

1

Gideon's army: the study of individual differences

In Old Testament times the Israelites were preparing for a war with the Midianites. Gideon was general of the Israelite army, and he had an unusual problem – too many volunteers. He needed to reduce their numbers, but not at random; he wanted to keep experienced and courageous soldiers. Gideon used two rough and ready tests. He first reminded his volunteers how great was their risk of death and injury: 'Whosoever is fearful and afraid, let him return and depart early.' About two-thirds of the volunteers changed their minds and went home. But 'the Lord said unto Gideon, The People are yet too many'. Gideon told the remaining volunteers to drink from the nearest stream. Those that lapped the water like a dog 'putting their hand to their mouth' passed the test, but 'all the rest of the people bowed down upon their knees to drink water', and failed, because experienced soldiers keep watch for enemies at all times.

Gideon used first a personality test, then an aptitude test. Neither was very good by psychologists' standards. Both are single item tests, so unlikely to give very reliable results. Gideon used one question, *Are you afraid of fighting a battle?* and one observation of fieldcraft (keeping watch while drinking). Gideon's personality test is the first recorded measure of what psychologists now call neuroticism or trait anxiety. The military have remained intermittently aware of individual differences and the need to assess them ever since. (Only intermittently: Cattell (1937) complained that the Army in 1914–18 'used some of the best brains from civilian life to stop bullets in front-line trenches'.) Modern methods of military selection are much more sophisticated than Gideon's two tests, but preserve the same distinction between *personality* (fearfulness) and *ability* (fieldcraft). This book deals only with personality; for a good review of ability, see McIntosh (1998), and of psychological testing, see Murphy and Davidshoffer (2005).

Definitions of personality

What personality is NOT. In the scientific sense of the word everyone has a personality and, in theory, all personalities are equal. No personality is better or

Gideon's army

worse than any other. The lay person talks of John 'having no personality' and Jill 'having a very bad personality'; the lay person means that John has little or no social presence, and that the way Jill behaves will make her and/or those around her unhappy.

What personality IS. Cattell (1965) defined personality very simply as 'that which permits a prediction of what a person will do in a given situation'. Definitions of personality conventionally exclude purely physical differences, such as height or strength, although these obviously affect personality. Most definitions (but not Cattell's) also exclude mental abilities. A neglected area of research is how intelligence may interact with personality; the aggressive but intelligent person can destroy enemies subtly and legally, while the aggressive but unintelligent person can think of nothing better than hitting them, so gets put in prison. Other exclusions are more specific: attitudes and interests, such as following professional football or voting Conservative, and demographic differences, mostly linked to social class.

Stagner (1961) distinguished *stimulus*, *response* and *intervening variable* definitions of personality.

Stimulus definitions focus on the impression the person creates on others. However, people create different impressions on different others at different times, so a stimulus definition implies people have multiple personalities. This tends to be the view of many social scientists, who see each person as a collection of roles: employee, parent, partner, neighbour, etc., without necessarily a central core. They follow the same line of thought dealing with issues like aggression, dealt with in Chapter 10, seeing behaviour as elicited by outside social pressures, and being suspicious of explanations in terms of what the person is like.

A typical *response* definition (Guthrie, 1944) said personality is 'those habits and habit systems of social importance that are stable and resistant to change'. Guthrie saw personality as true patterns of consistency in the individual's behaviour across a range of situations, but Chapter 3 shows that attempts to find such consistencies have proved only partially successful. Guthrie's definition slips in an element of fudge with the words 'of social importance', and begs a question by restricting it to 'habits and habit systems'.

Intervening variable definitions try to get below the surface, and away from the detail of what someone does, or how they appear to others; Allport's is typical: 'the dynamic organisation within the individual of those psychosocial systems that determine his unique adjustments to his environment'. The key words are 'organised systems within'. The important question – central to any account of personality – is, what form do these organised systems take? Are they traits, or factors, or habits, or motives, or complexes, or personal constructs, or cognitive affective units, or self-concepts? Most personality theories follow the intervening variable approach.

Tests, testing and discrimination

The first modern personality measure was the Woodworth Personal Data Sheet of 1917, intended, like Gideon's test before it, to weed out recruits to the army who could not stand the stress of battle. Since then personality tests of all types have proliferated, and continue to do so. Some examples will be presented in more detail, and some problems discussed in Assessment Boxes. A first point to note is that personality tests are designed to do something many are very wary of: to discriminate between people. But some dimensions of inequality are more equal than others; some are more or less tolerable to egalitarian thought, some are suspect, and a few are anathema. George Bernard Shaw remarked that the fact that one horse can run faster than another is a source of endless wonder to many and a source of great profit to a few, but offends hardly anyone. A belief in individual differences in mental ability, by contrast, is about as acceptable to most liberal opinion as a commitment to apartheid. Most personality measures occupy the middle ground: regarded with suspicion, but tolerated, at least until a biological basis is postulated.

Statistics Box 1.1 Mean and standard deviation

The psychologist who has tested 30 people for extraversion can calculate two statistics. The first is the average or mean, which is familiar to everyone. The second is more specialised. The standard deviation does two things: (1) it describes how one person compares with another, (2) it summarises the variability of the whole distribution. Standard deviation is usually abbreviated to SD.

The SD can be used to describe someone's height, without reference to any particular system of measurement. A man 6'2" high is 2 SDs above the mean. Anyone who understands statistics will know how tall that is, be the local units of height metres, feet and inches, or cubits.

A distribution is completely summarised by its mean and SD, so long as it is normal, i.e. bell shaped and symmetrical. (Distributions of some natural scores, like height, are normal; distributions of psychological test scores are made normal.)

Nature of the information collected

Personality assessment can use five qualitatively different types of information, listed in Table 1.1.

Gideon's army

Table 1.1 Five categories of qualitatively different information, used in personality research

self	information provided by person <i>self-rating, personality questionnaire, honesty test, projective test, interest questionnaire, interview, application form</i>
reported	information provided by other people <i>ratings by friends, family, peers' references</i>
demonstrated	the person performs a task or demonstrates a skill
(a) test	<i>emotional intelligence test, information test</i>
(b) behavioural	<i>group exercise, behavioural test</i>
recorded	the person has made a recorded achievement <i>college grades, criminal record</i>
involuntary	<i>graphology, drug use testing, polygraph, implicit association testing, psychophysiology</i>

Self-report evidence. Information that is provided by the person him/herself. Some self-reports are free form or unstructured, for example, some interviews or job application forms: 'Tell me about yourself in your own words'. Others are more structured, such as personality questionnaires (PQs), or structured interviews. Self-report data have some compelling advantages for the researcher; they are generally very cheap and very convenient. However, self-report also has two major drawbacks.

- (1) *Social desirability.* People may not always tell the truth about themselves, especially if they stand to gain something from the assessment: getting a job, getting a place on a course, getting out of prison. (Getting out of prison is possibly the most crucial event to depend on psychological assessment. Not getting a job is a disappointment, but one can always apply for another. 'Failing' the psychological test that secures release from confinement is much more final.)
- (2) *Self-insight.* People may not always know the truth about themselves. People may genuinely think they are good leaders, or popular, or creative, and incorporate this view of themselves into their PQ or interview. However, by any other criterion – test, others' opinion, or achievement – they lack the quality in question.

These problems make it important to confirm what people say about themselves by information from other sources.

Other report evidence. Information about the person is provided by other people, usually by ratings, occasionally using their own words. Other reports vary in

the degree of expertise involved. Some require no special expertise, such as peer ratings and the letter of reference. Others use experts, generally psychologists. Information about personality of young children often comes from parents or teachers. The advantages of other report data include getting information from people who know the person very well. The main disadvantage is that the others may not be willing to divulge what they know.

Demonstrated evidence. The person performs a task or demonstrates a skill. Demonstrated evidence of personality may be obtained by group exercises, where people can try to work together to solve problems, or by role-playing exercises, e.g. persuading your neighbour to make less noise. Demonstrated evidence cannot generally be faked. On the downside, demonstrated evidence tends to be much more difficult and expensive to collect. There is an interesting contrast between the areas of personality and intelligence here. The intelligence test is precisely that: a test, with right and wrong answers, and a time limit. It is not a self-report. Intelligence could be assessed by self-report, so why are these not used? Because it is assumed people either do not know, or will not say. Yet personality research is content to rely on self-reports to a large extent.

Recorded evidence. Some information can be characterised as recorded fact. The person has been voted most popular student of his/her year, or has sold a great deal of insurance, or has several convictions for serious assault. The information is recorded and verifiable.

Demonstrated and recorded information tends to have an asymmetric relationship with self- or other reported information. Evidence that someone cannot do something disproves the claim by him/her or by others that he/she can. However, the converse is not true: being told that someone cannot do something does not disprove demonstrated or recorded evidence that he/she can. To this extent, demonstrated and recorded evidence is superior to self- and other reported evidence, which implies that researchers should prefer demonstrated and recorded evidence.

Involuntary evidence. Some evidence is provided by people, but not from what they say, nor from things they do intentionally. The classic example is the polygraph, intended to assess the person's truthfulness from respiration, heart rate and electrodermal activity, not from the answers he/she gives. In fact, the polygraph is used to decide which of the person's self-reports to believe and which to classify as untrue. Another involuntary assessment is graphology, which seeks to infer someone's personality from the form of their handwriting (not from its content). *Implicit association* testing is widely used, especially to detect thoughts people might be reluctant to admit, such as dislike of minorities. If someone dislikes, e.g., people from Ruritania, they will be slowed down in a task that requires them to associate Ruritanians with anything desirable. The procedure is described in Chapter 14.

Gideon's army

It is highly desirable in any research to use at least two measures of personality, and to choose them from different types of evidence. So the researcher on child aggression will ideally get information from both the child's parents (other report evidence) and from observing the child's behaviour in the school playground (demonstrated behaviour evidence). Getting all your information from the same source can create problems. Consider, for example, a research project on the effect of parent behaviour on a child's personality, in which the parents tell the researcher how they deal with the child, and supply the information about the child's personality. They might be strongly influenced by their views about the 'right way' to bring up children, which might distort what they see. It would be better to get the information about personality from a PQ completed by the child, or from recorded data on aggressive acts. The technical name for this problem is *common method variance*. It is a particular problem with self-reports, which have a built-in tendency to agree or correlate (Statistics Box 1.2). Getting information from several different sources is, however, more time-consuming and expensive.

Statistics Box 1.2 Correlation

Height and weight are correlated; tall people usually weigh more than short people, and heavy people are usually taller than light people. Height and weight are not perfectly correlated; there are plenty of short fat and tall thin exceptions to the rule (Figure 1.1).

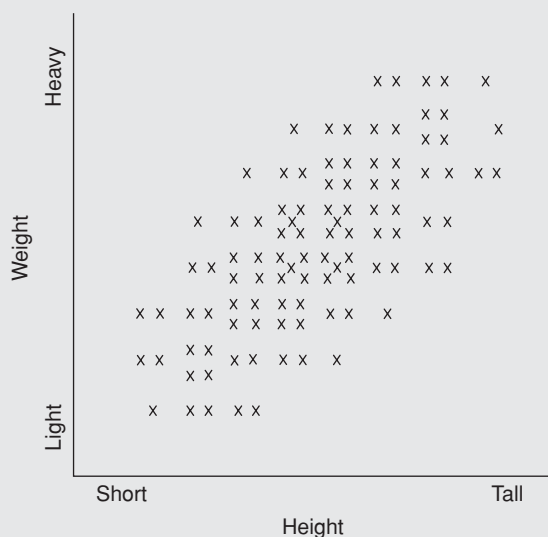


Figure 1.1 Height plotted against weight, showing a positive correlation of 0.75

The correlation coefficient summarises how closely two measures like height and weight go together. A perfect one-to-one correlation gives a value of +1.00. If two measures are completely unrelated, the correlation is $-(0.00)$. Sometimes two measures are inversely, or negatively, correlated: the older people are, the less fleet of foot they (generally) are.

Fair employment law. Most people know it is against the law to discriminate against certain classes of person when employing staff. These protected classes include women, ethnic minorities and disabled people. In the USA people over 40 are a protected minority, while in Europe discrimination on grounds of age is illegal throughout the age range. Most people think discrimination means deciding not to employ Mr Jones because he is black, or Ms Smith because she is female. Direct discrimination is illegal, but is not the main concern in assessing personality when selecting people for employment or education. The key issue is indirect discrimination, or *adverse impact*. Adverse impact means a selection system results in more majority persons getting through than minority persons. For example, some UK employers sift out applicants who have been unemployed for more than six months, on the argument that they will have lost the habit of working. The Commission for Racial Equality argued that this creates adverse impact on some ethnic minorities, because their unemployment rates are higher. Adverse impact assesses the *effect* of the selection method, not the *intentions* of the people who devised it. Adverse impact means an employer can be proved guilty of discrimination by setting standards that make no reference to ethnicity, gender or age. Adverse impact is a very serious matter for employers. It creates a presumption of discrimination, which the employer must disprove, possibly in court. This will cost a lot of time and money and may create damaging publicity. Selection tests that do not create adverse impact are therefore highly desirable, but unfortunately not always easy to find. Differences between male and female are often found in personality assessment; most questionnaire measures find that women report higher average levels of anxiety than men (Sackett and Wilk, 1994). Links with age are also found: for example, extraversion scores on the Eysenck Personality Questionnaire fall steadily between ages 16 and 70. The *d* statistic (Statistics Box 1.3) shows the difference is a large one. A selection programme that set cut-offs on personality tests, designed to employ non-anxious extraverts, would tend also to employ more young persons, and more men, which would give some 75% of the working population grounds for complaining of discrimination. (Note that the preceding text says 'women report higher levels of anxiety', NOT women are more anxious. One hypothesis about the difference suggests men will not admit anxiety, because it makes them appear weak.)

Gideon's army

Statistics Box 1.3 d statistic

The d statistic describes the size of a difference between groups of people. The section on fair employment noted that there are age differences in extraversion, and gender differences in reported anxiety. The d statistic indicates how big these differences are, by computing how many standard deviations separate the means. For age differences in extraversion, the computation is 16–20 age band mean by 61–70 age band mean divided by standard deviation of the whole data-set, which gives a value of 1.00. For male–female differences in anxiety, d is smaller at 0.44, meaning male and female averages differ by just under half a standard deviation. Sometimes the value of d is very small. The differences in average extraversion between British and Australian people has a d value estimated at 0.18. Values of d below 0.20 are considered small, values above 0.80 are considered large, and those in between, medium.

Reasons for studying personality

Psychologists have three main reasons for being interested in personality:

- to change people;
- to assess people;
- to gain scientific understanding.

The second and third are applied, whereas the first is theoretical. There is often a gap, in sympathy and understanding, between theoretical and applied psychologists. The former see the latter as technically incompetent, overconfident, either financially motivated or woolly idealists; the latter see the former as ivory tower academics, afraid of the real world, forever finding reasons never to put their ideas into practice, preferring to spend their time picking holes in the efforts of others.

Changing people. The clinical psychologist has a succession of clients coming through his/her door, needing help. The clinician must achieve sufficient understanding of each client to predict future progress, and choose a form of treatment. The clinical approach tends to emphasise motivation, and to see the client as a meaningful whole. It tends to the *idiographic* approach, which seeks to understand the individual as an individual, not as a point on a dimension. Some clinicians have to make vitally important, but very difficult, decisions: is it safe to release this person from prison? Is this person fit to work with children? Will this person kill him/herself?

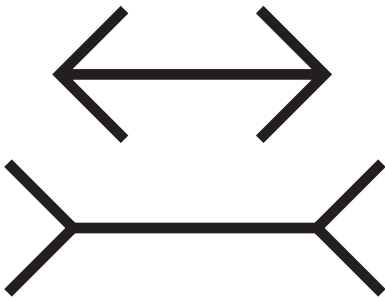


Figure 1.2 The Muller-Lyer illusion

Assessing people. Everyone who selects staff for employment, or students for university, or anyone for anything, is an assessment psychologist; the vast majority are amateurs, but some are trained professionals. The assessment psychologist needs no formal theory of personality, just a measure that predicts outcomes successfully. Constructing and testing new measures is a highly technical business, even though the theoretical complexities of selecting managers or police officers may be minimal. Research on assessing personality of job applicants is reviewed in more detail in Chapter 14.

Gaining a scientific understanding. People are not like machines: they do not all perform exactly the same. Psychologists should not complain about this, especially not personality psychologists, because differences between people in how they react and behave is the personality psychologist's domain. It does mean personality psychology is not like the hard sciences, where one single experiment on a chemical reaction or physical phenomenon is – in theory – sufficient.

Eysenck (1966) described a study testing the hypothesis that *massed practice* (numerous tests at very short intervals) would increase the strength of the Muller-Lyer illusion (Figure 1.2). The results were highly variable, so the small difference averaged across 50 participants was not statistically significant (Statistics Box 1.4). People who take part in psychological research used to be called *subjects*. Now they are usually called *participants* (Ps). Only 1% of the variation in Ps' susceptibility to the illusion resulted from the use of massed as opposed to spaced practice, meaning that massing or spacing made virtually no difference. It is a poor theory of visual illusions that accounts for only 1% of the variation in the effect, but Eysenck argues that any theory that ignores individual differences necessarily limits itself. A later study compared two rates of presentation in a serial learning task. Again people varied enormously; some did better when the interval between items was two seconds, some when it was four seconds, and some did equally well on either. But this time the Ps had completed the Maudsley Personality Inventory, which showed that extraverts and stable people were equally proficient at either speed, whereas introverts and anxious people made nearly twice as many errors

Gideon's army

when they had to work fast. Because people vary so much in how they react, how they see things, how they behave, it is always necessary to collect data from large numbers, and to use statistical analyses to see if any differences are real or could arise by chance. Eysenck suggests that assessing personality will help explain some of the apparent random variation.

Statistics Box 1.4 Statistical significance

Suppose a researcher was interested in height in different parts of Britain. The researcher has heard it said that people in Wales tend to be shorter. The researcher measures five Welsh people and finds an average height of 1.6 metres, then measures five people from Bristol who average 1.8 metres. Is this a sound piece of research? Clearly not, in quite a few ways. Statistical significance is one of these. How can we be sure a difference between two averages is real? It obviously depends on how big the difference is, and how many people the average is calculated from, and how much variation in height there is. Statistical significance testing calculates the likelihood of a difference between two averages being found by chance, given the size of the difference and the amount of variation there is. If the likelihood is less than 1 in 100, the researcher can conclude the difference is statistically significant, or roughly speaking, that there is a true difference.

Variation between people in psychological research is called *error variance* (Statistics Box 1.5). It is a nuisance to the experimental psychologist, because it means testing 10 or 20 or 50 people instead of one. If every human being behaved the same way, the experimental psychologist's task would be much easier. As Eysenck notes, most experimental psychologists treat individual variations as unsystematic, as error. A lot is random, but Eysenck argues that some can be accounted for by personality, especially by extraversion and anxiety. The experimental psychologist's nuisance is the individual psychologist's *raison d'être*. Eysenck proposes that general experimental psychology and individual psychology should integrate, and suggests analysis of variance designs can achieve this integration.

In practice the three themes of assessment, therapy and theory become interwoven. Approaches that start out as theoretical become applied; Cattell (Chapter 2) originally devised the 16PF personality questionnaire to determine the structure of human personality, but the measure is now widely used in personnel selection. By contrast, the California Psychological Inventory was written for use in selection and counselling, but has been used to develop theories of aggression (Chapter 10).