

Editorial: Hypothetically speaking

Christopher H Knight*

Faculty of Medical Sciences, University of Copenhagen, DK1870 Frb C, Denmark

The objective of this short Editorial is to briefly examine the hypothesis that scientific data is, in itself, ‘fully capable of advancing knowledge (of lactation science) in a significant way that could provide clear benefit to human society or (dairy) animals’, this being the basic requirement for publication in this Journal. Knowledge is variously defined as familiarity, awareness or understanding of someone or something, or more rigorously as an assured belief based on justification (for the piece of knowledge), truth (of the piece of knowledge) and belief (in the piece of knowledge). The ‘knowledge doubling curve’ (Buckminster Fuller, 1981) suggested that human knowledge doubled approximately once every century until 1900 but by the end of the Second World War was doubling every 25 years. Recent estimates suggest that clinical knowledge is now doubling every 18 months, and there is a proposal that future knowledge expansion will occur at the rate of doubling twice every day, fuelled by the emergence of the Internet of Things. In theory then, and especially if one adopts the familiarity/awareness definitions, human knowledge could expand very easily without the assistance of researchers dedicated to new discovery, simply from observation and connectivity. Please, take a moment to think about the implications and consequences of what has just been said. The knowledge base contributed by the Journal of Dairy Research is contained in around 5000 published articles accumulated from around 90 years of research. *Theoretically*, (if one accepts the 12 h doubling prediction) it would in future be possible to generate 20 000 articles in one day (I place considerable emphasis on ‘theoretically’!) *Hypothetically*, most of the knowledge contained in 20 000 articles would not be assimilated and would rapidly be forgotten by all apart from the machines that by then had learnt to read, store and later retrieve vast amounts of information. Herein lies a distinction that, whilst not based on strict etymology (there is debate about the exact meaning of ‘theory’ and ‘hypothesis’), is nevertheless useful for our purposes. *Theory* may be regarded as a reasonable expectation based on an analysis of accepted knowledge, *hypothesis* should be regarded as relating to a total unknown, but one that is open to being tested. In this case the test (human recall after reading 20 000 articles in one day)

is almost certain to support the hypothesis. Maybe, like me, you will have some knowledge of the Internet of Things, but not really understand what it is (its potential application in healthcare has been reviewed recently: Dimitrov, 2016). Galileo is often regarded as the father of modern science, but the first person to truly apply scientific evidence-gathering and analysis was generally considered to be Thucydides, who lived slightly more than 2000 years earlier, and the expression ‘Knowledge without understanding is useless’ is attributed to him. Is it useless to know that the bovine genome contains around 3 billion base pairs assembled into approximately 22 000 genes of which 14 000 are common to all mammalian species (Bovine Genome Sequencing and Analysis Consortium, 2009)? By providing ‘a window to ruminant biology’ (taken verbatim from the title) that data has enormous potential use and will doubtless provide major benefit, but in itself it is ‘simply’ a long sequence of base pairs. Is it useful to know that a certain percentage of the bovine population of a particular geographical area has a single nucleotide polymorphism at a particular locus on the genome? The answer to this question really depends on the starting point. If a large sequencing exercise threw up this polymorphism more or less at random then the usefulness is, at best, questionable. If, on the other hand, it was previously known that a proportion of this population exhibited a particular attribute, and there was some incomplete knowledge of the genomic regulation of that attribute, then the piece of research that subsequently finds the polymorphism becomes worthwhile. Another way of saying this would be ‘from where does the inspiration come?’ In the former case, the sequencing search is inspired simply by having a technology available, whereas in the second example the inspiration comes from the minds of one or more individuals, the research is hypothesis led. Perhaps reassuringly, it has been suggested that indexing the human brain (the ‘connectome’: Swanson & Lichtman, 2016) would require billions of petabytes of storage, in contrast to the paltry 5 M terabytes that is, approximately, today’s internet. The genome example is an obvious ‘data rich’ research category to use, but there are many others (consider the usefulness of exhaustive analysis of milk fatty acid profiles or Nth degree optimisation of any dairy product process). Data is essential to enable high quality research, and further data will flow from that research, but does it, in itself, provide benefit? I deliberately posed my hypothesis as one to be disproven,

*For correspondence; e-mail: chkn@sund.ku.dk

rather than proven. Statistical proof is, therefore, not possible, and furthermore I have now revealed that this Editorial is bad science because I was assuming an outcome before I tested it! So you, the readers, should be the arbiters. Is my hypothesis supported, or not? Take into account that, if supported, this hypothesis would render the need for further hypothesis redundant; good for machines that learn, not so good for authors wishing to publish in this Journal.

References

- Bovine Genome Sequencing and Analysis Consortium** 2009 The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **24** 522–528
- Buckminster Fuller R** 1981 *Critical Path*. New York: St Martin's Press, ISBN 0-312-17491-8
- Dimitrov DV** 2016 Medical internet of things and big data in healthcare. *Healthcare Informatics Research* **22** 156–163
- Swanson LW & Lichtman JW** 2016 From Cajal to Connectome and Beyond. *Annual Reviews of Neuroscience* **39** 197–216