Test User Conference, 1998, 4th-6th June, Brighton, UK. Published in Conference Proceedings, British Psychological Society, pp 11-20.

BPS Test User Conference Keynote Address

Author: Paul Barrett

Title: Science, Fundamental Measurement, and Psychometric Testing

Affiliations: Chief Scientist, The State Hospital, Carstairs, Scotland
Senior Research Fellow, Dept. of Clinical Psychology, University of Liverpool

Abstract:

This paper attempts to put forward the proposition that practically all of the measurement made within occupational psychology is of ambiguous scientific status. However, the definition of what constitutes scientific investigation and quantitative scientific measurement, required in order to permit the forwarding of this proposition, is very specific, and has been summarised comprehensively by Michell (1997). The distinction is further drawn between using a quantitative methodology for pragmatic purposes, and using the same methodology in the pursuit of the scientific explanation. The reason why such a proposition is of anything more than "academic" value is because if we are to make more accurate measurement within the domain of individual differences, and by extension occupational testing, then we need to explore both the underlying philosophy and investigative methods with which we might continue, or which we might adopt in the future. The main consequence of these arguments is that without possibly adopting an approach to investigative psychology that accords to the principles and axioms of scientific measurement, occupational selection and psychometric testing are unlikely to progress much further in terms of greater understanding and prediction of phenomena. By neglecting quantitative causal process attribution for so long, it is possible that our entire conception of individual difference traits is to some extent flawed. It is possible that we have relied too long upon the useful descriptive approximations of "trait" behavioural measurement, but in doing so, neglected the issue of "process". That is, precisely what mechanisms are responsible for a difference on a psychometric test say of extraversion? And how do these mechanisms work, and interact with the environment, in order to produce differential behaviour? However, from the reasoning below, it is also necessary to consider that many variables we are dealing with in the human being are simply not amenable to quantitative scientific analysis, but rather, can be conceived of as non-computable, complex emergent properties of a set of basic underlying cognitive operations. This paper aims to take the reader through these various arguments and debates – at stake is the answer to the question "as *a test user, what do you think is missing from the current generation of tests that might explain why the tests seem to predict behaviour so poorly ?*". This paper provides a snapshot of the kinds of thinking and research that are now being considered by myself, and Paul Kline, in the area of individual differences research.

Michell (1997) published an article in the British Journal of Psychology that overnight challenged the status of psychology as a science and applied psychologists as scientist-practitioners. Michell went as far as to conclude that many psychologists were suffering from a form of methodological thought disorder. The words "mythology", "delusional", and now "thought-disordered" have been used with increasing frequency amongst certain distinguished commentators on the practice and investigative methods within psychology. For example, the title of Robyn Dawes (1994) book "The House of Cards: Psychology and Psychotherapy built upon Myth" was the precursor to a devastating critique of the accuracy of statements made by many clinical psychologists concerning the effectiveness and utility of particular psychological therapies. Before the reader jumps to the conclusion that this book is largely concerned with psychotherapy, note the title: "**Psychology** and Psychotherapy"! Although there are some weaknesses in parts of the critique (Grohol, 1995), the overall arguments based upon much empirical evidence still hold. Ziskin, Faust and Dawes (1995) further challenged the professed accuracy of clinical judgement in a comprehensive evaluation of the role of the clinical (and by extension, the occupational psychologist) as an “expert”. To quote Ziskin et al, Volume 1, (p. 261-262) …

> *“the matter should be obvious. Even if there were not a massive body of research strongly suggesting that clinicians actually have limited capacity to manage complex information and often stumble over a few variables, it should be clear that the obstacles facing the clinician who hopes to integrate all of the data are not merely difficult hurdles, but impossible ones. It is hard to imagine a larger possible discrepancy between what is claimed to have been done and what is actually done and can be done, no matter how sincere these claims might be. A psychologist or psychiatrist, witnessing such a huge difference between self-belief and actual capability, a self-belief that is so implausible when carefully examined and so contrary to the available evidence, might have no hesitancy labelling the believing person delusional. Perhaps it would not be entirely unfair to say that a large percentage of clinicians do in fact evidence a shared myth about their own judgement capacities. As far as we can tell, there seems to be no plausible way one can legitimately support these beliefs on the basis of scientific evidence. Many clinicians, when confronted with the research on judgement capacity, will argue that their reasoning processes are as much an art as a science, and are not really amenable to fair scientific testing. We think that the lawyer can live quite well with this answer. First of all, it is obvious that if reasoning processes cannot be invalidated by scientific testing, neither can they be validated by it. Indeed, if this is not volunteered, we recommend questioning the expert as to whether psychiatry or clinical psychology is an art or a science. The closer to art and the further from science, the less the credibility..."*

This paragraph has especial relevance to those occupational psychologists and representatives of test publishers who argue that “clinical interpretation” of psychometric tests is either desirable or necessary. There are issues in those deceptively simple words that go to the very heart of what constitutes measurement in psychology.

Finally, in this depressing roll-call of substantive, empirically supported criticism, we must not forget Maraun (1996a, 1996b, 1997, and Jackson and Maraun, 1996) who has effectively undermined both the assumed property of factor analysis as a tool for inductive “discovery” of **latent** variables, and empirically demonstrated that the five factor model might simply be a methodological artefact of imposed constraints of the factor analysis model. Neither I suspect will we forget Kline’s (1998) book, “The New Psychometrics”, which is a considered review of the current status and future constitutive components of a new kind of psychometrics. I also detect this same sense of frustration in a recent paper published by another keynote speaker at this conference, Steve Blinkhorn (1997), who states (p. 183) ..

> *“It is hard to escape the conclusion that for all the technical advances and theoretical clarifications of the last half-century, test theory has contributed little of enduring value to the understanding of ability, aptitude, and temperament, or to more effective or credible measurement, that was previously unavailable.”*

It is to be noted that the most damning criticisms from Michell, Ziskin et al, and Maraun, are relevant to much of psychology in general, not just to the applied professions. Further, since both Ziskin et al and Maraun’s arguments seem to be concerned with the extent to which psychology actually derives credibility from its attachment to science and scientific measurement, it is then Michell’s treatise on what constitutes the practice of science, and scientific measurement that is of crucial relevance here. In this regard, it is useful to consider the extent to which current psychometric and psychological measurement in applied psychology is different from that within a quantitative science. If we are to avoid being provoked into rushing headlong into trying to practice science (along with the kind of measurement found in science) we need to evaluate the properties of what constitutes scientific investigation. So, in the next phase, we will first define precisely what is meant by a quantitative science. Then we can stand back and consider whether the augmenting of current practice with that based upon “scientific measurement” is actually optimal as a way forward.

First, let us explore Michell’s reasoning, and the nature of what constitutes the process of science. Michell’s key premise (1997, p.402) is ..

> *“If **science** is taken **realistically** (i.e. as the attempt to understand the ways of working of natural systems), and its successes leave us no reasonable alternative, then a major task for the philosophy of science is to specify the kind of place the world must be, in its most general features, for it to be possible that some scientific theories are **true** (where, by **true** is meant **absolutely true** i.e. things being just as stated in those theories). Applying this to quantitative science (exemplified paradigmatically in physics), the task is to specify the character which quantitative attributes must have if they are both measurable and interrelated continuously”.*

He then deduces that there are two tasks that are the constituent components of a quantitative science. First, a logically prior scientific task of experimentally investigating the hypothesis that the relevant attribute is quantitative, the *scientific* task. Secondly, the *instrumental* task of devising procedures to measure magnitudes of the attribute demonstrated to be quantitative. Michell then defines the axiomatic (i.e. an axiom is a self-evident statement, a proposition that is stipulated to be true for the purpose of a chain of reasoning), properties of quantitative scientific measurement, which essentially devolve into two major laws. That is, firstly, this kind of measurement must possess the property of ordinality, and secondly, the property of additivity. Appendix 1 provides the 9 axioms that formally define the property of quantitative measurement. It is also of use here to distinguish between *extensive* or direct, indirect, and implicit measurement. These three terms also serve to help define the kinds of quantitative measurement that can be made. Extensive measurement (Campbell, 1920, 1928) is concerned with the discovery of ordering and concatenation relations on the objects that directly reflect the quantitative structure of the variable involved. Examples of such extensively measured variables are: length, weight, duration (time), electrical resistance. Within the physical sciences, *indirect* measures of say velocity, acceleration, force, work, etc. are composed of extensive measures (weight, time, length) - where an extensive measure can be loosely defined as one where there is a more or less direct isomorphism between the numerical quantities of a scale, and the property of concatenation of an object being measured. However, as Luce and Tukey (1964) demonstrated, there were a set of axioms that could be derived such that the properties of ordinality and additivity could be inferred for a variable, even though the measures of variables used to make such an inference possessed only the property of ordinality. These axioms defined the property of conjoint measurement (see Appendix 2). A very useful summary of these three kinds of measurement can be found in van der Linden (1994). Examples of conjoint measurement within psychology are rare - one implementation is the Stankov and Cregan (1993) research that examines the hypothesis that intelligence (as proposed to be measured by a Letter Series task) could be considered a quantitative variable, measured conjointly by working memory capacity and motivation, thereby satisfying the 9 axioms given in Appendix 1 (however, measures using the Rasch measurement model also satisfy the constraints of conjoint measurement).

When we compare typical psychometric measures of personality and intelligence, generated using the classical test theory model against these axiomatic conditions of quantitative scientific measurement, we find that we can demonstrate ordinal relations with little difficulty, but that the properties of additivity (here defined specifically as the mathematical addition of our units of measurement, such as the items on a test, and having a one-to-one correspondence to the equal-interval magnitudes of the attribute being measured) is completely missing or at best, untested. This leads Michell to conclude that continuing with such measurement and attributing it to scientific measurement, is in fact, a *pretence* at science. For Michell, there is a clear distinction between the quantitative methodologies that we as psychologists may use to generate numbers, and the manner in which scientists will use these same methodologies as part of a scientific process of investigation.

One way to better conceptualise the issue is to stand back from Michell's arguments and take an even more broad approach as to what constitutes scientific investigation (given we accept Michell's arguments concerning the necessity for axiomatic quantitative measurement). Here we can use a very useful proposition from the philosopher Brian Haig (in preparation), who has isolated a fundamental task for science as "phenomena identification". Both qualitative and quantitative observations can be used here so as to initially identify phenomena. I would call this the first phase of the scientific process. Then, we need to engage in what I would call the second phase, that is, Michell's task of testing both the scientific hypothesis that the variables we hypothesise to define the phenomena actually do possess a quantitative structure, and performing the instrumental task of defining procedures to measure magnitudes of these variables. Finally, I would propose that science is also engaged in a third and final process of causal attribution. **That is, science concerns itself with not just phenomena identification and measurement, but with the understanding of the processes that cause the phenomena to occur**. Within this paragraph lies the problem for both psychology and especially occupational psychology. It is not that the psychometric concepts and constructs we use in individual differences psychology (personality, intelligence, motivation, needs, values etc.) lack pragmatic value, for obviously they do possess this quality. Rather, it is questionable to what extent they possess a scientific value. This is important when we wish to consider exactly what it is that we are measuring with a psychometric test. As Paul Kline has asked in his new book, what precisely is the **unit of measurement** of any current psychometric test? Consider when we make measurement of length, duration, weight, or speed, the units we use to express magnitudes of these attributes are meaningful and possess known properties. We know that 8 seconds is precisely two seconds longer in duration than 6 seconds. The "second" not only has specific measurement properties, but also is meaningful in that it conveys duration implicitly (although the name "second" might just as well be "blurg", as long as we maintain its properties as a unit of measurement of duration). However, when we use a scale that measures say "Independence", what is the unit? If we do not know this, can we ever hope to "understand" the processes by which we will explain the observed phenomena? Involved here is the very essence of the meaning of what it means to say that a person possesses a disposition toward independence, and that this disposition is being measured in a unit of say "indeps".

Having outlined the basis of what constitutes a quantitative science, the reader might reasonably ask "so what?" and "why should it matter to me, as a test user, whether the psychometric test is scientific or not – it is only ever an approximation to a person's attributes and styles, does it really matter if it is not *scientific*, given that it seems to have good practical value?". This is a good question. The answer is bound up with the answer to the very general question of "how are we to improve our measurement (and subsequent predictions) of individuals in an occupational setting". More concretely, as a test user, what do you think is missing from the current generation of tests that might explain why the tests seem to predict behaviour so poorly . An argument is that because psychometric tests have concentrated solely on the classification of self-report behaviours (and in some cases, behavioural performance measures), they are permanently limited in their capacity to exceed a certain level of prediction given they are dealing with a human cognitive system that interacts with the environment to a significant degree. Also, because the environments themselves are almost infinite in their range, solely working on trying to classify environments might also be expected to be of limited utility). I suggest that unless we begin to grapple with the issue of how the human cognitive system functions in order to permit us to observe the classifications so observed, we will never exceed this permanent limitation, simply because we are failing to measure the actual variables (and their processes) that are required to enhance our predictions. Further, this is complicated by the possibility that the classifications we are working with are themselves to some degree misleading. That is, the methods used for the classification have produced clusters of behaviours which conform to the statistical model used, but in fact are no more "real" than those found using a "competing" statistical model. The reason why "science" looks like a way forward is that the investigative method and processes has, within the realms of natural systems, produced knowledge unsurpassed by other means of investigation (whether of literary or mystic forms). It is therefore useful to enquire whether adopting these procedures in psychology, greater knowledge of human attributes will be acquired within psychology, and by extension will enable applied occupational tests to be created that exceed by several orders of magnitude, today's psychometric measures. It is for this reason that I am focussing on the distinction between science and psychometrics. The bottom line is that I would like to know just how much of an "edge" I am gambling for, if I drop research in conventional psychometrics issues.

I propose that occupational and clinical psychology actually exhausted the possible supply of dramatic and innovative test materials some 25 years ago. Essentially, since that time, we have been metaphorically re-arranging the psychometric furniture (items, names, and scales) with a series of increasingly sophisticated measurement methodologies, convincing ourselves that this was progress. In one sense it was, in that practical, utilitarian questions could be better answered (even if poorly). In another sense, one might also argue as Eysenck (1997) has done, that creating taxonomies is an essential component of phenomena identification and therefore, science. However, without ever making precise measurement and without ever seeking to understand the processes by which such taxonomically defined phenomena are seen to occur, it is unlikely that we can ever proceed beyond this ever more sterile cataloguing of phenomena. I conclude on the basis of the arguments above, that we have now reached (or are very near the vanishing point) the limit of the substantive utility in our measures that can be attained using the purely methodological approach to behavioural prediction. I note that Jensen (1997) has come to the same conclusion with regard to the further elucidation of the construct of intelligence.

Eysenck (1997) has set out a very clear model that encapsulates a paradigmatic approach to personality psychology. Figure 1 shows this model.

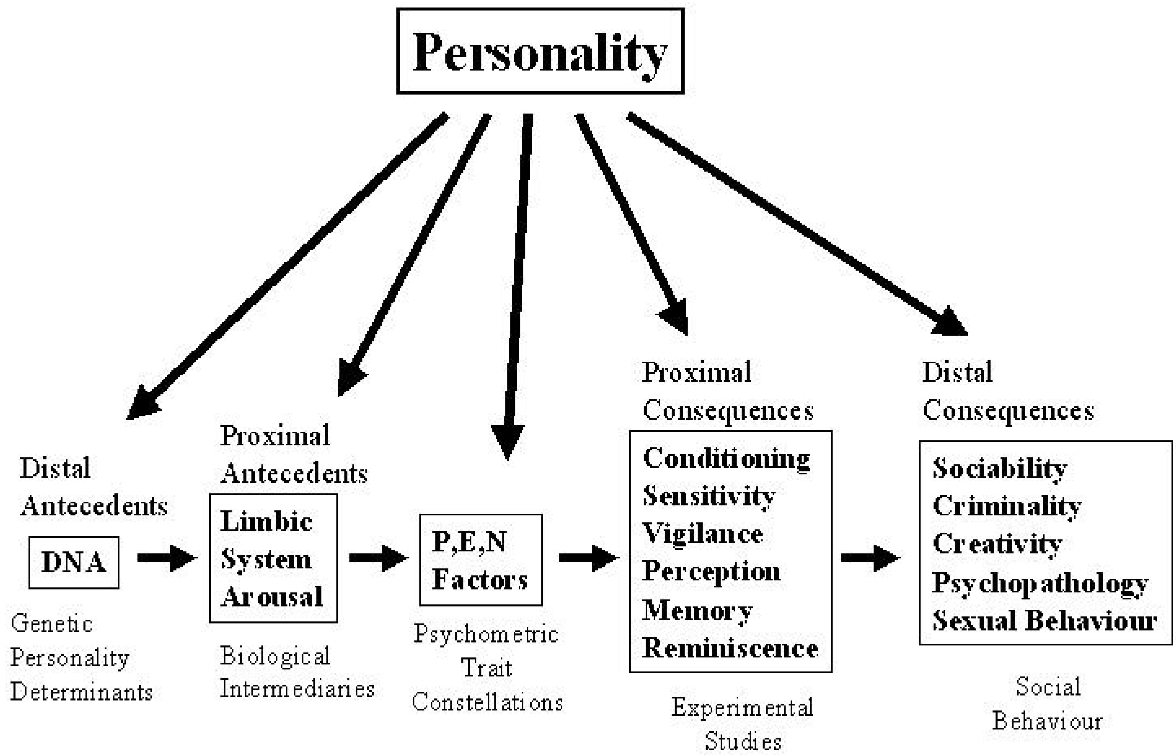

Figure 1: The Eysenck (1997) model of a personality paradigm.

Essentially, this model shows the central position of the personality factors, with antecedent causal variables and the consequent variables aligned either side of the factor representation. However, note the flow of causation from the Distal Antecedent to Distal Consequent. This is an implicit reductionist view.

I have taken another generalised perspective, solely from the standpoint of causal attribution. That is, I have tried to order investigative domains in terms of their proximity to a presumed absolute causal nexus (particle physics). Figure 2 below shows the essential features of this schematic model.

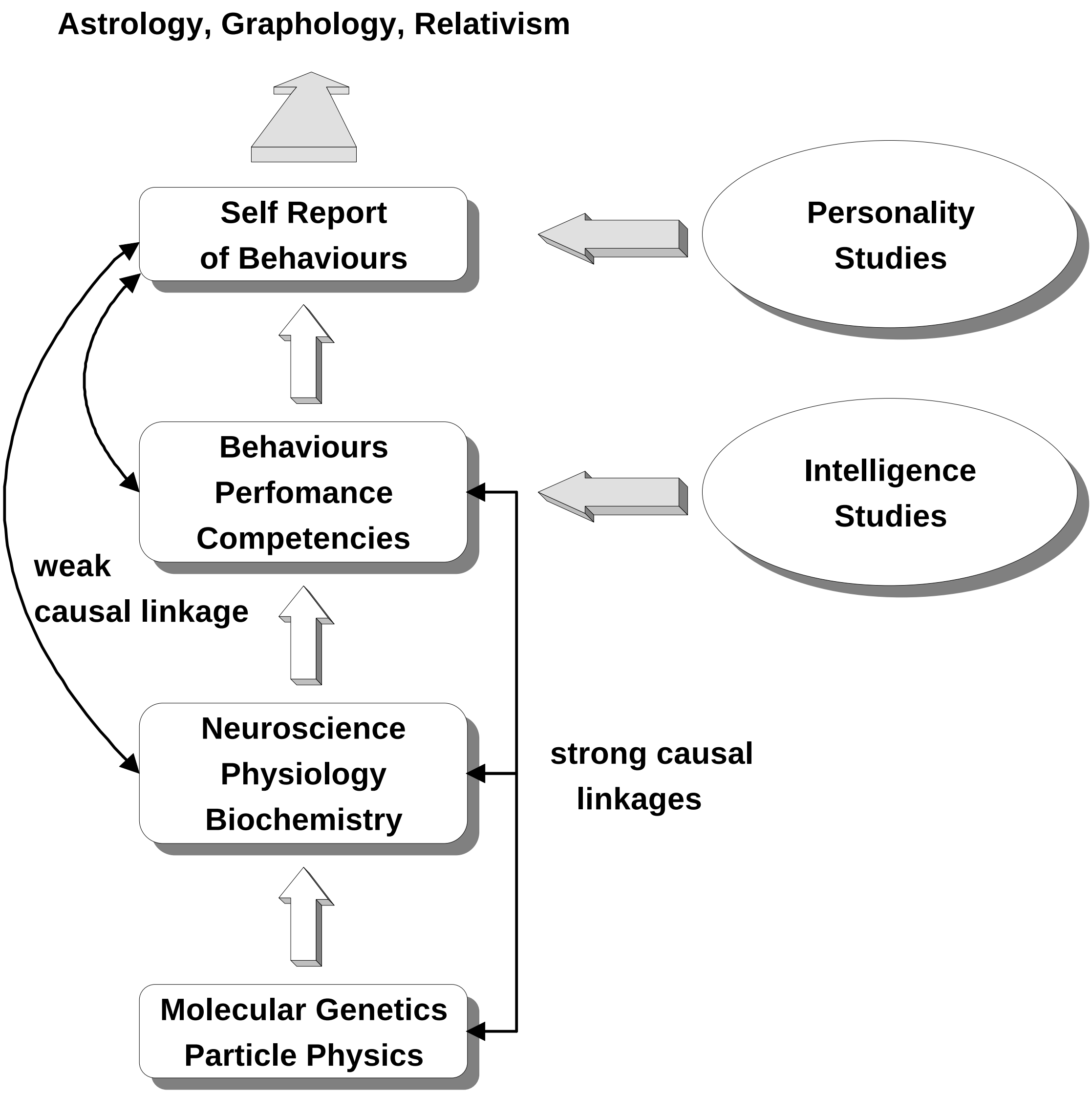

Figure 2: The Barrett schema of the Individual Differences research process.

The key feature of this truly reductionist model is that the further we begin to ignore the process of seeking causal explanation and measurement at a level below the one in which we make our initial measures, the more likely it is that we can create measures of identified phenomena that are mere observational or mathematical constructions. Even if we improve our measurement to conform to the properties as used by other scientists, it is not sufficient in and of itself unless we are also seeking to develop the causal models for those measurements. In this model, psychometric measures of individual differences are shown to be concerned with the measurement of attributes and features of behaviours, without ever seeking proper causal models for these concepts. The reason why intelligence measurement has pragmatic value, but almost no substantive causal theory attached to it is that the only focus for the measurement was an entirely utilitarian one. Likewise with personality measurement. Instead of accepting the fundamental work of Eysenck and Cattell back in the 1960s and early 1970s, then moving on into the scientific activity of seeking quantitative causal models for the proposed existence of the constructs, psychologists and psychometricians have been largely content (with a few exceptions) to simply continue trying to refine the measures and their relationships with "criterion" variables. Within individual differences psychology, some of us are now facing the unpalatable possibility that much of the work and thinking concerned with intelligence and personality might actually be misleading. That is, the very basis of the explanation of intelligence and personality measures as a series of factors/traits may be beginning to unravel, in the light of evidence from cognitive neuroscience, nuclear imaging, and developmental neurophysiology (Quartz and Sejnowski (in press), Posner and Raichle (1997)). Our developing knowledge of the spatial and temporal events associated with brain functioning (from positron emission tomography and functional magnetic resonance imaging) is beginning to show that personality, temperament, and cognitive abilities may well be indivisible from one another, given that there is evidence that the areas of the brain associated with say temperament are also those associated with memory and cognitive/memory processing (Fischer, Wik, and Fredrikson (1997)). The fact that we have carefully separated them in our psychometric tests does not preclude this possibility, because we have not been concerned with measuring the causal processes that generate the outcomes but have rather focussed almost exclusively on the mapping of the outcomes themselves into separable domains. Michael Eysenck's work on Anxiety as a cognitively defined construct (1997, 1998) has far reaching implications in this regard. So do Fahrenberg's (1988, 1992) and Myrtek's (1984) detailed studies demonstrating that there is almost no support whatsoever for a model of anxiety that is related to limbic arousability. I am not saying that the outcome behaviours cannot be differentiated and classified, but, that the biological and cognitive processes which are the causes of these various expressions of behaviours may actually share several fundamental and common precursors. The trait explanation is perhaps no more than a simple artefact of the factor analysis methodology, and as Maraun has indicated, wishful thinking. This is especially relevant since few, if any, fundamental behavioural "traits" have been demonstrated to possess quantitative structure. The trait model has convenience and utility, but is not a scientific explanation at all. A point made clearly by Eysenck (1997). I would propose that as a consequence of this, the measurement we continue to make in this style is never going to get beyond the 0.3-0.5 levels of criterion correlation (unless we severely constrain the criterion).

However, whilst it looks as though developing a research and thinking strategy that accords closely to scientific investigation might be one way forward in the quest to dramatically boost the success of selection testing (and psychology as a whole), there is the problem noted by Penrose (1994), that the serious possibility exists that much of human behaviour is non-computable by ordinary quantitative methods. That is, the human is not some "physical object" that obeys all the laws of physics (including quantum dynamics). Rather, there is some property (or properties) about us that is simply not amenable to measurement or description by any know quantitative methodology or mathematics. An easy example of this is the concept of "consciousness". We can build machines of enormous complexity, but they do not demonstrate "consciousness" as we might find in any example of the human race, including the youngest child. Once again, I hear some readers saying "what on earth has this to do with me as a test user, in June 1997, in Brighton??". Nothing in the immediate context of day-to-day testing, **except** I again ask you the question, a*s a test user, what do you think is missing from the current generation of tests that might explain why the tests seem to predict behaviour so poorly?* As a practitioner, you may or may not have opinions on this, but for someone like myself, a researcher into individual differences, it is the range of possible answers to this question that form the very basis of my working life. As practitioners drew applied benefits from other researchers such as Cattell, Guildford, and Eysenck, so it is hoped future practitioners will draw benefits of equal magnitude from researchers now probing beyond current practice and measurement methodologies (my own response to this question, by the way, will be presented in the talk itself!).

In conclusion, I hope you, the reader, can now begin to visualise the kind of areas and thinking that individual difference research is now beginning to address. What is driving this process is the gradual recognition that it is becoming harder and harder to conceive of generating tests and assessments that will dramatically improve our accuracy of prediction of individuals. As a spur to this process, several substantive professionals (noted above) have begun to question whether psychology’s approaches to its problems are in fact the best approaches. This has required a fundamental rethink on current strategies of research, and on the measurement processes involved in psychometric tests in general. As a test user, you use test scores on a day-to-day basis for the measurement of human attributes. Here I have tried to inform you of the kinds of questions and formulations of problems that are now becoming the focus of the next generation of individual differences research. Whilst these may not be relevant to immediate practice issues, I think they will surely become more relevant as the debate on these issues widens, and as empirical results begin to show their applied promise.

References

Blinkhorn, S.F. (1997) Past imperfect, future conditional: fifty years of test theory. *British Journal of Mathematical and Statistical Psychology*, 50, 175-185.

Campbell, N.R. (1920) *Physics, the Elements*. Cambridge University Press.

Campbell, N.R. (1928) *An account of the principles of measurement and calculation*. Longmans, Green, & Co.

Dawes, R. (1994) *House of Cards: Psychology and Psychotherapy built on myth*. The Free Press.

Eysenck, H.J. (1997) Can personality study ever be objective? In C. Cooper and V. Varma (eds.) *Processes in Individual Differences*. Routledge.

Eysenck, M.W. (1997) Anxiety and cognitive process. In C. Cooper and V. Varma (eds.) *Processes in Individual Differences*. Routledge.

Eysenck, M. W. (1998) *Anxiety and Cognition: A Unified Theory*. Psychology Press.

Fahrenberg, J. (1988) Psychophysiological processes. In J.R. Nesselroade and R.B. Cattell (eds.) *Handbook of Multivariate Experimental Psychology 2nd Edition*. New York: Plenum Press.

Fahrenberg, J. (1992). Psychophysiology of Neuroticism and Anxiety. In A. Gale and M.W. Eysenck (Eds.) *Handbook of Individual Differences: Biological Perspectives*. New York: Wiley

Fischer, H., Wik, G., Fredrikson, M. (1997) Extraversion, neuroticism and brain function: A pet study of personality. *Personality Individual Differences*, 23, 2, 345-352

Grohol, J.M. (1995) A review of Dawes' House of Cards: not dealing with a full deck. Psych Central Website: http://www.grohol.com/hocart.htm

Jackson, S.H. and Maraun, M.D. (1996) Whereof one cannot speak, thereof one must remain silent. *Personality and Individual Differences*, 21, 1, 115-118

Jensen, A.R. (1997) The neurophysiology of g. In C. Cooper and V. Varma (eds.) *Processes in Individual Differences*. Routledge.

Kline, P. (1998) *The New Psychometrics*. Routledge. (published July/August)

Luce, R.D. and Tukey, J.W. (1964) Simultaneous Conjoint Measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.

Maraun, M.D. (1996a) Meaning and Mythology in the factor analysis model. *Multivariate Behavioral Research*, 31, 4, 603-616

Maraun, M.D. (1996b) The claims of factor analysis. Multivariate Behavioral Research, 31, 4, 673-689

Maraun, M.D. (1997) Appearance and Reality: Is the Big Five the Structure of Trait Descriptors? *Personality and Individual Differences*, 22, 5, 629-647

McHenry , R. (1997) Quality standards for developing tests: an alternative to existing orthodoxy. *BPS Test User Conference Proceedings*, 59-63.

Michell, J. (1990) *An Introduction to the Logic of Psychological Measurement*. Lawrence Erlbaum

Michell, J. (1994) Numbers as quantitative relations and the traditional theory of measurement. *British Journal for the Philosophy of Science*, 45, , 389-406

Michell, J. (1997) Quantitative Science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 355-383 (Commentaries by Kline, Laming, Lovie, Luce, and Morgan, with a reply to all by Michell extend from pages 385-406 in the same issue)

Myrtek, M. (1984) *Constitutional Psychophysiology*. New York: Academic Press

Posner, M. W. and Raichle, M.E. (1997) *Images of Mind*. Scientific American Library.

Penrose, R.(1994) Shadows of the Mind. Oxford University Press.

Quartz, S. R., Sejnowski, T. J. (in press) The neural basis of cognitive development: A constructivist manifesto. *Behavioral and Brain Sciences Preprint Archive*. http://www.cogsci.soton.ac.uk/bbs/Archive/bbs.quartz.html,

Stankov, L., Cregan, A. (1993) Quantitative and Qualitative properties of an intelligence test: series completion. *Learning and Individual Differences*, 5, 2, 137-169

van der Linden, W. (1994) Fundamental Measurement and the Fundamentals of Rasch Measurement. In M. Wilson (ed.) *Objective Measurement: Theory into Practice Vol. 2*. Ablex Publishing Corp

Wright, B.D. (1997) *Fundamental Measurement for Psychology*. Memo 64, Rasch Measurement Transactions. Online publication at: (http://mesa.spc.uchicago.edu/memo64.htm)

Ziskin, J.Z., Faust, D., and Dawes, R. (1995*) Coping with Psychiatric and Psychological Testimony, volumes 1, 2, and 3*. Law and Psychology Press

Appendix 1

The nine uniformities of co-existence (J.S. Mill, 1848) reprinted in Michell (1990, p.52):

Let X, Y, and Z be any three values of a variable Q. Then Q is **ordinal** if and only if:

1. If $X \geq Y$ and $Y \geq Z$ then $X \geq Z$ (**transitivity**)
2. If $X \geq Y$ and $Y \geq X$ then $X = Y$ (**antisymmetry**)
3. Either $X \geq Y$ or $Y \geq X$ (**strong connexity**)

A relation possessing these three properties is called a simple order, so Q is ordinal if and only if ≥ is a simple order on all its values. All quantitative variables are simply ordered by ≥ , but not every ordinal variable is quantitative, for quantity involves more than order, it involves **additivity**. This is not a mathematical additivity (as in the operation of arithmetic addition) but instead represents an empirical concatenation operation. This concatenation operation can be arbitrary, but, if we assign numbers to measure quantitative variables, the rules of measurement will map this concatenation operation onto the mathematical operation of addition.

Additivity involves a ternary relation, symbolized as "X+Y=Z". Let Q be any ordinal ordinal variable such that for any of its values X, Y, and Z

4. $X+(Y+Z) = (X+Y)+Z$ (**associativity**)
5. $X+Y = Y+X$ (**commutativity**)
6. $X \geq Y$ if and only if $X+Z \geq Y+Z$ (**monotonicity**)
7. If $X > Y$ then there exists a value of Z such that $X=Y+Z$ (**solvability**)
8. $X+Y > X$ (**positivity**)
9. There exists a natural number n such that $nX \geq Y$ (where $1X = X$ and $(n+1)X = nX + X$) (**Archimedean condition**)

In such a case, the ternary relation involved is additive, and **Q is a quantitative variable**.

Appendix 2

Conjoint Measurement relates to situations of the kind $P=A+X$ or $P=A \times X$ (which can be represented logarithmically as an additive concatenation function). Its application is specifically for those instances where none of P, A, or X is already quantified. It requires that:

1. Variable P possesses an infinite number of values
2. $P=f(A,X)$ (where f is some mathematical function)
3. There is a simple order, ≥ , upon the values of P
4. Values of A and X can be identified (i.e. that objects may be classified according to the value of A and X they possess).

Such a system satisfying 1-4 is a conjoint system. Then if ≥ on P satisfies:

a) **Double Cancellation** (if certain pairs of values of P are ordered by ≥ , then the other particular pairs of values will also be ordered.
b) **Solvability**
c) the **Archimedean** condition

then:
5. P, A, and X are quantitative
6. *f* is a noninteractive function