Br. J. clin. Pharmac. (1984), **18**, 5S–9S

Performance and its measurement

D. E. BROADBENT
Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD.

1 Performance involves numerous sub-functions, and cannot be treated as a unit.
2 Several attempts have been made to separate functions impaired by different classes of drug: for example, high (monitoring) *vs* lower (performing) levels of function, early (perceptual) *vs* later (motor) functions, and strategic *vs* executive functions.
3 The practical implications are that one cannot predict effects of a drug from a single performance test. One must either simulate the practical situation exactly, or use a battery of analytical tests to form a profile of effects. The latter is the policy urged in this paper.

Keywords performance measurement psychotropic drugs

Introduction

As a starting point for the study of performance, it is useful to consider a technique originated by Cohen (1966) for an experiment on ethanol. The same method was later taken up by Brown *et al.* (1969) for looking at the effect of auditory distraction. In both cases the task was for a driver to steer a real vehicle through gaps of various sizes; Cohen (1966) used a bus, and Brown *et al.* (1969) an ordinary small car. As some of the gaps were too small for the width of the vehicle, the driver was told to go round rather than through any gaps that he felt he could not manage. In both experiments, there was no effect on the proportion of times the vehicle was successfully driven through a gap of any particular size. There was however an effect on the decisions to try; the distracted or mildly drunk driver tried to go through more gaps that were physically smaller than the vehicle.

These experiments show in a particularly clear form that it is too simple to treat 'performance' as a single variable. One has to distinguish the case in which people intend to do the right thing, but fail, and the case in which they intend to do the wrong thing, and succeed in the intention. The two experiments happen to show the latter, but one could easily meet cases in which the former was true. For example, if some drug produced a major muscular tremor but left all other functions intact, then the driver might quite well discriminate between possible and impossible gaps, but be unable to hold the car steady in going through them.

More generally we have to regard human performance as made up of a number of separate functions. Conditions that cause a deterioration in one of these functions may leave others unaffected. The hope of the psychologist, of course, is to separate performance into a number of such functions, each of which forms a natural module and can be lost or preserved as a whole. It would be desirable if each of these functions was affected by different drugs, but that cannot be guaranteed. It would also be desirable if one could find tests of each function that were pure, and uncontaminated by the others, and that may be slightly more hopeful. As yet, we do not have them, and some attempts to measure performance proceed in a kind of 'boot-strapping' manner, using the fact that some tests are sensitive to certain variables and other tests to others, as a way of identifying the hoped for modules of function.

There were practical reasons for concern about this problem, because it affects the choice of tests of performance when we have to assess the side-effects of some drug. This practical side will be considered at the end of the paper; in the meantime, let us look at some of the attempted separations of function.

The paced-unpaced distinction

A seminal paper in this field is that of Mirsky & Rosvold (1960). They noted that two closely

similar tasks showed different sensitivities to drugs and other treatments. One test was the continuous performance test (CPT), in which signals were presented at a speed controlled by the experimenter, and the other the digit-symbol substitution test (DSST), in which the subject determined the rate of work. This distinction seemed to be of importance in determining the tests sensitive to different conditions. The DSST was more sensitive than the CPT to effects of barbiturates; but the paced task was more sensitive to effects of sleeplessness or of tranquillizers such as reserpine or chlorpromazine. The distinction between paced and unpaced tasks had already been raised in a non-pharmacological connection by Broadbent (1953), who noticed that a paced task showed a rapid deterioration as a work period proceeded while an unpaced version of the same task did not, in terms of average speed. However, the upaced task showed increasing variability of speed as the person went on working; so that there were at the end of the work period a lot of pauses or very slow reactions, interspersed with faster bursts of work so that the average came out normal. Broadbent (1953) suggested that the paced task picked up the moments of inefficiency without allowing any speeding up to compensate; hence long work periods were more serious on paced tasks.

This analysis helps to explain why investigators of many environmental conditions have found great value in tests of 'vigilance'. In such tasks the person has to watch for signals happening at unexpected times (Mackworth, 1950). Tests in which signals happen only at specified known times, so that the person can make compensatory efforts, are sensitive to fewer stresses.

By the time of Broadbent (1971), this idea had elaborated into a concept of performance as organised at two levels. The lower level handled incoming stimuli and produced responses to them directly, while the higher level monitored the success of the performance, and compensated for any inefficiency that might develop. Hence, paced tasks or those sensitive to moments of inefficiency would be needed to show inefficiency at the lower level, while unpaced tasks might well be more sensitive to effects on the higher level. Tentatively, conditions such as sleeplessness, loud noise, or the administration of amphetamine or chlorpromazine were assigned to affecting the lower level, and we may call them Class 1 conditions. Factors such as time of day, the personality trait of introversion, or the administration of ethanol or barbiturates were supposed to affect the upper level and we may call them Class 2 conditions.

It is clearly too simple to separate tests into paced and unpaced, and expect them to discriminate drugs. Frowein & Sanders (1978) have shown a particular paced task to be *less* sensitive to amphetamine and *more* sensitive to barbiturates than an unpaced version of the task. It is fair to point out that their paced task still allowed the person to react to a signal after the next signal had arrived, so that it would not necessarily pick up brief inefficiencies. The simple distinction of paced and unpaced tasks cannot be maintained; but one could still think of Class 2 factors determining average speed of performance, and Class 1 determining variability.

Another line of evidence used by Broadbent (1971) was that variables in the same class may interact with each other to show reduction or exaggeration of effects on performance. For example, loud noise tends to cancel out the effects of sleeplessness on performance, and so notoriously does amphetamine. These are Class 1 variables. Amongst Class 2 variables, introverts show less impairment of work early in the morning than extroverts do; and also have higher thresholds for sedation by barbiturates. Since 1971, there have been two fresh studies of antagonistic effects that fit the scheme. One of these is that Hartley *et al.* (1977) reduced the effects of loud noise by chlorpromazine. A particularly striking demonstration is that of Revelle *et al.* (1980) that effects of caffeine are opposite in direction in introverts and extroverts. The interaction itself depends on the time of day at which it is measured. All these are interactions between variables in the same class as each other.

On the other hand, cancellation or potentiation effects are less reliable between variables in one class and variables in the other. They do happen: Colquhoun & Edwards (1975) report a reduction of noise effects by ethanol. But the interaction between alcohol and sleeplessness is inconsistent (Wilkinson & Colquhoun, 1968), and so is that between introversion and either sleeplessness or loud noise (Broadbent, 1971). The interaction of noise and time of day is opposite in direction in different individuals (Loeb *et al.*, 1982), and this is inconsistent with a single dimension of arousal/sedation (Broadbent, 1983). On the whole therefore the division into two groups by measuring paced/unpaced performance agrees well enough with that stemming from mutual potentiation or cancellation of effects.

The distinction of stages of processing

More recently Sanders (1981) has put forward a rather different way of separating the functions

that may be affected by various drugs. He uses a notion of Sternberg (1969), that the processing of an incoming signal by the nervous system proceeds through a series of stages such as coding the sensory input, searching memory for relevant information, deciding on a response, and executing the response. Factors affecting the time taken for one of these stages may well multiply each other's effects; when it is hard to recall each item in memory, the number of items needing to be recalled is more important than when it is easy to recall each item. On the other hand, factors acting at separate stages may well have additive rather than multiplicative effects; when it is hard to see the present stimulus, that will not alter the importance of the number of memory items. The idea then is to try the effects of drugs on tasks which remain the same in general but are altered in the difficulty of one stage alone. For example, one can have a choice reaction task in which the stimulus is easy to see, or alternatively is obscured by irrelevant dots. If this alters the sensitivity of the task to a drug, then Sanders (1981) would argue that the drug is likely to affect the initial stage of stimulus encoding. On the other hand, if the task remains the same in its visual component, one could alter the difficulty of the motor output. Suppose the basic task has four response buttons, North, South, East and West, and each button has a corresponding signal. Then one could make it harder by requiring the person to press a button to the South when a signal appears on the East, to press a button on the West when a signal appears on the South, and so on. Now if a drug has more effect on one version of the task than the other, Sanders would argue that the locus of action is in the mechanism that selects responses, rather than in that handling stimulus encoding.

Sanders and his team have in fact found that barbiturates and amphetamine show different sites of action by these criteria: amphetamine gives a larger effect when the response in the task is incompatible or unnatural, whereas the task most sensitive to the action of barbiturates is one with a stimulus that is hard to see.

Although the work of Monk *et al.* (1978) comes from a different theoretical background, it demonstrates that tests involving a heavy memory load are impaired late in the normal working day, while those with a light memory load are at peak performance. Adopting the logic of Sanders, one might suppose that one of the circadian rhythms affects the memory stage. A similar specific effect on memory appears in some recent results on ethanol. It is known that seeing a word produces two rather different kinds of after-affect; it makes it easier to see the word if it happens again with vision degraded in some way, and it also makes it easier to recognise the word if it is shown clearly visible amongst others that have not happened before. These two after-effects are affected by different variables and seem to represent different processes (Jacoby & Dallas, 1981; Tulving *et al.*, 1982). Ethanol impairs recognition memory but not the identification from partial visual information (Parker *et al.*, 1983). Thus we appear to have specific effects on sensory input, on memory, and on response selection. The logical basis of allocating drugs to stages in this way may be questioned (Broadbent, 1984); it is rather different from the use of two drugs or other treatments to cancel each other's effects. However, at the empirical level it is certainly clear that one can find pairs of tasks such that one task is more sensitive to one drug and the other task more sensitive to another drug. Further, there is no inconsistency as yet in the classification of drugs by the two approaches mentioned.

The choice of strategies

A third distinction has been made recently, not so much in the study of drugs as that of noise (Broadbent, 1983). When we test the performance of a human being, there are usually a number of different ways in which the task could logically be performed. If it is mental arithmetic, for example, the person could operate on the most significant digits first, memorise the result, then operate on the next most significant digits, and so on. Alternatively the same task could be done by starting with the least significant digits. Even in simple pattern discrimination, there is evidence that some people may perform by checking each feature of the stimulus successively until they have enough evidence to decide. Other people with the same task will match all aspects of the stimulus pattern in parallel and simultaneously against the possibilities (Cooper, 1980). Each such strategy has its advantages and disadvantages; for instance, if the task contains many repetitions of a single stimulus pattern, checking all the features of that pattern successively is bound to make performance slow, whereas if each pattern presented is a new one it might be very rational to check each feature singly. Thus, in performing a task the person has to choose a strategy as well as executing it once it has been chosen.

In the case of noise, a result that has been shown a number of times is that people find it easier to think of common examples of some category of word, when noise is present. Yet they find it harder to think of rare or uncommon examples (Eysenck, 1975; Millar, 1979; von

Wright & Vauras, 1980). This result has been replicated a number of times; but Smith & Broadbent (1982) showed that it could be reversed, with identical procedures, by using experimental subjects with and without prior experience at recalling examples of categories. In other words, the usual effect must depend on some prior characteristics of the person; probably, the strategy being used to perform the task.

There are a number of similar findings. They raise the possibility that noise and other conditions such as drugs may be affecting the mechanism that selects strategies rather than anything that executes the strategy. An effect of this kind would mean that the person affected would be bad at handling new problems, but having unimpaired perception, memory, or motor control, would appear normal in any test that was presented in such a way as to call out the best strategy easily. It would also mean that the familiar experimental design, in which each person is tested many times under various doses, could be misleading because the conditions of the very first test might determine strategy and therefore performance of that individual throughout all later conditions. In fact, results of that kind have been known since the pioneering findings of Welford *et al.* (1950), and have been particularly common in experiments on noise (Broadbent, 1957; Jerison, 1959).

A comparable emphasis on the importance of changes in strategy rather than in execution can be found in the work of Rabbitt (1979) on the performance of older people. It is probable that the development of better measures of performance, based on video games, will make it easier to test strategy shifts. In the meantime little work of a pharmacological aspect has been done from this point of view, and one can only raise the possibility that tests of specific functions may be missing something important. To adopt a term which this author first heard from A. N. Nicholson, we need tests of 'behavioural integrality'.

Practical implications

In the light of these findings, it is clearly unacceptable to take a measure of, say, reaction time and to assume that the results will have a bearing on the work of a stockbroker, a lorry driver, or a novelist who takes the particular drug under consideration.

There are two main alternative tactics one could adopt. First, one could try and simulate the actual task that the person is going to be doing, with as much accuracy as possible. Second, one could try to build up a profile of the effects of the drug by assessing it with a variety of tests.

At first sight, the technique of perfect simulation seems the most attractive. If this course is followed, there can be no niggling doubts whether perhaps the task involves some function that the test battery has not assessed. There is equally no danger that the interaction or combination of task elements may produce results that one would not expect from the isolated tests. It may well be quicker to run one experiment on a simulation than it would be to test a number of separate functions. Lastly, the visible similarity of the test to the real situation is likely to make it easy to persuade the general public that the results are to be believed.

However, these advantages presuppose that one is in fact able to simulate the real task perfectly. In practice, a task such as car-driving comes in a variety of forms; it is quite different for the driver of an articulated vehicle covering long distances on motorways, for the sales representative trying to locate a strange address from a fast car in heavy traffic, or for a mother collecting children from school and attempting to talk to them while avoiding pedestrians. The differences between these three tasks are along psychological dimensions that we know already to be of possible pharmacological importance.

Further, suppose we could manage to simulate many different kinds of driving. Suppose that in this way we could decide which kinds were safe and which dangerous for one particular drug. We should still be left uncertain whether the results could be taken as applying to aircraft pilots, operators of nuclear power plants, or pharmacists. If on the other hand we undertake the labour of establishing a test profile for a drug, then applying it to some new situation is a matter of analysing the situation to see which functions the task requires.

There may remain some doubts whether a particular function is needed by some task; does the work of a ship's officer require sustained alertness in unstimulating conditions, or coolness in highly stimulating ones? The answer in this case is probably both, on different occasions; for other tasks such problems may be harder to resolve.

Many of us, therefore, would prefer to be guided by the degree of danger in the job, and by the results of analytic tests; and to recommend that people are taken off dangerous jobs if analytic tests show an impairment of some function. This applies even if it is debatable whether that function is involved in the real job. Thus for example it would not be safe to relax pressure against ethanol consumption by drivers merely

because ethanol has (in the results of Cohen (1966)) no effect on the ability to drive through narrow gaps. Given the possible consequences of driver error, there is too high a probability that driving does involve the other functions known to be impaired by ethanol.

This however takes us into the realm of 'political' judgment rather than of science. The key points are that human performance involves many functions, and that drugs may selectively affect some of these even when others are unchanged.

Dr Broadbent is employed by the Medical Research Council.

References

Broadbent, D. E. (1953). Noise, paced performance, and vigilance tasks. *Br. J. Psychol.*, **44**, 295–303.

Broadbent, D. E. (1957). Effects of noises of high and low frequency on behaviour. *Ergonomics*, **1**, 21–29.

Broadbent, D. E. (1971). *Decision and stress*. London: Academic Press.

Broadbent, D. E. (1983). Recent advances in understanding performance in noise. Proceedings of *Fourth International Congress on Noise as a Public Health Problem*. ed. Rossi, G. (in press).

Broadbent, D. E. (1984). The Maltese Cross. A New Simplistic Model for Memory. *Behavioural and Brain Sciences,* (in press).

Brown, I. D., Tickner, A. H. & Simmonds, D. C. V. (1969). Interference between concurrent tasks of driving and telephoning. *J. appl. Psychol.*, **53**, 419–424.

Cohen, J. M. (1966). *A new introduction to psychology,* pp. 182–183. London: Allen & Unwin.

Colquhoun, W. P. & Edwards, R. S. (1975). Interaction of noise with alcohol on a task of sustained attention. *Ergonomics*, **18**, 81–87.

Cooper, L. A. (1980). Recent themes in visual information processing: a selected overview. In *Attention and performance VIII*, ed. Nickerson, R. S., pp. 319–345.

Eysenck, M. W. (1975). Effects of noise, activation level, and response dominance on retrieval from semantic memory. *J. exp. Psychol.*, **104**, 143–148.

Frowein, H. W. & Sanders, A. F. (1978). Effects of amphetamine and barbiturate in a serial reaction task under paced and self-paced conditions. *Acta Psychologica*, **42**, 263–276.

Hartley, L., Couper-Smartt, J. & Henry, T. (1977). Behavioural antagonism between chlorpromazine and noise in man. *Psychopharmacology,* **55**, 97–102.

Jacoby, L. L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *J. exp. Psychol. Gen.*, **110**, 306–340.

Jerison, H. J. (1959). Effects of noise on human performance. *J. appl. Psychol.*, **43**, 96–101.

Loeb, M., Holding, D. H. & Baker, M. A. (1982). Noise stress and circadian arousal in self-paced computation. *Motivation and Emotion*, **6**, 43–48.

Mackworth, J. F. (1950). Researches in the Measurement of human performance. *MRC Special Report Series No. 268*, HM Stationery Office.

Millar, K. (1979). Word recognition in loud noise. *Acta Psychologica*, **43**, 225–237.

Mirsky, A. F. & Rosvold, H. E. (1960). The use of psychoactive drugs as a neuro-psychological tool in studies of attention in man. In *Drugs and behaviour,* eds Uhr, L. & Miller, J. G., pp. 375–392. New York: John Wiley.

Monk, T. H., Knauth, P., Folkard, S. & Rutenfrantz (1978). Memory based performance measures in studies of shiftwork. *Ergonomics*, **21**, 819–826.

Parker, E. S., Schoenberg, R., Schwartz, B. L. & Tulving, E. ((1983). Memories on the rising and falling blood-alcohol curve. (Abstract). *Bulletin of the Pschonomics Society*, **21**, 363.

Rabbitt, P. M. A. (1979). Current paradigms and models in human information processing. In *Human stress and cognition*, eds Hamilton, V. & Warburton, D. M., pp. 115–140. Chichester: Wiley.

Revelle, W., Humphreys, M. S., Simon, L. & Gilliland, K. (1980). The interactive effect of personality, time of day, and caffeine. A test of the arousal model. *J. exp. Psychol. Gen.*, **109**, 1–31.

Sanders, A. F. (1981). *Machine pacing and occupational stress*, eds Salvendy, G. & Smith, M. J., pp. 57–64. London: Taylor & Francis.

Smith, A. P. & Broadbent, D. E. (1982). The effects of noise on recall and recognition of instances of categories. *Acta Psychologica*, **51**, 257–271.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donder's method. In *Attention and performance II.*, ed. Koster, W. G., pp. 276–315.

Tulving, E., Schacter, D. L. & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *J. exp. Psychol: Learning, Memory & Cognition,* **8**, 336–342.

von Wright, J. & Vauras, M. (1980). Interactive effects of noise and neuroticism on recall from semantic memory. *Scand. J. Psychol.*, **21**, 97–101.

Welford, A. T., Brown, R. & Gabb, J. E. (1950). Two experiments on fatigue as affecting skilled performance in civilian aircrew. *Br. J. Psychol.*, **40**, 195–211.

Wilkinson, R. T. & Colquhoun, W. P. (1968). Interaction of alcohol with incentive and with sleep deprivation. *J. exp. Psychol.*, **76**, 623–629.