4

Determinants of Health-Related Behaviours: Theoretical and Methodological Issues

STEPHEN SUTTON

INTRODUCTION

The term *health behaviour* (or *health-related behaviour*) is used very broadly in this chapter to mean any behaviour that may affect an individual's physical health or any behaviour that an individual *believes* may affect their physical health.

This chapter focuses on determinants of health behaviours. More specifically, it focuses on what we will refer to as 'cognitive' determinants, as specified by theories of health behaviour or 'social cognition models' as they are sometimes called. After briefly considering some of the more important distinctions and dimensions of health behaviours and the definition and measurement of target behaviours, we provide an extensive discussion of research designs that are used – or could be used – to investigate the cognitive determinants of health behaviours. Then, a classification of theories of health behaviour is presented, followed by a detailed discussion of one particular theoretical approach, the theory of planned behaviour (TPB: Ajzen, 1991, 2002b).

Theories of health behaviour acknowledge that health behaviours may be influenced by numerous biological, psychological, and social factors, but they specify only a limited subset of cognitive determinants that are assumed to be most proximal to the behaviour. For a more complete explanation of particular health behaviours, it is necessary to extend the theories to include other relevant determinants. To this end, we outline a broader theoretical framework, drawing on the 'social ecological framework' (Emmons, 2000; Green, Richard & Potvin, 1996; McLeroy, Bibeau, Steckler & Glanz, 1988; Stokols, 1992, 1996) and ideas from multilevel modelling (Bryk & Raudenbush, 1992; Duncan, Jones & Moon, 1998; Hox, 2002). We conclude by making a number of recommendations to guide future research in this area.

The chapter presents a generic approach to explaining health behaviours, focusing on theoretical and methodological issues. Although a number of different examples of health behaviours are used, we do not attempt to review the determinants of particular health behaviours.

QUESTIONS ADDRESSED IN RESEARCH ON HEALTH BEHAVIOURS

Most health psychological research on health behaviours attempts to explain *between-individual*

variation in particular health behaviours using theories of health behaviour. Such research typically addresses questions like: Why do some people engage in regular physical activity while others do not? Why do people differ in the frequency with which they engage in physical activity? Why do some women accept an invitation to go for breast screening while other women do not? Why do some adolescents try smoking while others remain non-smokers? This chapter focuses on determinants of between-individual variation because this is the dominant approach in the field of health behaviour research. However, it is also important to study determinants of *within-individual variation*: why does an individual's behaviour vary over time or across different settings? For example, why does a woman attend for her first breast screen but not for subsequent screens? Why does a smoker smoke more on some days or in some situations than in others?

There are other important questions concerning health behaviours that we do not consider in this chapter. One such question concerns the extent to which health behaviours cluster together (e.g., Røysamb, Rise & Kraft, 1997). For example, do smokers have generally less healthy lifestyles than non-smokers? Are people who attend for one kind of screening test more likely to attend for another kind of screening test?

Health behaviours are extremely diverse. In the next section, some of the more important distinctions and dimensions are briefly discussed (see also Carmody, 1997).

DIMENSIONS OF HEALTH BEHAVIOURS

Positive and Negative Behaviours

A distinction is often made between positive and negative health behaviours. Examples of positive, 'healthy', 'healthful' or 'health-enhancing' health behaviours are taking regular exercise, going for annual health checks, eating at least five portions of fruit and vegetables a day, and using a condom with a new sexual partner. Negative, 'unhealthy', 'risky', 'health-compromising' or 'health-impairing' health behaviours would include, for example, smoking, drinking heavily, driving too fast, and eating a diet high in saturated fat. In many cases, this is simply a matter of framing. All these behaviours can be thought of as dichotomies that have a positive alternative and a negative alternative, for example going for annual health checks versus not going for annual health checks, or smoking versus not smoking.

Behavioural Stages

Individuals show varying and complex patterns of changes in particular health behaviours over the life course. Take smoking for example. Some adolescents may try a cigarette while others remain never-smokers. Some of those who try smoking may continue to experiment while others never smoke again. Some of those who experiment may become regular smokers whereas others may stop smoking. Some of those who become regular smokers may try to quit while others do not. Some of those who try to quit may succeed while others relapse. Some of those who relapse may make further attempts to quit.

Such a process can be simplified by defining *behavioural stages* such as adoption or initiation, maintenance, cessation and relapse, conceived of as a series of dichotomous dependent variables. Behavioural stages are not the same as the stages specified by stage theories of health behaviour, some of which are defined in terms of non-behavioural variables such as intentions, but can be analysed in similar ways, for example by estimating the transition probabilities: given that an adolescent tries smoking, what is the probability that he or she will become a regular smoker within a specified period of time? The determinants of different stages may differ (Rothman, 2000). For example, the factors that influence whether or not adolescents try smoking may differ from the factors that influence whether or not those who experiment progress to becoming regular smokers.

Health Behaviours versus Illness Behaviours

If a person who has suffered a heart attack takes up regular exercise, perhaps on the advice of his doctor, this could be referred to as an

illness behaviour, because the person has a medically diagnosed illness or condition. If the same person had started exercising before he had his heart attack, the term *health behaviour* would be more appropriate. Even though the behaviours may be defined in exactly the same way, the determinants of illness behaviours may be different from the determinants of health behaviours. In this chapter, we use the term 'health behaviour' to cover both cases.

Many other types or dimensions may be important, for example 'detection' behaviours versus 'prevention' behaviours (Weinstein, Rothman & Nicolich, 1998), 'public' behaviours (e.g., going jogging) versus 'private' behaviours (e.g., exercising at home), and behaviours that involve the use of health services (e.g., going for a mammogram) versus those that do not (e.g., breast self-examination).

Although we have suggested that different types of health behaviours *may* be influenced by different factors, it is more parsimonious to start from the assumption that all types of health behaviour are influenced by the same limited set of proximal cognitive determinants, that is, that the same theory of health behaviour can be used to explain different behavioural stages, health and illness behaviours, detection and prevention behaviours, and so on.

DEFINING AND MEASURING BEHAVIOUR

In any study of the determinants of health behaviour, it is important to start by defining the behaviour of interest as clearly as possible (Ajzen & Fishbein, 1980; Fishbein et al., 2001). Following Ajzen and Fishbein (1980), behaviours can be defined in terms of four components: action, target, time and context. The action component is a necessary part of the definition of any behaviour. The target component is usually necessary, though not always. Time and context are optional; they enable the definition of behaviour to be as specific as required. For example, consider the definition 'eat breakfast tomorrow'. Here, 'eat' is the action, 'breakfast' is the target (alternative targets would be 'a bowl of cereal' or 'lunch') and 'tomorrow' is the time component. No context is specified in this example. As an illustration of the importance of context, consider the following definitions:

1 using a condom the next time I have sex
2 using a condom the next time I have sex with my regular partner
3 using a condom the next time I have sex with a new sexual partner.

The first definition omits a potentially important contextual factor, namely type of partner, and is therefore probably too general for most purposes. The other two definitions each specify a context. These may be considered to be quite different behaviours both from a public health viewpoint and from the viewpoint of an individual who has both types of sexual partner.

Such behaviours are often measured as dichotomies: for example, 'Did you use a condom? Yes/No.' This implies that the person has a choice between two mutually exclusive and exhaustive alternatives: performing the behaviour or not performing it. This approach enables the simplest possible application of a given theory of health behaviour: participants' cognitions are assessed with respect to performing the behaviour. Occasionally, their cognitions with respect to *not* performing the behaviour are measured as well, but the usual assumption is that these will be the complement of the first set of cognitions and therefore provide no additional information. For example, if a person states that they are extremely likely to perform a given behaviour, it is assumed that, if asked, they would say that they were extremely unlikely not to perform it. This *complementarity assumption* seems plausible for intentions but questionable in the case of other cognitive variables. For example, a smoker may believe that his chances of developing lung cancer are 'quite high' if he continues to smoke. If we had only this information, we might assume that he would be motivated to quit smoking. However, he may also believe that his chances of developing lung cancer are 'quite high' if he stops smoking (because 'the damage has already been done'). Ideally, then, relevant cognitions should be measured with

respect to *both* alternatives (performing and not performing the behaviour). An alternative approach is to phrase measures explicitly in terms of changes or differences (Weinstein, 1993).

In some cases, it is more realistic to define more than two alternatives. For example, in a study of contraceptive use, one could ask: 'The last time you had sex, did you use (a) a condom, (b) the pill, (c) another method of contraception, or (d) no method of contraception?' As an alternative to this multiple-choice format, respondents could be asked a series of separate questions (e.g., 'Did you use a condom? Yes/No', 'Did you use the pill? Yes/No' etc.). In this example, it is possible for respondents to have used more than one method, for example a condom and the pill; the alternatives are not strictly mutually exclusive.

Some behaviours are defined so generally that they are best thought of as *behavioural categories*. Behavioural categories cannot be directly observed. Instead they are inferred from single actions assumed to be instances of the general behavioural category. Ajzen and Fishbein (1980) give the example of dieting. Dieting may be inferred from specific behaviours such as eating two instead of three meals a day, not eating desserts, drinking tea and coffee without adding sugar, taking diet pills, and so on. There are two approaches to assessing behavioural categories. The first is simply to ask respondents questions like, 'Are you currently dieting?' In this case, a definition of 'dieting' should be provided, unless the aim is to explore different interpretations of this term. The second approach is to ask about a number of specific behaviours and use these to create an index of dieting. These approaches have different implications for the measurement of intention and other proximal determinants.

In many situations, we may be interested not in whether or not a behaviour is performed but in the magnitude and/or frequency of a behaviour. For example, in a study of drinking, we could ask people how often they have an alcoholic drink (frequency) and how much they usually drink on such occasions (quantity). Such behavioural criteria pose problems for theories of health behaviour because they imply multiple alternatives, and it is not practicable to assess relevant cognitions with respect to all possible values of frequency and/or quantity. A common strategy is to convert the behaviour into a dichotomy. For example, one could use the official definition of 'safe' drinking and ask respondents about their cognitions with respect to exceeding the safe drinking limit. For further discussion of the problems created by magnitude and frequency measures, see Courneya (1994; Courneya & McAuley, 1993).

It is also important to distinguish between behaviours and *goals* (Ajzen & Fishbein, 1980, use the term *outcomes*). Losing weight is a goal not a behaviour. Attaining this goal may be influenced in part by behaviours such as eating low-fat foods or jogging 2 miles every day. But goals may be influenced by factors other than the person's behaviour. Losing weight depends on physiological factors such as metabolic rate, as well as on behavioural factors. If the aims are to predict and explain a goal or behavioural outcome, the actions that lead to the goal have to be identified and measured along with other, non-behavioural factors.

Most studies of health behaviours use self-report measures of behaviour. The limitations of self-reports are well known (Johnston, French, Bonetti & Johnston, 2004, Chapter 13 in this volume; Schwarz & Oyserman, 2001; Stone et al., 1999), but in many cases there will be no feasible alternative. Sometimes it may be possible to use 'objective' measures of behaviour, but these usually have limitations too. For example, records of attendance for breast screening may be inaccurate and may miss women who go for screening at other centres; electronic monitoring of tablet use may increase adherence; and biochemical measures of tobacco smoke intake are sensitive only to recent intake. Nevertheless, objective measures of behaviour may be more predictive of relevant health outcomes than are self-report measures. It is conceivable that theories of health behaviour may predict self-reported behaviour quite well but may be less effective in predicting the behavioural measures that are most strongly related to important health outcomes.

PREDICTION VERSUS EXPLANATION

It is important to distinguish between prediction and explanation (Sutton, 1998). A key aim of research on health behaviours is to identify the determinants (causes) of particular health behaviours. Researchers do this by developing theories that identify potentially important determinants of behaviour and specify the causal pathways by which they influence behaviour and by conducting empirical studies that enable the effect sizes to be estimated, usually in the form of regression coefficients. Ideally, we want to obtain *unbiased* and *precise* estimates of the causal effects of the putative determinants of a particular health behaviour in a given target population.

If the aim is prediction rather than explanation, we do not need to concern ourselves with identifying the determinants of behaviour or with specifying causal processes (although a causal model may suggest suitable predictor variables). We are free to choose convenient predictors and weights. Any predictor that 'works' can be included in the regression model. Past behaviour, for instance, is often a strong predictor of future behaviour, even though its theoretical status as a determinant of behaviour is contentious (Sutton, 1994). Similarly, it does not matter if relevant causal variables are omitted from the model.

Prediction can be useful without explanation, particularly when the time interval allows interventions to be applied. For example, it would be useful to be able to predict who is at risk for becoming a problem drinker. Identification of high-risk individuals may enable an early intervention to be made. Thus prediction enables interventions to be targeted to high-risk groups. However, an understanding of the factors that lead some people but not others to develop a drinking problem (explanation) would be even more useful because it would have implications for the nature and content of the intervention programme; it would tell us not only who to target but also what to do to them. Although prediction and explanation are not the same, the first is necessary for the second; models that do not enable us to predict behaviour are unlikely to be useful as explanatory models.

RESEARCH DESIGNS FOR INVESTIGATING COGNITIVE DETERMINANTS OF HEALTH BEHAVIOURS

The most commonly used designs for studying the cognitive determinants of health behaviours are between-individuals cross-sectional studies and prospective studies with two waves of measurement and relatively short follow-up periods of days, weeks or months. We use the relationship between attitude and behaviour to illustrate these designs, where it is hypothesized that attitude influences behaviour and where attitude represents any variable whose values may change over time within an individual and which may be influenced by behaviour (i.e., where there may be *reciprocal causation*); other examples would be self-efficacy, intention, risk perceptions, worry, attributions and illness representations. (See Weinstein, Rothman & Nicolich, 1998, for an illuminating discussion of the use of correlational data to examine the relationship between risk perceptions and behaviour.)

Before discussing between-individuals designs and possible causal models, we consider within-individuals designs, because ultimately we want to draw inferences about within-individual causal processes. Between-individuals designs may or may not be informative about the processes that occur at the within-individual level.

Within-Individuals Designs

Consider a study in which a person's attitude and behaviour with respect to a particular health behaviour are measured on a number of occasions, say once a month over a 12-month period. Figure 4.1 shows one possible relationship between attitude and behaviour. Across occasions, higher levels of attitude are associated with higher levels of behaviour. (For simplicity, we assume that there is no trend in

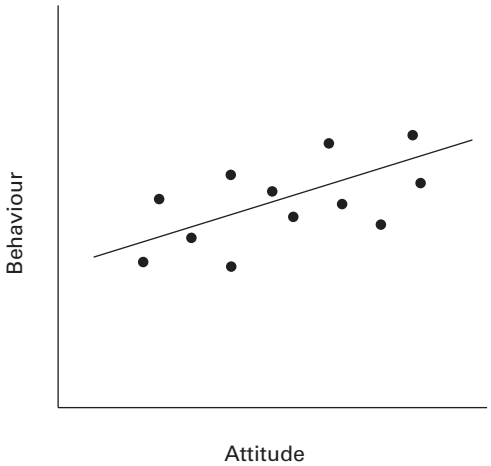

Figure 4.1 *Relationship between attitude and behaviour across 12 occasions for one individual*

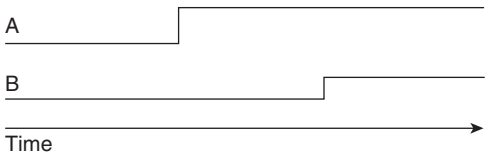

Figure 4.2 *Timeline showing causal relationship between attitude (change) and behaviour (change) for one individual*

behaviour over time, that is, that the individual's behaviour varies across occasions but there is no systematic tendency for behaviour to increase or decrease over time; of course, the analysis could be extended to fit a time trend, if there was reason to expect one.) If we are prepared to make a number of strong assumptions, we can interpret the slope of the regression line as an estimate of the causal effect of attitude on behaviour for this particular individual. We can use the regression line to estimate the likely effect on the person's behaviour if we were to intervene to change their attitude. We can think of such an intervention 'sliding the person up their own regression line'.

Note that at the within-individual level, many variables that may influence attitude and behaviour at the between-individual level are automatically controlled. For example, gender, genetic make-up and childhood experiences are fixed and cannot possibly account for the correlation between an individual's attitude and behaviour over time. Similarly, age changes slowly over the time period in question and personality variables are likely to change very little.

Possible within-individual causal relationships between attitude (or other variables) and behaviour can be represented by timeline diagrams. Figure 4.2 depicts a simple timeline diagram for one individual, showing one possible causal relationship between attitude and behaviour. The person's attitude is initially stable, then increases by a certain amount, and then remains at this higher level. The increase in attitude is followed after a certain time lag by an increase in behaviour. (Such a lag could be built into the example above by relating attitude at one time point to behaviour at the next time point.) The change in behaviour does not lead to a change in attitude (no reciprocal causation), at least within the time period shown. The main question of interest is how much behaviour change is produced by a unit increase in attitude, holding other relevant variables constant.

Mixed Designs

Now consider a study in which repeated measures of attitude and behaviour are measured in a sample of individuals. This is a *mixed* or two-level design, incorporating both within-individuals and between-individuals components. Given such data, it is possible to estimate both the effect of attitude on behaviour *within* an individual and the effect of attitude on behaviour *between* individuals. An appropriate statistical approach is random-effects or multilevel regression analysis (Bryk & Raudenbush, 1992; Hox, 2002). Figure 4.3a shows one possible pattern of results. For simplicity, only three individuals are shown, with low, medium and high levels of attitude, and we assume that the within-individuals regression lines (shown in bold) have a common slope (it is possible to fit different slopes for different individuals).

The within-individuals slope and the between-individuals slope will not necessarily be equal. In Figure 4.3a, the within-individuals slope is shallower than the between-individuals slope, but other patterns are possible. It is even possible for one slope to be positive (higher attitude going with higher behaviour) and the other to be negative (higher attitude going with lower behaviour), though this would seem implausible in the present case.

Where the two slopes differ, this can be interpreted in causal terms as indicating that an individual's behaviour on a particular occasion is influenced not only by their attitude on that occasion but also by their characteristic level of attitude, as indexed by their mean attitude across occasions. Now we have two different estimates of the causal effect of attitude on behaviour. As outlined above, the within-individuals regression lines can be used to estimate the causal effect of a change in attitude on a particular occasion on behaviour for each individual and hence the likely impact of an attitude-change intervention. By contrast, the between-individuals regression line can be used to estimate the likely effect on a person's behaviour (not just their behaviour on one occasion, but their *characteristic* level of behaviour across occasions) of increasing their *characteristic* level of attitude by a certain amount. Here we are 'sliding the person up the between-individuals regression line' in order to estimate the new level of behaviour corresponding to their new level of attitude.

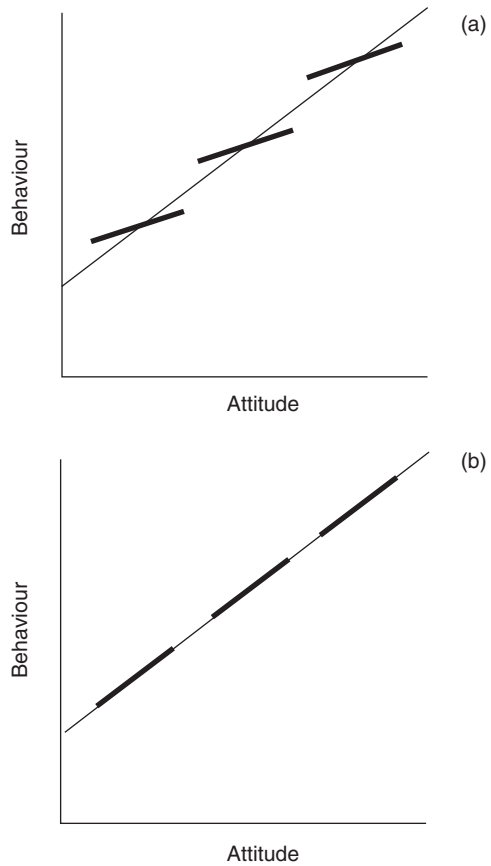

Figure 4.3 *Two different patterns of within-individuals (bold) and between-individuals relationship between attitude and behaviour*

However, there are two problems with this interpretation. First, it is not clear how we would shift a person's characteristic level of attitude, as opposed to their attitude on a particular occasion. Second, the inference is based on comparing different individuals. What we are saying, in effect, is that if we could increase person 1's characteristic level of attitude so that it was the same as that of person 2, we would expect person 1's behaviour to be the same as person 2's. However, there may be many differences between individuals other than their characteristic level of attitude that may partly explain why one individual has a characteristically high level of behaviour and another person has a characteristically low level; for example, differences in childhood experiences, gender, age and personality. Some of these variables may be correlated with attitude level. Thus, the between-individuals slope may give a misleading estimate of the effect of the hypothetical intervention because of uncontrolled differences between individuals. Put simply, we can't turn person 1 into person 2 just by changing their attitude. Of course, it is possible to measure variables that may influence behaviour across individuals and take them into account in the analysis to adjust the between-individuals slope, but it is unlikely that we will be able to anticipate and measure all the important influences.

There will also be relevant causal variables that are uncontrolled in the within-individuals

analysis – variables that vary over time, are correlated with attitude over time and that influence an individual's behaviour. Nevertheless, the within-individuals analysis has the advantage of ruling out variables that are stable over time. In sum, where the within-individuals and between-individuals regression slopes differ, there are reasons for believing that the former may give a better estimate of the likely effect of an attitude-change intervention.

Figure 4.3b shows another possible pattern of results from a two-level study. As in Figure 4.3a, the within-individuals regression lines have a common slope but this time they coincide with the between-individuals regression line. This means that we can predict an individual's behaviour on a particular occasion from their attitude on that occasion, but that knowledge of their characteristic level of attitude (or knowing which individual the observation belongs to) does not provide any additional information. Given appropriate assumptions, this can be interpreted in causal terms as follows. Differences between individuals in their behaviour on a particular occasion are influenced by differences in their attitude on that occasion, but differences in their characteristic level of attitude have no additional impact. Or, to put it differently, between-individual differences in behaviour on a particular occasion can be explained in terms of within-individual causal processes without needing to invoke additional, or different, causal processes operating at the between-individual level. In this happy situation, we can make a valid cross-level causal inference from the between-individuals relationship to the within-individuals relationship (or vice versa). See Sutton (2002a) for a detailed example that uses timelines to illustrate this case for the relationship between attitude and intention.

It is possible that attitude and behaviour are relatively stable over time within individuals. If so, we could simply compute the mean level of attitude and the mean level of behaviour for each individual (to give a more reliable estimate of each individual's characteristic levels of attitude and behaviour than would be given by a single measure of attitude and behaviour on one occasion), and then carry out a between-individuals analysis. However, the problems of estimating the likely effect of an attitude-change intervention (assuming that it were possible to change an individual's attitude given its stability over time) would still apply: we would still be inferring within-individual effects from comparisons between individuals. Similar problems arise with variables such as personality that are known to be highly stable.

Within-individuals designs and mixed designs are extremely rare in this field. A number of studies using the TPB have obtained repeated measures of cognitions and behaviour with respect to different target behaviours at the same time point (e.g., Trafimow & Finlay, 1996; Trafimow et al., in press), but only one study to date has obtained a sufficient number of repeated measures of cognitions and behaviour over time with respect to the same target behaviour to allow a within-individuals analysis (Hedeker, Flay & Petraitis, 1996). This study used the theory of reasoned action (TRA; Fishbein & Ajzen, 1975), but the measures of attitude and subjective norm departed widely from Ajzen and Fishbein's (1980) recommendations. Hedeker et al. obtained measures of cognitions and behaviour on four occasions; at least six would be preferable.

We need more studies of social cognition models that use within-individuals designs or mixed designs. However, since the vast majority of studies in this field use between-individuals designs, the remainder of this section focuses on these.

Between-Individuals Designs

Figure 4.4 shows six possible true causal models defined at the between-individual level. (Similar models could be specified at the within-individual level.) In model I, attitude influences behaviour. The small arrow pointing to behaviour represents the aggregate of all other causes of behaviour apart from attitude. This is often referred to as the *error* or *disturbance* term. The disturbance term can be thought of as comprising two components: a systematic component consisting of other

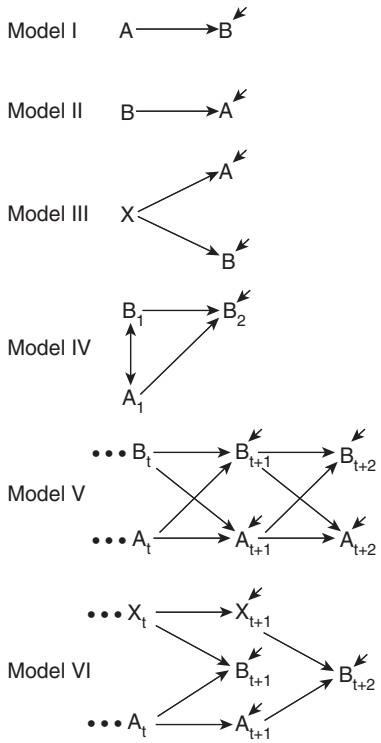

Figure 4.4 *Six possible true models of the relationship between attitude and behaviour*

important causes of behaviour apart from attitude; and a random component consisting of tens, hundreds, or even thousands of minor, independent, and unstable causes of behaviour. The latter causes would be impossible to specify in practice. (An alternative justification for assuming a random component is that 'there is a basic and unpredictable element of randomness in human responses which can be adequately characterized only by the inclusion of a random variable term': Johnston, 1984: 14.) Although, individually, the effects of these 'random shocks' are assumed to be small, in aggregate their effect may be quite large. Conceivably, the majority of the variance in behaviour could be explained by the random component, and only a minority by the major causal factors.

We have assumed no correlation between these other causes and attitude (indicated by the absence of a two-headed arrow between them). This assumption can be restated as saying that all variables that influence behaviour and that either influence attitude or are correlated with attitude because of common causes are specified in the model (or, more simply, that 'all relevant variables are included in the model'). This assumption is necessary in order to be able to interpret the regression coefficient for attitude as an unbiased estimate of the causal effect of attitude on behaviour. Unfortunately, this assumption is arbitrary and untestable (Clogg & Haritou, 1997; Sutton, 2002a).

Model II shows behaviour influencing attitude. In model III, a common cause (or causes), *X* influences both attitude and behaviour. In this model, attitude does not influence behaviour, and behaviour does not influence attitude. Thus, the correlation between attitude and behaviour is entirely due to *X*. Of course, the true model may include all these effects: attitude influences behaviour and vice versa (reciprocal causation) and other variables, *X* influence attitude and behaviour.

Cross-sectional designs

In a cross-sectional design, attitude and behaviour would be measured at the same time point for each of a sample of individuals. (Unlike the mixed design outlined above, we would typically have only a single measure of attitude and behaviour on each individual.) If we assume that model I is the true model, we can regress behaviour on attitude (i.e., conduct a regression analysis in which *B* is the dependent variable and *A* is the independent variable) to obtain an estimate of the causal effect of attitude on behaviour, in the form of the regression coefficient. (Note that the standardized coefficient is equal to the correlation in this simple case.) If model I is in fact the true model, and a number of other assumptions hold, then the coefficient will be an unbiased estimate of the true causal effect of attitude on behaviour. Thus, in order to draw causal inferences from cross-sectional data, we have to assume the truth of the hypothesized causal model, and our inference (the estimate of the size of the causal effect) is conditional on this and other strong assumptions. (This is also

true of other observational designs: causal inferences are always conditional on the truth of the hypothesized causal model.) Note that the observed correlation between attitude and behaviour in a cross-sectional study can be interpreted as having been generated by causal processes that have occurred *in the past* prior to the time of measurement.

Prospective longitudinal designs

Suppose we measure attitude at time 1 and behaviour at time 2 in a sample of individuals, and regress behaviour on attitude. We refer to this as *design 1*. This appears to be a stronger research design than the cross-sectional design outlined above because the correlation or regression coefficient between attitude and behaviour cannot include a component due to behaviour at time 2 influencing attitude at time 1. The temporal ordering of the measurement of attitude and behaviour rules out this explanation. However, as shown below, the correlation between attitude and behaviour may still be due to behaviour influencing attitude.

In prospective designs, causal lag needs to be considered. The causal lag is the time it takes for a change in attitude to produce a change in behaviour. Ideally, the length of the follow-up period should be approximately equal to the hypothesized length of the causal lag (Finkel, 1995; Sutton, 2002a). If the follow-up period is shorter than the causal lag, a change in attitude that occurs just prior to time 1 will not have produced its effect on behaviour by time 2; thus, the effect of recent changes in attitude will be missed at the later time point. If the follow-up period is too long, a change in attitude that takes place during the follow-up period may produce a change in behaviour within the same period. In both cases, the value of attitude at time 1 and the value of behaviour at time 2 will be mismatched. If the causal lag is very brief (as it may be, for example, in the case of the effect of attitude on intention), a prospective design is not appropriate and a cross-sectional design should be used (though, if feasible, an experimental study would be preferable – see below).

Regardless of the length of the causal lag, if behaviour is extremely stable over the follow-up period (in the sense that individuals show little change over time relative to one another), it makes little difference whether a cross-sectional or a prospective design is used; both should yield similar estimates of the effect of attitude on behaviour. This is likely to be the case, for example, when a quantitative aspect of behaviour such as frequency of exercising is assessed on two occasions over a short period of time (2 weeks, say). Here, the correlation between behaviour at time 1 and behaviour at time 2 is likely to be very high, indicating high stability, and whether the investigator uses behaviour at time 2 or behaviour at time 1 as the dependent variable in the analysis is likely to make little difference to the results.

The observation that behaviour is frequently quite stable (as indexed by a high correlation) over short time periods provides one rationale for extending the prospective design discussed above by including behaviour at time 1. In this design (which we call *design 2*), B_2 is regressed on A_1 and B_1. The assumption here is that the stability of behaviour arises from a causal influence of prior behaviour on future behaviour, that is, both attitude at time 1 and behaviour at time 1 influence behaviour at time 2 (see model IV in Figure 4.4). The two-headed arrow between A_1 and B_1 indicates that the correlation between attitude and behaviour at time 1 is treated as given: the model does not specify how it came about. Unlike a cross-sectional study, which analyses the current correlation between attitude and behaviour in terms of past causal processes, this design focuses on causal processes that are assumed to operate *over the follow-up period*. If model IV is true, then design 2 will give unbiased estimates of the two causal effects (and design 1 will give a biased estimate of the effect of A_1; given positive correlations between A_1 and B_1 and between B_1 and B_2, the effect of A_1 will be overestimated).

In design 2, the coefficient for B_1 is the estimated causal effect of B_1 on B_2, holding A_1 constant. (This is often referred to as the *stability coefficient*.) The coefficient for A_1 is the estimated causal effect of A_1 on B_2, holding B_1 constant.

This coefficient also has an interpretation in terms of behaviour *change*: it estimates the causal effect of A_1 on change in behaviour between time 1 and time 2 ($B_2 - B_1$), holding B_1 constant (Finkel, 1995). This can be shown as follows. The initial model is

$$B_2 = \beta_0 + \beta_1 A_1 + \beta_2 B_1 + \varepsilon$$

where β_0 is the intercept, β_1 and β_2 are the unstandardized coefficients for A_1 and B_1 respectively, and ε is the error term. Subtracting B_1 from both sides gives:

$$B_2 - B_1 = \beta 0 + \beta_1 A_1 + (\beta_2 - 1) B_1 + \varepsilon$$

The coefficient for A_1 is identical in the two equations. Thus, one could estimate β_1 by computing the change score and regressing it on A_1 and B_1. In practice, it is more convenient to regress B_2 on A_1 and B_1.

Model IV in Figure 4.4 can be thought of as being embedded in model V, in which attitude and behaviour influence each other over time. The model assumes that over each time interval, behaviour is influenced by prior behaviour and prior attitude, and attitude is influenced by prior attitude and prior behaviour; in other words, attitude and behaviour have *cross-lagged* effects. If model V (and hence model IV) is true, design 1 will yield a biased estimate of the effect of A_1 on B_2. For example, it can be seen from model V that the correlation between A_{t+1} and B_{t+2} will include a component due to the effects of B_t on A_{t+1} and B_{t+2} (via B_{t+1}). Thus, although design 1 rules out an effect of B_2 on A_1, it does not rule out the possibility that the correlation between A_1 and B_2 is partly or wholly due to the effect of relatively stable behaviour on attitude. Design 1 is of particular interest because it represents, in minimal form, a design that is commonly used to test theories of health behaviour, in which cognitive variables at time 1 are used to predict behaviour at time 2 without controlling for initial behaviour. By contrast, if model V is true, design 2 will yield the appropriate estimate of the causal effect of A_1 on B_2.

An alternative to controlling for prior behaviour statistically is to control for it by *stratification* or *restriction* (Weinstein, Rothman & Nicolich, 1998). If we select people with equivalent behaviour at time 1, there can be no effect of behaviour at time 1 on behaviour at time 2 and we can rule out a causal effect of behaviour on attitude as a possible explanation of the observed correlation between attitude at time 1 and behaviour at time 2. Note that, like design 2, this design focuses on the effect of attitude on behaviour over the follow-up period and can also be interpreted in terms of the effect of initial level of attitude on behaviour *change*.

Design 2 assumes that behaviour is partly determined by prior behaviour. However, it is difficult to explain how past behaviour can *directly* influence future behaviour (Sutton, 1994). More generally, the idea that a variable can directly cause a later version of itself (*autoregression*) is problematic (Allison, 1990; Liker, Augustyniak & Duncan, 1985; Stoolmiller & Bank, 1995). Stoolmiller and Bank give the following example:

> consider a simple experiment designed to study growth of money left in bank accounts. Suppose we deposit a range of sums of money in each of several banks paying a range of interest rates on deposits. In an AR [autoregressive] model with the initial sum of money and the interest rate as predictors of the amount of money at time 2, we would typically find that initial amount was a very strong predictor of money at time 2. But clearly if we … isolate the money at time 1 away from all suspected causal forces (e.g., in a shoebox under the bed), we will find at time 2, much to our dismay, that the money has failed to grow. Despite the fact that AR effects would be large in the bank example, they are not true direct causal effects. Interest causes money to grow, not initial amount of money … To discard interest as a predictor of change because it failed to compete with initial amount of money in an AR model would be an error. (1995: 271)

An alternative explanation for the stability of behaviour is that behaviour is stable because its underlying causes are stable. Model VI in Figure 4.4 shows an additional variable *X* which, like attitude, has a lagged causal effect on behaviour. In this model, there is no direct causal effect of prior behaviour on behaviour.

Behaviour will nevertheless be stable over time to the extent that its causes (*A* and *X*) remain stable over time. (Note that, although this model assumes that behaviour is not directly influenced by prior behaviour, it still assumes that *X* and *A* have autoregressive effects.)

Assuming model VI is true, the simplest observational design for estimating the effects of *A* and *X* on *B* would be to assess *A* and *X* at one time point and *B* at a later time point and then regress *B* on *A* and *X*, *without* controlling for prior behaviour. This is a version of design 1. If data were available from three time points (*X* and *A* measured at time 1; *X*, *A* and *B* measured at time 2; and *B* measured at time 3), this analysis could be done twice. However, a more sophisticated approach would be to use a *first difference model*, in which difference scores are calculated and change in behaviour is regressed on change in attitude and change in *X* (Liker et al., 1985; see also Allison, 1990). An advantage of this method is that the effects of unmeasured, stable causes of *B* are automatically controlled for. For example, suppose behaviour was also influenced by personality variables whose values did not change over the time period in question. If a first difference model were used, omitting such variables from the analysis would not bias the results. This method will work only when the values of *A* and *X* change for a substantial portion of individuals over time. If high stability was expected, the investigator could consider introducing an intervention between time 1 and time 2 in order to produce differential changes in *A* and *X* across individuals.

Returning to models that assume that behaviour is influenced by prior behaviour, an extension to design 2 is the cross-lagged panel design in which attitude and behaviour are both measured at two time points. As in design 2, B_2 is regressed on A_1 and B_1 but, in addition, B_2 is regressed on A_1 and B_1. If model V in Figure 4.4 is true, such a cross-lagged regression analysis estimates the effect of attitude on behaviour and the effect of behaviour on attitude, assuming that both effects have the same causal lag which is approximately equal to the length of the follow-up period. Note that an alternative analysis using the cross-lagged correlations is not recommended when the aim is to estimate causal effects (Campbell & Kenny, 1999; Rogosa, 1980). If the cross-lagged regression analysis is done using a structural equation modelling program, an estimate of the residual covariance or correlation between A_2 and B_2 can be obtained. This is the portion of the covariance or correlation between these two variables that cannot be explained by A_1 and B_1. If this residual term is substantial, one interpretation is that there are important variables that influence both attitude and behaviour that have been omitted from the analysis. This would cast doubt on the validity of the estimates of the causal effects in the model (Hertzog & Nesselroade, 1987).

Although we have considered only very simple models in the preceding discussion, many other models are possible. For example, we have assumed that attitude has a lagged effect on behaviour (and vice versa). In some cases, it may be more plausible to postulate 'synchronous' or almost instantaneous causal effects. For example, in the motivational model discussed by Weinstein, Rothman and Nicolich (1998), an increase in precautionary behaviour is assumed to lead promptly to a decrease in perceived risk.

Again, variables such as attitude may act as *predisposing factors* that increase the likelihood that other variables influence behaviour. For example, consider a study in which we select a sample of 12-year-olds who have never smoked, measure their attitude toward smoking, and then use this to predict whether or not they try smoking a cigarette by age 14. Attitude may act as a predisposing factor in the sense that an event such as being offered a cigarette may be more likely to lead to smoking among adolescents who hold a positive attitude.

Health behaviour researchers are recommended to carefully consider possible plausible causal models and to draw timeline and path diagrams to represent them before selecting an appropriate research design and analysis approach.

Experimental Designs

None of the observational designs outlined above can rule out the possibility that an

observed correlation between attitude and behaviour is partly or wholly due to omitted variables (as in model III in Figure 4.4). Where possible, health behaviour researchers should consider conducting experimental studies in which the determinant of interest (attitude in this example) is manipulated independently of other potential causes, with random assignment of participants to experimental conditions. In principle, such designs allow strong causal inferences to be drawn. More than one explanatory variable may be manipulated orthogonally in a factorial design. If repeated measures are obtained, experiments can be analysed at the within-individual level as well as at the between-individual or group level.

Of the theories of health behaviour in common use, only protection motivation theory (Rogers & Prentice-Dunn, 1997) has been subjected to extensive experimental testing. In view of the strong assumptions required to draw causal inferences from observational data, health psychology researchers should make much more use of randomized experiments to test predictions from theories of health behaviour. Such experiments should use proper randomization methods rather than arbitrary or alternate assignment of participants to conditions. The huge advantage of randomization is that, if it is done properly, it guarantees that any baseline differences between groups must be due to chance (Shadish, Cook & Campbell, 2002).

Intervention Studies

Such analytic or theory-testing experiments should be distinguished from intervention studies, which may also employ randomization to conditions. In intervention studies, the aim is usually to change a number of potential explanatory variables simultaneously in order to maximize the effect of the intervention on behaviour. Such studies can provide information about the extent to which the intervention effect on behaviour, if one is obtained, is mediated by the hypothesized explanatory variables (Baron & Kenny, 1986; Kenny, Kashy & Bolger, 1998). Mediation analyses are partly experimental and partly observational. In particular, interpretation of the relationship between the hypothesized mediating variable and the dependent variable requires the same assumptions as the analysis of other observational data (Sutton, 2002a).

THEORIES OF HEALTH BEHAVIOUR

Theories of health behaviour can be classified by range of application (general, health-specific, and domain- or behaviour-specific) and formal structure (stage versus non-stage theories) (Sutton, 2003; see Armitage & Conner, 2000, for a related classification scheme). General theories, such as the TPB (Ajzen, 1991, 2002b) and its predecessor the TRA (Fishbein & Ajzen, 1975), are those that, in principle, can be applied to a wide range of behaviours, not simply health-related ones. Health-specific theories like the health belief model (Strecher & Rosenstock, 1997) are specific to health-related behaviours. Behaviour- or domain-specific models have a still narrower range of application. For example, the AIDS risk reduction model (Catania, Kegeles & Coates, 1990) was developed to understand STD-preventive behaviour such as condom use.

Stroebe argues that general models should be preferred for the sake of parsimony: 'it is not very economical to continue to entertain specific theories of health behaviour unless the predictive success of these models is greater than that of general models of behaviour'(2000: 27). If we can use a single theory to explain why some young people use condoms consistently with new sexual partners while others do not, why some people engage in regular exercise more often than others, and why some people recycle their newspapers whereas others throw them away, this is much more useful and economical than developing different theories for each of these three behaviours. The argument, then, is that general theories should be preferred to health- or behaviour-specific theories unless the latter can be shown to be better in some important way. This suggests a strategy of always starting with a general theory and only modifying it if absolutely necessary when applying it to a new behaviour or behavioural domain.

The second important distinction is between stage and non-stage (or *continuum*) theories (Sutton, in press; Weinstein, Rothman & Sutton, 1998). These two types of theories have different formal structures and different implications for intervention. Stage theories assume that behaviour change involves movement through a sequence of discrete stages, that different factors are important at different stages, and therefore that different (*stage-matched*) interventions should be used for people in different stages. The best known stage theory is the transtheoretical model (Prochaska & Velicer, 1997), which has been applied to a wide range of health-related behaviours. The version of the model that has been used most widely in recent years specifies five stages: precontemplation, contemplation, preparation, action, and maintenance. Although the transtheoretical model is the dominant stage theory in the field of health behaviour, it suffers from serious conceptual and measurement problems and cannot be recommended in its present form (Sutton, 2001, in press). Another stage theory that is attracting increasing interest is the precaution adoption process model (Weinstein & Sandman, 1992), which is a health-specific model. Health behaviour researchers who are thinking of using stage theories need to be aware that they are complex and difficult to test (Sutton, 2000, in press; Weinstein, Rothman & Sutton, 1998).

This chapter focuses on continuum or non-stage theories. Each of these theories specifies a small set of proximal cognitive determinants of behaviour. (Note that some theories also include variables that, strictly speaking, cannot be described as cognitive determinants, for example skills and actual behavioural control.) The causal relationships specified by such theories can be represented in the form of a path diagram with behaviour on the far right. Figure 4.5 shows three prototypical representations. The variables *X*, *Y* and *Z* are the hypothesized cognitive determinants; *B* is behaviour. In Figure 4.5a, both *X* and *Y* are assumed to influence behaviour directly. In Figure 4.5b, the effect of *X* and part of the effect of *Y* are mediated by *Z*; *Z* is a *mediator* or *intervening variable* (Baron & Kenny, 1986). Thus, *Y* has direct and indirect effects on behaviour whereas *X* has only an indirect effect. In Figures 4.5a and b, the variables combine *additively*. By contrast, in Figure 4.5c, *X* and *Y* interact to influence behaviour. The effect of *Y* on behaviour depends on the level of *X*, that is, *X moderates* the relationship between *Y* and *B*. (Interactions are symmetric, so, alternatively, *Y* can be regarded as the moderating variable.) Figure 4.5c does not show the nature of the interaction. Where theories of health behaviour specify interactions, they are nearly always of the multiplicative or synergistic type, in which, to put it simply, the effects of two variables together are greater then their sum.

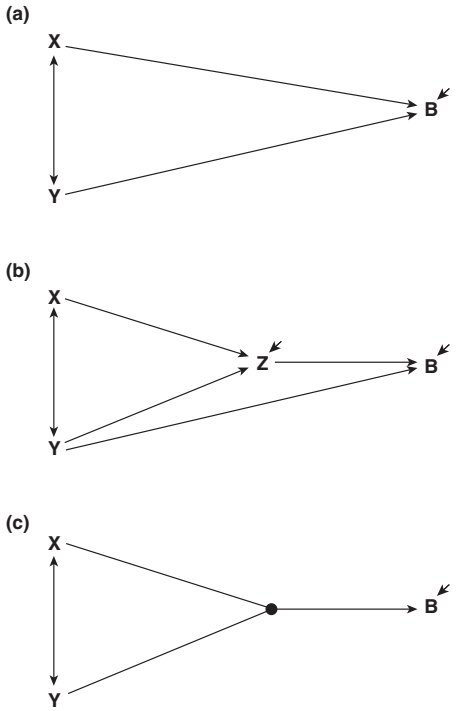

Figure 4.5 *Three prototypical representations of the causal relationships specified by theories of health behaviour (X, Y and Z are explanatory variables; B is behaviour)*

As in the preceding section, the small unlabelled arrows represent the errors or disturbances. Algebraic equations can be used as an alternative to the diagrammatic representation (although, as Pearl, 2000, points out, path diagrams contain more information than equations). Figures 4.5a and c can each be represented by a single equation, but Figure 4.5b

requires two equations because there are two endogenous (dependent) variables.

Although theories of health behaviour are sometimes described as 'static' models, they are dynamic in the sense that they specify causal relationships between variables whose values may change over time and may be deliberately changed through interventions. For example, the theory depicted in Figure 4.5b says that, if *Y* is held constant, a change in *X* will produce a change in *Z*, which in turn will produce a change in behaviour. Thus, theories of health *behaviour* are also theories of health *behaviour change*. However, with the exception of Bandura's (1997, 1998) social cognitive theory, the theories do not tell us *how* to change the variables on the far left (the exogenous variables) (Sutton, 2002c).

In Figure 4.5, the exogenous variables *X* and *Y* are linked by a two-headed arrow. This indicates that these variables *may* be correlated with each other but that this correlation is not explained by the theory; it is treated as given. The implicit assumption is that the correlation must be due to other variables external to the theory (common causes). If it seems plausible that the correlation is due to one variable influencing the other, then the two-headed arrow should be replaced by a single-headed arrow, thus making one of the exogenous variables endogenous. This apparently minor change can have major implications for estimates of the effect size and for interventions (Sutton, 2002c).

None of the diagrams in Figure 4.5 shows arrows from behaviour back to the explanatory variables. Most theories of health behaviour acknowledge that behaviour may influence cognitions, as well as vice versa, but few theories explicitly incorporate such feedback effects.

Ideally, theories of health behaviour should specify the causal lag for each of the causal relationships in the theory. Causal lag is an important consideration in deciding on an appropriate research design. An alternative approach is to try to estimate the causal lag empirically (Finkel, 1995).

Theories of health behaviour generally assume *linear* (straight-line) relationships. Indeed, none of the social cognition models in common use specifies curvilinear relationships. All the theories are specified (often implicitly) at the between-individual level and they are nearly always tested at the between-individual level. The main aim of research in this area can be characterized as 'putting numbers on the paths', where these numbers represent unbiased and precise estimates of the causal effects for particular health behaviours in particular target populations in the form of regression coefficients and their standard errors or confidence intervals.

We will not attempt to describe the main theories of health behaviour, even in outline form. The reader is referred to the book by Conner and Norman (1996, in press) for a detailed exposition of the major theories. (See also Sutton, 2002b, and Weinstein, 1993, for comparisons of theories.) The position taken in the present chapter is that there are too many theories of health behaviour and that this is hindering progress in the field. Progress would be more rapid if research efforts were concentrated on a small number of theories. This chapter therefore focuses on one theory, the TPB (Ajzen, 1991, 2002b), which can be argued is a prime candidate for guiding future research on health behaviour, for the following reasons: (1) it is a general theory; (2) the constructs are clearly defined and the causal relationships between the constructs clearly specified; (3) there exist clear recommendations for how the constructs should be operationalized (Ajzen, 2002a); (4) the theory has been widely used to study health behaviours (Ogden, 2003) as well as many other kinds of behaviours; and (5) meta-analyses show that it accounts for a useful amount of variance in intentions and behaviour (but see the later discussion of percentage of variance explained).

Although the next section focuses on the TPB, many of the points made apply equally to other theories of health behaviour.

THE THEORY OF PLANNED BEHAVIOUR

The TPB is shown in Figure 4.6. According to the theory, behaviour is determined by the strength of the person's *intention* to perform

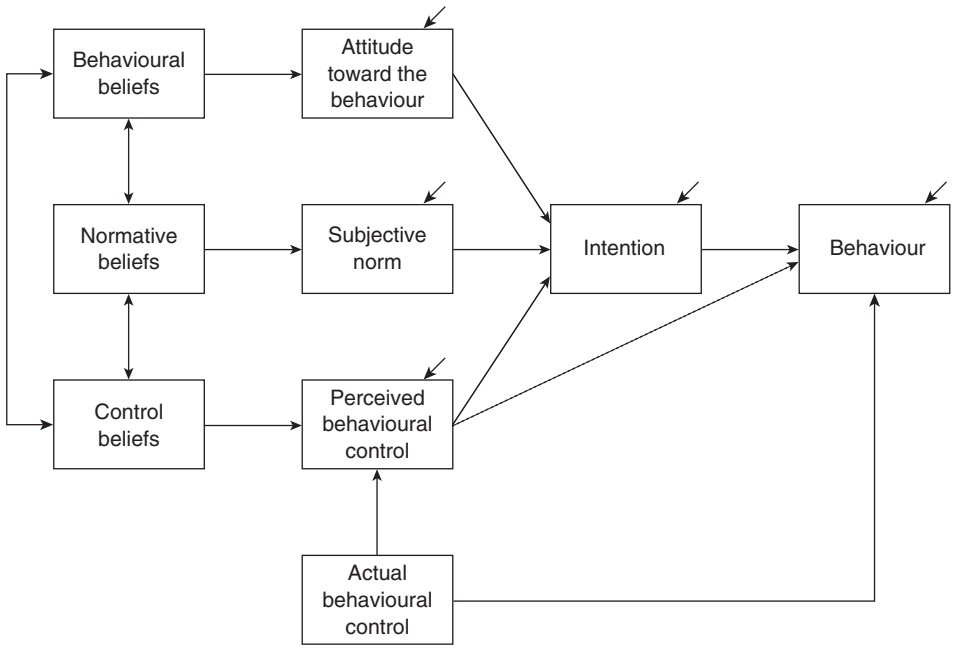

Figure 4.6 *The theory of planned behaviour*

that behaviour and the amount of *actual control* that the person has over performing the behaviour. According to Ajzen (2002b), intention is 'the cognitive representation of a person's readiness to perform a given behavior, and … is considered to be the immediate antecedent of behavior', and actual behavioural control 'refers to the extent to which a person has the skills, resources and other prerequisites needed to perform a given behavior'. Figure 4.6 also shows an arrow from perceived behavioural control to behaviour. *Perceived behavioural control* refers to the person's perceptions of their ability to perform the behaviour. It is similar to Bandura's (1997) construct of self-efficacy; indeed Ajzen (1991) states that the two constructs are synonymous. Perceived behavioural control is assumed to reflect actual behavioural control more or less accurately, as indicated by the arrow from actual to perceived behavioural control in Figure 4.6. To the extent that perceived behavioural control is an accurate reflection of actual behavioural control, it can, together with intention, be used to predict behaviour.

The strength of a person's intention is determined by three factors: their *attitude toward the behaviour*, that is, their overall evaluation of performing the behaviour; their *subjective norm*, that is, the extent to which they think that important others would want them to perform it; and their perceived behavioural control.

Attitude toward the behaviour is determined by the total set of *accessible* (or *salient*) *behavioural beliefs* about the personal consequences of performing the behaviour. Specifically, attitude is determined by $\sum b_i e_i$, where b_i is belief strength and e_i is outcome evaluation. Similarly, subjective norm is determined by the total set of accessible *normative beliefs*, that is, beliefs about the views of important others. Specifically, subjective norm is determined by $\sum n_j m_j$, where n_j is belief strength and m_j is motivation to comply with the referent in question. Finally, perceived behavioural control is determined by accessible *control beliefs*, that is, beliefs about the presence of factors that may facilitate or impede performance of the behaviour. Specifically, perceived behavioural control

is determined by $\sum c_k p_k$, where c_k is belief strength (the perceived likelihood that a given control factor will be present) and p_k is the perceived power of the control factor (the extent to which the control factor will make it easier or more difficult to perform the behaviour).

According to the theory, changing behaviour requires changing these underlying beliefs and/or actual behavioural control (Sutton, 2002c).

The *principle of correspondence* (Ajzen & Fishbein, 1977; Fishbein & Ajzen, 1975) or *compatibility* (as it was renamed by Ajzen, 1988) states that, in order to maximize predictive power, all the variables in the theory should be measured at the same level of specificity or generality. This means that the measures should be matched with respect to the four components of action, target, time and context (see earlier section 'Defining and Measuring Behaviour'). Most researchers who use the TPB recognize the importance of using compatible measures, though the principle is frequently violated in empirical applications of the theory. Researchers who use other theories seem largely unaware of the principle. The rationale given for the principle is a pragmatic one: it improves prediction. Presumably, however, there is also a *theoretical* rationale for the principle, namely that, by measuring the TPB variables at the same level of specificity, we are matching cause and effect (Sutton, 1998).

Although the TPB holds that all behaviours are determined by the same limited set of variables, each behaviour is also substantively unique, in two senses (Fishbein, 2000). First, for a given population or culture, the relative importance of attitude, subjective norm and perceived behavioural control may vary across different behaviours. For example, some behaviours may be influenced mainly by attitude, whereas other behaviours may be influenced mainly by subjective norm. Ogden (2003) points out that many studies using the TPB find no role for one or other of the three putative determinants of intention and therefore that the theory 'cannot be tested'. However, this represents a misunderstanding of the TPB. If at least one of the components is found to predict intention in a given study, this is consistent with the TPB. Nevertheless, it is a weakness of the theory that it does not specify the conditions under which intention will be mainly influenced by attitude, subjective norm or perceived behavioural control.

The second sense in which each behaviour is substantively unique is that, for a given population or culture, the behavioural, normative and control beliefs that underlie attitude, subjective norm and perceived behavioural control respectively may also differ for different behaviours. In the same way, for a given behaviour, the relative importance of attitude, subjective norm and perceived behavioural control, and the content of the underlying beliefs, may vary across different cultures or populations.

The TPB is a general theory. In principle, it can be applied to any target behaviour without needing to be modified. For example, in applying the theory to a health-related behaviour, there should be no need to add a variable representing risk perceptions. If beliefs about the health risks of the behaviour (or its effect on reducing risk) are salient to a substantial proportion of the target population, this should emerge in an elicitation study that uses open-ended questions to elicit accessible beliefs (Ajzen, 2002a; Ajzen & Fishbein, 1980; for an example of an elicitation study, see Sutton et al., 2003).

Like other theories of health behaviour, the TPB is a *causal model* and should be treated as such. It says, for instance, that if you hold constant a person's subjective norm, perceived behavioural control and actual behavioural control and you change their attitude toward the behaviour, this will lead to a change in their intention (assuming that attitude is a determinant of intention for the behaviour in question in this target group), and this in turn will lead to a change in their probability of performing the behaviour (assuming that the behaviour is at least partly under the person's control).

The TPB is often depicted without actual control in the path diagram and, to date, has always been tested without measuring actual control. In this case, the direct path from perceived behavioural control to behaviour is causally ambiguous (Sutton, 2002a, 2002c). As

already pointed out, one rationale for this direct link is that perceived behavioural control can often be used as a substitute for actual control (Ajzen, 1991). Although actual control influences behaviour, it is argued that it is difficult to measure and is less interesting psychologically than perceived control. Perceived control can be used as a proxy for actual control to the extent that people's perceptions of control are accurate. According to this rationale, the direct link between perceived behavioural control and behaviour is not a causal path, and changing perceived behavioural control would not lead to behaviour change directly. (It could lead to behaviour change indirectly, of course, via a change in intention.) In order to change behaviour directly, it is necessary to change actual control.

However, Ajzen suggested a second rationale for the direct link between perceived behavioural control and behaviour: 'holding intention constant, the effort expended to bring a course of behavior to a successful conclusion is likely to increase with perceived behavioral control. For instance, even if two individuals have equally strong intentions to learn to ski, and both try to do so, the person who is confident that he can master this activity is more likely to persevere than is the person who doubts his ability' (1991: 6). Note that this effect is held to be mediated by 'effort' and 'perseverance', neither of which are constructs in the theory.

Putting these two rationales together, this means that if we observe an independent predictive effect of perceived behavioural control on behaviour in an observational study in which actual control is not measured, this may be due partly to a causal effect of perceived behavioural control on behaviour and partly to a correlation induced by actual behavioural control influencing both perceived behavioural control and behaviour (Sutton, 2002a, 2002c). More generally, failing to measure and control for the effects of actual behavioural control will lead to biased estimates of the causal effects of perceived behavioural control and intention on behaviour, unless it can be assumed that perceived control is an accurate reflection of actual control (i.e., that perceived and actual control are perfectly correlated and this correlation arises from a direct causal effect of actual on perceived control).

Although Figure 4.6 shows an arrow going directly from actual control to perceived control, this is inconsistent with the theory's assumption that the effects of any variable on perceived control must be mediated by control beliefs. The absence of arrows, either one- or two-headed, between actual control and behavioural and normative beliefs respectively can be interpreted as indicating zero correlations and no direct causal influence in either direction. However, to date, Ajzen has not discussed these possible relationships. If actual control were related to one or both of these variables, again this would have implications for the interpretation of regression analyses from which actual control was omitted.

A further complexity concerns the interaction between perceived behavioural control and intention on behaviour that was postulated by Ajzen and Madden (1986). Ajzen (2002b) states it as follows: 'Conceptually, perceived behavioral control is expected to moderate the effect of intention on behavior, such that a favorable intention produces the behavior only when perceived behavioral control is strong.' He also notes that, 'In practice, intentions and perceptions of behavioral control are often found to have main effects on behavior, but no significant interaction' (see also Conner & Armitage, 1998). This interaction derives from an interaction between intention and *actual* control (and so would be predicted to occur only in situations in which perceptions of control are accurate). In particular, intention is expected to have a stronger influence on behaviour, the greater the degree of actual control the person has over the behaviour. As Ajzen (2002b) puts it, 'successful performance of the behavior depends not only on a favorable intention but also on a sufficient level of behavioral control'. For simplicity, this interaction is not shown in Figure 4.6.

Extensions of the TPB

There have been numerous attempts to extend the TPB by adding variables such as anticipated

regret, moral norm and self-identity (Conner & Armitage, 1998). For the sake of parsimony and theoretical coherence, candidate variables should be provisionally accepted as official components of the theory only if a number of conditions are satisfied. First, there should be sound *theoretical* reasons for believing that a given candidate variable influences intention or behaviour independently of the existing variables, that is, that the variable has a direct causal effect on intention or behaviour. In some cases, it is possible that the proposed additional variable is already captured by one of the existing variables.

Second, in order to retain the existing structure of the TPB, the proposed new variable should have an expectancy-value basis like attitude, subjective norm and perceived behavioural control; in other words, the new variable should be determined by accessible beliefs that are specific to the target behaviour. This would seem to rule out some variables, for example self-identity. This also means that the expectancy-value basis of *descriptive norm* (the belief that significant others are or are not performing the target behaviour), which Ajzen (2002a) has proposed as a subcomponent of subjective norm in the latest version of the theory, needs to be specified. This requirement, that any additional variable is homologous to the existing variables, also implies that including too many additional variables in the theory would make it unwieldy to use in practice. Furthermore, additional open-ended questions for eliciting accessible beliefs would need to be devised for use in pilot studies. This has not yet been done for descriptive norm (Ajzen, 2002a).

Third, measures of a proposed new variable should be shown to have discriminant validity with respect to measures of the existing components, in other words to be measuring something different from measures of the existing variables.

Finally, the new variable should be shown to predict intention and/or behaviour independently of the existing components in studies in which the latter are well measured in accordance with published recommendations. It is likely that there are many false positive findings in the literature because the existing components are not always optimally measured. Of course, if the aim is simply to improve the predictive power of the theory rather than to specify additional determinants of intention, only the last of the requirements set out above is relevant.

How Well Does the Theory Perform?

There have been remarkably few experimental tests of the TPB or its predecessor the TRA (Sutton, 2002a). The vast majority of studies have used observational designs. Table 4.1 summarizes the findings from meta-analyses of research using the TPB in terms of the multiple correlation R and its square (which can be interpreted as the proportion of variance explained) for predicting intention and behaviour. Also shown is an effect size index called f^2 that is recommended by Cohen (1988, 1992) for use in power analysis where the statistical test involves multiple correlations.

With the exception of Ajzen (1991), all the meta-analyses explicitly or by implication restricted the analysis of prediction of behaviour to prospective studies in which intention and perceived behavioural control were measured at time 1 and behaviour was measured at time 2, that is, they used a version of what we referred to earlier as design 1. The meta-analyses differed in a number of ways, including the selection criteria for the studies. However, there is not space here to give a detailed comparison and critique of the reviews or to map the degree of overlap between them. Instead we focus on the 'headline' figures to gain an impression of the predictive utility of the theory. The findings for both intention and behaviour show reasonable consistency. For intention, the multiple correlations range from 0.59 to 0.71 (between 35 per cent and 50 per cent of variance explained). Prediction of behaviour was lower, as expected, with the multiple correlation ranging between 0.51 and 0.59 (between 26 per cent and 35 per cent of the variance explained).

Godin and Kok (1996) found differences between different kinds of behaviours with respect to how well the theory predicted

Table 4.1 *Summary of effect sizes[a] from meta-analyses of the theory of planned behaviour*

	Predicting intention (BI) from AB, SN and PBC				Predicting behaviour from BI and PBC			
Meta-analysis	k[b]	R	R^2	f^2	k[b]	R	R^2	f^2
Ajzen (1991)	19	0.71	0.50	1.00	17	0.51	0.26	0.35
Godin & Kok (1996)[c]	76	0.64	0.41	0.69	35	0.58	0.34	0.52
Sheeran & Taylor (1999)[d]	10	0.65	0.42	0.72	–	–	–	–
Albarracín et al. (2001)[d]	23	0.71	0.50	1.00	23	0.53	0.28	0.39
Armitage & Conner (2001)	154	0.63	0.39	0.64	63	0.52	0.27	0.37
Hagger et al. (2002)[e]	49	0.67	0.45	0.82	35	0.52	0.27	0.37
Trafimow et al. (2002):[f]								
PBC as perceived difficulty	11	0.66	0.44	0.79	9	0.59	0.35	0.55
PBC as perceived control	11	0.59	0.35	0.53	9	0.58	0.34	0.52

BI = behavioural intention; AB = atttitude to behaviour; SN = subjective norm; PBC = perceived behavioural control.

[a]Effect sizes are given in terms of the multiple correlation R, R^2, and $f^2 = R^2/(1 - R^2)$. According to Cohen (1988, 1992), an f^2 value of 0.35 is 'large'.

[b]k is the number of datasets.

[c]Restricted to studies of health-related behaviours.

[d]Restricted to studies of condom use.

[e]Restricted to studies of physical activity.

[f]Restricted to studies that included measures of both 'perceived difficulty', defined as 'the extent to which the person believes that performing the behaviour would be easy vs. difficult or the level of confidence about performing the behaviour' (p. 11) and 'perceived control', defined as 'the extent to which the behaviour was perceived to be under or outside one's control or was "up to me"' (p. 11).

intentions and behaviour. For example, for behaviour, the theory worked better in studies of HIV/AIDS-related behaviours than in studies of 'clinical and screening' behaviours. However, these results were based on small numbers of studies, and possible confounds such as sample characteristics and differences in how the TPB variables were measured were not examined. Godin and Kok's review needs to be updated and extended.

Should we be encouraged or discouraged by these results? The answer depends on the standard of comparison. One possible standard is the ideal maximum of 100 per cent. Clearly, the theory does not perform well by this standard. In practice, however, the maximum percentage of variance that can be explained in a real application is often substantially less than 100; one reason for this will be discussed later in the chapter. There are other more realistic standards of comparison. Another possible benchmark is provided by the effect sizes that are typically found in the behavioural sciences using a diverse range of outcomes and predictors. According to Cohen's (1988, 1992) operational definitions, the effect sizes in Table 4.1 are 'large' for both intention and behaviour. In evaluating the predictive performance of the TPB, it is important to remember that it is highly parsimonious, at least when direct measures rather than indirect (belief-based) measures of its constructs are used. Thus, although it explains no more than 50 per cent of the variance in intention, on average, it achieves this level of performance with only three predictors. In addition, although percentage of variance explained is widely used as a measure of effect size, it tends to give a rather pessimistic impression (Rosenthal & Rubin, 1979; Sutton, 1998).

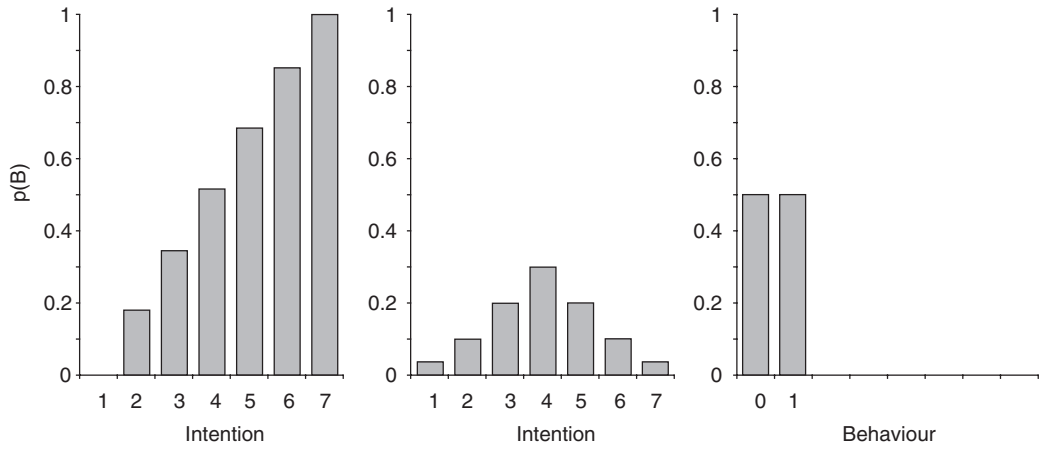

Figure 4.7 *Example showing a perfect linear relationship between a seven-point measure of intention and a dichotomous measure of behaviour and the distributions of the two variables*

Reasons for Poor Prediction

There are a number of important methodological and measurement reasons why theories such as the TPB often have lower predictive power than we would prefer (Sutton, 1998). For example, consider the simple case of predicting behaviour (measured at time 2) from intention (measured at time 1). Intention is often measured using a seven-point semantic differential rating scale. But the measure of behaviour is often a dichotomy, that is people either perform the behaviour or they don't; indeed, this is the classic application of the TPB. However, when predicting a variable with two categories from a variable with seven categories, it is not possible to obtain a perfect correlation or 100 per cent of the variance explained (unless respondents treat the intention measure as if it consisted of only two categories). It is possible to obtain a perfect correlation between the variables only if the distributions match, which means that they have to have an equal number of response categories.

Figure 4.7 shows a hypothetical example. The graph on the far left of the figure shows a *perfect* linear relationship between the seven-point intention measure and the probability of performing the behaviour. So no one who scores 1 on the intention scale performs the behaviour whereas everyone who scores 7 does so. The middle graph shows the distribution on the intention measure, which is approximately normal, and the graph on the right shows the distribution on the behaviour measure, showing an even split: half the people perform the behaviour and half do not.

In this example, intention explains *less than 20 per cent* of the variance in behaviour. The problem is that the two measures have a different number of response categories and therefore the distributions cannot match. For maximum prediction, we need to make sure that the number of response categories is equal. Note that this problem has nothing to do with the use of dichotomous measures *per se* or with skewed distributions. If we had a dichotomous measure of intention, we could in principle explain 100 per cent of the variance in a dichotomous measure of behaviour. This would be the case even if the distributions were highly skewed. So if we had 100 people in the sample and only one of them performed the behaviour while 99 did not, it would still be possible to explain 100 per cent of the variance in behaviour if the sample were equally skewed on the measure of intention and if the one person who performed the behaviour was also

the one person who intended to do so. (This example also shows that the fact that a variable has a small (but non-zero) observed variance cannot on its own explain why it does not correlate highly with other variables.) The same argument would apply if we had a seven-point measure of intention and a five-point measure of behaviour, although the effect would be less dramatic.

There is an important theoretical issue here as well. Theories like the TPB do not explain how intention, which is conceived of as a continuous variable, translates into a binary outcome – performance or non-performance of the behaviour.

Other reasons for poor prediction include random measurement error in the measures of intention and/or behaviour, violation of the principle of compatibility, and lack of stability in intentions (Sutton, 1998). Such measurement factors help to explain both the intention–behaviour 'gap' (Sheeran, 2002) and why theories of health behaviour often do not explain as much variance as we would like them to.

Undue Emphasis on Amount of Variance Explained

Investigators naturally want to maximize the amount of variance explained by a theory and, other things being equal, would usually prefer a model that explains more variance to one that explains less. However, it can be argued that undue emphasis is placed on the total amount of variance explained by theories such as the TPB. First, from the standpoint of assessing the potential of a theory as the basis for interventions, the proportion of variance in behaviour explained by the theory as a whole is less relevant than the proportion explained by the variables on the far left of Figure 4.6. This is because, according to the theory, it is not possible to intervene *directly* to change intention, for example. Interventions have to be applied to the exogenous variables, that is, to the beliefs that are assumed to underlie attitude, subjective norm and perceived behavioural control. It is therefore important to estimate the percentage of variance in behaviour explained by the variables on the far left, or the *effective variance explained* (Sutton, 2002c). If the effects of the far left variables on behaviour are completely mediated by the other variables in the theory, the effective variance explained will be lower, often much lower, than the variance explained by the theory as a whole.

The effective variance explained can be estimated in a single study simply by regressing behaviour on the variables on the far left, omitting the hypothesized mediating variables from the regression model. Where a primary study or meta-analysis reports the correlations among the exogenous variables and between the exogenous variables and behaviour, the effective variance explained can be computed by entering the correlation matrix into a regression or structural equation modelling program. For example, from Table 3 in Albarracín, Johnson, Fishbein and Muellerleile's (2001) meta-analysis of the TPB applied to condom use, it can be calculated that behavioural beliefs, normative beliefs and perceived behavioural control together explained 13.3 per cent of the variance in condom use, which is substantially lower than the variance explained by intention and perceived behavioural control (Table 4.1). The unique variance explained by each of these components was 2.9 per cent, 1.7 per cent and 1.3 per cent respectively. (A direct measure of perceived behavioural control was used because very few studies assessed control beliefs.)

Why have we apparently lost all this explained variance? One way to look at this is to ask what we gain by including intention in the theory. By adding intention, we gain in terms of explanation, because we have specified a potential mechanism by which attitude, subjective norm and perceived behavioural control (and their underlying beliefs) influence behaviour. We are 'filling in the causal chain'. We also gain in predictive terms, because intention adds to the prediction of behaviour over and above attitude, subjective norm and perceived behavioural control. But this gain in predictive power is not helpful for the purposes of producing behaviour change because of causal dilution. The only way we can change intention, according to the theory,

is by changing attitude, subjective norm and perceived behavioural control, and the only way we can change these variables is by changing their underlying beliefs. But behavioural beliefs, for example, may not completely determine attitude; attitude, subjective norm and perceived behavioural control do not completely determine intention; and intention does not completely determine behaviour. Although intention may add to the prediction of behaviour over and above attitude, subjective norm and perceived behavioural control, the additional predictive power provided by intention is useless for the purposes of behaviour change because it arises from *unspecified causes* of intention, as represented by the small arrow pointing to intention in Figure 4.6. Because they are unspecified, they cannot be targeted in an intervention. So we cannot exploit the additional predictive power provided by intention.

The same argument would apply to any theory that specifies a causal chain. Of course, if an exogenous variable has a *direct* effect on behaviour as well as indirect effects, this may offset the dilution effect. An alternative way of gauging the intervention potential of a theory is to use the path-analytic calculus (Heise, 1975; Kenny, 1979) to calculate the total effect of each of the variables on the left-hand side on behaviour, controlling for other relevant variables. Either the standardized or the unstandardized regression coefficients could be used. The total effect can be interpreted as an estimate of the effect on behaviour of increasing the variable on the left by one unit, while holding constant the other variables on the left-hand side.

The example outlined above of effective variance explained should be treated as illustrative only. Because of the way that indirect measures of attitude and subjective norm are computed in applications of the TPB, the analyses involved product terms or multiplicative composites. A problem that affects such analyses is that the correlations between a multiplicative composite and other variables may vary depending on the particular scoring schemes used for its components (Bagozzi, 1984; Evans, 1991; French & Hankins, 2003). (Although the correlations are difficult to interpret, the multiplicative relationship between, for example, behavioural belief strength and outcome evaluation on attitude can be tested using standard approaches for testing interactions in multiple regression: Aiken & West, 1991; Sutton, 2002c; for an example, see Sutton, McVey & Glanz, 1999).

The second reason for arguing that undue emphasis is placed on the amount of variance explained is that a regression model that explains more variance in behaviour is not necessarily more valid than one that explains less variance. The validity of a regression model depends on the validity of its underlying assumptions, not on the proportion of variance it explains. Sutton (2002a) gives several hypothetical examples, including one in which close to 100 per cent of the variance is explained but the estimates of the causal effects of the predictor variables are seriously biased, and another showing that unbiased estimates of causal effects can be obtained even if the regression model does not explain a large proportion of variance in the criterion.

Is the TPB Too 'Rational'?

Theories like the TPB may be criticized for providing an unrealistically rational explanation of behaviour. However, the term 'rational' has several different meanings. Behaviour as explained by TPB can be regarded as rational in some ways but not in others. On the one hand, the theory holds that a person's behaviour will tend to be consistent with their accessible beliefs. Such consistency can be regarded as rational in one sense of the word. Furthermore, the TPB assumes that beliefs are combined in a systematic way such that a person's attitude towards a given behaviour, for example, is a mathematical function of the belief strengths (subjective probabilities) and outcome evaluations (utilities) of the accessible behavioural beliefs. The function derives from the expected value and expected utility models of 'economic man', which have a long history of use as normative models of decision-making (Edwards, 1954).

On the other hand, people's accessible beliefs may be incomplete and incorrect or influenced by strong emotions. For instance, a person may erroneously believe that a particular health behaviour is doing them good when in fact it is not. Thus, intentions and behaviour may be based on information that is incomplete and incorrect. Furthermore, although some decisions may involve conscious deliberation and careful weighing up of pros and cons, in many cases the processes involved in the formation and modification of beliefs, attitudes and intentions may be largely automatic (Ajzen & Fishbein, 2000; Fishbein & Ajzen, 1975). For example, a person's attitude toward a particular behaviour may be automatically updated when new information about the behaviour is received, and this attitude may be automatically elicited and guide behaviour in relevant situations. (However, although it seems plausible that automatic processes control the formation and change of beliefs, attitudes and intentions, for most health-related behaviours it seems less plausible to suggest that behaviour itself is automatically elicited.)

A BROADER THEORETICAL FRAMEWORK

The TPB, like other social cognition models, does not rule out other causes of behaviour. Many other factors such as socio-demographic, cultural and personality factors may influence behaviour, but these are assumed to be distal factors, in other words to be farther removed from the behaviour than the proximal factors specified by the theory. Thus, the TPB divides the determinants of behaviour into two classes: a small number of proximal determinants, which are specified by the theory (i.e., are internal to the theory); and all other causes, which are left unspecified but which are assumed to be distal and to influence behaviour only via their effects on the proximal determinants. In this sense, the TPB is sometimes said to be *sufficient*.

There are a number of ways in which external factors may impact on the internal variables and on behaviour. First, external factors may influence the beliefs that are assumed to underlie attitude, subjective norm and perceived behavioural control. For instance, people in non-manual occupations may have a *greater number* of accessible beliefs; for example, when asked to list the advantages and disadvantages of performing a given behaviour, people in non-manual occupations may list a greater number of advantages. External factors may also influence the *content* of accessible beliefs. Compared with people in manual occupations, those in non-manual occupations may hold different kinds of accessible behavioural beliefs; for example, beliefs about the health consequences of the behaviour may be more common among people in non-manual occupations. Neither of these effects would necessarily lead to a difference in behaviour between the two groups. Only if the total belief scores differed between the two groups would we expect a difference in behaviour. For instance, if $\sum b_i e_i$ was higher, on average, among people in manual occupations, then, assuming that attitude was an important determinant of intention for the behaviour in question and that the behaviour was under volitional control, we would expect to observe differences in behaviour between the two groups. Differences in one component of the theory may offset differences in another component. For instance, people in manual occupations may be higher on $\sum b_i e_i$ but lower on $\sum sn_j mc_j$; with the result that there is no difference in behaviour between the two groups.

Second, external factors may influence attitude, subjective norm or perceived behavioural control *directly* without influencing the underlying beliefs. Such effects would be inconsistent with the assumptions of the theory. For instance, the theory holds that attitude is completely determined by $\sum b_i e_i$; in the same way, subjective norm and perceived behavioural control are held to be completely determined by normative beliefs and control beliefs respectively.

Third, external factors may influence intention directly without influencing behavioural, normative, or control beliefs, and hence without influencing attitude, subjective norm or perceived behavioural control. Again, such an effect is inconsistent with the assumption that intention is completely determined by attitude,

subjective norm and perceived behavioural control. However, if the assumption is relaxed to allow other determinants of intention (e.g., anticipated regret, moral norm), this provides possible pathways by which an external factor could influence intention directly without influencing the official components of the theory.

Fourth, external factors could influence actual control. Figure 4.6 shows actual control influencing perceived behavioural control and behaviour. Thus, there are three distinct pathways by which an external factor could influence behaviour via actual control: (1) from actual control to perceived control to intention to behaviour; (2) from actual control to perceived control then directly to behaviour; and (3) from actual control directly to behaviour.

Fifth, external factors could influence behaviour directly, that is, they could bypass *all* the components of the theory. Conceptually, it is difficult to distinguish this mechanism from the preceding one. Actual control is a nebulous concept that could be thought of as including – or being influenced by – almost every factor that has a systematic influence on behaviour apart from those factors whose effects on behaviour are mediated by the other components of the theory.

Finally, external factors could moderate one or more of the causal relationships in the theory. For example, for a given behaviour, attitude may be a more important determinant of intention among those in manual occupations compared with those in non-manual occupations. Or the size of the causal effect of intention on behaviour may differ for the two groups.

A strategy for guiding future research on the determinants of health behaviour is to continue to use the TPB as a model of the proximal determinants of a given behaviour and to specify external factors that are hypothesized to influence the components of the theory or to influence behaviour directly, that is to develop theories that relate external factors to the theory's components. In effect, this is extending the causal model representing the TPB to the left, specifying the more distal causes of a particular behaviour and the mechanisms by which they influence the components of the theory and behaviour. (The causal model could also be extended to the right by including physiological sequelae of the behaviour and relevant health or disease outcomes, thus 'integrating psychology and epidemiology'; see Hardeman et al., submitted.)

The number of potential external factors or distal causes of a particular behaviour is huge. Other psychological variables such as personality factors may affect health behaviours (Contrada & Goyal, 2004, Chapter 6 in this volume). Variables such as age and sex may influence health behaviours and can be thought of as being on the far left of a complex causal model that has the TPB variables and behaviour on the far right. Health behaviours may also be influenced by biological factors. For example, there may be genetic influences on smoking. Like sex and age, genetic factors will also be located on the far left of the causal model.

All the potential causes of health behaviours mentioned so far are located at the (between-) individual level. However, there are numerous other factors that may influence health behaviours that can be summarized by the label 'social' factors. Several different classifications of such social factors have been proposed by theorists who have contributed to the development of what is known in the fields of health promotion and public health as the 'social ecological framework'. For example, McLeroy and colleagues distinguished the following sets of determinants of health behaviour:

> (1) Intrapersonal factors – characteristics of the individual such as knowledge, attitudes, behaviour, self-concept, skills, etc. This includes the developmental history of the individual.
> (2) Interpersonal processes and primary groups – formal and informal social network and social support systems, including the family, work group, and friendship networks.
> (3) Institutional factors – social institutions with organizational characteristics, and formal (and informal) rules and regulations for operation.
> (4) Community factors – relationships among organizations, institutions, and informal networks within defined boundaries.
> (5) Public policy – local, state, and national laws and policies. (1988: 355)

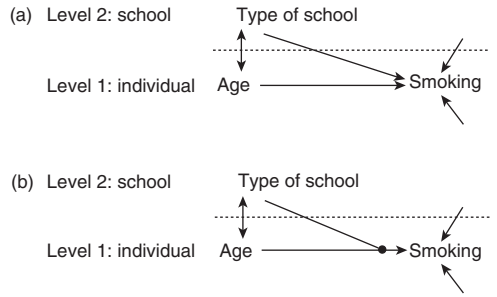

Figure 4.8 *Two-level path diagrams showing (a) a main effect of school type and (b) a cross-level interaction*

To many psychologists, the social ecological framework may seem vague and difficult to operationalize. However, multilevel modelling (Bryk & Raudenbush, 1992; Duncan et al., 1998; Hox, 2002) provides a way of operationalizing the framework and of organizing the social causes of health behaviours. From this perspective, the last four items on the above list refer to higher-level units or entities of which the individual is a 'member' and whose characteristics may influence the individual's health-related behaviours either directly or indirectly.

To give a simple concrete example, an adolescent's smoking behaviour may be influenced by the characteristics of the school they attend as well as by individual factors. This is shown in Figure 4.8a. The variables below the dotted line are (between-) individual-level variables including the dependent variable. So, the age of the student (an individual-level variable) is shown as influencing the likelihood that he or she smokes. However, the dependent variable may also be influenced by characteristics of the school, for example whether the school has a strict non-smoking policy, whether it is an independent (fee-paying) or a state school, and the proportion of students who receive free school meals (an index of deprivation). Figure 4.8a shows type of school directly influencing the likelihood that the individual student smokes. Type of school is located above the dotted line to indicate that it is a school-level (level 2) variable. Two error terms are shown in the diagram, one originating from level 2, the other from level 1.

School characteristics may be correlated with individual characteristics. In this example, type of school is shown (by the two-headed arrow) as potentially being correlated with age. In other words, the students who attend one type of school may be older, on average, than those who attend another type of school. The arrow from type of school to smoking is interpreted to mean the causal effect of type of school on smoking, controlling for possible differences in age between different types of school. Older students may be more likely to smoke, and students who attend types of school that cater for older students may be more likely to smoke for this reason. However, the path diagram indicates an *independent* causal effect of type of school on smoking. This can be labelled a *contextual* causal effect, meaning an effect that cannot be accounted for by the *compositional* effect of different types of schools having different kinds of students (in this case, students of different ages). More generally, arrows originating from level 2 indicate contextual causal effects – effects of level 2 variables that cannot be accounted for by the compositional effects of different kinds of individuals being associated with different kinds of level 2 units.

Some causal effects of level 2 variables on the individual-level dependent variable may be mediated by other individual-level variables. To give a simple example, whether or not a school has a strict non-smoking policy may influence an individual student's attitudes to smoking which in turn influences the likelihood that they smoke. Causal pathways may involve more than one variable at level 2 and more than one variable at level 1. One school-level variable may influence another school-level variable which influences one individual-level variable which in turn influences another individual-level variable.

Level 2 variables may also interact with level 1 variables to influence the dependent variable (see Figure 4.8b). For example, the effect of age on smoking may differ in different types of school (e.g., there may be a weaker effect of age on smoking in schools that have a strict non-smoking policy). This is a *cross-level interaction*. Another way of putting this is to say that

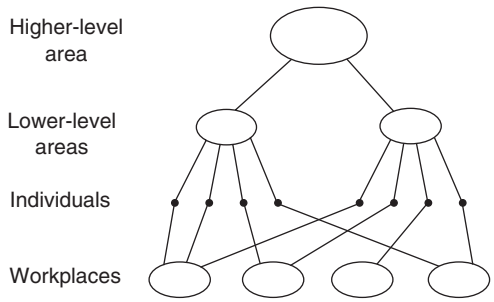

Figure 4.9 *Imperfect hierarchical relationship between individuals and higher-level units*

type of school moderates the effect of age on smoking (or that age moderates the effect of type of school on smoking).

This simple two-level example can be extended to three or more levels. For example, a student's smoking may be influenced by characteristics of the class of which they are a member as well as the school that they attend. Individual, class and school form three hierarchical levels. School-level variables may influence individual-level variables directly or through class-level variables.

However, in general, the levels of factors that may influence an individual's health behaviour do not form a pure hierarchy. For example, individuals' health behaviour may be influenced by characteristics of the neighbourhood in which they live, and, for those who are in work, by characteristics of the workplace. Typically, employees at a given workplace will live in many different neighbourhoods, and the residents of a given neighbourhood will work in many different workplaces. Thus, while both neighbourhood and workplace are at a higher level than the individual, the levels do not form a pure hierarchy (see Figure 4.9). Instead, individuals are nested within a cross-classification of workplaces by neighbourhoods (Rasbash & Browne, 2001). In this example, an individual's health behaviour may be influenced by individual-level variables, workplace-level variables and neighbourhood-level variables, operating additively or interactively. Interactions may be *within-level* (e.g., interactions between two or more workplace-level variables) or *cross-level* (e.g., interactions between workplace-level and neighbourhood-level variables).

For example, some workplaces may have a non-smoking policy whereas others do not. This variable (has a non-smoking policy, yes/no) is a characteristic of the workplace: it is a workplace-level variable rather than an individual-level variable. Such a higher-level variable may influence the smoking behaviour of employees. In other words, the prevalence of smoking may be lower among employees at workplaces that have a non-smoking policy than it is among employees at workplaces that do not have such a policy, and this difference may be a *consequence* of the presence or absence of such a policy. Thus, a workplace-level variable may influence the behaviour of an individual employee, an example of a cross-level causal effect. There may also be cross-level interactions. For example, the effect of, say, age (an individual-level variable) on the smoking status of the individual may be moderated by characteristics of the workplace: a higher-level variable modifying the causal relationship between two lower-level variables.

Note that there are two kinds of workplace-level variables. The first are characteristics of the workplace that are not derived from the characteristics of the individuals who are employed there, for example, the presence or absence of a non-smoking policy. The second kind is derived by aggregating the characteristics of the individual employees. For example, smoking prevalence among employees at a worksite is a characteristic of the worksite and not of an individual employee, but it is obtained by combining the smoking status (1 = current smoker, 0 = current non-smoker) of all the employees at the worksite.

An individual's cognitions and behaviour may also be influenced by characteristics of the geographical area in which they live. For example, the number of parks and open spaces in an individual's neighbourhood may influence the frequency with which they walk for pleasure. Again, the number of parks and open spaces is a characteristic of a neighbourhood or other geographical area, not of the individual who lives in that area. Geographical areas may form a perfect set of nested levels. For example, in

the UK postcode system, sectors are nested within districts, which in turn are nested within areas. There are characteristics attached to each of these levels that may in principle influence the cognitions and behaviour of an individual resident. In general, in a multilevel system, variables at the lowest level (the individual level in this case) may be influenced by variables at a higher level, either directly (i.e., by bypassing intermediate levels) or indirectly (i.e., by influencing variables at intermediate levels). (Of course, cross-level influence may also flow from lower level to higher level, but we are making a simplifying assumption here that only downward influence occurs.)

Thus, the social factors that influence a particular health behaviour at an individual level can be thought of as being located in a complex system of higher levels, some of which may form perfect hierarchies but others of which do not. This framework is individual-level with respect to the dependent variable, that is, the health behaviour of interest, but it is multilevel with respect to the explanatory variables. (It is also possible to define dependent variables at higher levels than the individual. For example, why do some worksites have a non-smoking policy while others do not?)

CONCLUSIONS AND RECOMMENDATIONS

We conclude with a number of recommendations to guide future research in this area. Two important issues are how to manage complexity at both the theoretical and the empirical levels, given multiple theories, multiple causes, multiple behaviours, and multiple target populations, and how to ensure that research findings are cumulative.

First, there are too many theories of health behaviour. More rapid progress would be made in the field if research focused on a smaller number of theories. As mentioned earlier, for the sake of parsimony, general theories are preferable to health-specific or domain-specific theories, although it is acknowledged that a general theory may need to be modified when applied to a particular behaviour. Theories should also be clearly specified with clear definitions of constructs and clear specifications of the causal relationships between them. Health psychology researchers should avoid using theories that are not fully specified. The common practice of 'picking and mixing' components from several different theories (or what Bandura, 1998, calls 'cafeteria style research') should be discouraged. Any study that claims to test or extend a given theory should use a complete version of that theory and should try to make sure that each of its components is measured well. More studies that directly compare two or more different theories would be valuable (Weinstein, 1993); for two examples of empirical comparisons of theories, see Quine, Rutter and Arnold (1998) and Bish, Sutton and Golombok (2000). Some of these aims will be difficult to achieve in practice, though funding initiatives that require researchers to use one or two particular theories would be one possible mechanism.

Second, the field would benefit from greater standardization of measures. This is likely to be facilitated by the creation of a web resource that defines the major theoretical constructs employed in health behaviour research and lists common measures of these constructs, as planned by the US National Cancer Institute's 'Improving Theories' project (see http://cancercontrol.cancer.gov/brp/health_theory_index.html).

Third, more studies are needed that test social cognition models using within-individuals designs in which repeated measures of cognitions and behaviour with respect to the same target behaviour are obtained on a number of occasions. These would allow comparisons between causal effects estimated from within-individuals data and between-individuals data. Where these differ, the former are likely to provide a better estimate of the effects of changing cognitive variables through intervention.

Fourth, wherever possible, predictions from social cognition models should be tested using randomized experiments in which the explanatory variables are manipulated orthogonally and the data are analysed at the within-individual level as well as the between-individual level.

Finally, to properly investigate the effects of what we have called 'social' variables requires multilevel designs in which data are obtained from a sufficient number of units at a higher level than the individual (e.g., neighbourhoods, workplaces, schools) as well as from individuals within units. Ideally, such studies should include a complete implementation of a social cognition model such as the TPB. This would enable mediation analyses to be conducted to examine the extent to which the effects of higher-level variables on behaviour are mediated by cognitive variables. (In the case of the TPB, measurement burden could be reduced by initially using only direct measures of the constructs, that is, by omitting measures of beliefs. If the effects of higher-level variables on behaviour were found to be mediated by the TPB variables measured directly, then subsequent studies could conduct a more fine-grained analysis of mediation using indirect as well as direct measures.)

The data requirements for multilevel studies are formidable (Hox, 2002). For example, in a two-level design, it is recommended that data are obtained from at least 30 higher-level units. Thus such studies are likely to be larger and more expensive than most studies that are conducted in health psychology. In such research, priority should be given to investigating the effects of smaller, more proximal units that are likely to be more meaningful to the individuals who belong to the unit. For instance, the characteristics of a person's neighbourhood are likely to have larger effects on their health behaviour than the characteristics of the region in which they live. Similarly, characteristics of the individual's family or household are likely to have a greater influence than those associated with their network of acquaintances. Where it is difficult to obtain information about higher-level units, one shortcut is to assess social variables at the individual level. For example, instead of objectively measuring the local availability of open spaces where people can walk for pleasure, a study could assess the *perceived* availability of such open spaces. (Where both 'objective' and subjective measures were included, it would be possible to examine the extent to which the effects of objective measures were mediated by their corresponding subjective measures.) Another shortcut is to aggregate variables from the individual level to a higher level. Both the selection of levels (e.g., should we study neighbourhoods or districts?) and the selection of variables to be measured at these levels should be guided by theories that attempt to explain how these variables may influence individuals' cognitions and behaviour. The challenge is for health psychologists to develop and test such theories in collaboration with scientists from other disciplines.

ACKNOWLEDGEMENTS

This chapter is based partly on keynote addresses delivered to the European Health Psychology Society Annual Conference at the University of Leiden, The Netherlands, August 2000, and the 43rd Annual Congress of the German Psychological Association, Humboldt University, Berlin, Germany, September 2002.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Ajzen, I. (1988). Attitudes, personality, and behavior. Buckingham, UK: Open University Press.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.

Ajzen, I. (2002a). Constructing a TpB questionnaire: Conceptual and methodological considerations. Retrieved from http://www.people.umass.edu/aizen.

Ajzen, I. (2002b). The theory of planned behavior. Retrieved from http://www.people.umass.edu/aizen.

Ajzen, I., & Fishbein, M. (1977). Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888–918.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Ajzen, I., & Fishbein, M. (2000). Attitudes and the attitude–behavior relation: Reasoned and

automatic processes. *European Review of Social Psychology*, *11*, 1–33.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intention, and perceived behavioral control. *Journal of Experimental Social Psychology*, *22*, 453–474.

Albarracín, D., Johnson, B. T., Fishbein, M., & Muellerleile, P. A. (2001). Theories of reasoned action and planned behavior as models of condom use: A meta-analysis. *Psychological Bulletin*, *127*, 142–161.

Allison, P. D. (1990). Change scores as dependent variables in regression analysis. In C. Clogg (Ed.), *Sociological methodology 1990* (pp. 93–114). Oxford: Blackwell.

Armitage, C. J., & Conner, M. (2000). Social cognition models and health behaviour: A structured review. *Psychology and Health*, *15*, 173–189.

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, *40*, 471–499.

Bagozzi, R. P. (1984). Expectancy-value attitude models: An analysis of critical measurement issues. *International Journal of Research in Marketing*, *1*, 295–310.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology and Health*, *13*, 623–649.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bish, A., Sutton, S., & Golombok, S. (2000). Predicting uptake of a routine cervical smear test: A comparison of the health belief model and the theory of planned behaviour. *Psychology and Health*, *15*, 35–50.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: Guilford .

Carmody, T. P. (1997). Health-related behaviours: Common factors. In A. Baum, S. Newman, J. Weinman, R. West & C. McManus (Eds.), *Cambridge handbook of psychology, health and medicine* (pp. 117–121). Cambridge: Cambridge University Press.

Catania, J. A., Kegeles, S. M., & Coates, T. J. (1990). Towards an understanding of risk behavior: An AIDS risk reduction model (ARRM). *Health Education Quarterly*, *17*, 53–72.

Clogg, C. C., & Haritou, A. (1997). The regression method of causal inference and a dilemma confronting this method. In V. R. McKim & S. P. Turner (Eds.), *Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences* (pp. 83–112). Notre Dame, IN: University of Notre Dame Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of Applied Social Psychology*, *28*, 1429–1464.

Conner, M., & Norman, P. (Eds.) (1996). *Predicting health behaviour: Research and practice with social cognition models*. Buckingham, UK: Open University Press.

Conner, M., & Norman, P. (Eds.) (in press). *Predicting health behaviour: Research and practice with social cognition models* (2nd edn.). Buckingham, UK: Open University Press.

Contrada, R. J., & Goyal, T. M. (2004). Individual differences, health, and illness: The role of emotional traits and generalized expectancies. In S. Sutton, A. Baum & M. Johnston (Eds.), *The Sage handbook of health psychology*. London: Sage.

Courneya, K. S. (1994). Predicting repeated behavior from intention: The issue of scale correspondence. *Journal of Applied Social Psychology*, *24*, 580–594.

Courneya, K. S., & McAuley, E. (1993). Predicting physical activity from intention: Conceptual and methodological issues. *Journal of Sport and Exercise Psychology*, *15*, 50–62.

Duncan, C., Jones, K., & Moon, G. (1998). Context, composition and heterogeneity: Using multilevel models in health research. *Social Science and Medicine*, *46*, 97–117.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380–417.

Emmons, K. M. (2000). Health behaviors in a social context. In L. F. Berkman & I. Kawachi (Eds.), *Social epidemiology* (pp. 242–266). New York: Oxford University Press.

Evans, M. G. (1991). The problem of analyzing multiplicative composites: Interactions revisited. *American Psychologist*, *46*, 6–15.

Finkel, S. E. (1995). *Causal analysis with panel data*. Thousand Oaks, CA: Sage.

Fishbein, M. (2000). The role of theory in HIV prevention. *AIDS Care*, *12*, 273–278.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fishbein, M., Triandis, H. C., Kanfer, F. H., Becker, M., Middlestadt, S. E., & Eichler, A. (2001). Factors influencing behavior and behavior change. In A. Baum, T. A. Revenson & J. E. Singer (Eds.), *Handbook of health psychology* (pp. 3–17). Mahwah, NJ: Erlbaum.

French, D. P., & Hankins, M. (2003). The expectancy-value muddle in the theory of planned behaviour – and some proposed solutions. *British Journal of Health Psychology*, *8*, 37–55.

Godin, G., & Kok, G. (1996). The theory of planned behavior: A review of its applications to health-related behaviors. *American Journal of Health Promotion*, *11*, 87–98.

Green, L. W., Richard, L., & Potvin, L. (1996). Ecological foundations of health promotion. *American Journal of Health Promotion*, *10*, 270–281.

Hagger, M. S., Chatzisarantis, N. L. D., & Biddle, S. J. H. (2002). A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: Predictive validity and the contribution of additional variables. *Journal of Sport and Exercise Psychology*, *24*, 3–32.

Hardeman, W., Sutton, S., Griffin, S., Johnston, M., White, A., Wareham, N. J., & Kinmonth, A. L. Use of causal models in the development of theory-based behaviour change programmes for trial evaluation: Integrating psychology and epidemiology. Paper submitted for publication.

Hedeker, D., Flay, B. R., & Petraitis, J. (1996). Estimating individual influences of behavioral intentions: An application of random-effects modeling to the theory of reasoned action. *Journal of Consulting and Clinical Psychology*, *64*, 109–120.

Heise, D. R. (1975). *Causal analysis*. New York: Wiley.

Hertzog, C., & Nesselroade, J. R. (1987). Beyond autoregressive models: Some implications of the trait–state distinction for the structural modeling of developmental change. *Child Development*, *58*, 93–109.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Erlbaum.

Johnston, J. (1984). *Econometric methods* (2nd edn.). New York: McGraw-Hill.

Johnston, M., French, D. P., Bonetti, D., & Johnston, D. W. (2004). Assessment and measurement in health psychology. In S. Sutton, A. Baum & M. Johnston (Eds.), *The Sage handbook of health psychology*. London: Sage.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, 4th edn., pp. 233–265). New York: McGraw-Hill.

Liker, J. K., Augustyniak, S., & Duncan, G. J. (1985). Panel data and models of change: A comparison of first difference and conventional two-wave models. *Social Science Research*, *14*, 80–101.

McLeroy, K. R., Bibeau, D., Steckler, A., & Glanz, K. (1988). An ecological perspective on health promotion programs. *Health Education Quarterly*, *15*, 351–377.

Ogden, J. (2003). Some problems with social cognition models: A pragmatic and conceptual analysis. *Health Psychology*, *22*, 424–428.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge: Cambridge University Press.

Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American Journal of Health Promotion*, *12*, 38–48.

Quine, L., Rutter, D. R., & Arnold, A. (1998). Predicting and understanding safety helmet use among schoolboy cyclists: A comparison of the theory of planned behaviour and the health belief model. *Psychology and Health*, *13*, 251–269.

Rasbash, J., & Browne, W. (2001). Modelling non-hierarchical structures. In A. H. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 93–105). Chichester, UK: Wiley.

Rogers, R. W., & Prentice-Dunn, S. (1997). Protection motivation theory. In D. Gochman (Ed.), *Handbook of health behavior research: Vol.1. Determinants of health behavior: Personal and social* (pp. 113–132). New York: Plenum.

Rogosa, D. R. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*, 245–258.

Rosenthal, R., & Rubin, D. B. (1979). A note on percent variance explained. *Journal of Applied Social Psychology*, *9*, 395–396.

Rothman, A. J. (2000). Toward a theory-based analysis of behavioral maintenance. *Health Psychology*, *19*, 64–69.

Røysamb, E., Rise, J., & Kraft, P. (1997). On the structure and dimensionality of health-related behaviour in adolescents. *Psychology and Health*, *12*, 437–452.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and

questionnaire construction. *American Journal of Evaluation*, *22*, 127–160.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Sheeran, P. (2002). Intention–behavior relations: A conceptual and empirical review. *European Review of Social Psychology*, *12*, 1–36.

Sheeran, P., & Taylor, S. (1999). Predicting intentions to use condoms: A meta-analysis and comparison of the theories of reasoned action and planned behavior. *Journal of Applied Social Psychology*, *29*, 1624–1675.

Stokols, D. (1992). Establishing and maintaining healthy environments: Toward a social ecology of health promotion. *American Psychologist*, *47*, 6–22.

Stokols, D. (1996). Translating social ecological theory into guidelines for community health promotion. *American Journal of Health Promotion*, *10*, 282–298.

Stone, A. A., Turkkan, J. S., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (Eds.) (1999). *The science of self-report: Implications for research and practice*. Mahwah, NJ: Erlbaum.

Stoolmiller, M., & Bank, L. (1995). Autoregressive effects in structural equation models: We see some problems. In J. M. Gottman (Ed.), *The analysis of change* (pp. 261–276). Mahwah, NJ: Erlbaum.

Strecher, V. J., & Rosenstock, I. M. (1997). The health belief model. In A. Baum, S. Newman, J. Weinman, R. West & C. McManus (Eds.), *Cambridge handbook of psychology, health and medicine* (pp. 113–117). Cambridge: Cambridge University Press.

Stroebe, W. (2000). *Social psychology and health* (2nd edn.). Buckingham, UK: Open University Press.

Sutton, S. (1994). The past predicts the future: Interpreting behaviour–behaviour relationships in social psychological models of health behaviour. In D. R. Rutter & L. Quine (Eds.), *Social psychology and health: European perspectives* (pp. 71–88). Aldershot: Avebury.

Sutton, S. (1998). Predicting and explaining intentions and behaviour: How well are we doing? *Journal of Applied Social Psychology*, *28*, 1317–1338.

Sutton, S. (2000). Interpreting cross-sectional data on stages of change. *Psychology and Health*, *15*, 163–171.

Sutton, S. (2001). Back to the drawing board? A review of applications of the transtheoretical model to substance use. *Addiction*, *96*, 175–186.

Sutton, S. (2002a). Testing attitude–behaviour theories using non-experimental data: An examination of some hidden assumptions. *European Review of Social Psychology*, *13*, 293–323.

Sutton, S. (2002b). Psychosocial theories of health behavior. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences*. Oxford: Elsevier.

Sutton, S. (2002c). Using social cognition models to develop health behaviour interventions: Problems and assumptions. In D. Rutter & L. Quine (Eds.), *Changing health behaviour: Intervention and research with social cognition models* (pp. 193–208). Buckingham, UK: Open University Press.

Sutton, S. (2003). Using theories of behaviour change to develop and evaluate sexual health interventions. In J. Stephenson, J. Imrie & C. Bonell (Eds.), *Effective sexual health interventions: Issues in experimental evaluation* (pp. 51–66). Oxford: Oxford University Press.

Sutton, S. (in press). Stage theories of health behaviour. In M. Conner & P. Norman (Eds.), *Predicting health behaviour: Research and practice with social cognition models* (2nd edn.). Buckingham, UK: Open University Press.

Sutton, S., French, D. P., Hennings, S. J., Mitchell, J., Wareham, N. J., Griffin, S., Hardeman, W., & Kinmonth, A. L. (2003). Eliciting salient beliefs in research on the theory of planned behaviour: The effect of question wording. *Current Psychology*, *22*, 234–251.

Sutton, S., McVey, D., & Glanz, A. (1999). A comparative test of the theory of reasoned action and the theory of planned behavior in the prediction of condom use intentions in a national sample of English young people. *Health Psychology*, *18*, 72–81.

Trafimow, D., & Finlay, K. A. (1996). The importance of subjective norm for a minority of people: Between-subjects and within-subjects analyses. *Personality and Social Psychology Bulletin*, *22*, 820–828.

Trafimow, D., Sheeran, P., Conner, M., & Finlay, K. A. (2002). Evidence that perceived behavioural control is a multidimensional construct: Perceived control and perceived difficulty. *British Journal of Social Psychology*, *41*, 101–121.

Trafimow, D., Sheeran, P., Lombardo, B., Finlay, K. A., Brown, J., & Armitage, C. J. (in press). Affective and

cognitive control of persons and behaviors. *British Journal of Social Psychology.*

Weinstein, N. D. (1993). Testing four competing theories of health protective behavior. *Health Psychology*, *12*, 324–333.

Weinstein, N. D., Rothman, A. J., & Nicolich, M. (1998). Use of correlational data to examine the effects of risk perceptions on precautionary behavior. *Psychology and Health*, *13*, 479–501.

Weinstein, N. D., Rothman, A. J., & Sutton, S. R. (1998). Stage theories of health behavior: Conceptual and methodological issues. *Health Psychology*, *17*, 290–299.

Weinstein, N. D., & Sandman, P. M. (1992). A model of the precaution adoption process: Evidence from home radon testing. *Health Psychology*, *11*, 170–180.