

Wonders Without Number: The Information Economy of Data and its Subjects

Alan F. Blackwell

Manuscript accepted for publication in AI & Society

Many of my computer science students, and even some teaching colleagues, struggle to recognise the epistemological distinction between the words *quantitative* and *objective*. As they work on their research dissertations, inventing the software technologies that will become the basis of the next generation of mobile apps, social media start-ups, and internet infrastructure, they are cautioned that their design work must be evaluated quantitatively. This advice is taken very seriously, even where the goals of the project might be health (quantified), empathetic emotion (quantified), creative arts (quantified), or personal trust and security (naturally, quantified). The conflation of quantification with objectivity can lead to faintly ridiculous research conventions.

If a computer science student were to ask a research volunteer to describe how some new wearable technology made them feel, the verbal response, no matter how subtle or insightful, is likely to be denigrated as purely qualitative or perhaps “anecdotal” data. The recommendation for a student designing such technology is that they should instead ask the volunteer to supply a number, expressing how they feel on a scale of 1 to 7. Aggregations of such *Likert scale* values are preferred as scientific evidence, despite their clear inadequacy by comparison to plain speech, because the numbers are seen as objective – a ludicrous claim considering the perfectly plain fact that the “feelings” being studied are by definition subjective.

It is this strategy of using numbers to avoid human subjectivity that makes computer science and engineering attractive to many young people, with the promise that messy ambiguities of social and emotional life might be resolved through immersion in quantitative study. Those who as children have struggled to understand social nuance, or whose belief systems lead them to expect strictly defined boundaries of classification and behaviour, seem particularly likely to choose such areas of study. And of course, lack of nuance is celebrated by prominent technology entrepreneurs and other extremists who serve as role models to such students. In order to accumulate wealth, every attribute to be valued must ultimately be quantified. If not, how would it be possible to define the conversion rate to dollars?

When viewed from outside, these tendencies of thought are sometimes attributed to the fundamentally binary nature of digital data storage and processing, as the underlying reason for loss of nuance and ambiguity, rather than the desires and motivations of those who build the systems. Although there may once have been a time when computers could be observed at the binary level, decades of evolving complexity mean that this is now the wrong level of abstraction at which to develop critiques of the digital.

Instead, we need to consider the drivers of iteration in terms of Shannon's information theory, which measures in bits the extent to which data is surprising rather than repetitiously redundant¹. Importantly, we must recognise that information has value and cost, and is acquired from individuals at real cost to those persons. Repetition is not a service, because no new information is created. But the consumption of mindlessly repeated statements is costly to every person who must filter noise from knowledge. The commercial dynamic of today's online monopolies is driven by the mundane profit of delivering repetition while consuming attention.

By one analysis, our collective investment in informationally efficient infrastructure has led inevitably to a commercial imperative that rewards iteration rather than understanding. The consequence of reshaping knowledge to fit such an infrastructure has been, as the other contributions to this issue clearly demonstrate, an epistemological shift away from (informationally costly) discourse and consensus to the cheaper alternatives of measurement and quantitative aggregation that are laughably characterised as "artificial intelligence" when in fact they serve only to make us all more stupid.

[911 words]

¹ I refer here to information in the engineering sense introduced by Claude Shannon to describe the relationship between signal and noise, quantifiable in bits to measure meaningful order as opposed to entropy. Information in this sense is a physical quantity, directly related to mass, energy and space in the work of Heisenberg, Hawking and others, although these relations are not yet widely taught in school curricula. The understanding of information as surprise, relative to the human observer, was understood from the outset and explained in Shannon and Weaver's *Mathematical Theory of Communication* (1949). In everyday use of computers, the quantifiable surprising-ness of information can be contrasted with redundancy, and relates to factors such as file compression (files that are more redundant and less surprising contain less information so can be compressed), and predictive text (more surprising text is harder to predict, and can only be offered by larger and more expensive language models). Information as surprise can be related to Bayes theorem, as the amount of information that is gained from an observation, by comparison to the observer's prior knowledge. I have discussed these issues at greater length, for audiences in the humanities and social sciences, in *Objective Functions: (In)humanity and Inequity in Artificial Intelligence*. *HAU: Journal of Ethnographic Theory* 9(1) (2019), 137-146. and *What does digital content mean?* Umberto Eco and *The Open Work* In J. Bardzell, S. Bardzell and M. Blythe (Eds). *Critical Theory and Interaction Design*, MIT Press (2018), pp. 167-185. A comprehensive mathematical treatment, with more detailed explanations of all the aspects alluded to in this footnote, can be found in David MacKay's definitive text *Information theory, inference and learning algorithms*. Cambridge University Press (2003).