In Praise of Abundance: Why Individuals Matter in Design Science

David Wastell

In praise of abundance: why individuals matter in Design Science.

David Wastell
Nottingham University Business School
Nottingham, UK
david.wastell@nottingham.ac.uk

Abstract. The Platonic quest for universal principles dominates the mainstream of IS research, typically relegating individual differences to the error term as pet theories and derived hypotheses are put to the statistical test. In design science, this neglect of "particulars" is especially egregious as it wastes valuable information about individuals and their interactions with technology. I present a case study of the design of adaptive automation, which shows how critical such information can be when designing complex IT-based systems. The obsession with theory has gone too far, I conclude; it is time to fight back against the tyranny of universals.

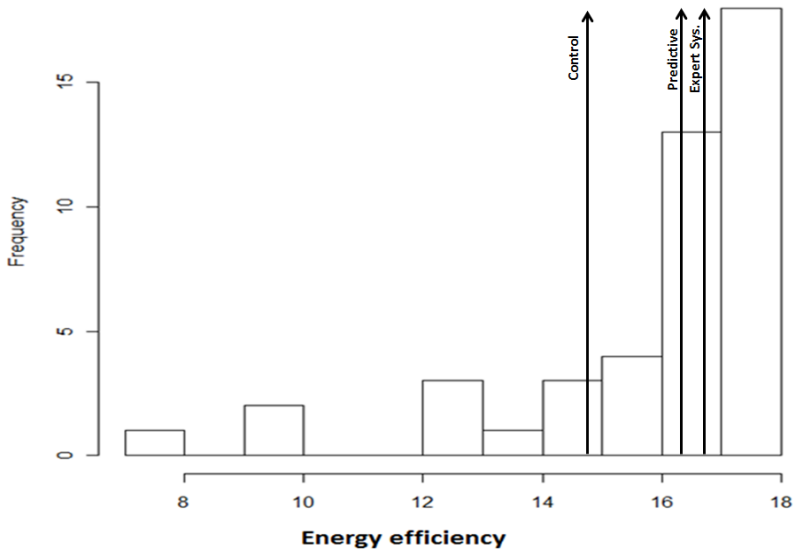
Keywords: Universals, particulars, theory, Design Science, Plato, individual differences, unit-treatment additivity

1 Prologue: The tyranny of universals

Dirt is matter out of place – Mary Douglas

Consider the histogram in figure 1, taken from a study presented at the Working Conference of IFIP WG8.6, in Madrid in 2008 (Wastell et al., 2008). The diagram depicts the performance of 45 participants in a design experiment on decision-support tools for the users of complex systems, in this case a domestic heating system. The main dependent variable was energy-efficiency. The experiment compared three conditions: no support (control), a predictive aid and an expert system (in addition to the predictive aid). The means for these three conditions are superimposed on the histogram: control (14.4), predictive aid (17.3) and expert system (17.6); and the ANOVA summary table for the analysis is shown below the graph.

From the ANOVA, the conclusion was drawn that both forms of decision support enhance performance, but that the expert system does not provide additional support above the predictive aid. On the surface, this is a neat, unexceptionable example of design science in action; in its use of ANOVA to examine the effects of an independent variable, it follows to the letter the conventions of the incumbent IS paradigm. But standing back, we may take a more questioning view, not just of this experiment, but of the epistemological settlement in which it reposes. This is my somewhat daunting task in this philippic.



Source	SS	df	MS	F	prob
Decision tool	46.3	2	23.2	5.10	0.0104
Error	190.7	42	4.5		

Fig. 1 Results of experiment on decision aids

My line of attack is primarily ontological, beginning with a "frontal assault" on the assumption of "unit treatment additivity" upon which ANOVA fundamentally depends. This axiom assumes that the response (y_{ij}) of any "experimental unit" (i.e. human subject) *i* receiving treatment *j* is simply the sum of the unit's individual unique response (irrespective of the treatment) y_i plus an invariant treatment effect t_j. In other words:

$$y_{ij} = y_i + t_j$$

This seemingly innocuous equation is deeply problematic. For the present experiment it assumes that there is a mathematical abstraction denoted by y_i, which corresponds to an individual's intrinsic energy efficiency performance, an in-built measurable property like their height. It further assumes an abstract treatment effect, t_j, which is identical for every individual, e.g. that the predictive aid improves everyone's performance by the *same amount* and that this increment is *additive*. Let me re-emphasize, neither y_i *or* t_j are in the real world, they are metaphysical inventions, ideal types. Exposed and held up to critical examination, they are deeply questionable, if not absurd. In what sense is the quality of "energy efficiency" meaningful as some

constant, intrinsic attribute of an individual? Why should all subjects respond to the same treatment by the same amount? Why is the operation additive; why not some other mathematical form, multiplicative for instance? Why not, indeed!

Statistically, ANOVA further assumes that the variation between individuals follows a normal distribution. This is more metaphysics, though empirically testable in this instance. The test does not turn out well: it is patently clear from the figure that the data are highly skewed. 31 of the subjects (approx. 70%) gain a score of over 16, the remaining 14 scatter in a long trail of declining performance. The overall mean[1] of 15.8 is empirically meaningless; it sits quite detached from the bulk of the data, in no man's land. It is clear from the graph that most people are about as good as each other in performing the task. Drilling down reveals the true picture, that the minority who struggle with the task are concentrated in the control group. 5 subjects in this group perform very badly, though the majority do as well as the subjects in the aided conditions. This is by no means a simple treatment effect. It suggests that the decision aids are useful but only for the minority of individuals who find difficulty with the task.

These metaphysical reflections beg the obvious question. Why is ANOVA so universally used when it makes such untenable assumptions? The answer: because it provides an expedient way of testing the statistical significance of hypotheses, and thus provides the orthodoxy that most follow. Beyond this critique of the standard ANOVA analysis, I wish to make a further point. Not only is the conventional analysis metaphysically implausible, it is also deeply wasteful. So much critical information has been thrown away, about the particulars of individual variation. Yes, the treatment effect is significant, but it only explains a small proportion of the variation. The partial eta ($\eta 2$) of 19.5 is hardly a cause for self-congratulation; it means that 80.5% of variation goes unexplained. From a design science point of view, such information really matters. We need to know what works and what does not, why people differ and how technology can be adapted to individuals. This knowledge is critical especially if large investments are to be made in new technology to increase productivity. For the conventional behavioural scientist, in his platonic quest for universal laws, we can understand why particulars are a nuisance, to be consigned to the dustbin of the error term. But for the design scientist, particulars and universals should be of at least equal priority. That is the argument of this paper, to celebrate the ideographic over the nomothetic, and the heuristic over the hypothetico-deductive. I shall use my recent research on adaptive automation to prosecute the case.

2 Case study – adaptive automation

Over recent years, automation has become a salient area of design research (e.g. Parasuraman and Wickens, 2008) as technological advancement has enabled an increasing number of tasks to be completed by machines that were previously the preserve of humans. Automation designs that flexibly adapt to the needs of the human operator have attracted considerable research interest. Adaptive automation (AA)

[1] The mean itself is a dubious metaphysical abstraction, if regarded as the estimated property of an unseen, inferred (and therefore unreal) "population", rather than a simple descriptive way of denoting the mid-point of a group of numbers.

conveys the idea that tasks can be dynamically allocated between the human and the machine according to operational requirements, with changes in task allocation being based on the human operator's current functional state (Inagaki 2003, Kaber and Endsley 2004). Changes in task allocations are often described in terms of a shift in the level of automation (LOA), drawing on models of automation as proposed by several authors, most notably the seminal model of Sheridan and Verplank (1978) which distinguishes 10 LOAs, ranging from full manual control (LOA1) to full control by the automatic system (LOA10). The rationale for adaptive automation is the potential for balancing out variations in operator workload and the research literature distinguishes between two main types of adaptive automation: implicit and explicit (Tattersall and Hockey 2008). In the *implicit control mode*, the machine decides which LOA is the most appropriate; in the *explicit mode*, this decision is under the jurisdiction of the human. Overall, when compared to static automation, there seem to be benefits of adaptive automation with regard to operator performance, including reductions in mental workload, although there is a considerable degree of inconsistency in the literature (e.g., Inagaki, 2003; Kaber & Riley, 1999; Sauer, Kao & Wastell, 2012).

The work featured here is drawn from long-term programme of research on adaptable automation involving colleagues at the university of Fribourg, Switzerland. There are three distinctive characteristics of this work: first, that we have used the same computer-based simulation in all the studies; second, its psychophysiological nature; third, that task performance has been assessed under adverse working conditions (created by an external stressor, white noise) as well as the optimal circumstances of the typical laboratory experiment. Four studies have been reported to date. The first experiment compared the benefits of static versus adaptable automation, (Sauer, Nickel and Wastell, 2012); although a *preference* for higher levels of manual control emerged, no advantages were found in terms of performance or mental workload. The second experiment investigated different modes of explicit AA: where the operator was completely free to choose, when a prompt was given, and when a decision was forced (Sauer, Kao, Wastell and Nickel, 2012). No salient differences were found between these different regimes. The third experiment compared two modes of implicit adaptive automation (based on decrements in task performance or the occurrence of high demand events) versus explicit AA, where the operator was free to make the change (Sauer, Kao and Wastell, 2012). The results for performance suggested no clear benefits of any automation mode, although participants with explicit control adopted a more active system management strategy and reported higher levels of self-confidence. In the most recent experiment, the effect of system reliability was assessed (Chavaillaz, Sauer and Wastell, 2012). Three levels of automation reliability was compared: interestingly, although unreliability undermined trust, no effects were found on the actual choice of automation level.

The present study returns to the central issue of paper three, the feasibility of performance-based adaptive automation. This type of AA is based on a comparison between current operator performance and a normative criterion. Although an obvious case can be made for using direct measures of primary task performance to provide this criterion, a strongly advocated alternative is to focus on indirect measures of mental workload, using secondary task methodology. Regarding the latter, models of human performance suggest that performance on secondary tasks is more sensitive to

variations in operator workload (Hockey, 1997). This was the approach adopted in that study. It was assumed, on an *a priori* theoretical basis, that this was the optimal approach; but as we have seen, it failed to confer an advantage. Whether such an approach was the best one, or even feasible, was not evaluated empirically. It is therefore difficult to interpret the above null result. Can it be taken to mean that performance-based AA is not beneficial, or does it simply mean that there was a problem with the particular version implemented in that study. Here we explore this issue in depth, using an heuristic, idiographic approach. The data for this investigation will be taken from the high reliability condition of our last experiment (Chavaillaz et al., 2012).

3 Method

A PC-based simulation environment, called AutoCAMS 2.0 (Cabin Air Management System) has been used in all our experiments. AutoCAMS provides a model of a complex process control task, namely the life-support system of a space shuttle. The simulation involves five critical parameters (CO_2, O_2, pressure, temperature, and humidity) reflecting the air quality in the shuttle cabin. When functioning normally, automatic controllers ensure that these parameters remain within a defined target range. When a problem develops, a diagnostic support system is available to provide assistance; it is called AFIRA (Automated Fault Identification and Recovery Agent) and provides five different levels of support, ranging from LOA1 (full manual control) to LOA5, where AFIRA proposes a diagnosis of the fault and an automatic procedure for repairing it.

The main interface is shown in figure 2. Operators have four tasks to accomplish. They are asked to diagnose and fix any system disturbances as fast as possible, and to maintain the stability of the system throughout the experimental session by manual control if necessary. In addition to these two primary tasks, operators had to perform two secondary tasks: a prospective memory task, for which they had to record periodically the current level of the N_2 tank, and an annunciator acknowledgement task, which requires them to click a symbol which appears at irregular intervals (on average about 30 s) to indicate the connection between ground control and the space shuttle.

Thirty-nine participants took part in the original, full study (10 females, 29 males), with an average age of 22.8. The present analysis focused on the 13 subjects in the high reliability condition, in which the automatic systems worked perfectly throughout. Subjects attended the laboratory for two sessions, training and testing, separated by a one-week interval. The testing session lasted approximately 2.5 h (with a 15-min break) and consisted of a sequence of two blocks. Each block was 39 min long and contained five fault scenarios.

4 Results

All control actions performed by the operator and each change in the system are automatically recorded by AutoCAMS for further analyses. The purpose of the present study was to go back to these original log files of individual interactions in order

to test the assumption that secondary task performance provides a reliable and effective basis for detecting changes in mental workload. The connection task was used for this purpose as it had shown a greater sensitivity to workload manipulations in the original study (i.e. it showed a stronger effect of fault difficulty); moreover, it required a more frequent response, giving a finer grained level of temporal resolution. MATLAB programmes were written to process these log files to enable the construction of synoptic graphs giving a detailed record of connection task performance across the experimental session. Figure 3 provides an example; the figure also shows the occurrence of 2 faults, when we may presume that mental demands are objectively higher.

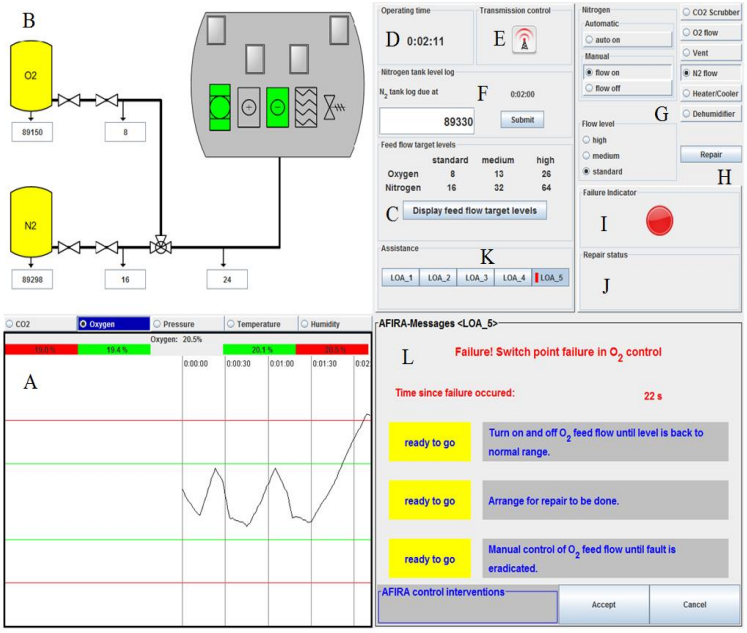


Fig. 2. Main interface of AutoCAMS. Components shown include: (A) history graph for key parameters, (B) functional schema of the cabin (with O_2/N_2 flow meter readings), (C) target levels of flow, (D) system clock, (E) connectivity check icon (secondary task: probe detection), (F) N_2 tank level logging facility (secondary task: prospective memory), (G) manual control panel, (H) repair facility, (I) subsystem failure indicator, (J) repair status (indicates type of repair in progress), (K) control panel of support system, and (L) the support system information display (AFIRA).

For the first analysis, the sensitivity of connection task (CT) reaction time (CTRT) was appraised by examining its temporal profile in response to fault states. For each fault, 3 observation points before the fault and 7 points after fault were extracted. The

serial positions (SP) before the fault were designated -2, -1 and 0, and the seven points after the fault, 1 to 7. Note that SP0 indicated the last CT task before the fault began, and SP1, the first CT after its commencement; this was because the CT task and the occurrence of faults was not exactly synchronised. To reduce the effect of outlying values, a logarithmic transformation was carried out, as is customary for reaction time data. Following this, CTRTs for the same serial position were averaged, giving an overall time profile for each participant. Such profiles will show clearly whether CTRT provides a sensitive and reliable indicator of the additional task demands imposed by the fault. If there is a sharply defined increase in CTRT, this suggest it could form the basis for effective performance-based AA; if there is no change in RT, or if it is inconsistent across individuals, this suggests it would not be useful. Figure 4 shows the time profiles for 4 individuals.

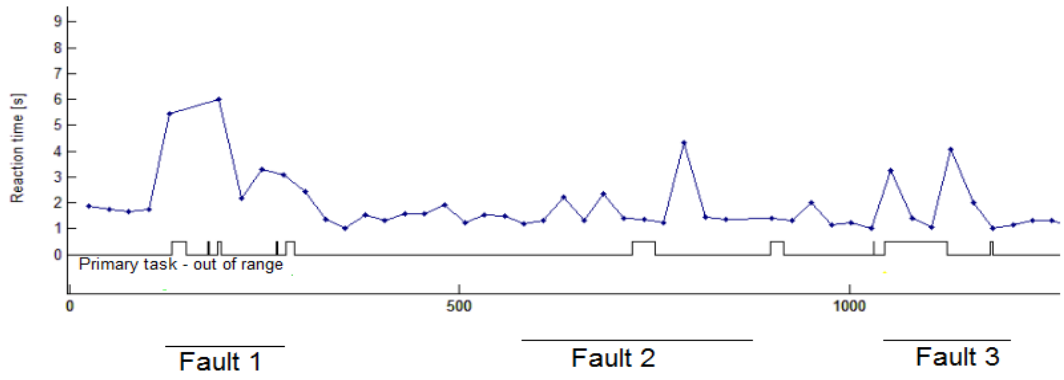


Fig. 3 Example data for 1 subject showing primary and secondary task changes across time and in relation to the occurrence of faults.

That there was considerable variability in time profiles is shown clearly by figure 4. Examining the whole group of 13 subjects, almost all showed a tendency for CTRT to increase, though the time course and amplitude of the trend varied considerably. The average peak time was 95 seconds into the fault. Only 5 subjects showed a marked increase (> 0.1 log units) immediately after the fault (SP1); for the remaining 8 subjects, the increase was less than this, and in 4 cases it was less than 0.05, or stayed constant.

Although far from convincing, this provided some *prima facie* evidence that CTRT is responsive to task demands, i.e. to the additional workload putatively associated with fault handling. Whether a reliable detector could be built is another matter. In order to appraise the how well CTRT could perform in this role, a crude "signal detection" analysis was also carried out. A simple algorithm was designed, similar to that deployed in experiment three of our prior work. CTRT values at any instant were compared to the average across the whole experimental session; an anomalous CTRT was held to have occurred when a certain threshold was exceeded based on the stand-

ard deviation of CTRT during the baseline period. Two levels of sensitivity were compared: 0.5 and 1 standard deviation from baseline. Three parameters were of interest: the number of Hits, i.e. the detector accurately identified raised workload when a fault was present; the number of False Alarms, i.e. raised workload was detected, but no fault was present; and the number of Misses, i.e. a fault was present, but the detector did not pick up any augmented workload. Two performance indices were derived from these parameters:

Accuracy – Percentage of detections that were correct, i.e. hits/(hits + FAs)
Reliability – Percentage of faults accurately detected , i.e. hits/(hits + misses)

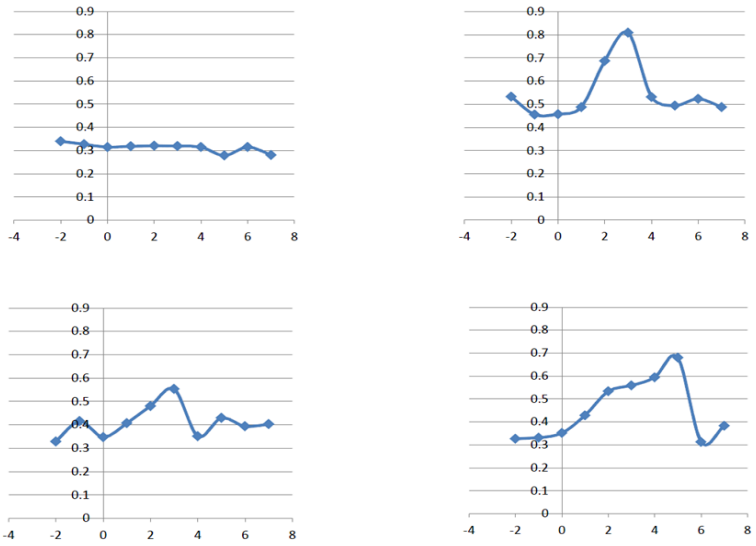


Fig. 4 Time profiles for found representative subjects

Results are shown in table 1. Detection performance is poor. At the 1SD criterion, although accuracy was high (few false alarms), half the faults were missed, producing a reliability of only 50%. Lowering the criterion had the expected effect of reducing the misses and hence increasing reliability, although only to 68%, but at the same time accuracy drastically declined to less than 50%.

As well as secondary task performance, primary task performance was also examined as a potential detector of workload. The following algorithm was used: an abnormal load was inferred when any system parameter (oxygen, pressure, CO2 etc.) was out of its prescribed range for more than 10 seconds. The results for this detector are also shown in the table. It will be seen that this crude approach performed much better in terms of reliability and accuracy. It is notable that the average latency for

picking up a change in workload was 50.3 seconds, which also compares favourably with the time profile of the secondary task.

Table 1 – Detection performance of the various algorithms averaged across the 13 subjects

Detector	HITS	MISSES	FA	Accuracy	Reliability
1 SD, secondary task	5	5	1.4	77.8%	50%
0.5 SD, secondary task	6.8	3.2	7.2	48.8%	68.3
Primary task	9.5	0.5	1.3	87.9%	95%

5 Discussion

> *The universals of Plato are tyrants which 'annihilate' particulars.. (Feyerabend, 1999*)

This discussion is structured in two parts; first, we will reflect on the significance of the specific results of the experiment, and their implications for the design of adaptive automation and for experimental research in this area. I will then return to main theme of the paper, the need to pursue design research in a heuristic mode, which mixes the ideographic and the nomothetic, and to cast off the Platonic yoke.

The results of our experiment are unequivocal. An axiomatic "human factors" principle has been discredited, the assumption that secondary task performance provides a more reliable method for appraising changes in mental workload than primary task performance. This axiom underpinned our choice of the secondary task as the basis of the adaptive automation regime evaluated in experiment 3 of our research programme. Secondary task performance does decline during times of known stress (fault handling) but it is apparent that primary performance is impaired too, indeed it deteriorates more[2]. To the central question of whether secondary task performance affords a reliable basis for AA, the answer is a resounding "no". It was impossible to identify in the graphs of individual subjects, a consistent pattern of secondary task degradation. The detector analysis strongly supported this general conclusion, and indeed showed the primary task to provide a mechanism which worked reliably and accurately enough to support a more feasible algorithm. There is a powerful cautionary tale in this story regarding the limits and seductive lure of theory and the need to challenge dogma. Donald Hebb, in my undergraduate psychology textbook, put it pithily: "theory, like rum, is a good servant and a bad master – this book has emphazised the importance of theory *and the importance of not believing it"* (Hebb, 1972).

[2] In fact, the data of the original experiment show this effect too: two levels of fault difficulty were present; the results showed the effect of novel vs. practised faults was stronger for the primary (F=17.5) than for either secondary task, where a significant effect was only obtained for the connection task (F=4.15).

It is clear from the present study that the use of performance-based methods of adaptive automation is problematic. Alternative methods are available, including the use of psychophysiological measures. Heart rate variability (HRV), for instance, tends to decrease when workload increases (Wickens & Hollands, 2000). The advantage of these methods is that they do not pose additional workload for operator to manage. However, the reliability of such methods is yet to be shown and they are costly to deploy. And this brings us to another critical point. What is important in a design context is whether a technology will work in practice. This imposes a much higher level of proof than theoretical research. It is not enough to demonstrate statistical significance, that some feature or manipulation has an effect; after all, the power of an experiment can arbitrarily be increased simply by running more subjects. But an intervention which requires 100 subjects before it reveals itself is not an effect which is likely to have any interest in a practical setting. Before making expensive investments in technology, the cost-benefit equation must be favourable; there must be significant gains in productivity, sufficient to justify the investment. We wish the majority, if not all, individuals to improve their productivity, a condition nearly met in our opening vignette which showed the value of predictive aids in managing heating systems; a sizeable minority of individuals were performing below par, and the provision of the predictive aid certainly seemed to help, unlike the additional assistance afforded by the expert system.

The goals of what Hevner et al. (2004) call the behavioral science approach to IS research differ significantly from the design science paradigm. For Hevner et al (2004), the aim of "behavioralists" is to develop psychosocial theories which "explain or predict organizational and human phenomena" surrounding the application of technology. Design science, in contrast, seeks to develop a corpus of practically-oriented knowledge through which "the design, implementation, management, and use of information systems can be effectively and efficiently accomplished" (ibid, p. 77). In management science, Van Aken (2005) makes a similar distinction between two forms of research: *Mode 1*, knowledge for the sake of knowledge, aimed at explanation and description; and *Mode 2* which is multidisciplinary and aimed at solving problems. Van Aken goes on to argues for the recasting of management research in the mould of Design Science, rather than conventional explanatory science.

The primary rationale for design experiments, such as the present, is that they generate data from which we can learn to design better. Realistic simulations ("microworlds") like CAMS provide "a valuable tool in the arsenal of design science… for generating realistic behavioral data, testing ideas and developing theory" (Wastell, 1997). But if we are to capitalize fully on this potential, a break with the Platonism of conventional behavioral science is needed. In our search for universals, particulars have been relegated to the error term. Abundance has been conquered (Feyerabend, 1999), but at a price; the wanton waste of important information.

In an early paper using CAMS, I first made the case for the relevance of idiographic analysis in the context of design (Wastell, 1997). The goal of the design experiment, I argued, was heuristic, "theory generating, not theory testing … to use the rich but controllable environment of the microworld to explore complex behavioural phenomena under quasi-controlled [and] ecologically realistic conditions". The idiographic analysis in that earlier study provided some fascinating insights into the difficulties of controlling a system that was complex enough to present a serious chal-

lenge. Some subjects succeeded very well but others manifestly struggled. The weak subjects showed a number of common characteristics: some reacted by withdrawing and adopting too narrow a focus; others responded by taking too much on, throwing themselves into excessive manual control. I likened this dichotomy to the typology of "pathological coping behaviours" (encystment and "thematic vagabonding") observed by another investigator, Dietrich Dörner, with a similar passion for particulars rather than universals. Such an idiographic analysis is not merely an anthropological curiosity; it could have practical value too, suggesting alternative options for implementing adaptive automation. Perhaps a qualitative pattern-matching strategy attuned to detecting symptoms of encystment or vagabonding, might well provide a more effective approach than the measurement of simple quantitative properties of performance. This is something to be explored in future research.

Finally, on a statistical note, it is surely time to give up the black magic of the orthodox ANOVA. There are other ways of testing the null hypothesis, without its dubious ontological and statistical baggage. Testing the null hypothesis simply means evaluating the probability that the difference between the three groups of the heating experiment could have arisen by the chance allocation of subjects to groups. A simple randomisation test would accomplish this[3], without making any statistical or metaphysical assumptions. It provides a direct test of what was operationally done in the experiment – i.e. individuals were actually randomly assigned to three groups and we have compared their average performance. What we want to know is simply whether the performance of the three groups represents a genuine effect. The mean is just one way of characterizing the overall performance of the group, but it is just that, a humble "real world" summary statistic, not a mysterious "population" estimate, hovering spectrally in the background, like Quetelet's *homme type*.

6 Coda: down with Plato!

> *I have sat down with the Entities at table,*
> *Eaten with them the meal of ceremony,*
> *And they were stamped with jewels, and intuned God's ordered praises.*
> *But now the Activities hand me to the dancing,*
> *Brown naked bodies lithe and crimson-necklaced,*
> *And tambourines besiege the darkened altars, In what God's honor?*
>
> Two Methods – Elizabeth Sewell (1960)

Doing experiments on people is an odd business. Consider for a moment, the experiment from the perspective of the "subject". Answering the call for participation, you turn up at the laboratory. In this "strange situation", you are instructed what to do, but given little more information about why you are really here. You are not told to conform to the norm, but you're expected to carry out the task like everyone else, to behave like *l'homme type*. But (unlike the man at the back of the mob in "The Life of

[3] Such a test was carried for the present dataset. Interestingly, 10,000 replications generated the equivalent of 65 F values greater than 5.1, i.e. a two-tailed probability of 0.0065. This is actually more significant than the ANOVA result, whilst making no assumption of normality.

Brian") *you are an individual* and you can only tackle the task in your own way, making sense of what is going on and doing your best. But your individual efforts, however heroic or perfunctory, are of no interest to the experimenter; in his scientific arrogance, only the treatment mean matters, the rest is error, silent error. How very odd is that, you might think – he might at least have asked me what I thought of the experiment. It was about the design of a system, and I had quite a few ideas which might have helped improve it… I did get a lot of training, but I still wasn't sure what to do. But you were not paid to think - you have been infantilized, treated like a guinea pig! But it is the experimenter that wears the motley. Better not to do experiments *on* people; perhaps, better to work with them, with users "as partners and co-producers of design knowledge, rather than passive guinea pigs" (Wastell et al., 2008). In that latter paper, I commented ironically that the current practice of design science, by aping the scientific method, was not itself following well-established precepts of effective design work, e.g. prototyping and user participation. I ended on the chastening thought that, had we worked more collaboratively and iteratively with our users, we may well have produced not only better decision aids (and not spent time and effort on an expert system which users clearly did not like) but more robust theory too.

And a (nearly) final thought. Although my reflections have been directed at the design research, they apply to research in the conventional mode, i.e. behavioural research (Hevner) or mode 1 research, to use Van Aken's terminology. Again, the argument is the same – what a waste of information not to look into particulars to try to understand patterns of individual variation, seduced instead by our Platonic infatuation with universals. Instead of boasting that a significant correlation has been found confirming a cherished hypothesis, we should be more humble. We like to think that a correlation of 0.3, for instance, is impressive. But a correlation of 0.3, means less than 10% of the variation is explained by our hypothesis. Our ignorance (90%) thus exceeds our knowledge by nearly an order of magnitude.

Now the final paragraph in which too much is crammed, but here goes. Any scientific endeavour involves the design of an information system, i.e. a sociotechnical system of people and technology for capturing, processing, and making sense of data. The technology may be very sophisticated, such as the Large Hadron Collider, or mundane, such as a paper-based questionnaire survey. The subject of IS research is IS… so the IS researcher uses IS to study IS; what else was our experiment but an IS. In our perennial angst about the proper subject of our field, some have argued that the technological artifact is what gives IS research its distinctive identity. But this is absurd – it arbitrarily removes from our purview any IS not based on computers! For me the considerations raised in this paper are not esoteric debates about how best to carry on our research[4]; they directly relate to our core business – how best to design an information system. The matter of what technology to use is secondary; epistemolo-

[4] And we certainly worry about how to do research: hence the "method wars" which rumble on to this day, amongst those with time on their hands! But because we tend to see these issues as ones of research methodology, rather than IS design, we limit the applicability of our expertise to our research practice, rather than applying it to the design of information systems in general.

gy[5] is at the heart of IS, i.e. considerations of the best means of producing valid knowledge about the world. This applies whether the IS has been developed by an IS researcher to study IS; or it has been deployed by an organisation to manage customer relations. Research expertise is, at bottom, IS design expertise. What then are the implications of my argument for the practical business of developing "real world" IS. Actually in the real world, the particular fares better; there is greater concern with understanding individual variation (e.g. customer segmentation in marketing) though again there is the same hierarchy of knowledge, the same Platonic tendency to endow statistics such as means and correlation coefficients with superior prestige, because they are held to reveal general, universal principles. This reverence has got to stop. Surely it is time to leave the ceremonial table, to shake the tambourine and join the naked dancers!

7 References

Feyerabend. P. (1999). *The conquest of abundance*. University of Chicago Press.

Chavaillaz, A., Sauer, J., and Wastell, D. (2012). System reliability, performance and trust in adaptable automation, *Submitted for publication to Applied Ergonomics*.

Dörner, D. On the difficulties people have in dealing with complexity. In J. Rasmussen et al. (Eds), *New technology and human error*. Wiley, New York, 1987.

Hebb, D.O. (1972). *Textbook of Psychology*, 3rd Edition. Saunders: Philadelphia.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004). Design science in information systems research. *MIS Quarterly,* 28 (1), 75-105.

Hockey, R. (1997). Compensatory Control in The Regulation of Human Performance under Stress and High Workload: A Cognitive-Energetical Framework. *Biological Psychology, 45*, 73-93.

Inagaki, T., 2003. Adaptive Automation: Sharing and trading of control. In: E. Hollnagel, ed. *Handbook of cognitive task design*. London: Lawrence Erlbaum Associates, 147–169.

Kaber, D.B. and Endsley, M.R., 2004. The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5, 113–153.

Kaber, D.B. and Riley, J.M., 1999. Adaptive automation of a dynamic control task based on secondary task workload measurement. *International Journal of Cognitive Ergonomics*, 3, 169–187.

Parasuraman, R. and Wickens, C.D., 2008. Humans: Still vital after all these years of automation. *Human Factors*, 50 (3), 511–520.

Sauer, J. Nickel, P. and Wastell, D. (2012) Designing automation for complex work environments under different levels of stress. *Applied Ergonomics*, In Press.

Sauer, J., Kao., C, and Wastell, D. (2012) A comparison of adaptive & adaptable automation under different levels of environmental stress. *Ergonomics*, 55, 1-

Comment [w1]:

[5] Of course, to mention epistemology in the context of management is to risk ridicule as an "other-worldy" egg-head, but there have been noble attempts to raise the standard , most notably Stamper (1985).

Sauer, J., Kao., C-S, Wastell, D., and Nickel, P. (2012). Explicit control of adaptive automation under different levels of environmental stress. *Ergonomics,* 54, 755-766

Tattersall, A.J. and Hockey, G.R.J., 2008. Demanding work, technology, and human performance. In: N. Chmiel, ed. *Introduction to work and organizational psychology: A European perspective*. Malden Mass: Blackwell, 169–189.

Sewell, E. (1960). *The Orphic Voice: Poetry and Natural History*, Routledge and Kegan Paul.

Sheridan, T.B. and Verplank, W.L., 1978. *Human and computer control of undersea teleoperators*. Arlington: Office of Naval Research.

Stamper, R. (1985). Management epistemology: garbage in, garbage out. In: Methlie, L.B. and Sprague, R.H. (Eds). Knowledge Representation for Decision Support Systems. Springer, North-Holland, pp. 55-77.

van Aken, J. E. (2005). Management research as a design science: articulating the research products of mode 2 knowledge production in management. *British Journal of Management*, 16, 19-36.

Wastell, D.G. (1997). Human-machine dynamics in complex information systems: the "microworld" paradigm as a heuristic tool for developing theory and exploring design issues. *Information Systems Journal*, 6, 245-260

Wastell, D.G., Sauer, J. and Schmeink, C. (2008). Homeward bound: ecological design of domestic information systems. *IFIP Advances in Information and Communication Technology*, 287, 273-290.

Wickens, C. D., & Hollands, J. G. (2000). Attention, Time-Sharing, and Workload. In N. Roberts (Ed.), *Engineering Psychology and Human Performance* Upper Saddle River, New Jersey: Prentice-Hall, pp. 439-479.