Chapter 3

Research in behavioural genetics

Research in behavioural genetics

Introduction

3.1 There are different ways in which researchers can study the contribution that genetic factors make to human behaviour. First, there are observational studies, which involve assessing and comparing relatives such as twins or siblings, families and adopted children. This type of research is called quantitative genetics because it aims to examine the extent to which variation in a trait is influenced by genetic factors in a population. It uses statistical methods to examine and compare groups of people, without focusing on particular genes. More detail about research in quantitative genetics is provided in Chapter 4. Secondly, researchers can try to identify differences in genes that contribute to trait variation in characteristics or traits between individuals. This type of research is called molecular genetics and its application to behavioural research is explained in Chapter 5. Thirdly, researchers can use animals to try and examine the effects of particular genes on behaviour. Chapter 6 describes this type of research. The focus of research in behavioural genetics is summarised in Box 3.1.

Box 3.1: What does research in behavioural genetics study?

Research in behavioural genetics examines the effects of genotype and environment on a range of phenotypic traits such as anxiety, intelligence, sexual orientation and antisocial behaviour.

Genotype

An individual's genotype is his or her entire complement of DNA.

Phenotype

An individual's phenotype consists of all his or her measurable or observable properties and characteristics aside from his or her genes. These could include characteristics such as hair colour, height and IQ score. Researchers in behavioural genetics often include such diverse traits as marital status, taste in music and religious beliefs as part of the phenotype.

Environment

An individual's environment is to be understood very broadly. It includes everything that influences an individual's phenotype, apart from his or her genotype. Environmental factors include where a person lives and how many siblings he or she has, but also biological factors such as to which chemicals a person might have been exposed to before and after birth.

3.2 Before examining the various types of research and their advantages and limitations, it is important to have a clear understanding of what is meant by the suggestion that genes influence, affect or contribute to human behaviour. The remainder of this chapter attempts to do this by addressing the following questions:

- What are genes and how do they work? (See Box 3.2).
- What is genetic variation?
- What do we mean by human behaviour in the normal range?
- How might genes influence human behaviour in the normal range?
- How could the behaviour of an individual be predicted from information about his or her genotype?

Box 3.2: How do genes work?

The human genome contains the genetic information required to build the human body. This information is held in code on tightly coiled threads of deoxyribonucleic acid (DNA). A DNA molecule consists of two strands that wrap round each other to resemble a twisted ladder – the famous double helix. Each strand of DNA is made up of a string of units called nucleotides, or bases. There are four different bases: adenine (A), thymine (T), guanine (G) and cytosine (C). These bases pair together – A with T, and C with G. Each base pair forms a rung of the ladder. The way these pair together causes the strands to coil up into the spiral twisted ladder (see Figure 3.1). It also allows the DNA to replicate, or copy itself.

Figure 3.1:
The structure of DNA. The base pairs, A-T and C-G, form the rungs of the ladder.

A gene is a segment of DNA that contains the instructions for making a specific protein (or sometimes ribonucleic acid (RNA)). Each set of three base letters, for example ACG, is a code providing the instructions to assemble a protein. A gene may contain anything from a few hundred to over a million base pairs. Genes are assembled into chromosomes, long strands of DNA large enough to be seen down the microscope. A chromosome contains between a couple of hundred and several thousand genes, arranged in a specific order end-to-end, with sections of spacer DNA, which does not code for any genes, in between. Humans have 22 pairs of chromosomes plus the sex chromosomes (XX in the female, XY in the male). One set is from the mother and one from the father. Together, these 23 pairs make up the human genome. It is estimated that each human has about 30,000–40,000 genes, and around six billion base pairs of DNA.

Proteins carry out the work of a cell. They are made of various combinations of 20 chemical building blocks, called amino acids. The sequence of the gene determines the order that these blocks assemble together, and hence which protein is made (see Figure 3.2). Different proteins have different specialised functions, such as making muscle, binding oxygen from the air, transmitting nerve impulses, and breaking down food substances. Many proteins are enzymes, with the specialised function of synthesising, breaking down or altering other chemical molecules. Some of the products of genes, and some of the substances made by these products, are 'messengers' exported by cells to have effects on other cell types. For example, hormones are made in specialised endocrine glands, and can stimulate or suppress the functions of other cells in distant organs.

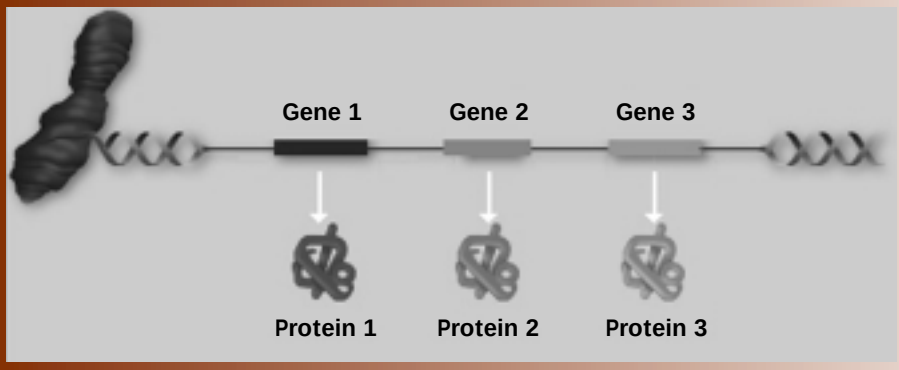

Figure 3.2: Genes make proteins

The human body consists of many different types of cell, each with a specialised function, for example skin cells and liver cells. These cells rely on different proteins to perform their specialised jobs. Nearly every cell in the body has the same genetic material. But what makes cells differ from each other is not which genes they have, but which genes are active in that cell. A protein, or part of a protein, is only produced when its corresponding gene is active. The production of proteins is also called gene expression. A gene is active when RNA, the intermediate between DNA and protein, is being synthesised. It is not yet completely understood what determines whether a gene is active or inactive, though the function of some genes is (through their encoded proteins) to turn other genes on and off. It is increasingly recognised that secondary modifications to the chemical structure of the original gene and its association with specific proteins are important factors in this process. The environment of the cell may influence these so-called epigenetic effects, which can be stably inherited through cell divisions and even through generations.[*,†]

[*] Pennisi, E. (2001). Behind the scenes of gene expression. *Science* **293**,1064–7. For an account of the complex etymology of the term 'epigenetics', see Wu, C.-T. & Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science* **293**,1103–5.

[†] Images in this box are reprinted with permission from Roche Genetics. http://www.roche.com/pages/rgg/science-gengen-cdrom%5b2%5d+jpg_page1.html (11 June 2002).

What is genetic variation?

3.3 A comparison of the DNA sequence of corresponding chromosomes between two people chosen at random would show that the DNA sequences were almost, but not precisely, identical. On average, one in every 1,300 positions along the sequence will have different bases present at the corresponding positions. For example, some people might have an 'A' base whereas others have a 'G' base at a particular position. These two alternative possibilities are termed alleles. If the rarer of the two alleles is present in at least 1% of chromosomes in a population, it is termed a polymorphism. The simplest type of variation, where a single letter is substituted for another (as in the example above), is called a single nucleotide polymorphism (SNP). Continuing with this example, an individual's DNA sequence at a particular point could be AA, GG or AG, because the chromosomes occur in pairs. In the first two situations, the person is called a homozygote because both letters in the pair of chromosomes are the same. In the third, the person is called heterozygote because both the letters at this position are different. Most people are heterozygous at

over 20 million different sites in their genome. Included in this total would be a significant proportion of SNPs, estimated to number about 11 million.[1]

3.4 Much of this variation occurs in the stretches of DNA situated between the genes and probably has no important effect on the organism. However, variation occurring near to, or within, genes could affect either the amount of protein made in a particular cell, or the sequence of amino acids in the protein. If this variation is found to correlate with a particular behaviour, or other trait, it is termed a susceptibility allele. The difficulty for the researcher is in sifting out, from the millions of polymorphisms in the human genome, the smaller number, possibly thousands, that actually contribute to variation between individuals, and which are presumed to underlie the contribution of genetic factors to differences between people.

3.5 Such 'genetic variation' arises in the first place because of damage to DNA or mistakes in copying DNA during replication. This process is called mutation. Mutation can occur in any cell, but is of particular concern when it affects eggs or sperm, as this allows the variant alleles to be passed on to future generations. However, new alleles that have major adverse effects will be eliminated rapidly from a population, because individuals carrying those alleles are less likely to reproduce. How, then, can particular alleles become common in the population and yet influence genetic susceptibility? Combinations of the following processes influence the frequency of alleles:

- *Age of onset of the trait*. An allele whose major effect occurs after the age of reproduction will be subject to very little selection because individuals carrying the allele will already have had the potential to transmit it to their offspring by the time that the effect becomes apparent.

- *Chance factors in the context of weak selection*. The rapid growth of the human population from relatively small numbers of individuals, and chance factors influencing reproductive success, could together enable mildly harmful alleles to reach a significant frequency in the population (this is termed genetic drift).

- *Strength of selection in relation to genotype*. Selection of an allele at the level of the population will be strong if the trait manifests in the heterozygote (dominance) but weaker if it manifests only in the homozygote (recessivity). A special case is when the heterozygous state has a survival benefit over either homozygote (this is termed heterozygote advantage).[2]

- *Different selection in different environments*. Alleles may be beneficial in some environmental circumstances and harmful in others. For example, they may protect from starvation but predispose to obesity. Environmental variation may then give rise to a balanced polymorphism between two alleles. Whether an allele is considered beneficial or harmful will depend on the context.

---

[1] Kruglyak, L. & Nickerson, D. A. (2001). Variation is the spice of life. *Nat. Genet.* **7**, 234–6. The remainder consists of single nucleotide variations present on less than 1% of chromosomes and other types of variation such as simple sequence repeats. The identification of these SNPs has been a significant focus of the Human Genome Project: already over 2.7 million are known (figure from dbSNP, available at: http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi (3 July 2002)) and it is likely that virtually all these SNPs will be identified within the next few years.

[2] A well-known example is the sickle cell mutation of ß globin. The homozygous state ($ß^A/ß^A$) causes the blood disorder sickle cell anaemia, but the heterozygous state ($ß^S/ß^A$) protects against malaria and has a survival advantage over the normal ($ß^A/ß^A$) state in malaria-infested countries. The two alleles ($ß^A/ß^S$) are maintained in the population because their net effect is quite neutral. The deleterious effects become apparent in disease-based studies and beneficial effects may only be discovered later from population-based studies.

- *Pleiotropy*. Susceptibility alleles may affect different traits, which are subject to different selection pressures. For example, homozygosity for the Ɛ2 allele of a blood protein called apolipoprotein E predisposes an individual to developing excessive fat levels in the blood, but also seems to protect the brain against the development of Alzheimer's disease.[3]

- *Genetic hitch-hiking*. When selection of an allele occurs, the adjacent segment of DNA is passively selected with it, only later becoming separated by rare recombinations.[4] This, and other factors, such as the mixing of populations, gives rise to linkage disequilibrium, which means that alleles may appear to be associated with variation in a phenotype without themselves causing that variation.

3.6 It is important to note that genetic variation in the normal range is usually neither good nor bad. Genetic variation causes people to have different natural hair colours or different blood groups, but this is not to say that a particular hair colour or blood group is 'better' than another. In the case of genetic mutations that cause diseases such as cystic fibrosis or Huntington's disease, it might be reasonable to say that these mutations are deleterious. With most forms of genetic variation, all one can say is that differences exist, not that they are deleterious or advantageous.[5]

What is meant by normal variation in human behaviour?

3.7 Many human traits are not viewed as either present or absent, but rather as continuously distributed measures which each individual in the population will show to a greater or lesser extent (for example, height, blood pressure, aggressiveness and intelligence, as measured by IQ test scores). These characteristics vary from person to person in a population and this variability is known as population variation. When the frequency of the effect is plotted against the magnitude, many of these continuously distributed characteristics show a bell-curve distribution that is known as a normal distribution (see Figure 3.3).

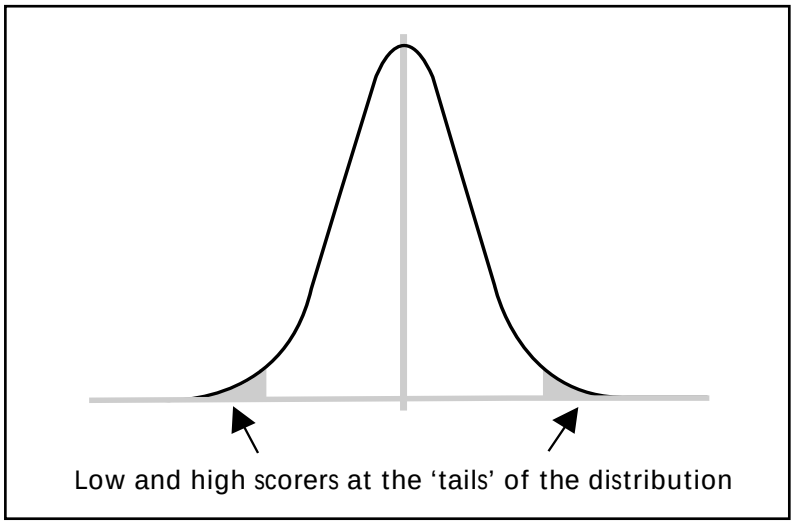

Figure 3.3: A normal distribution curve

[3] Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D. (2001). *The Metabolic and Molecular Base of Inherited Disease*. 8th ed. New York: McGraw-Hill. pp. 2835–62 and 5875–99.

[4] Recombination is the process by which chromosomes are broken and the fragments rejoined in new combinations, and is a vital aspect of reproduction and inheritance.

[5] Some traits that are genetically influenced may have both positive and negative aspects in an individual. For example, it has been hypothesised that people with manic depression are also more creative (for an exposition of this theory see Jamison, K. R. (1996). *Touched With Fire: Manic-Depressive Illness and the Artistic Temperament*. New York: The Free Press). Alternatively, the desirability of a trait may be perceived differently in different individuals. For example, high scores on the Psychoticism dimension of personality tests have been shown, in non-psychotic individuals, to be associated with high creativity.

3.8 Most individuals score somewhere in the middle of the population distribution, but a certain percentage will form the 'tails' of the distribution (i.e. very high or low scorers). Statisticians can quantify the 'spread' of scores in this distribution and this statistic is known as variance. Extremes of variation can be associated with an increased risk of adverse outcomes, for example high blood pressure is associated with an increased risk of a stroke, but they are not necessarily associated with illness. Such extremes might have a number of different explanations. In the case of high blood pressure, this could be due to excessive intake of salt in the diet, a combination of dietary and weak genetic factors, or a single mutation causing failure of the kidney to eliminate salt. In the field of behavioural traits, extreme scores for traits such as neuroticism or aggression may be associated with an increased risk of adverse outcomes such as mental illness.

'A gene for X'?

3.9 Having understood that there are differences in both the genotypes and the phenotypes of individuals, the next challenge is to understand how genetic variants might be related to the variations in behavioural and personality traits. It is common to hear of research that claims to identify a 'gene for aggression' or a 'gene for homosexuality'. But how could our genes *cause* us to act in a particular way? What is really meant by saying 'a gene for X'?

3.10 We all have a working knowledge of the system of causal relations that enables each of us to function in our complex world. The accumulation of this knowledge begins early in life. Once a child reaches a certain age he will push an object off the tray of his high chair and observe it falling to the floor. On discovering the effects of gravity, he will repeat the experiment over and over again to test the discovery beyond all reasonable doubt. Unfortunately, the concepts of causation that are established early in life are too rudimentary to serve well as the basis for scientific theories.

3.11 A useful way to understand what is meant by a *cause* is to think of it as a factor that increases the chance that an event occurs. So, when the child pushes its toy off its highchair, various factors combine to make the toy fall to the floor, including the movement and force of the child's hand, the weight of the toy and the effect of gravity. Together, these factors make it extremely likely that the toy will fall on the floor: they raise the chance of this happening to a very high degree. It is easy to assume that there is a one-to-one correspondence between cause and effect; that one cause leads to one effect. However, as the example of the child's toy shows, there are usually a number of causal factors that come together to make certain events more likely. These two points, the existence of more than one causal factor in a particular situation, and the idea that causes raise the chance of something happening without necessarily making that outcome inevitable or 100% likely to occur, are important in understanding how genetic factors can influence behaviour.

3.12 There are some genetic mutations that do seem always to be present when disease occurs, such as the genetic mutations that are associated with cystic fibrosis and Huntington's disease. In these diseases, an alteration in a particular gene can be said to have caused the disorder. Without the genetic mutation, the individual would not have the disorder, which means that the genetic mutation is a necessary condition for that disorder. In addition, for a very small subgroup of single-gene disorders, for example Duchenne's muscular dystrophy, the genetic mutation is also a sufficient condition: just possessing the mutation makes it certain that you will have the disorder. However, it is worth noting that even in apparently clear cases, the connection between the gene alteration and the disorder is not

simple. With Huntington's disease, once the mutated gene is identified in an individual, all one can say is that the disorder will appear. Knowledge of the genetic mutation does not enable one to say how serious the disease will be for a particular individual, the precise symptoms they will experience, or exactly when it will arise in their lifetime.

3.13 Diseases in which there is a clear connection between a particular genetic mutation and a disorder are comparatively rare. Most disorders, such as breast cancer, heart disease and diabetes, are much more complicated. Often, a genetic mutation or susceptibility allele will cause a disease only in the presence of certain environmental factors, such as stress or a particular diet. The relationship between genetic and environmental factors in causing disease can be illustrated by a disorder called alpha-1 antitrypsin deficiency. If an individual with this genetic disorder takes up smoking, he or she is much more likely to develop a form of a chronic lung disease called emphysema, because of a genetic mutation which prevents the production of a normal form of a protein in the lung that protects against inflammation. However, if the individual never smokes, it is probable that he or she will never develop emphysema and may be able to have a relatively healthy life.

3.14 The connection between genes and diseases is far from straightforward, and the relationship between genes and behaviour is even more complicated. It is often difficult to establish which genes contribute to a trait and how they do so because:

- More than one genetic factor usually contributes to a particular trait.
- These multiple genetic factors may interact with each other and have different effects depending on which other factors are present in the individual's genotype.
- As well as genetic factors, many non-genetic (environmental) factors may contribute to the manifestation of a trait.
- These environmental factors may also interact with each other.
- The genetic factors may affect which environmental factors have an effect. (This is called gene–environment interaction: see Box 3.3).
- Conversely, environmental factors may affect which genetic factors have an effect.
- Certain genetic and environmental factors may go hand in hand. (This is called gene–environment correlation: see Box 3.3).
- A protein may be modified after it has been produced from a gene, and this can alter its function.
- Genes do not have a continuous effect in our bodies. They may be turned on and off, both during our overall development and within the lifetime of an individual cell.

So, while it might be correct to say that a particular genetic variant is part of the cause of a particular trait, or that it is one causal factor, it will seldom be the only cause, nor is it likely to be either a necessary or sufficient condition for the trait to be manifested. Furthermore, even if particular genes that contribute to a trait can be identified, this is only a small part of the story. There is still a need to understand the very indirect pathway

between a gene, a particular protein and an individual scoring highly on an IQ test or having an aggressive personality. Our understanding of these causal pathways is at an even earlier stage than our understanding of which genes influence behavioural traits, which is itself extremely limited.[6]

3.15 In an effort to overcome this difficulty, some research focuses on 'intermediate' traits that are related to the phenotype in question but are in some sense 'nearer' the biological mechanisms and can be more objectively measured.[7] For example, in research on cognitive ability, such intermediate traits may include electrophysiological measures of brain function, behavioural measures of the speed at which information is processed or behavioural measures of the capacity of working memory in an individual (such as performance on tasks of verbal and spatio-visual working memory).[8]

Box 3.3: The relationship between genes and the environment

Gene–environment correlation

It is sometimes assumed that genetic and environmental influences act independently and additively: that separate influences add up in a linear manner to make a given outcome more likely. However, genes and environmental factors can be correlated, or interdependent. Children not only inherit genes from their parents but are also exposed to environments that are influenced by their own and their parents' genetic make-up. Thus, for example, sociable parents not only pass on genes to their children but may also provide an environment that encourages the development of sociability in their children. This is known as passive gene–environment correlation. A sociable child may actively seek out situations that serve to further increase sociability (active gene–environment correlation) or evoke responses from others that increase sociability. In both cases, the existence of the genetic variant is linked to the presence of a particular type of environment.

Gene–environment interaction

Genetic and environmental factors may interact non-additively to influence characteristics. That is, the impact of environmental factors may differ depending on a person's genetic makeup. For factors that are correlated, detecting interaction using statistical methods can be difficult. Large samples are needed. Nevertheless, studies of twins have shown that the impact of life events on, for example, depression varies depending on an individual's genetic susceptibility.*

* Kendler, K. S. et al. (1995). Stressful life events, genetic liability and onset of an episode of major depression in women. *Am. J. Psychiat.* **152**, 833–42; Silberg, J., Rutter, M., Neale, M. & Eaves, L. (2001). Genetic moderation of environment risk for depression and anxiety in adolescent girls. *Brit. J. Psychiat*. **179**, 116–21. See also Kendler, K. S. (2001). Twin studies of psychiatric illness: an update. *Arch. Gen. Psychiatr* **58**, 1005–14 for a review of gene–environment interaction in research in behavioural genetics.

[6] Rutter, M. & Silberg, J. (2002). Gene–environment interplay in relation to emotional and behavioural disturbance. *An. Rev. Psychol*. **53**, 463–90.

[7] These intermediate traits are sometimes called endophenotypes.

[8] See for example, de Geus, E. J., Wright, M. J., Martin, N. G. & Boomsma, D.I. (2001). Genetics of brain function and cognition. *Behav. Genet*. **31,** 489–95**.**

Describing human behaviour

3.16 Any description of a human action can be set at a level that includes information about the biological characteristics of the individual. For example, the following descriptions could all correctly refer to the same act:

- The man's brain sent messages to his leg muscles.
- The man's leg muscles contracted and then relaxed.
- The man moved his leg.
- The man kicked the dog.

3.17 Thus, it could be said that the movement of the man's muscles caused him to kick the dog, or that the movement of his leg caused him to kick the dog. If one broadens a description far enough, it will certainly include information about the physiological characteristics of the individual, since these will be involved in any human action. The important question is: which way of describing or understanding the act is the most useful? If genetic factors are one aspect of causal explanations of human behaviour, what importance should be accorded to them? The answer is likely to depend on what use the questioner wants to make of the information:

> 'It is a well known fact that we describe as the cause of an event that particular condition by which we hope to control it.'[9]

In the example above, if one wants to admonish the man for kicking the dog, an explanation at the neural or physiological level is unlikely to be relevant. We return to this important issue in the last section of this Report, in the context of moral and legal responsibility.

Predicting human behaviour from genetic information

3.18 Even if it is not known precisely how a genetic variant contributes to a behavioural trait, it might be possible to predict how likely it is that individuals with that genetic variant will display the trait in question. Here, it is important to differentiate between predicting the future development of a phenotypic trait or specific behaviour, and measuring a phenotypic trait that is already established in an individual and can be observed. For example, if there were a genetic variant, or group of genetic variants, known to be associated with lower or higher intelligence, it would be possible to measure the genotype of a baby and to make some prediction of the IQ that the baby will have as an adult. Alternatively, measuring that genotype in an adult might enable the current IQ of the adult to be estimated. A third scenario for the predictive use of genetic information would be to predict the likelihood of the future occurrence of a specific act linked to a behavioural trait, for example an act of aggression.[10]

3.19 However, in whatever context the term prediction is being used, it is highly unlikely that individual genetic variants will often be accurate predictors of behavioural traits. It is not

[9] Barton Perry, R. (1926). *General Theory of Value*. Cambridge, MA: Harvard University Press. p. 394.

[10] It is pertinent to note here that non-genetic influences are often used in attempts to predict behaviour, for example correlations between family environment and antisocial behaviour. These are discussed further in Chapter 14.

known how many genes will account for the genetic influence on a trait that is normally distributed, even if that genetic influence appears to be substantial. Behavioural traits are complex, and are likely to be the result of the expression of many different genes, which interact with each other and with the environment. No single gene is likely to account for more than a small proportion of the total variance of a given trait. Furthermore, even if it were possible to identify all the genes that contribute to the heritable component of the trait, the predictive power would still be very limited. In view of existing evidence from studies of monozygotic (MZ) twins, it appears that such genetic predictions of behavioural traits might be able to account for at most 50% of the variance (see the reviews of the evidence in Chapters 7–10 for details of particular studies). This would still leave at least half the variance in trait scores unpredictable. Whether this environmental portion of the variance will become predictable depends upon future advances in understanding which variables give rise to environmental variance. This is as yet virtually unexplored territory.

Conclusion

3.20 The complexity of human behaviour and the difficulties in understanding how genes are involved may seem overwhelming. There is wide agreement that genes do have an indirect effect on behaviour. However, some commentators have suggested that any attempt to understand the processes by which genes influence behaviour will certainly fail. We disagree. We consider that it is neither a theoretical nor a practical impossibility to identify genes that contribute to behavioural traits and to understand some of the mechanisms by which they do so. However, we note that terminology such as ‘a gene for X’ or ‘a set of genes for X’ is very misleading because it fails to convey the complexity of the role of genetic factors in causal explanations of human behaviour. Genes determine which proteins are made. They do not determine which behavioural or personality traits an individual possesses. Furthermore, the product of an individual gene will only very rarely be directly related to a complex behavioural characteristic. It will normally interact with many other genes and with many non-genetic factors, which means that the predictive capabilities of tests for any single or small number of genes will in general probably be quite limited. Nonetheless, the proteins that genes make and the way these affect our bodies and brains will be one part of an explanation of human behaviour.