for these types of usage, Big Data seems to digitally extend, accelerate and perfect the scope of established quantitative foci. In which case, too, is the hype about Big Data (and the hype is the real concern of this presentation) a deepening of the claim that only social relations numerically expressed can form the object of scientific enquiry? Is large number in social inquiry being afforded even greater epistemic primacy than before? This may indeed be the case with important consequences for the academic reproduction of anthropological research.

On the second question: sitting on the Steering Committee of the ESRC Doctoral

Training Centre, I've become sensitive to the vicissitudes of government policy on what direction the social sciences should take, indexed by the proscribed allocation of research monies. To this end, sequentially, research policy since c.2000 has stressed (i) the increasing primacy of 'thematic priorities'; then (ii) the importance of social impact and collaboration with non-academic bodies; then (iii) AQM (Advanced Quantitative Methods); and now (iv) Big Data. In fact, an ever-increasing percentage of the research budget in the social sciences is now being diverted towards quantitative research designed to support non-academic social policy and practice. Big Data is the biggest, newest component in this quantitative colonisation (of anthropological ground especially) and, consequently, it begins to play a conspicuous role in the marginalisation of qualitative methodologies. Moreover, it is likely that much of the funding for specifically ethnographic projects will be ethnographies of Big Data and other contexts in which actors, organisations, and projects quantitatively compute social relations on the basis of large data sets i.e. much funded ethnography may well morph as meta-quantitative social scientific research.

C. AIMS

My aim here is not to attempt either an ethnography of Big Data or a full methodological discussion of its forms and possibilities (though a comprehensive ethnography of Big Data is surely overdue). Rather, my aim is to conduct a brief ideal-typical comparison of the cosmological place and significance of the processes that underlie number and the relations it indexes. The exercise scratchily highlights by way of theoretically-informed comparison (i) the worldly conditions that promote number as non-data in non-modern social fields, and then (ii) underpin the transformation of number into data (both big and small) in different social worlds.

The aim of this anthropological exercise is to show that situating number cosmologically does indeed help us re-affirm that (i) there can be number without data, and that (ii) when/where data did emerge, it emerged historically on the back of transforming comprehensions of God, creation and the causal interrelations of these prime-moving modalities. And that, if we can accomplish this basic contrast, we can ask in more detail about the novelty of Big Data not only in terms of its methodological possibilities (a non-anthropological angle), but in terms of the sort of consensually projected cosmos its correlations (and other possible derivations) occupy, and the type of strategic relation it (like God elsewhere) establishes in relation to creative and regenerative process. It may be possible to show here that, in cosmological relief, Big Data as a correlational tool continues a familiar modernist agenda (of knowing things with high degrees of certainty) but now in conditions of some epistemic adversity: against the philosophical and cultural grain, so to speak, with severe limitations within the contexts it seeks to authoritatively change.

A cosmological approach to social theory and its methodology may well seem outlandish but, in effect, it's not hard to show that cosmological motifs – on

various scales - have long been central to the classic sociological and anthropological imagination of modern society (recall Weber's in and out-worldly ethics and the in-worldly construction of modern rationality amidst chaotic irrationality at large; Marx's thermodynamic characterisation of internally generated crisis in 'All that's solid melts into air"; Durkheim's typification of Society as the spherical crystallisation of a self-transcending solidarising force in Nature; Sahlins and Taussig's variable cosmologies of capitalism; and finally, Deleuze & Guattari's revaluation of the social as a plane constituted amidst the sort of fundamental dissociations and relations characteristic of inflating universe. If we inform methodical studies of contemporary social relations with similar cosmological insight, what can we propose methodologically for the binary pair Big Data and ethnography?

C. **THE COMPARISON**

1. *No Number, Number No Data*

Two relevant issues arose in field-research I conducted in Fiji, several years ago. The first concerned land in village-chiefdoms (*koro*) and whether or not land was indigenously thought of as property (in which case, it would be most likely described numerically in some way) or whether or not, totally different perceptions of land prevail(ed) that defied metric description. In fact, the land in question was of two sorts: (a) land that fanned out from the village-chiefdom boundaries through the forest hinterland into the hills, and (b) land just outside the village which was used as vegetable and fruit gardens for everyday use and for small-scale market trade.

Land in the forest was distributed amongst patriclans according to their narrated migration patterns. This distribution was on the basis not of boundaries that separated each tract off from the land of other clans, but on the basis of sacred centres (Old Villages that were – and still are - also burial grounds). A clan's land radiated outwards from this venerable centre until it intersected with the outer perimeter of similar land at points that people knew (e.g. a certain tree; a creek; a rock). In people's mind's eye, land in the forest and hills was distributed as a result of migration and subsequent use: it was not the result of any wilfully transcendental act of division or partition of primordial unity. None such division is narrated until Colonial Commissions enacted precisely such a parcellisation of Fijian lands. Consequently, seen indigenously, tracts of land possessed centres but no area, limits but no boundaries; or at least no boundaries in any Euclidian sense. Consequently, too, attributing planar finitude to these estates was impossible and measuring the area was culturally illogical. The colonial government collected this areal data on the basis of their own projections: and, in fact, indigenous groups in the chiefdoms do hold books containing the official information which they would take to the courts in cases where they stand to lose or gain. Otherwise these books remain part of a clan's

hidden patrimony and the local dynamics of land follow more closely enduring pre-colonial geometry than its legal successor.

By contrast, garden land *is* converted into plots, and these finite plots are fenced off, creating boundaries that are both clear and enforced. The principle of accession is that each and every household receives a plot (on marriage) to cultivate from an uncultivated part of the aggregate village stock of garden land (which is itself fended off from the forest of clan estates). No household goes without. As a household grows demographically, it is entitled to take and enclose more garden land, proportional to its need. Crucially, none of this land can be rented out or sold or even loaned gratis. Its boundaries are therefore real enough in the lifetime of the household that has been allotted it. However, should a household leave the village or have no issue; the fences are taken down, leaving the plot to dissolve into the garden reserve. The function of boundaries therefore – both at the level of the village as a whole and the emergent households – is to physically inscribe and economically reproduce the relative autonomy/separateness of entities on the land, rooting/grounding this domestic separateness as an identifiable node on the collective body of the primordial population. Given then the role of boundaries in establishing a certain kind of person (individualised, amoeboid, as a well-defined extrusion of an expanding whole, tied somatically to land which is an equal part of the entity), measuring plot boundaries and determining area numerically makes no sense. These parcels of land come and go, pulse size-wise in tandem with the human groups that are their bodily extension and social incarnation in the chiegdom

A deeper insight into these processes is given by a consideration of certain of the chiefdom's rituals: or, in fact, a recurrent element of them all.