Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Behavioural Processes 72 (2006) 191–194
BEHAVIOURAL PROCESSES
www.elsevier.com/locate/behavproc

Preface

SQAB 2005: Complexity and generalizability

Sometimes we have posed the following question to Ph.D. students during their doctoral examination: "Assume that you have run an experiment in which 20 data points have been collected, and developed two models. Model A has 10 parameters and accounts for 99% of the variance in the data. Model B has two parameters and accounts for 85% of the variance. Now you plan to replicate the experiment. If you use the parameters estimated from the original data, which model do you expect will account for more variance in the replication data, and why?"

The answer, of course, is Model B. Despite its nearly perfect fit to the original data, Model A contains a relatively large number of parameters and their estimated values will almost certainly have been influenced by unsystematic or chance variation, which will probably not be duplicated exactly in the replication. By contrast, the simpler Model B is more likely to have captured a real (and hence repeatable) relationship in the original data. In the literature on statistical model selection, this is known as the *complexity–generalizability tradeoff* (Pitt et al., 2002). The more complex the model, in terms of its ability to describe a larger set of potential outcomes, the less likely it is that the model's predictions will generalize. Given this tradeoff, optimal policy is to select the model, which is just sufficiently complex to capture the systematic variation in the data, but no more, in other words, Occam's razor.

Modelling has always been at the heart of the SQAB enterprise, and there are some interesting implications of the complexity–generalizability tradeoff when one considers the changing technological environment in which we conduct our research. When the first SQAB was held in 1978, the most common microprocessor CPU of the time for personal computers (the 8086) had 29,000 transistors on a single silicon chip. By 2004, the Itanium 2 microprocessor contained 592,000,000 transistors, an increase of more than 20,000 times in just 26 years. The exponential growth in chip density (with the number of transistors per CPU doubling approximately every 2 years) is known in the computer industry as Moore's law, and illustrated in Fig. 1. Increases in hard disk storage for personal computers have been even more dramatic. According to Kryder's law, shown in Fig. 2, hard disk capacity doubles approximately every 13 months. If this rate of increase is maintained over the next two decades, it has been suggested that an individual could purchase a $100 device which could store, in principle, the entire creative works of every human who had ever lived as well as a real-time video capture of the individual's lifetime! It is clear that we are entering a brave new world, if we are not there already.

One application that has been enabled by the massive growth in computing power is data mining, which has been defined as "the science of extracting useful information from large data sets or databases" (Hand et al., 2001). For many, data mining has acquired sinister, Orwellian overtones, with major examples in the USA being the warrantless electronic spying by the National Security Agency (NSA) and the Pentagon's total information awareness (TIA) program. But data mining is also widely used by companies in order to develop sophisticated and detailed models of consumer behavior, for example, what factors predict an individual's decision to switch from one bank to another. Whether or not 'Big Brother' has finally arrived, one could argue that these applications of data mining are fundamentally non-scientific because they are only concerned with predicting 'useful' outcomes, not with understanding phenomena in general.

Data mining may be a particularly salient example, but overall, the increase in desktop computing power represents a corresponding increase in the potential complexity of behavioral models, as well as the data sets to evaluate those models. Our heuristic for decision making remains unchanged (although not unchallenged): $p < .05$. But because statistical power increases with the square root of N, we have an ever greater ability to detect smaller and smaller relationships that meet the criteria for 'significance'. This places even more responsibility on scientists to make informed, strategic judgments about data, and not to rely on algorithms or formal procedures for decision making. The search for optimal data mining procedures may be scientific, but science is not data mining. Science involves the detection and identification of empirical regularities, and their organization and explanation in terms of laws, models, and theories. It goes beyond mere description of the data to achieve a deeper level of understanding.

Most of the articles in this special issue with the proceedings of the SQAB 2005 conference would not have been possible with the computer technology that was available when our first meeting was held in 1978. But they all illustrate, in various ways, the strategic decision making that is essential for good science.

Psychophysics has always been a central concern for the quantitative analysis of behavior, and two articles continue this tradition. Machado and Arantes describe results of an exper-

0376-6357/$ – see front matter
doi:10.1016/j.beproc.2006.03.010

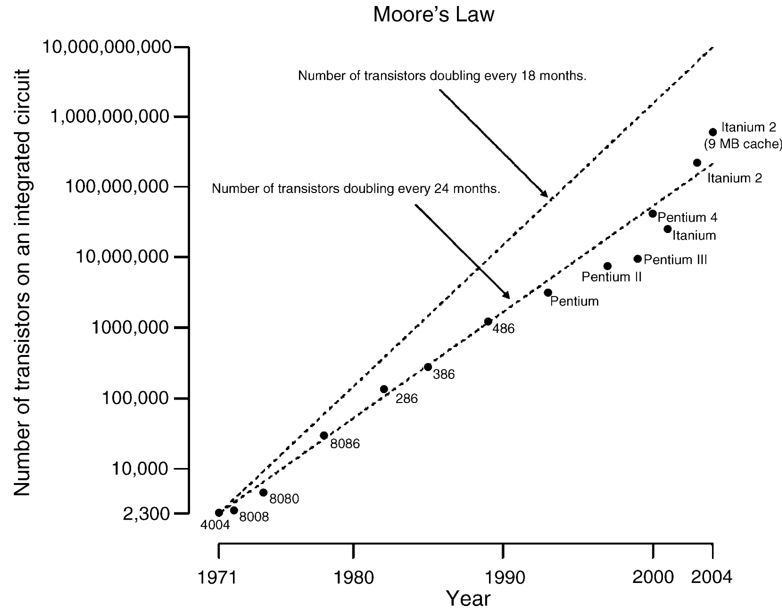

Fig. 1. Exponential increase in density of microprocessor integrated circuits for personal computers (Moore's law). Shown are the number of transistors per integrated circuit for Intel microprocessors released between 1971 and 2004. Data points fall closely on the bottom dashed line, indicating that transistor density has doubled approximately every 24 months. Note: ordinate scale is logarithmic. Figure is in the public domain and downloaded from http://www.wikipedia.org/.

iment using an ingenious procedure in which pigeons were initially trained to make two temporal discriminations: Choose red after a 1 s signal and green after a 4 s signal; and choose blue after a 4 s signal and yellow after a 16 s signal. After pigeons had learned this 'double bisection' task, they were trained on a new discrimination involving the blue and green stimuli and signals between 1 and 16 s in duration. For one group of pigeons, blue choices were reinforced after 1 s signals and green choices were reinforced after 16 s signals, whereas a second group was trained with the reverse mapping. Machado and Arantes show that two prominent models of timing, scalar expectancy theory (SET) and learning to time (LeT) make contrasting predictions for the rate at which the two groups learn the new discrimination. Results supported the predictions of LeT, and suggest that 'what is learned' in a temporal discrimination includes not only which stimulus to choose after a particular delay (i.e., learning about S+) but also which stimulus *not* to choose (i.e., learning about S−).

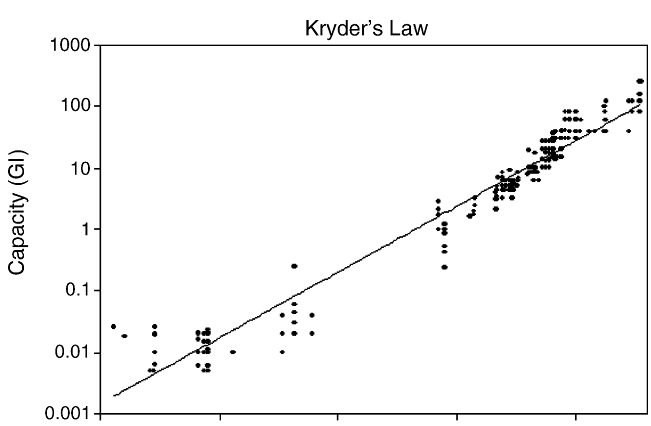

Fig. 2. Exponential increase in density of hard drives for personal computers (Kryder's law). Shown are the size (in gigabytes) of hard drives released by various manufacturers between 1979 and 2004. The linear trend corresponds to a doubling of disk capacity approximately every 13 months. Note: ordinate scale is logarithmic. Figure is in the public domain and downloaded from http://www.wikipedia.org/.

W.A. Roberts describes an experiment based on a simple temporal bisection task, which investigates whether pigeons' representation of time is logarithmic or linear. Of course, this question is related to one of the most fundamental debates in all of psychophysics: Fechner versus Stevens. In Roberts' experiment, pigeons were trained to respond to a red key if the duration of a preceding houselight was between 1 and 8 s, and to respond to a green key if the duration was between 9 and 16 s. There were some revealing asymmetries: near the midpoint, pigeons were more accurate at higher than lower values (i.e., performance was better on 9 and 10 s than 7 and 8 s), whereas at the extremes, pigeons were more accurate at lower values (performance was better at 1–4 s than 13–16 s). Roberts shows that these results follow naturally from predictions based on a logarithmic scale, specifically, the increasing compression in the time scale as values increase, compared with a linear scale. It will be interesting to see whether models based on linear representations (e.g., SET) can provide a satisfactory account of Roberts' data.

An analogy between natural selection mechanisms in organic evolution and the acquisition and maintenance of operant behavior was emphasized by Skinner and in many subsequent treatments. The analogy is based on three features: Variation, selection, and retention. Behavioral variants may be differentially selected by consequences and the new variants retained, perhaps to be further modified through behavior–consequence interactions. In this way, for example, completely new behaviors never seen in the initial repertoire