# Latent Trait and Latent Class Models in Survey Analysis: Case Studies in Public Perceptions of Biotechnology

**Sally Rebecca Stares**

**London School of Economics & Political Science**
**University of London**

**Thesis submitted for the Degree of Doctor of Philosophy**

I confirm that the work presented in this thesis is my own.

Sally R. Stares

Sally Stares

# Abstract

In latent variable models the existence of one or more unobserved (latent) variables is posited to explain the associations between a set of observed (manifest) variables. These models are useful for analysing attitudinal survey data, where multiple items are used to capture complex constructs such as attitudes, which cannot be directly observed. In such research they are most commonly applied in the form of factor analyses based on linear regression models. However, these are inappropriate when observed items are categorical, which is often the case with attitudinal surveys. Latent trait and latent class models, based on logistic models, are then more suitable. In this thesis I demonstrate how they can be employed to address common challenges in attitudinal survey research.

The case study data illustrating these challenges are from the Eurobarometer survey on public perceptions of biotechnology, fielded in 2002 in fifteen European countries. Using these data I investigate the viability of cross-nationally comparable measures of three central constructs in studies of public perceptions of biotechnology: attitudes towards applications of biotechnology, knowledge of biology and genetics, and engagement with science and with biotechnology. The analyses aim to capture these complex constructs, taking account of 'don't know' responses by including them as categories of nominal observed items, and exploring the comparability of measures of these constructs cross-nationally by assessing the similarity of measurement models between countries.

The results of these analyses are informative in three ways: substantively, adding to our knowledge of people's representations of biotechnology; methodologically, increasing our understanding of how the survey items function; and practically, informing future questionnaire design. I also formulate a taxonomy of issues and choices in attitudinal survey research as a conceptual framework through which to discuss more broadly the potential value of latent trait and latent class models in survey research in social psychology.

## Acknowledgements

I have been very fortunate to carry out research for this thesis funded by a full studentship from the UK Economic and Social Research Council. I am grateful to my supervisors and to Professor Colin Mills and Dr Matthew Mulford for supporting my application for this award.

I would like to express my sincere thanks to my supervisors, Professor George Gaskell and Dr Jouni Kuha. Professor Gaskell has been a great source of inspiration and guidance throughout my studies. Dr Kuha has been a very patient statistics supervisor to a non-statistician, and has kindly written several functions in S-PLUS to deliver particular pieces of output, including model fit statistics, which have played a very important role in my empirical analyses.

I am extremely grateful, too, for the comments, advice, and support I have received from senior academics, colleagues and friends: Dr Kavita Abraham, Dr Nick Allum, Dr Martin Bauer, Tom Chivers, Dr Stella Creasy, Katrin Hohl, Dr Susan Howard, Dr Robbie Hudson, Professor Patrick Humphries, Dr Jonathan Jackson, Dr Nicole Kronberger, Dr Niels Mejlgaard, Professor Mark Reiser, Dr Anne Katrin Schlag, and Dr Ricky Wong. Finally I would like to thank my family, especially my parents, for their endless support.

# Table of Contents

# List of Tables

# List of Figures

# 1 Introduction

In this thesis I investigate the viability of cross-nationally comparable measures of three constructs of central importance in studies of public perceptions of biotechnology: attitudes towards applications of biotechnology, knowledge of biology and genetics, and engagement with science and with biotechnology. These empirical studies are a vehicle through which to demonstrate and comment on the use of latent trait and latent class statistical models in analyses of social surveys, specifically those concerning attitudes and opinions. The case study topic, public perceptions of biotechnology, is an area of research that could profit greatly from the insights afforded by these models: substantively, adding to our knowledge of people's representations of biotechnology; methodologically, increasing our understanding of how survey items function; and practically, informing the design of future questionnaires.

Latent trait and class models are types of latent variable models. These models have been highly developed by statisticians, but are often not used perspicaciously by social psychologists analysing attitudinal social surveys. Typically indeed, social psychologists try to capture attitudes and similar constructs in ways that are only weakly informed by a rich heritage of literature on methodologies for attitude measurement. There are a number of possible reasons for this. The key theoretical reason is that many social psychologists regard the classic methodologies as unattractive, since they are based on an individualistic and reductionist conception of attitudes. Latent variable models do not imply such a theoretical position – neither inherently nor historically. Indeed, as freestanding statistical tools they have no necessary relationship with any single theoretical conception of the attitude. But they can be interpreted in a way which moves us a little closer to the 'social' than was allowed by the classic attitude measurement techniques.

Latent variable models provide a way of analysing a set of survey items in combination, and summarising the relationships between them. The rationale for doing this in attitudinal survey research begins from the established fact that responses to survey questions on attitudes, knowledge and other psychological constructs do not *only* return information about those constructs. They also contain information about idiosyncrasies of question wording, response options offered, and a range of other contextual influences on the respondent (e.g. Schuman & Presser, 1981). These influences can distort single-item indicators of attitudes and the like. In cases where respondents have not considered the topic very deeply before being questioned and thus make a judgement on the spot, this distortion is likely to be exacerbated (Converse, 1964). Besides these technical problems, it is often the case that the constructs of interest are themselves too complex to be feasibly captured by single survey questions. For both of these reasons there is a wide consensus among researchers of attitudes (e.g. Eysenck, 1954; McKennell, 1977) that when trying to learn what people think about an issue, more reliable and valid results are obtained by deriving an answer jointly from several questions than from taking a single item as fully and straightforwardly informative.

Latent variable models begin with a set of items designed to capture a construct of interest – for example, attitude towards genetically modified (GM) food. Finding that responses to these items are associated with each other, it is assumed that the associations are a function of some underlying, general variable characterising perceptions of GM food. This general variable cannot be observed directly: its existence is inferred from the associations found between the observed responses. It is therefore a hypothesised, *latent* variable, presumed to lie beneath the observed or manifest survey responses. This latent variable may be conceived of as continuous or discrete; if discrete, its categories may be ordered or unordered. Discrete latent variables are described in terms of their categories or 'classes'. Continuous latent variables are termed 'traits' or 'factors', depending on the conventional terminology of the particular model used: essentially, in factor analysis the observed variables are treated as continuous, and in latent trait analysis they are treated as categorical. Returning to the example then, the underlying variable can be thought of as defining a space which characterises perceptions of GM food. That space can be described either in terms of a number of mutually exclusive ideal types of attitudes (classes), or in terms of one or more dimensions of attitudes (traits or factors), which may be correlated with or independent of each other.

A latent variable model contains two parts. One part is the 'measurement' model, which describes the relationships between the latent variable(s) and the observed items. The other part is referred to in this thesis as the 'structural' model; this describes the distribution(s) of the latent variable(s), and if there is more than one latent variable, the relationship(s) between them. The measurement model is used to describe the content of the latent space. Items strongly associated with a latent variable are central to its definition, while those that are weakly associated are more peripheral. The structural model is used to describe the shape of the latent space. For example, in a latent class model we would consider how many general types of attitude there are and what proportions of people could be said to adopt each one. In a latent trait model we consider whether attitudes are dispersed along a single continuum (say, positive to negative) or whether the latent space is multidimensional, and with what frequency people are found at various locations in the latent space.

The depiction of such a space returned by a latent variable model of cross-sectional survey data is the product of between-individual, rather than within-individual, analyses. It does not describe a single representation of a concept, shared identically in the minds of all respondents. Instead it is something closer to a collective representation – that is, a representation resulting from the combined input data from all respondents. Albeit those of a strong social constructionist persuasion object to survey methods for other reasons, latent variable models may be applied in a less individualistic and more social spirit than is generally supposed. On theoretical grounds they might, then, be more popular with social psychologists than they currently are.

The models are also of considerable use to those who are less concerned with the theoretical side of social psychology and more with substantive research questions. They enable us to take account of some common features of attitudinal survey data which are not addressed in a very satisfactory way in standard survey analyses. The case study data used in this report, on public perceptions of biotechnology, illustrate these challenges well. In fact, the empirical studies are intended to be of interest not only to those working in the substantive research area of the Public Understanding of Science (PUS), but also, as examples, to those analysing other kinds of attitudinal survey data, especially cross-national data sets.

The case study data are from the Eurobarometer survey 58.0 on perceptions of biotechnology, administered in fifteen European Union countries in 2002, with random samples of around 1,000 respondents in each country (Gaskell, Allum, & Stares, 2003). Eurobarometer surveys on biotechnology have been fielded at three-yearly intervals since 1991, and the data from them have been widely used and quoted by many actors who have an involvement or interest in the PUS research field. PUS researchers regularly combine information from sets of survey items to derive measures of constructs, both in order to report their distributions and to use them in analyses with other variables. The empirical chapters of this thesis are studies of three such constructs which are of central theoretical importance in PUS: types of positive and negative attitudes towards GM food and towards therapeutic cloning; levels of knowledge about genetics and biology; and types and levels of engagement with science and technology and with biotechnology.

The studies address two key features of the data which to date have received relatively little attention when analysed. The first of these key features is that response options to items in the Eurobarometer on biotechnology comprise a small number of discrete categories: either dichotomous items or polytomous Likert-type response scales (e.g. 'strongly agree' to 'strongly disagree') with up to five response options. Factor analysis is statistically unsuitable for these data, since the normal linear factor model specifies a linear regression model for the relationship between the latent variable(s) and observed items. Latent trait and latent class models, however, specify a *logistic* rather than a linear model (binary or multinomial for binary or polytomous observed items, respectively), making them well suited to the task at hand. They are particularly useful when a high number of 'don't know' (DK) responses are returned, as is the case with the Eurobarometer on biotechnology, and other surveys of unfamiliar topics. To use factor analysis in such a scenario typically means either ignoring these responses altogether, or recoding them as a middle category, which involves a somewhat heroic leap of interpretation. Latent trait and class models represent an improvement over factor analysis for analysing such data simply because they provide a way of modelling observed items as nominal: that is, with unordered categories. As such, there is no need to ignore DK responses in latent trait and class models. Rather than recode DK responses as middle response options, the analyses can be used specifically to discover whether, statistically speaking, they do in fact function as middle categories.

Latent trait models have already been employed in PUS research, specifically, by Jon Miller and colleagues for building scales of knowledge about science and biotechnology (e.g. Jon D. Miller, 1998; Jon. D. Miller & Kimmel, 2001; Jon D. Miller & Pardo, 2000). However, there remains an opportunity to build on their analyses in a number of ways, most notably by addressing directly the second important feature of the Eurobarometer data: its cross-national coverage. This second feature is much more evidently interesting and difficult than the first, from the perspective of a lay audience. It goes without saying that questionnaire items administered in several different languages and cultural contexts may carry varying meanings for respondents at these varying vantage points. Sensitive cross-national comparisons should therefore attend as carefully as possible to the question of whether they are comparisons of like with like. This must realistically be a question of *to what extent* rather than *whether*, since equally *within* languages and cultural settings, people bring their own frames of reference to bear on the surveys in which they participate. It is an empirical question at what point varying interpretations of questions become so diverse as to make comparisons meaningless and misleading.

This empirical question demands both qualitative and quantitative responses, and latent variable models can make a useful contribution to a quantitative response. Returning to our example, imagine that we pooled the data from fifteen European countries and, running a latent trait model, found a single dimension depicting attitudes towards GM food, which, using information from the measurement model, we interpreted as running from general positive attitudes at one end of the continuum to negative sentiments at the other end. The question then arises whether the same interpretation of the latent space is found if the models are run separately within each country – for us this means, is the measurement model the same for every country? If it is, then statistically speaking we are dealing with approximately the same attitude continuum for every country, and we can plausibly make comparisons of the distributions of positive and negative attitudes between them. In technical terms, this means we fix the measurement model to be the same for each country, and allow the structural model to differ between them. If the measurement models are very different between countries, then the interpretation of the latent variable is likewise different, and this should serve as a warning against making simple comparisons between countries – comparisons of *unlike* attitude traits. Although this is a straightforward analysis, it has not been used in any published studies of PUS data, to my knowledge. Of course other analyses and critiques, both quantitative and qualitative, have been published on the subject of comparative analyses in PUS, but none producing the particular insights that can be gained with latent variable models.

The empirical studies of three constructs each address these two features of the data: the categorical (mostly nominal) nature of the data, and the problem of cross-national comparisons. The analyses are used to glean as much information as possible about the ways in which the items fit together – or do not fit together – to represent the constructs we wish to model. The outcomes of the studies can be divided into three types: substantive, methodological and practical. The substantive findings should be of interest to the PUS community generally. The methodological insights into item functioning may be of interest to PUS survey analysts, and more broadly to others analysing attitudinal survey data. Finally, the practical implications are directed towards those designing future Eurobarometer surveys on biotechnology, and potentially to others fielding similar surveys on this topic. These implications concern the successes or difficulties found in using the items analysed in the studies to represent the constructs of interest.

Chapter 2 provides an outline of the substantive context for the case studies: the PUS research field and the task of gauging public perceptions of biotechnology using survey data. It explains the importance of obtaining accurate information on public opinion of this controversial technology, and describes three key challenges involved in analysing survey data on this topic: how to capture complex constructs, how to analyse 'don't know' (DK) responses, and how to make valid cross-national comparisons. The main objections to the use of surveys for PUS research are also outlined. This chapter also introduces the three constructs on which the empirical studies are based: that is, positive and negative attitudes towards two applications of biotechnology, knowledge of biology and genetics, and engagement with science and with biotechnology. For each, a brief account is given of their significance for PUS, the ways in which they are usually measured using survey data, and the particular angle to be adopted in the corresponding empirical chapter.

Chapters 3 and 4 set out the theoretical tools for the empirical studies. Chapter 3 gives a brief sketch of the historical background of approaches to the concept of the attitude and its measurement, and presents a scheme or taxonomy of themes arising in attitudinal research, based on a review of relevant literature regarding measurement in social psychology. This forms a conceptual framework to aid the discussion of the latent variable models. In Chapter 4, latent trait and latent class models are introduced, with a formal statistical specification followed by a more conceptual discussion, using the taxonomy of themes from Chapter 3.

The three empirical studies are presented in Chapters 5, 6 and 7. Chapter 5 focuses on models of positive and negative attitudes for two applications of biotechnology: GM food and therapeutic cloning. These biotechnologies, one agri-food and one medical, tend to elicit systematically different reactions from the public, so an informal comparison of models of attitudes towards them is of particular interest. The task for the chapter is to use latent class models to characterise types of support and opposition for these biotechnologies. Chapter 6 takes knowledge of biology and genetics as its theme, using latent trait models to investigate the properties of a scale of knowledge, with a particular interest in the insights gained by including DK responses as separate response categories. Chapter 7 investigates engagement, both with science and technology in general, and with biotechnology in particular. This construct is a newer topic of interest in PUS research and as such there is no clear steer from the literature on the best approach to modelling it, therefore both trait and class models are explored in this study.

The empirical studies thus comprise one study focusing on classes, one on traits, and one informally comparing the two. Each begins with a detailed examination of the British sample, using a selection of variants on the model in question to gain as much information as possible about patterns of association and item functioning in the data. The second half of each study extends the analyses to the full fifteen country samples, comparing measurement models between countries, and distributions of latent variables where appropriate. The final goal of each chapter is to find a well fitting representation of the construct being explored, for which the measurement model can be constrained sufficiently between countries to permit meaningful cross-national comparisons. Building on these models, Chapter 8 presents four simple loglinear analyses to explore the relationships between these measures and basic socio-demographic variables.

Chapter 9 reviews the empirical findings from the analyses, presenting them thematically in terms of substantive and methodological results, and practical implications for the design of future Eurobarometer surveys. It then offers some comments on the models used, including the performance of fit statistics and their use in model selection, and ways in which the analyses may be taken forward in future studies. The chapter concludes with a reflection on the potential value of these models in attitudinal survey research. The challenges involved in analysing these data are not specific to the Eurobarometer, nor to perceptions of biotechnology as a topic. Indeed the empirical studies are intended as examples that illustrate the use of latent trait and class models for attitudinal survey research more broadly: they provide us with useful diagnostic and substantive information about sets of survey items, and reconnect us with a rigorous approach to attitude measurement that has become uncommon in contemporary survey research in social psychology.

# 2 The substantive context of public perceptions of biotechnology

This chapter sketches the substantive scene in which the case study analyses of the thesis are set. It begins by outlining why public perceptions of biotechnology are of interest at all and to whom, and continues with a brief description of the academic research area, Public Understanding of Science (PUS), in which studies of them are located. PUS is broad in scope, involving actors in a range of academic and non-academic fields, with a range of interests. But its core substance can be characterised in terms of a small number of key themes, which are assigned different emphases over time and between schools of thought within the field. Running across these varying interests is a continuing concern with surveys as research tools: their potential contributions to knowledge, and the drawbacks associated with their use. Such features are not peculiar to the substantive topic; they apply more generally to many applications of social survey research.

The three particular challenges for PUS surveys, introduced in Chapter 1, are elucidated further in Section 2.3: the task of capturing complex constructs, how to treat 'don't know' responses in analyses when they are returned in high numbers, and how to conduct sensitive analyses of cross-national data. These are recurring themes in the empirical studies in this thesis, which are themselves framed around three constructs of key interest to PUS researchers: attitudes towards applications of biotechnology, knowledge of biology and genetics, and engagement both with science in general and with biotechnology in particular. Section 2.4 introduces these concepts, explaining how they have been operationalised in surveys to date, and how each empirical chapter will build on existing approaches to their measurement.

Before embarking on these studies it is apposite to address the main arguments opposing the use of surveys in PUS. In addition to commonly held and general reservations about the shortcomings of surveys as research tools, some vociferous anti-survey stances are to be found within the field. Some PUS researchers would abandon them altogether – and with them, the research questions which cannot feasibly be addressed with any other research method: that is, questions concerning the distributions of attitudes, beliefs, and other key psychological constructs, within and across populations. I take issue with the absolutism of this anti-survey school of thought in PUS, although I would not portray surveys as a panacea – far from it. Surveys are not a suitable method for every research question. Where they are appropriate, they can always be developed and improved. There are notable calls in the published literature for more critical methodological attention to be given to the measurement of constructs and concepts for PUS with survey data. This thesis is a response to such calls.

## 2.1 The significance of biotechnology for the public, and of the public for biotechnology

The most basic reason for studying public perceptions of biotechnology is that biotechnology has the potential to transform our lives, even the nature of life itself – while in its turn, the public potentially has the power to determine whether it will do so. The use of the singular term 'biotechnology' belies the diversity of its application, however. In agriculture and food production, the so-called 'green' biotechnologies, it includes genetic modification of crops for a range of purposes, such as increasing their resistance to pesticides, producing their own insecticides, improving their taste or increasing their shelf-life. Industrial, 'white' applications of biotechnology use living cells and enzymes for a range of purposes, replacing traditional chemical processes that were less efficient or that produced more waste. A prime example of white biotechnology is the use of plants as an alternative to conventional sources for the manufacture of plastics and fuels. Medical, 'red' applications include genetic interventions to fight diseases, cloning human organs for transplant, a range of pharmaceutical applications, including the production of medicines and the study of genetic influences on responses to them (known as pharmacogenetics), and a host of uses, both medical and non-medical, for human genetic data. It should be noted, for clarification, that in speaking of biotechnology in this thesis, I refer to what is sometimes termed 'modern biotechnology', that is, technologies involving modifications at the level of the gene, dating back to the discovery of recombinant DNA in the early 1970s. A few sources still use the term to refer also to any technology making use of biological systems in general (see Torgersen et al., 2002 for a brief outline of terminology), which would make for quite a different discussion.

Just as biotechnologies can be classified by area of application, they can be grouped, informally, by the levels of controversy surrounding them. On one hand are a range of universally endorsed applications, such as using biotechnology to produce insulin to treat diabetes, or for purposes of environmental repair or bioremediation – for example, to clean up oil slicks. On the other hand are a cluster of biotechnologies which spark considerable media attention and public debate. For example, medical applications involving the use of stem cells, or any described in terms of 'cloning' tend to be contentious, and agri-food applications have had a high political profile over the last decade. Indeed, levels of controversy vary not only between applications but also over time and between interested parties. In the early 1970s, when Cohen and Boyer's development of recombinant DNA technology realised the potential of 'genetic engineering', as it was known then, debates about the safety and ethics of such a technology took place mainly amongst scientists and interested professionals. It was not government officials, nor civil society, nor the public, but *scientists* who in 1974 called for a voluntary moratorium on the use of certain recombinant DNA experiments, pending an international conference to consider the implications and regulation of this new powerful technology (Berg et al., 1974). Although the proceedings of this conference, held at Asilomar 1975, were reported in the press, biotechnology retained a relatively low public profile for the next few years.

It was only in the late 1990s that public controversy burgeoned (Bauer & Gaskell, 2002). Two significant events served to ignite the debate. In 1996 a shipment of soya from the US arrived in Europe – a cargo containing, for the first time, a fraction of genetically modified (GM) soya crops. Other GM products had recently been enjoying passable sales – for example, Zeneca's tomato puree, clearly labelled with its GM content, had been successfully sold in the UK. But the absence of labelling of the GM content in this soya catalysed an outburst of debate about the safety and desirability of GM products, and about the trustworthiness of food producers and regulators. The following year, scientists at the Roslin Institute in Edinburgh announced the birth of Dolly the sheep, cloned from an adult sheep cell, fuelling outcry over the potentially fantastical implications of human reproductive cloning.

These events marked a watershed in the development and reception of biotechnology among the public, especially in Europe (Grabner, Hampel, Lindsey, & Torgersen, 2001). Partly as a result of public protests over the soya imports, the European Union

(EU) adopted an unofficial moratorium on the approval of GM crops between 1998 and 2004, in spite of its political consequences. The moratorium was ruled illegal by the World Trade Organisation in 2006, in response to a complaint from the US, Argentina and Canada, and a host of third party countries exporting GM crops. But Europeans have not been alone in their reservations about GM crops and food. Osgood (2001) charts the global scale of public and civil society reactions to the impacts of plant biotechnologies, across different economic and geographical contexts. In high income countries, salient objections centre around the perception of GM crops as artificial products, with the potential to spread into and contaminate their unspoilt, natural environs; the widely used term 'Frankenfoods' is redolent of the freakish, unanticipated consequences of human ambition run wild. In developing countries, salient objections are grounded in the immediate consequences of GM technologies on the success and development of agriculture, which itself is more immediately a matter of survival for many people. Multinational firms such as Monsanto promote the potential of GM technology for overcoming natural hazards to crop success. But their corporate concerns for protecting their investments are a cause for consternation. Under widespread pressure, Monsanto very publicly declared a halt on developing proposed 'terminator gene' technology for preventing crops from reseeding (Shapiro, 1999), which would have the dual effect of protecting their investments, and perpetuating a relationship of dependence between farmers and seed producers.

Different concerns and emphases are present in debates surrounding medical applications of biotechnology. For example, 'therapeutic cloning', which will be addressed in the empirical chapter on attitudes in this thesis, involves using cloned embryonic stem cells as host cells from which tissues and organs can be grown for transplant. The principal controversy over the use of embryonic stem cells concerns the moral implications of destroying embryos in this process. Debates about this technology therefore contain a strong religious element, often mirroring arguments surrounding abortion. Correspondingly, the regulation of stem cell research varies cross-nationally, in a way that echoes the religious or cultural climate in the country. For example, in the UK, scientists may draw on stocks of unused embryos created in the course of in-vitro fertilisation processes, for use in stem cell research. In the US, by contrast, where objections based on religious principles are voiced loudly, sources of embryonic stem cells for federally funded research are strictly limited to a number of stem cell lines created before 2001. Internationally, regulation remains a site of difference and disagreement between governments. For example, the UN Declaration on Human Cloning (UN, 2005) is strongly supported by the US administration, but is rejected by UK on the grounds of its absolutism and concomitant prevention of progress in the field of medical biotechnology.

Perceptions of these technologies vary, then, between applications and between cultural contexts. In Europe, public opinion is generally held to be negative towards GM crops and food, but positive towards therapeutic cloning; in the US, the opposite pattern is found (Gaskell, Thompson, & Allum, 2002; National Science Board, 2004). In both applications of biotechnology, governments draw on public opinion data in the course of policy formation – be it in the context of trade disputes, international treaties or the development of infrastructure for scientific research and for the science industry. It is not only politicians who make reference to public opinion: scientists in industry and academia; academics in other disciplines such as social scientists and political scientists; civil society and interest groups; and the mass media, all quote survey data on public perceptions to defend their positions and illustrate their arguments. The general influence of public opinion on the development and use of biotechnology is now widely recognised (Pühler, 2002), and the study of public perceptions of biotechnology is becoming an increasingly important exercise.

## 2.2 The Public Understanding of Science (PUS) research field

Studies of public perceptions of biotechnology fit within a much broader area of research commonly known as the Public Understanding of Science (PUS). It is quite a task to delineate the boundaries of this field, even at the most basic level – not least because there is no absolute consensus on the definitions of the terms 'public', 'understanding' and 'science' themselves (Gregory & Miller, 1998). Debates surrounding the meanings of these terms are complex, and it is not of great relevance for this thesis to rehearse them here, other than to qualify that for the most part, the acronym PUS is recognised as referring to science and technology as a broad category. Although some would stress that this thesis concerns PUST studies, with 'Technology' added at the end, PUS remains the more widely used short-hand term, and as such I will use the latter.

To give a working definition of PUS – blunt, but adequate for the purposes of this report – I would say that PUS research concerns the representations of science and technology created and held by the public, along with their knowledge of science and technology and their engagement with it, and the range of values they bring to bear in understanding and reacting to it, such as ethical and moral concerns, political convictions and religious values. Studies of the public's perceptions of science and technology often focus on the relationships between particular perceptions and particular socio-demographic characteristics, such as age, sex, level of education, and a number of group-level social identities – perhaps foremost, nationality. PUS is often concerned with the roles of *actors* in the field of science and technology; notably the academic scientific community, industrial science, politicians, interest groups and the mass media. The dynamics between these science actors, and between these and the public, form the framework for PUS activities. Indeed 'activities' may be a more apt word than 'research' to describe a field whose normative components are characterised by some as a 'movement' (e.g. Bauer, 2003).

PUS as a movement for the popularisation of science has a very long history (Bauer, 2003). The modern period of PUS is commonly thought of as spanning the last fifty years, approximately. For example, Miller (1992) identifies 1957 as the year marking the beginnings of significant work on PUS in the USA, in the form of the US National Association of Science Writers' survey of attitudes towards science and technology (Withey, 1959). In Britain, PUS activity is said to have begun in earnest with the Royal Society's 1985 report (Bodmer, 1985), and the subsequent formation of the (now disbanded) multi-agency Committee on the Public Understanding of Science, from members from the Royal Society, Royal Institution, and British Association for the Advancement of Science. In addition to these and many other national research programmes, cross-national studies have played a prominent role in the field, particularly those employing survey research. The European Commission sponsored Eurobarometer surveys began to address topics of science and technology as early as 1977.

The field of PUS is highly politically charged, in two senses. Firstly is the sense in which it is directly affected by and directly impacts on the development and uptake of biotechnology. The fact that many historical accounts of PUS research select particular government reports or government-commissioned reports as milestones in its trajectory is testament to sense in which PUS is often Political, with a capital 'P' (e.g. S. Miller, 2001). But there is also the sense in which PUS is always political with a small 'p', in that it is shaped to such a great extent by normative concerns and motivations. Bauer, Allum, & Miller (2007) describe three such dominant themes, corresponding to three phases which characterise the recent history of PUS.

From the 1960s to mid 1980s, they explain, the focus of attention in PUS was on 'Scientific Literacy'. Research efforts were directed towards assessing whether members of the public held sufficient knowledge of science to be able to engage in debates and decision-making in science policy. Knowledge of science was assessed by means of small sets of items in social surveys. According to the well known 'deficit model' of the public's relationship with science, the layperson was presumed to be deficient in knowledge, and formal education was seen as the route by which this shortfall should be addressed, as a prerequisite for enabling science policy to be made on democratic terms.

A shift in focus occurred in the mid-1980s, and Bauer et al. identify this period, from 1985 until the mid-1990s, as the heyday of 'Public Understanding'. In this research programme a deficit on the part of the public was still perceived, but not so much in terms of knowledge as in terms of *attitudes* – specifically, positive attitudes towards science and technology. Knowledge remained an important accompanying focus, however, based on the premise that increased knowledge about science should correspond to increased enthusiasm for it. But in research terms, the focus of attention shifted to measuring attitudes towards science and towards science actors, through a wide range of research tools, both quantitative and qualitative. The lasting influence of this theme in the research field is demonstrated by the fact that 'PUS' is still a widely recognised acronym which may be used to refer to almost any research or activities concerned with the relationship between the public and science.

The most recent phase in PUS, from the mid-1990s until the present day, is the project known as 'Science and Society'. This new paradigm also centres on a deficit, but this time one which lies with science *experts*. Granted, the public may not be as positive about science or know as much about science as the scientific community would like. But perhaps the scientific community is likewise deficient in its understanding of the public, and perhaps it is therefore responsible to some extent for public alienation from science. In this phase of PUS, public *engagement* with science is key, and in many current initiatives, Bauer et al. (2007) note that social scientists are required to take on the role of 'angels' – mediating and bolstering dialogue between scientists and the public. Favoured research methods in this phase have been largely based around participation projects and forms of action research, rather than traditional surveys.

The different directions and emphases of these three phases are not incommensurate; Bauer (2003) portrays them as representing not so much a linear path as a multi-way expansion of research questions and concerns. These questions and concerns share a common methodological problem: how to operationalise the concepts they seek to understand. First was the question of how to measure the public's knowledge, then how to measure their attitudes, and finally how to measure their engagement with science (Bauer et al., 2007). This is of course a simplification of a diverse research field. Nevertheless, Bauer et al (2007) contend that among the many research needs, questions and obstacles in PUS is an enduring need for good survey indicators of key constructs. Three of these key constructs will be investigated in the empirical studies in this thesis. They will be outlined in some detail in Section 2.4, following the general introduction to social surveys in PUS, in the next section.

## 2.3 Social surveys in PUS research

Surveys have enjoyed a continuous and influential presence in PUS research since its early days (von Grote & Dierkes, 2000). Withey's 1957 survey, for example, is a milestone often cited as marking the start of modern PUS research in the US (Jon D. Miller, 1983). Following this landmark study, surveys continued to be used in earnest from 1979 via the US National Science Board's biennial Science Indicators survey. In the UK, significant survey research originated in a study by Durant and colleagues in the late 1980s (Durant, Evans, & Thomas, 1989), and continued with several surveys either partially or completely dedicated to PUS topics, including modules in the British Social Attitudes series, a survey sponsored by the Wellcome Trust, and a number by government departments and government-commissioned bodies. These two examples are from a large number of countries in which surveys covering PUS topics have been fielded. For comprehensive reviews of surveys on public perceptions of biotechnology, see Hamstra (1998) and Sturgis and Allum (2006).

From an early point in the use of national public opinion surveys, efforts have been made to replicate questions in different countries. For example, a small set of items relating to knowledge of science and technology have been used recurrently in many country-specific surveys (see Section 2.4.2). More systematic comparative surveys have become the dominant reference points for cross-national comparisons of public opinion. The European Commission's Eurobarometer series provides the most significant contribution to this area of enquiry. These surveys are fielded at regular intervals in all EU countries, often repeating certain sets of survey items from wave to wave, providing a valuable source of trend data with a broad coverage of countries (Durant et al., 2000). Eurobarometer surveys have included modules on biotechnology in 1991, 1993, 1996, 1999, 2002 and 2005. My empirical studies use data from the 2002 survey, fielded in the then fifteen member states, with samples of approximately 1,000 per country, selected using multi-stage stratified random sampling.

Standard social surveys on perceptions of biotechnology which use methods of probability sampling are expensive to field, and yet remain somewhat under-analysed. Perhaps this is partly explained by fondness of the mass media for reporting headline percentages, rather than detailed analyses of the relationships between variables. Arguably, however, biotechnology surveys do not always lend themselves well to such opinion poll-style reporting. Three key features of these surveys make the application of careful multivariate analyses an important part of understanding the information they contain. These features reflect three significant challenges which are commonly encountered in survey research – to a greater or lesser degree, depending on the subject matter and coverage of the survey. The following sections outline these challenges, particularly with the Eurobarometer in mind, since this is the focus for the empirical chapters.

### *2.3.1 Capturing complex constructs*

The psychological constructs in which PUS researchers are interested are by definition not directly observable, and are often complex. In order to investigate attitudes towards GM food, for example, it would be inadvisable to simply pose the question, 'what is your attitude towards GM food?'. Such a general question would be cognitively difficult for respondents, and would elicit a range of types and amounts of information from different people. The conventional and better approach is to prompt respondents by asking a number of specific, focused questions, and analysing answers to them in combination to reach a representation of the construct of interest. So in order to investigate attitudes towards GM food, we might ask respondents whether they think it is risky – that is, ask them to evaluate an object (GM food) in terms of an attribute (riskiness). Or we might ask them about relevant behaviours – for example, have they ever bought GM food? Or would they consider doing so, hypothetically? Having teased out more manageable, discrete elements of the construct with these specific questions, however, does not make the survey unproblematic. Individual items of this type can *also* be challenging for both respondents and survey analysts, albeit in different ways.

A key challenge of biotechnology as a survey topic is the fact that it is unfamiliar to the average respondent – although some applications are more widely known than others, of course. For example, in the most recent Eurobarometer on biotechnology, fielded in 2005, 20 per cent of the population of the 25 member states had not heard of GM food before, while 73 per cent had not heard of pharmacogenetics (Gaskell et al., 2006). Low levels of familiarity about a topic raise questions about the meaningfulness of people's responses to survey items about it. However, such questions are not limited to surveys of unfamiliar subjects. In a sense, they are simply a special case of a more general and longstanding concern in survey research – that of trying to understand the precise nature of the relationship between respondents' perceptions and opinions, and their answers to questionnaire items.

Chapter 1 pointed out in simple terms the problem that the responses given to individual survey questions are the product of a host of influences, some internal and some external to the survey researcher's focus of interest. Empirical research shows that on repeated asking (over time or in different experimental conditions), an individual's responses to the same question can vary a good deal, especially if he or she does not already hold a firm conviction on the issue. Paul Lazarsfeld famously uncovered this effect in a panel survey of political opinions, in which a startling amount of oscillation between preferences at the individual level was concealed beneath stable distributions of opinions at the aggregate level (Lazarsfeld, Berelson, & Gaudet, 1948). His finding led to a number of competing explanations of survey responses, most notably Converse's (1964) 'black-and-white' model, which distinguishes between those respondents who hold fixed, 'true' opinions – whose responses to questions change little if at all between panel waves – and those whose responses vary, as if randomly, and which are termed 'non-attitudes'. Converse's model resonates well with the languages of psychometrics and educational testing, where the term 'measurement error' is used to describe responses which do not logically tally with the underlying 'true' abilities or attributes measured in tests.

The terms 'non-attitude' and 'error' may appear to the non-psychologist to entail unfortunate negative connotations. So it is important to note that 'error' essentially simply means departure from an expected pattern, and that 'non-attitudes' simply denote unstable response patterns – neither of which should be classed as invalid responses. Converse stressed that people with few or no pre-considered views on a topic do not actually answer survey questions randomly. Rather, in fashioning their responses they draw on a range of ideas, beliefs and considerations available to them at the time – something analogous to an 'apperceptive mass' (Converse, 2000). The precise way in which they draw on and process this information is itself the subject of a number of alternative theories.

Zaller and Feldman (1992), for example, suggest that a survey respondent forms his or her answer to a particular question at the moment of asking by averaging across a sample of ideas and considerations. These ideas may be diverse, or even logically speaking dissonant, and some ideas will be more salient than others. Saliency might be attributable to factors external to the questionnaire, such as media coverage of the survey topic (see Bauer, 2000 re the media in relation to biotechnology), or personal experiences. Some considerations might be salient as a result of factors internal to the questionnaire; perhaps powerful imagery or significant sub-text contained in question wording, or the framing effects of the order in which questions are posed. In PUS surveys, for example, Gaskell, Wright, and O'Muircheartaigh (1993) show how questions testing respondents' knowledge of biotechnology affect responses to subsequent questions asking about their interest in the topic.

When data suggest that some considerations have been salient for respondents because of the characteristics of the survey instrument rather than the substantive topic, the over-sampling of such considerations is commonly termed a 'response effect'. A wealth of literature documents this phenomenon (see e.g. Dillman, 2000). As well as the framing effects resulting from question ordering, response effects can be induced by the way in which a question is posed. For example, social desirability bias occurs when respondents answer a question so as to please the interviewer, or conform to social norms. Over-reporting perceived 'good' behaviours, such as taking exercise or eating healthy foods, is an example of this. Response effects can often be induced by the response options offered. For example, selecting from a list of options can involve vulnerability to memory biases (primacy and recency effects), and some responses per se can be more commonly chosen in certain types of respondents. For example, 'acquiescence bias' is the tendency to give positive responses (to 'agree', or say 'yes', or 'true').

According to Zaller and Feldman's theory, saliency is positively related to accessibility, with more recently considered issues being more accessible. Respondents who are more involved with the subject matter of the survey should then have easier access to considerations pertinent to it, and their sample of considerations should contain fewer contextual factors than those of respondents who are unfamiliar with the substantive topic. This implies that we should expect contextual factors, including response effects, to have a significant presence in survey data on unfamiliar topics such as biotechnology. However, it remains to be shown conclusively that response effects really occur more frequently in such a scenario. While some survey researchers find that selected types of response effect are related to issue involvement (Bishop, 1990), others investigating a range of response effects find no evidence of a relationship between the occurrence of these and respondents' certainty, intensity of attitudes and the personal importance of the survey topic to them (Krosnick & Schuman, 1988). Response effects, then, should be a matter of concern for all social surveys – not just those on unfamiliar topics.

### *2.3.2 'Don't know' (DK) responses*

Given that biotechnology is an unfamiliar topic, it is no surprise to learn that surveys about it tend to return high rates of DK responses. Recalling the earlier examples of GM food and pharmacogenetics from the 2005 Eurobarometer, when asked, 'Do you think GM food should be encouraged?' 16 per cent of respondents across the 25 EU member states responded 'don't know'. On being posed the same question in relation to pharmacogenetics, 33 per cent said they didn't know.

Albeit 33 per cent is a high DK response rate, as social surveys go, it is rather lower than the 73 per cent of respondents who said they had not heard of pharmacogenetics before taking part in the survey. Apparently a good proportion of these respondents have formed their judgements ad hoc, for the purposes of answering the survey question, rather than revealing stable, previously considered preferences. The willingness of respondents to make a spontaneous evaluation of an unfamiliar topic is a well known phenomenon. Gill (1947) found that 70 per cent of a sample of respondents could be persuaded to offer a non-DK response when asked their opinion on a fictitious 'Metallic Metals Act'.

The way in which a survey is administered can dramatically affect DK response rates: they can be increased when DK is explicitly offered as a response alternative, and can be reduced by pressing unsure respondents to say what their opinion would be if they had to choose, or had to give their best guess in response to a factual question. Advice from survey researchers is mixed as to whether to disallow, allow or encourage DK as an answer. There are a number of reasons for retaining it as a valid response option. Firstly, in the light of the preceding discussion of response variability and quality, it can be argued that forced responses are compromised in quality, or at the least, not worth the extra cost of obtaining them (Sturgis, Allum, Smith, & Woods, forthcoming).

Secondly, it might be contended that DK is meaningful response. On one hand, it could be seen as an important signal to the survey designer that an item does not work very well. For example, Coombs and Coombs (1976-1977) suggest that a DK response may be due to a mismatch between respondent and survey instrument, with the respondent being unable to map his or her true position on to one of the response options offered – perhaps because that position lies between or across several categories, or because the categories themselves are too vague, or because they are simply not relevant. On the other hand, a DK response might be interpreted as a substantively meaningful statement on the part of the respondent. For example, it may not denote lack of a suitable available answer so much as an explicit expression of self-doubt (Bauer & Joffe, 1996), specifically in relation to the topic in question, or generally in relation to oneself. In more extreme interpretations, DK responses may denote disengagement, apathy or alienation from the topic (Jodelet, 1996), or more radically, disengagement from the research instrument, regarded as a means of the colonisation of the life-world (Turner & Michael, 1996).

To investigate the 'meaning' of DK responses in the Eurobarometer would require other, qualitative studies, and is not the focus of this thesis. The point of departure is that it is the policy of the designers of the Eurobarometer on biotechnology to allow DK responses, and that survey participants choose to give this response in relatively high numbers. The question for this thesis is how to treat them in analyses.

When there are very few DK responses, so few that we cannot confidently say anything useful about them (for example, in terms of associations with other variables), it is generally regarded as acceptable to exclude DK responses from analyses, even though one would probably be optimistic in thinking that they were randomly distributed. Respondents might be deleted listwise from analyses, or some method used to impute missing values, or the estimation method used for the analysis may be able to deal with missingness in some other way (see e.g. Full Information Maximum Likelihood in Chapter 4).

Where there are high rates of DK responses, however, discounting them is cause for greater concern. Listwise exclusion of respondents giving DK answers then results in a considerable loss of information, and quite probably the removal of a section of respondents who differ systematically from those who give substantive responses. Indeed, in an analysis of Eurobarometer data, Bauer (1996) found that DK responses to biotechnology items were more likely to be found amongst respondents who were older, less affluent, less highly educated, and who consumed less news media. So DK can in some instances be couched as a response effect, attributable jointly to the nature of the survey and the respondent, similarly to effects such as acquiescence bias.

An alternative, common approach to dealing with DK responses is to recode them for the purposes of analysis. For example, in the case of an ordinal Likert response scale with an odd number of categories, it is typical to interpret DK as a neutral response, and to recode it as a middle category. The motivation for doing this is strong if the researcher wishes to analyse the variable as if it were measured at the ordinal or interval level. In the context of this thesis, a particular concern is the common practice of recoding or dropping DK responses in order to apply factor analysis to the data. Using factor analysis for ordinal data is a separate concern that will be discussed in Chapter 4. Putting to one side the model applied to the data, the practice of recoding DKs as middle response categories is rarely accompanied by any empirical justification.

A simple way to improve survey analysis, where high DK rates are returned, is to treat DK as a separate category in a nominal variable. This is the approach adopted in the empirical studies in the thesis.

### *2.3.3 Cross-national comparisons*

The challenge of cross-national comparisons is clearly not pertinent to all social survey research. However, it is undoubtedly increasing in importance, with recent years witnessing a considerable growth in the development of cross-national survey programmes. In the area of social and political attitudes, the European Commission's Eurobarometer surveys have been fielded biannually since 1973 in all member states, with additional 'flash' Eurobarometers used on an ad hoc basis to gauge immediate public reactions to topical issues. In the context of an enlarging EU, the coverage of these surveys has been expanded, with barometers for central and eastern Europe, and candidate countries. The Eurobarometer has inspired similar programmes in other regions: the Latinobarómetro, launched in 1995; the Afrobarometer, launched in 1999; the Asianbarometer, covering east Asian countries since 2001; and the Asiabarometer, fielding surveys in southern, southeast and central Asia since 2003. Other notable contributions to the study of social attitudes have been made by the European Values and World Values surveys, since 1981, and the International Social Survey Programme, since 1982, as well as more recently by the European Social Survey, first fielded in 2002.

Cross-national studies of social attitudes may be approached from a broad range of theoretical perspectives. In social psychology such studies would more usually be described as cross-*cultural*, with a corresponding de-emphasis on the political framework of the nation state. In more politically oriented academic disciplines, a great deal of literature has been devoted to the 'comparative method', and a number of classifications have been proposed to clarify the different concerns and emphases entailed in cross-national comparative studies. For example, Kohn (1987) distinguishes between those in which the nation is the object of the study; those in which the nation provides the context in which to understand some social phenomenon; those in which the nation is a unit of analysis in a larger scheme; and those in which the nation is a component of a transnational system. The types of analysis carried out using Eurobarometer and similar social survey data tend to fall under the second of these types, with varying emphases given to the role of the country in explaining attitudes, values and institutions (Arts & Halman, 2002). In the standard Eurobarometer biotechnology reports, for example, the primary and immediate objective is to describe differences and similarities between countries in terms of their populations' attitudes towards biotechnology. These patterns carry considerable political and economic significance, and are the subject of a great deal of interest from a wide range of science actors. In addition, however, a notable strand of comparative research in PUS is given to explaining such similarities and differences and their trends over time, in terms of transnational patterns linked to processes of industrialisation and secularisation (e.g. Durant et al., 2000), reminiscent of similarly broad theories of changes in social values (e.g. Inglehart, 1990).

Regardless of the particular theoretical orientation adopted, all analyses employing data from comparative survey programmes such as the Eurobarometer share the challenge of how to make valid comparisons between countries (see e.g. Harkness, van de Vijver, & Mohler, 2003). 'Valid' broadly means, in this context, comparing like with like – that is, to what extent do items take on the same meanings in the countries on which they are being compared? Or more realistically, to what extent are meanings sufficiently similar across countries to enable comparisons to be made between them?

There are many simple ways in which item meaning can vary. If a survey is administered in a number of different languages, the comparability of translations is immediately called into question. The relationship between language and meaning is well known to be complex – it is a theme with a long history in philosophy and psychology. Arguably some concepts simply do not translate between languages at all (Elder, 1976). Linguistic issues are commonly addressed by teams of researchers representing the different languages in which a questionnaire is administered, alongside thorough back-translation tests. But language is, of course, only one element of culture which can contribute to differences in item meaning.

'Item bias' is the broad term used to denote scenarios in which items apparently take on different meanings between groups of respondents. If the content of some items has varying cultural significance, *construct* bias can result when these items are combined into a composite measure. Indicators of political activism, for example, would be expected to vary markedly cross-nationally. In addition, however, the survey instrument itself may play a role in obscuring cross-national comparisons. For example, acquiescence bias is observed more often in respondents from certain cultural groups (Smith, 2003). Culturally specific response styles are given the general term 'method bias' by Van de Vijver and Leung (1997).

Identifying and allowing for cross-national differences in the relationship between the survey and the respondent – be it in terms of item meaning or response style – is the key challenge in cross-national survey research. There are many ways in which the task of identifying and dealing with these differences can be addressed. Traditionally, cross-national survey research should draw heavily on supporting qualitative data on the cultures and languages involved in the survey. Another approach, not widely used in survey research to date, is to exploit the potential of statistical models for identifying items which 'function' in different ways between groups. The empirical studies in this thesis will employ statistical models precisely for this purpose. The details of these are reserved for Chapter 4, when the technical specifications of the models are presented.

## 2.4 Three social psychological constructs central to PUS research

The three constructs investigated in the empirical chapters correspond to the central variables of interest in the three phases of PUS research described by Bauer et al. (2007). My empirical work begins with the task of measuring general (positive and negative) attitudes towards biotechnology, specifically on two applications: GM food and therapeutic cloning. Recall that attitudes were the focus of interest in the Public Understanding paradigm. Next is the measurement of knowledge about biology and genetics, of central interest in the Science Literacy era, and still a somewhat contentious issue. The final construct explored is engagement with science and technology – a key theme in the current Science in Society turn. The data for these analyses are all from the same survey wave, to allow some simple analyses of associations between the constructs, presented in Chapter 8.

There are many other phenomena which could have been chosen for this report – for example, perceptions of risk are a significant ongoing theme in PUS; the role of religious, moral, and political values are increasingly of interest; and trust in science actors is arguably of equal importance to engagement in the Science and Society paradigm. The three constructs were chosen partly because of their clear substantive importance, and partly, given the available data, because they entail a balanced selection of methodological questions and concerns arising in PUS survey analyses, which can all be addressed effectively with latent trait and class models. The sections below describe, in non-technical terms, some of the ways in which each construct has been operationalised using survey data to date, drawing focus towards the most promising approaches suggested by the literature. These will be taken as the points of departure for each of the empirical chapters.

### *2.4.1 Positive and negative attitudes towards biotechnologies*

The first construct of interest is very broad, and not surprisingly, over the years a plethora of questions have been posed to respondents on this subject. In their review of surveys, Sturgis and Allum (2006) list 817 different questions used to gauge attitudes towards the 'red' biotechnologies alone. These range from questions asking respondents to speculate on its future (will life be made better or worse by it, in the long term?), to behavioural questions (would you consume it?) to evaluative questions (do you approve of it?).

A useful initial distinction to make when discussing this topic is between those questions which ask for opinions of biotechnology in general, and those which ask about specific applications of biotechnology. For the former, a common approach, which has been used for example in the Eurobarometer since 1991, is to ask respondents whether each of a number of technologies will improve our lives in the foreseeable future, or make them worse, or make no difference either way. This 'technological optimism' measure provides an interesting indicator for comparisons with other technologies over time – with biotechnology losing favour amongst the European public during the 1990s, hitting a nadir in 1999 but steadily recovering support subsequently (Gaskell et al., 2006). However, it makes no pretence to capture anything beyond respondents' basic intuitions with reference to a very broad class of technologies, with which, as I noted above, the average respondent is rather unfamiliar.

An alternative approach to capturing general attitudes is to calculate a composite indicator from a set of questions covering a range of aspects of a selection of different technologies. With this strategy, it is the *researcher* who produces a single indicator of attitudes, by averaging over a range of considerations on behalf of the respondent. This approach has been criticised quite heavily (see e.g. Pardo & Calvo, 2002), mainly for attempting to capture too heterogeneous a phenomenon in a single measure. The phenomenon is so diverse that any single scale will always be open to the charge of bias, due to the questions included in and excluded from the scale. General measures of this kind tend to be considered of limited use in application too: in analyses they generally do not exhibit good scaling properties, and have less predictive validity than measures capturing attitudes towards single applications of biotechnology (Evans & Durant, 1995).

It is not surprising that measures of general attitudes suffer from problems such as these. A good deal of PUS research demonstrates that people tend not to view every biotechnology in the same way, but discriminate quite sharply between them. A striking and recurring finding from the literature is that 'red' applications of biotechnology receive notably more support than 'green' applications amongst Europeans (e.g. Bauer, 2005). But it is not simply the case that people judge medical biotechnologies more favourably than plant biotechnologies; it is apparent that they also judge them *differently*. These different structures of attitudes towards different applications of biotechnology have become a longstanding interest in PUS.

To give an example, the role that perceptions of risk play in people's evaluations of biotechnology has been a continuing conundrum. It has been widely assumed in the expert community that public aversion to GM food, for instance, is motivated by alarm regarding the possible hazards associated with it. Science communication programmes have therefore focused on allaying such fears. But research suggests that is it not so much the presence of risk as the absence of benefits that underlies people's broadly negative reactions to GM food (Gaskell et al., 2004). Indeed, to take a contrast, therapeutic cloning tends to be perceived as risky, but broadly supported nonetheless – it is apparently worth the risk (Gaskell et al., 2003). So perceptions of high risk do not go straightforwardly hand in hand with negative overall attitudes. In fact it seems that the layperson's approach to risk assessment is quite different from the cost-benefit calculus of the risk analyst. Some theorists reason that affective evaluations play a significant role in cuing assessments of the riskiness of a technology (Slovic, Finucane, Peters, & MacGregor, 2002). Others contend that it is more appropriate to conceive of risk perception simply as one of the many elements comprising a social attitude, rather than as phenomenon sui generis (Allum, 2002; Eiser, 2001). In this formulation, risk is amongst a number of potentially relevant considerations, in Zaller and Feldman's terminology, which contribute to overall positive or negative evaluations of a technology. For some people risk might be a salient consideration, whilst for others, moral and ethical concerns may take a more prominent role in their evaluations. In this vein, Hviid Nielsen, Jelsoe, and Ohman (2002) distinguish between two classes of biotechnology sceptics, based on 'modern green' and 'traditional blue' resistance. The former reason against biotechnology by balancing risk against benefits in a utilitarian framework, while the latter's opposition is framed around moral value judgements, including traditional worries about tampering with nature, and 'playing God'.

Despite the focus on attitudes in the Public Understanding project of PUS, Pardo and Calvo (2002) note that the measurement of attitudes has been the subject of very little explicit critique – barring the considerable critique from those who object to its measurement altogether (see Section 2.5 below). Their own methodological assessment of one of the widely used scales of general attitudes is very critical in tone. A more enthusiastic contribution is made by Pardo, Midden, and Miller (2002), who use a factor analysis to create a two-dimensional model from a set of items in the Eurobarometer which ask respondents to judge several biotechnologies individually on four criteria: to what extent they are risky, useful, and morally acceptable, and to what extent they should be encouraged, overall. In Pardo et al.'s model, which combines respondents' assessments of six biotechnologies together, one dimension is given to judgements of risk, while the other criteria form a general scale of support and opposition.

The analysis of these items has been approached in an alternative way by Gaskell and colleagues, who use them to define typologies rather than scales of support and opposition. In these typologies, which they term 'logics' of support and opposition, they distinguish between 'risk-tolerant' and 'risk-relaxed' respondents. Creating a typology separately for each application of biotechnology, these items have been used to track and compare public perceptions of different biotechnologies since 1996 (see e.g. Gaskell et al., 2000; Gaskell et al., 2006).

These 'logics' of support and opposition have perhaps been used more consistently over recent years than any other composite measure of specific attitudes, and they exhibit good predictive validity in a number of analyses (Gaskell et al., 2006). They have not been subjected to any close methodological scrutiny, however. A careful re-analysis of these items would seem to be a particularly useful way to proceed with the project of finding a good measure of attitudes. This is not only because these particular items seem to be useful, but also because Pardo and Calvo (2002) specifically suggest, as a more general point, that searching for clusters rather than adhering to the convention of constructing scales, would be a fruitful approach to attitudes. Following their lead, the logics will form the point of departure for creating models of positive and negative attitudes, in Chapter 5. To carry forward the interesting distinction between red and green biotechnologies, judgements of therapeutic cloning and GM food will be modelled separately, and compared.

### *2.4.2 Knowledge of science in general, and of biology and genetics in particular*

Knowledge of science and technology in general, and of knowledge of biotechnology in particular, have been central constructs in PUS from its early days; Withey's 1957 survey included items on knowledge of science, for example. The enduring question attached to this construct is its relationship with attitudes: specifically, is greater knowledge associated with more positive attitudes, as scientists assume must logically be the case?

The precise formulation of the construct 'knowledge', or 'literacy' as it is sometimes controversially called, has been the topic of considerable debate within the PUS community. This concept has been dismantled along functionalist lines to clarify what exactly it is that experts think the public ought to know about science. Shen (1975) is one of several researchers to propose a classification of different *types* of literacy, distinguishing between practical, cultural and civic scientific varieties. It is the last which is of most interest to science actors: this is the kind of knowledge required for understanding science reports in popular media, and hence for appreciating and potentially contributing to debates in the public sphere on scientific issues.

Deconstructing *this* concept, in turn, other researchers have proposed a distinction between different types of civic scientific literacy. In the US, Jon Miller theorises that civic scientific literacy comprises three key elements (Jon D. Miller, 1983). Firstly, it involves possessing a basic understanding of scientific terms, such as would enable one to understand science stories and issues reported in the media and other sources – for example, knowing what a gene is and what it does. Secondly, it involves an understanding of the nature of scientific process, including for example the peer-review system and the logic of experimentation. Thirdly, it entails an appreciation of the societal- and individual-level impacts of science. In the UK, researchers have proposed a fourth element, contending that any understanding of scientific process is incomplete without an acknowledgement of the people taking part in that process. This fourth element of scientific literacy is thus an appreciation of science as a social institution: an awareness of its internal and external politics (Bauer, Durant, & Evans, 1994; Bauer, Petkova, & Boyadjieva, 2000).

Of these four elements, the first two have been more successfully, or at least more often, operationalised. Miller found that the third element – appreciation of the impact of science – was too culturally specific to be feasible for inclusion in cross-national survey analyses (Jon D. Miller, Pardo, & Niwa, 1997). Bauer's study of the fourth – perceptions of the institution of science – spanned such great cross-cultural differences in science institutions that for some items in his scales the correct answer was dependent on the country in which it was asked (Bauer et al., 2000). Such a feature would be counterintuitive to attitude measurement theorists, as well as to a lay audience reading reports employing such scales.

Of the first two elements, questions on the *content* of science have been much more widely used than questions on the *process* of science. And of the former, close-ended questions have been more widely used than open-ended questions. The latter, though informative, are rather more difficult and costly to field and analyse. A standard set of closed-ended items on science and technology has been developed and used by Jon Miller in the US (Jon D. Miller, 1998), and John Durant and colleagues in the UK (Durant et al., 1989). Known as the 'Oxford scale' in the UK, or the 'science literacy scale' in the US, this comprises a set of usually around ten statements which respondents are asked to identify as true or false. Such statements include, for example, 'The centre of the Earth is very hot', and 'All radioactivity is man-made'. The method by which the scale is constructed from these items varies, but the items are by far the most widely used in the literature (Pardo & Calvo, 2004).

Miller's approach to measuring knowledge with these variables is based around a binary distinction between the scientifically literate and illiterate. Having constructed a scale of knowledge, he converts it into a dichotomy by setting a threshold for the number of correct answers that would denote 'literacy'. This critical threshold varies slightly according to the data at hand. For example, in Miller and Pardo's (2000) comparative study of the EU, US, Canada and Japan, those in the first three regions who answered two thirds of items correctly were classed as being scientifically literate. For Japanese respondents the bar needed to be set slightly lower, due to the technical characteristics of the scale created from the items put to them. The threshold was reset at 60 per cent due to a break in the distribution of the data at this level. So the arbitrary threshold is derived sometimes externally and sometimes internally to the data. In other and more recent studies, notably those by Durant and his UK colleagues, the concept of literacy has been largely rejected in favour of the more neutral label 'knowledge'. The relative comparison of possessing more or less knowledge is seen as less derogatory than the absolute judgement of being literate or illiterate. And a continuous measure is, speaking practically, convenient for correlation based analyses.

The Oxford scale, while widely used, has also been widely criticised. Most criticisms relate to the content of the items. In a similar vein to the critique of the general attitude scales, it is argued that ten or twelve items cannot possibly capture such a heterogeneous field of knowledge. This is a misinterpretation of the logic of the measure, which, following surveys of political knowledge, posits that a person's comprehension of the few facts asked about in the survey is likely to be indicative of his or her general scientific knowledge (see e.g. Converse, 2000). The items are to be viewed as a *sample* of facts from a wider domain, rather than as a set of exhaustive facts which test all important elements of knowledge.

Further objections to the scale contend that the domain of knowledge from which the items are sampled is biased towards the kind of text-book knowledge learned in formal, Western education, and that it thus favours respondents who are closer to that education system: notably, younger and more privileged Western respondents (Peters, 2000). This seems to be borne out in Pardo and Calvo's (2004) methodological analysis of the scale created by summing the number of correct answers to the Oxford items. They find that the scale has smaller variance (and hence lower reliability, in terms of Cronbach's alpha) in more industrially advanced countries. The cross-cultural transferability of the scale is thus an open question – where cultures might be defined within as well between countries (see e.g. Raza, Singh, & Dutt, 2002 for a further perspective on this issue).

To maintain consistency with the approach adopted for the analyses of attitudes, the empirical study of knowledge in this thesis will depart from the Oxford scale, which has already received a good deal of attention and critique. Although I will make brief reference to the Oxford scale, the focus will be on a set of items specifically concerning biotechnology. This scale follows the same format as the Oxford scale, but contains only statements relating to biology and genetics, such as 'It is the mother's genes that determine the sex of the child', and 'The cloning of organisms results in genetically identical offspring'. The full set of items will be presented in Chapter 6. A meta-analysis by Allum et al. (forthcoming) suggests that this focused measure is more strongly related to attitudes towards biotechnology than is the Oxford scale, and therefore likely to yield interesting findings, especially in the final analyses of the relationships between the three constructs, presented in Chapter 8.

This biotechnology scale, which is used widely in the Eurobarometer, has not been the subject of very much methodological critique; certainly not of the kind published with reference to the Oxford scale. So an analysis of these items might be especially valuable for the designers of the Eurobarometer on biotechnology. A notable feature that these items share with the Oxford scale is a high rate of DK responses – perhaps more consistently high across biotechnology items than is the case with the Oxford science items. So investigating possible interpretations of DK responses will constitute an important element of these analyses.

### *2.4.3 Engagement with science and technology, and with biotechnology*

Engagement with science and technology broadly, or with biotechnology specifically, enters into the PUS arena in quite a different guise from the concepts of attitudes and knowledge. In the current Science and Society programme, the approach to engagement is not to measure it and then investigate the ways in which it is correlated with other variables, but to *create* it. Social researchers are recruited as 'angels' (Bauer et al., 2007) to facilitate public consultations, rather than as onlookers to survey the scene of public opinion. In a sense, we might then say that it would be tautological to try to also *measure* engagement with survey data. However, arguably there remains a need for quantitative indicators of engagement. They may be used as a tool in the evaluation of such engagement programmes; Rowe, Horlick-Jones, Walls, and Pidgeon (2005) administered questionnaires to participants as part of their evaluation of the UK Government's 'GM Nation?' public consultation exercise, and explicitly discuss the need for reliable and valid measures to use in their assessments. They may also be applied as a population-wide tool, for gauging the general climate for participation among the public (Bauer et al., 2007).

In contrast with the first two constructs, then, there is a notable absence of discussion and critique in PUS about indicators of engagement, both in terms of the form and of the content of such a measure. It is not the case, however, that it is entirely absent from quantitative analyses in PUS. A number of composite indicators can be cited in the literature which might be identified as more or less closely related to the idea of engagement.

Jon Miller, for example, writes about the 'attentive public' for science, or for biotechnology, drawing on a model from political science (Almond, 1950). He creates a three-category ordinal classification of levels of attentiveness. To be part of the attentive public for an issue is to be interested in it, to feel informed about it, and to seek or be exposed to information about it via various media sources. To be interested but to feel uninformed, and also practically speaking *be* uninformed, is to be part of the 'interested' public. Those with any other combination of characteristics are classified as belonging to the unengaged class of the 'residual' public. Miller uses this typology in relation to science in general (e.g. Jon D. Miller & Pardo, 2000) and to biotechnology in particular (e.g. Jon. D. Miller & Kimmel, 2001). In the latter study, Miller and Kimmel also use an ordinal measure of 'awareness' of biotechnology, as a combination of two criteria: having heard of biotechnology before and having talked about it with others, occasionally or frequently. The resulting variable takes five categories, from having neither heard nor spoken about biotechnology before, to having both heard about it and spoken about it frequently with others.

In a later study, Pardo, Midden, and Miller (2002) define a typology of 'informedness' about biotechnology on the basis of a combination of binary criteria: awareness versus lack of awareness (i.e. having heard of biotechnology before, or being completely unfamiliar with the topic), and high versus low knowledge (i.e. above or below a certain threshold on a scale of knowledge of biotechnology). A slightly different angle is adopted in a study by Evans and Durant (1995) who define 'interest in science' as a combination of a number of items asking respondents for ratings of their level of interest and their consumption of science-related media.

Finally, in a latent class analysis of the latest Eurobarometer on biotechnology, fielded in 2005, Gaskell et al. (2006) create a measure specifically termed 'engagement' with biotechnology. They define four types of respondents: the 'attentive' public have high levels of awareness and knowledge about biotechnology; the 'active' are aware of biotechnology and are likely to have taken part in public meetings on the subject; 'spectators' report lower levels of exposure to biotechnology; and the 'unengaged' give negative responses to all indicators of engagement.

Since the measurement of knowledge is a separate study in this report, it will not be included in analyses of the concept of engagement. Beyond this limitation, the analyses will only be confined according to the available variables in the survey. In the 2002 Eurobarometer there are three sets of plausible items, posed at various points in the questionnaire. Details of these will be given in Chapter 7. In one set, respondents are asked to rate their levels of interest and informedness with regard to science and technology. The two other sets relate to biotechnology specifically, and ask about behaviours: some ask whether respondents have heard of or talked about biotechnology before (capturing the awareness dimension of the construct), while two ask respondents whether they would hypothetically be willing to engage with the topic, either through media consumption or more actively through participation in public meetings. So the items cover a range of facets of engagement: the general (science) and the particular (biotechnology), affect, past behaviours and potential willingness. The analyses will therefore be fundamentally exploratory in nature, adopting the simple objective of investigating the associations between them, and characterising types or levels of engagement accordingly. Without any firm guidance from the literature on whether types or scales would be more appropriate models for this construct, both latent class and latent trait models will be considered.

## 2.5 Critiques of survey research in PUS

The abundance of survey data on biotechnology is not universally celebrated by PUS researchers – in fact a number of high profile figures in the field have criticised survey research rather vociferously. Whilst some objections to its use are rehearsals of well known concerns regarding survey research in general, some hinge on a putative connection between survey methods and particular theoretical positions on the substantive subject matter. And while some critiques portray surveys as mildly deficient, some ascribe them an altogether more insidious character.

Milder objections to the use of surveys remind us of the drawbacks of imposing the rigid framework of a questionnaire on the data collection phase. Participants may express their opinions only through the standardised format of the survey structure; they have no power to steer the course of the survey interview, as would be possible in an unstructured interview or focus group meeting; and they must usually respond to questions by choosing one from a set of pre-specified answers. Thus, survey researchers miss the opportunity to gain new insights into the topic from their respondents – the voice of the respondent is restricted, by design. However, survey researchers would argue that the disbenefits of this are counterbalanced by the resulting standardised and highly structured data set, which is amenable to much more powerful statistical analyses than are possible with the more heterogeneous and loosely structured information obtained using qualitative data collection techniques.

There is always a risk, not only of leaving important avenues of enquiry unexplored, but also of asking questions and offering response categories that lack relevance or meaning for respondents. One can try to insure against this during the questionnaire design phase by drawing on existing research on the topic, and by employing focus groups, cognitive interviews and pilot studies to develop and test questions. The use of such supporting materials is crucial to the success of a survey; the chances of it misfiring are all the higher if it is written in a vacuum. One of the greatest obstacles in survey research is precisely the fact that meanings can vary between researchers and respondents, as well as among respondents. We may then report results in way that does not fully resonate with respondents' understandings. This is freely admitted among discerning survey researchers – to the point where it goes without saying. Unfortunately the tacit nature of this scepticism means that critics often assume survey researchers have a heroic confidence in the interpretations they give to their data (see e.g. Wynne, 2001).

It is not unusual for such misperceptions to exist between colleagues adopting different methodologies in their shared research field. In PUS, however, the division between qualitative and quantitative research has sometimes been acrimonious, at one point manifesting itself in the UK as a real wall of silence between researchers working within the same research programme, with common funding (Bauer, 2003). Quantitative researchers often use the metaphor of a map to describe the characteristics of surveys (see e.g. Durant et al., 2000): they provide a broad but shallow overview of a landscape, highlighting particular features of interest in a standardised and largely predetermined way. Other, usually qualitative, research tools are vital for providing the local colour that brings the map to life, to enhance its interpretation. Many qualitative researchers dislike this description, insisting that surveys and qualitative methods are fundamentally opposed, epistemologically. From a strong social constructionist perspective, surveys construct and objectify the phenomena they are studying. In PUS surveys, it is contended, the meanings of the concepts of 'science', 'the public', and 'understanding' are not open to scrutiny, but are treated as unproblematic objects defined by the survey designers. As surveys are increasingly presented as sources of facts about science and the public, they fulfil a function for politicians and scientists as 'ontological ordering devices' (Jasanoff, 2000a, p.85). Specifically, they perpetuate an accepted image of an ignorant public, distanced from an omniscient science, with this distance bolstering the authority of science experts (Felt, 2000).

By contrast, qualitative methods are championed as a means of deconstructing such hegemonic representations and exposing their political nature; qualitative research reveals a questioning public, in a crisis of trust with science (Jasanoff, 2000a). At its most extreme, this account divides PUS into two camps, with survey researchers on the side of the scientists and politicos, and qualitative researchers on the side of the people. To polemicise the field so dramatically is bemusing to PUS survey researchers, for whom the categorical distinction that PUS 'can be seen *either* as an objective phenomenon to be measured, monitored, and, if possible, manipulated, *or* as a social construct to be interpreted for the light it sheds on science-society relations in democratic societies' (Jasanoff, 2000b, p.39, my italics) is itself a false objectification. The further claim that survey research has a necessary connection to the first perspective and qualitative research to the second (e.g. in Irwin & Wynne, 1996), is likewise perplexing to those who conduct surveys in good faith, in full recognition of the shortcomings of this method.

Positing an inextricable link between certain methods and certain knowledge interests is both unfounded and unhelpful, and obscures the valid and important critiques of substantive theories in PUS proposed by these qualitatively oriented researchers (Bauer et al., 2007). Sturgis and Allum (2004) explicitly demonstrate that theoretical debates need not be tethered to methodological disputes, by using survey data to test a hypothesis relating to the deficit model, inspired by PUS survey critics. These 'contextualists', as they are termed by Sturgis and Allum, offer valuable critiques of PUS survey items that could contribute to survey design, if they were framed in a more positive way. For example, they criticise the content of science knowledge scales as being too close to the text book, and failing to cover types of science knowledge which are relevant for the layperson. Applying items with irrelevant content would result in an underestimation of the public's level of scientific understanding (Irwin & Wynne, 1996). In cross-national context, Peters (2000) has pointed out the possible bias of the Oxford items towards the Anglo-Saxon curriculum, as an artefact of the educational backgrounds of the questionnaire designers. Moreover, testing people's knowledge in the format of the Oxford scale may not be appropriate at all for the modern information society, where the key skill is not knowing a collection of facts but knowing where to find them if one needs them, and appreciating their significance (Jasanoff, 2000b). Finally, Jasanoff reminds us of an essential critique which all researchers should keep in mind: the danger of verificationism: 'Publics and their understanding, in short, are only imperfectly captured in studies designed to characterize them' (Jasanoff, 2000a, p.88).

Whilst acknowledging the limitations of social surveys for PUS, it must be conceded that surveys are indispensable for answering certain research questions. In particular, there is no alternative to probability sample surveys for gauging distributions of public opinion within the population of a country, or in a range of countries. And it is difficult to imagine a scenario in which distributions of public opinion on biotechnology were not of significant interest and importance to a number of science actors. This is by no means to say that surveys are a panacea for PUS research. They have clear limitations, and other research methods, both quantitative and qualitative, are also needed to understand this complex field.

## Conclusion

In this chapter I have sought to explain why we should be interested in public perceptions of biotechnology, and why social surveys play an important, though not exhaustive, role in studying such opinions. There are specific challenges to address in biotechnology surveys; challenges that are not unique to PUS, but which are arguably especially problematic in this research field. Biotechnology is an unfamiliar topic, heightening concerns regarding how to capture complex constructs with survey data, in the light of existing knowledge about response variability, and the possible distortions of response effects. DK response rates are often high, so that discounting or recoding them in some way would be a cause for considerable concern. And finally, cross-national comparisons need to be approached particularly cautiously with this relatively new and unfamiliar topic area; in multiple cultural settings we potentially multiply the range of meanings attributable to survey items. All of these challenges make good analyses of survey data difficult. Of course, some would abandon the survey project altogether in principle, but this would leave us without any way of gauging distributions of opinions, which are surely crucial for a consultative approach to science policy making.

A wealth of survey data already exist on this subject – moreover, good quality data, from large probability samples in a range of countries and over a period of some years. They carry great potential for informative studies. However, their potential is currently not fully exploited, and there is scope for improvement in their analysis and possibly their future design. The Eurobarometer, which is the most comprehensive source in terms of coverage of time points and countries, is designed for a wide usership rather than for theoretically sharply focused hypothesis-testing. Published analyses of Eurobarometer data include sometimes single indicator and sometimes multiple indicator representations of the constructs that are central to PUS, and little attention has been given to the statistical properties of these measures, especially in the context of cross-national comparability. The few existing methodologically oriented analyses of measures in PUS surveys have tended to be critical. For example, Pardo and Calvo (2002) complain that many high profile PUS publications use 'conceptually fuzzy scales and indicators that fall short of the standards generally applied in other areas of social-scientific research' (ibid., p.162).

One of the aims of this thesis is to contribute to the establishment of robust measures of key constructs in PUS, as called for by Durant et al. (2000). I hope to demonstrate that latent variable models are a valuable tool for this purpose. They are of course of limited value in isolation; the development of good measures requires careful theoretical consideration from experts in the substantive discipline. As an interested non-expert in PUS, my contribution is by way of exploratory rather than confirmatory analyses, following Pardo and Calvo's (2002) recommendation. Taking for each study a starting point suggested by the literature, my aim is to use latent trait and class models to glean as much useful information as possible from the data, which might be used to inform future PUS survey design and analysis.

The second aim of the thesis is to elucidate the value of latent trait and class models in attitudinal survey research generally, and to comment on their potential contribution to survey methods in social psychology. Latent variable models are one of many possible techniques we could use to scrutinise survey data and create summary variables, and social psychology is rich in scholarship on the subject of how to measure complex constructs such as attitudes, knowledge and engagement. In social psychology the measurement of these different constructs can be approached within the same broad framework commonly termed 'attitude measurement'. The next chapter sets the practical challenge of creating measures of constructs for PUS in the theoretical context of attitude measurement in social psychology.

# 3 The theoretical context of modelling attitudinal data

This chapter introduces the key themes in social psychology which inform the empirical studies in the thesis. It begins with the concept of the attitude, describing the range of social psychological perspectives historically taken on it, and the position adopted in this report. The diversity of interpretations given to this fundamental concept has substantial implications for the ways in which researchers have sought to capture it in empirical data, making the measurement of attitudes and related constructs an often controversial subject. To give an impression of the range of literature on this topic, documented in detail by many authors (cf. Eagly & Chaiken, 1993), a short introduction to classical approaches to attitude measurement follows. These approaches parallel debates in social psychological theories of measurement more generally; some central ideas from this broader field of enquiry are outlined to complete the overview of significant social psychological literatures for the thesis.

For a number of reasons these literatures appear to remain somewhat distant from contemporary attitudinal survey research. In an effort to bring them closer, the second half of the chapter presents a synthesis of relevant themes found in them, in the form of a taxonomy of issues and choices which arise in empirical attitudinal research. Some of these issues and choices go unnoticed in survey research because convention renders them invisible. But highlighting such issues enables us to understand the characteristics of particular research methods more clearly. The taxonomy therefore provides a conceptual framework which will inform the presentation of latent variable models and the discussion of the empirical studies later in the thesis.

## 3.1 Theoretical approaches to psychological constructs: attitudes, social representations, opinions and perceptions

In this report I use the term 'attitudinal survey research' to refer broadly to the study of what might be described, colloquially, as finding out what people think and feel about a given topic. This includes what they know about it, how they feel about, and any relevant aspects of their past or potential future behaviour in relation to it. This inclusive approach to a range of psychological constructs is consistent with the theoretical frame of reference of 'social representations' (Moscovici, 1961), which will be elucidated later in this section. It is also, pragmatically speaking, appropriate, since the measurement models developed by early psychologists (described in Section 3.2) are collectively known as 'attitude measurement' techniques, although they were often applied to other psychological constructs.

I also use the terms 'attitudes', 'social representations' and 'opinions' interchangeably throughout the thesis, to refer to elements of 'what people think and feel' about a topic. This runs counter to the popular contention among European social psychologists that each term entails an inextricable link to a different theoretical position; most notably, attitude to an individualistic perspective and social representation to an constructivist outlook. But a number of authors have documented the changing theorisations of the concept of the attitude in social psychology, challenging such narrow classifications of terminology. For example, in the first significant publication on attitudes, Thomas and Znaniecki's (1918-1920) studies of the *Polish Peasant*, attitudes were theorised as individual-level reflections of group-level values. Values in turn were orientations carrying a common meaning for some social group. So the original conception of the attitude was clearly social, even sociological (Jaspars & Fraser, 1984).

Subsequent studies and scholars who came to dominate the field in the mid-twentieth century turned the concept towards the individual level. Doise, Clemence, and Lorenzi-Cioldi (1993) attribute this particularly to the seminal works on attitude measurement by Thurstone, Guttman and Likert, which redirected attention away from orientations shared between individuals and towards differences observed between them. Gordon Allport's well-known definition of an attitude as a 'mental and neural state of readiness' in relation to some object or issue embodies the conception of attitudes as within-individual phenomena (G. W. Allport, 1935, p.8). From this point in the discipline the model of the attitude was variously dissected and integrated into elaborated schemes of related constructs. For example, Rosenberg and Hovland formalised the separation of cognitive, affective and behavioural components of the attitude (Rosenberg & Hovland, 1960). Ajzen and Fishbein extracted the behavioural component from the model to develop theories of attitude-behaviour relations, bringing in associated constructs such as beliefs, evaluations and intentions (e.g. Ajzen, 1991; Ajzen & Fishbein, 1980). Other theorists modelled attitudes and related constructs in systems of consistency (Festinger, 1957), congruence (C. E. Osgood & Tannenbaum, 1955) and balance (Heider, 1958), to name just a few examples. The subtext of classical attitude research is that attitudes have an ontological status as solid at the individual level as that of Durkheim's social facts at the collective level.

Critics of these individualistic theories adopt the alternative term 'social representation' and emphasise their distance from mechanistic models of the attitude. They oppose the classical reification of the concepts of the individual, the attitude and the attitude object. The alternative social representation perspective begins from some of Durkheim's ideas, but takes a different direction. Durkheim spoke of 'collective representations' as values, ideas and norms found at the group level and constituting the relatively stable conceptual framework within which people understand their social surroundings. In a contemporary twist Moscovici (1961) exchanged the term 'collective' for 'social' to create a concept of more fluid representations reflecting the dynamism of late modernity; representations that are shared but continually constructed and reconstructed through social interaction, and continually open to contestation and change. Social representations are not single ideas or values, but *systems* of values, ideas and practices (Moscovici, 1973) which provide the individual as a member of a group with sets of organising principles for understanding and communicating within that world. Thus the term 'social representation' comes to be associated with a social constructionist epistemology.

The term 'opinion' tends to be associated more closely with the classic concept of the attitude than with the freer concept of social representations. Conventionally they are held to share with attitudes the property of being directed specifically towards some well-defined object, and less stable than the more general and deeply held beliefs and values. When opinions from a broad population are sought, the term 'public opinion' is used, and another realm of research and literature is invoked. A topic of interest across a range of social sciences, it is most strongly rooted in political science, and as such entails connotations of normative ideals: from its utilitarian role as an indicator of what is desired by the greatest number to democratic principles of communication and debate in the public sphere. This long tradition of political thought gave way to empiricism around the beginning of the twentieth century (Lazarsfeld, 1957), and it is perhaps from this point that the range of meanings ascribed to the term multiplied. In a paper in the first issue of *Public Opinion Quarterly*, Floyd Allport (1937) aptly describes the many connotations and misuses of the term even at that time – in common parlance, in journalism and in social scientific research.

Whilst acknowledging Allport's fine example of clarifying terminology, I do not adhere to any tight definitions of attitudes, opinions, public opinions, etc. in this thesis. Where I do use these terms, their common sense meanings are implied. The intention is not then to invoke cognitivist notions of the attitude or political notions of public opinion. The less loaded term 'perception' is often used as an alternative, in a spirit of neutrality. However, where I use the language of social representations in the report, it *is* to allude informally to this theoretical framework, since it has many characteristics which recommend it over traditional attitude theories for PUS research (Farr, 1993). Biotechnology is an unfamiliar subject for the average person; as such, it does not constitute an unproblematic attitude object as is required in the classical models of the attitude. In social representations theory, however, the process by which unfamiliar topics are assimilated and integrated into the social world is a central focus, so there is a specific vocabulary for explaining the reception of biotechnology among people over time.

This emphasis on representation as *process* is strong in the theory – and while providing potentially valuable contributions to PUS research, also provides fuel for the anti-survey school within the field, in addition to their objections to surveys outlined in Chapter 2. The methodological critique of existing PUS surveys from a social representations viewpoint is that they are limited to providing only snap-shots of perceptions and no insights into processes of representation, such as might be achieved with qualitative data.

That said, surveys have been championed by some researchers of social representations for their value in exploring representations as *structures* (Doise et al., 1993), which is of central interest in this thesis. A key contribution of social representations for PUS research is to eschew on theoretical grounds a distinction between knowledge and affect. The social process of taking an attitude towards an object implies having some knowledge of it, whilst gaining knowledge about that object involves adopting an attitude or position in relation to it. Through such positioning, aspects of attitudes and knowledge are strongly related to social identities. Identities are mostly spoken of in the plural, because it is assumed that each person possesses a range of social identities, each of which could be linked to a different representation with regard to the same object – biotechnology, for example. These varying social identities enable apparently logically dissonant representations to coexist; a phenomenon termed 'cognitive polyphasia' (Moscovici, 1961). Again, these elements of a social representations approach on the one hand give fuel to the contextualist critiques of survey research in PUS, which object to using responses to out-of-context survey questions as indicators of properly fluid concepts.

Speaking of social identities requires us to speak of groups. Here also, social representations theory provides useful theoretical contributions for survey research in PUS. A group can be defined by a shared representation, and a representation can be defined by the group that holds it. This logical circularity has been censured by some (Potter & Litton, 1985), but I would contend, following Gaskell (1994), that it opens up a valuable conceptual path in survey research, freeing researchers from the straitjacket of only defining groups for comparison a priori, in terms of socio-demographic characteristics. The more flexible and creative definition of groups allows room for fuzzy boundaries and overlaps between groups and representations. Abric (1993) proposes that the structure of a social representation comprises a core of shared meanings and a periphery of variation between people – an particularly appealing idea for researching perceptions of an unfamiliar topic such as biotechnology.

Whilst the theory and language of social representations greatly facilitate the task of talking about people's perceptions of biotechnology, it is not the only theory that could be used. Likewise, Chapter 4 will recall Doise et al.'s (1993) insightful accounts of the resonance between latent variable models and social representations research, but it must be noted that this social psychological framework has no *necessary* or tight connection to the data or statistical models used in this thesis. For this reason I use the language of social representations along with more commonplace terms such as attitude, perception and opinion, to avoid giving the impression that the models privilege one social psychological perspective over another. As the next sections illustrate, models and methods in attitudinal research are not uncontroversial, so some care needs to be taken to avoid attaching a social psychological theory too closely to a statistical model, where such a close link is actually unwarranted.

## 3.2 Classical attitude measurement

The traditional attitude study is quantitative, using numbers, in more or less formal ways, to represent patterns of responses. This approach is rooted in the work of nineteenth century psychophysicists, notably Weber and Fechner (see e.g. Michell, 1999, for a review), who derived mathematical formulae to depict the relationship between actual and perceived magnitudes of differences between pairs of physical stimuli. The formulae they derived produced the first clearly delineable units of analysis for studying perceptions. The project of quantifying psychological attributes then became the dominant drive in the discipline until relatively recently. Danziger (1979) describes the 'positivist repudiation of Wundt' at the turn of the twentieth century, which kitsched Wundt's distinction between experimental Physiologische Psychologie and Völkerpsychologie, the latter of which Wundt argued was not suitable for quantitative investigation. Michell (2003) claims that this 'quantitative imperative' remains a strong impulse today, while others (e.g. Farr, 1996) would qualify such a diagnosis, distinguishing between two broad streams of research in modern social psychology. On one side of the Atlantic is the North American tradition, home to the traditional models of the attitude and heavily quantitative and experimental, whilst on the other is a European tradition which is more qualitative and interpretative, and home to approaches such as that of social representations.

The classic methods of attitude research developed in this North American tradition. The quantitative approach par excellence was Louis Thurstone's application of psychophysical methods to perceptions of social rather than physical stimuli. In his law of comparative judgement (Thurstone, 1927a) he reasoned that any judgement between two stimuli implies the construction of an individual judgement for each stimulus separately, i.e. a basis by which to make a comparison. To take Thurstone's example, if a person is asked to decide which of two types of crime – say, arson and burglary – is the more serious, the process of making the judgement will implicitly involve placing each crime somewhere on a scale of seriousness. If he is asked to do this several times, or to compare each crime with another type, or if other people are asked to make similar judgements, then the place at which each crime is positioned on the scale will vary – it will have a distribution (for his purposes, a normal distribution). The aim of scaling for Thurstone was to reconstruct the scale of seriousness of crimes from these distributions, i.e. to determine the order the items take, from least to most serious, and the relative distance between each one: perhaps some crimes share very similar levels of seriousness, whilst others are more strongly differentiable. The reconstruction (i.e. estimation) could take various forms, depending on which method of data collection was used, which other details of the model were fixed, and what kinds of assumptions were made. The method of paired comparisons (Thurstone, 1927b) employed a group of judges to compare pairs of stimuli; a less labour-intensive method was that of equal-appearing intervals (Thurstone & Chave, 1929) where judges would rate items on a scale rather than make comparisons; the method of successive intervals (Saffir, 1937), combined features of these two.

Rensis Likert is another of the founders of scale construction, though known less for his scaling technique per se (Likert, 1932), and more for his item response format: the five- or seven-point response categories – for example, from 'strongly agree' to 'strongly disagree' – widely used today. Likert himself placed great emphasis on constructing robust scales by checking their unidimensionality and scrutinising the discrimination power of the items used for them. His method employed judges to rate a number of items with reference to some concept, followed by a careful item analysis with a number of goals. Firstly, the analysis would identify items around the neutral, central point of the scale, to be discarded from the question set, since these would not provide any means of discriminating between subjects' viewpoints. Correlation analyses, anticipating Cronbach's (1951) coefficient alpha, and comparisons of average scale scores of judges positioned at high and low points on the posited scale, were used to select items with maximum discrimination power. So item discrimination was crucial, rather than the position of items on a scale (in contrast to Thurstone's approach).

The third distinctive approach to attitude measurement was Louis Guttman's strictly cumulative scale, modelled on dominance relations rather than the proximity relations used in Thurstone's and Likert's methods (cf. Coombs, 1964, elaborated in Section 3.4.4 in this chapter). Guttman's scalogram model is specified such that when items are rank ordered, then for example answering 'yes' to item 5, or dominating item 5, implies having answered 'yes' to items 1, 2, 3 and 4. Accordingly, if we know a respondent's scale score, we should be able to say exactly to which items he responded positively and negatively. In this sense it is 'interlocking', placing both items and respondents on a continuum. Guttman's theory was, strictly speaking, deterministic, but in practice some room was allowed for error, with a reproducibility coefficient calculated to denote the extent to which a respondent's entire response profile was reproducible from his scale score, on average. One of the first examples of this scale was seen in the *American Soldier* studies (Stouffer et al., 1950), in a scale of the frequency with which US soldiers experienced physical symptoms of fear. The process of constructing the scale showed that, for example, if a soldier reported having experienced a severe symptom of fear whilst under fire, such as vomiting, then he would also have reported having experienced all the less severe symptoms – feeling sick at the stomach, shaking or trembling, etc.

To glance ahead to Chapter 4 – paralleling these studies in attitude measurement, latent variable models were being developed, although initially not with attitude research in mind. For example, in the context of rather controversial intelligence studies Thurstone introduced exploratory factor analysis in its modern form (Thurstone, 1935) following Spearman's (1904) original exposition of this idea (see also Gould, 1981/1996). Lord and Novick (1968) and Rasch (1960) introduced key Item Response Theory (IRT) models, in the context of educational testing. Lazarsfeld and Henry (1968) introduced the idea of latent class models, but these were taken up more in sociology than in psychology.

## 3.3 Measurement theory in social psychology

The broad literature on measuring social psychological phenomena can be read against the background of a much older discussion reaching beyond social psychology, regarding what exactly it is that constitutes measurement itself. Definitions of measurement are numerous – from the vague: 'the assignment of numerals to objects or events according to rules' (Stevens, 1946, p.677), to the more precise: 'the attempt to discover real numerical relations (ratios) between things (magnitudes of attributes)' (Michell, 1999, p.17), and the abstract: 'process by which ... infinitely varied observations are reduced to compact descriptions or models that are presumed to represent meaningful regularities in the entities that are observed' (Judd & McClelland, 1998, p.181). The common thread running through them, however, is a conviction in the importance of measurement as part of the process of understanding a phenomenon. Again, memorable quotations abound on this point; for example, 'Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality' (Thorndike, 1918, p.16).

The difficulty surrounding the use of numbers as a means of understanding phenomena hinges around the question, 'what do the numbers mean?'. A fine introduction to this problem is given in a short paper by Lord (1953), in which he tells the story of a professor who calculates means and variances of sets of numbers on football shirts, despite the numbers being purely labels and having no intrinsic meaning with regard to the football players wearing them. Lord's point is that making use of properties of the numerical world is inappropriate when there is really no link between these and the properties of the empirical reality they are taken to represent. As Hand (2004) clarifies, it is perfectly sensible *mathematically* to compute the mean of the shirt numbers assigned to a football team ('since the numbers don't remember where they came from, they always behave just the same way, regardless', Lord, 1953, p.751), but this mean tells us nothing about the *reality* it is representing.

Adherents to the 'representational' measurement approach argue that the use of numbers requires a formal and tested link between the phenomenon studied and the numbers used to represent it. This approach owes its formal development to theorists such Helmholtz, Holder and Russell, working at the turn of the twentieth century (Hand, 2004). It is best known to psychologists via *Foundations of Measurement* (Krantz, Luce, Suppes, & Tversky, 1971), which sets out two key elements in any representational measurement: an empirical relational system (ERS) and a numerical relational system (NRS). An ERS is a subject of study or observation which consists of objects or elements and specified relations between them, involving one or more attributes of interest. Deriving a measurement of an ERS amounts to finding a NRS that maps on to the ERS. In practice, finding a complete mapping is unlikely (Stevens, 1959), but it should in principle be homomorphic – that is, essentially structure-preserving. The set of possible NRSs we might use to represent an ERS are related to each other in terms of admissible transformations, a concept well known by psychologists in the form of Stevens' classification of levels of measurement: nominal, ordinal, interval, and ratio. Conjoint analysis (Luce & Tukey, 1964) is the prototypical method based on the idea of representational measurement, in which additive scales are derived from order relationships between objects on a set of attributes.

In 'pragmatic' measurement, by contrast, the link between the ERS and NRS is looser. In Hand's classification, the essential feature of pragmatic measurement is the attention given to practical convenience and utility in the resulting variables – to adjustments and modifications made to the variables that have nothing to do with the ERS in question (Hand, 2004). Whilst representational measurement should involve the minimum possible element of creativity, pragmatic measurement can, at its extreme, constitute a procedure for *defining* variables. In this sense it echoes an operationalist perspective (Bridgman, 1927) in which an attribute of interest is defined by the operations by which it is measured. It is not vital here to go into details of the difference between pragmatism and operationalism in measurement, or between pragmatic measurement and other terms used by other authors to denote similar approaches, such as 'index measurement' (Dawes, 1972) and 'psychometric measurement' (Judd & McClelland, 1998). For the discussion here the important distinction is between representational and pragmatic approaches. The former implies demonstrating a tight link between an ERS and its corresponding NRS, while the latter implies a looser connection, and a measurement strategy governed significantly by the practical requirements of the researcher. These are ideal types of course, and most measurement exercises will involve elements of both approaches (Hand, 2004).

In contemporary quantitative psychological research, some still regard representational measurement as the only proper way to carry out empirical research. For measurement theorists such as Michell (1999) and Barrett (2002), strict representational measurement of concepts (using conjoint analysis and related methods) must be established before they are used in any quantitative analyses of substantive research questions. Very few researchers using attitudinal survey data adhere to a representational line such as this. They are much more likely to adopt a looser conceptualisation of measurement, and a pragmatic approach. A further alternative but unfortunately common position regarding measurement is to ignore the concerns about measurement entirely. All three perspectives are open to criticism from those engaged in more interpretative approaches to social psychology. Broadly speaking the first is seen as unrealistic, the second as undisciplined and the third as irresponsible. It is partly as a result of trying to find my own position amongst these disparate outlooks that the taxonomy presented here emerged. Its purpose is described in the next section.

## 3.4 A taxonomy of choices in researching attitudes

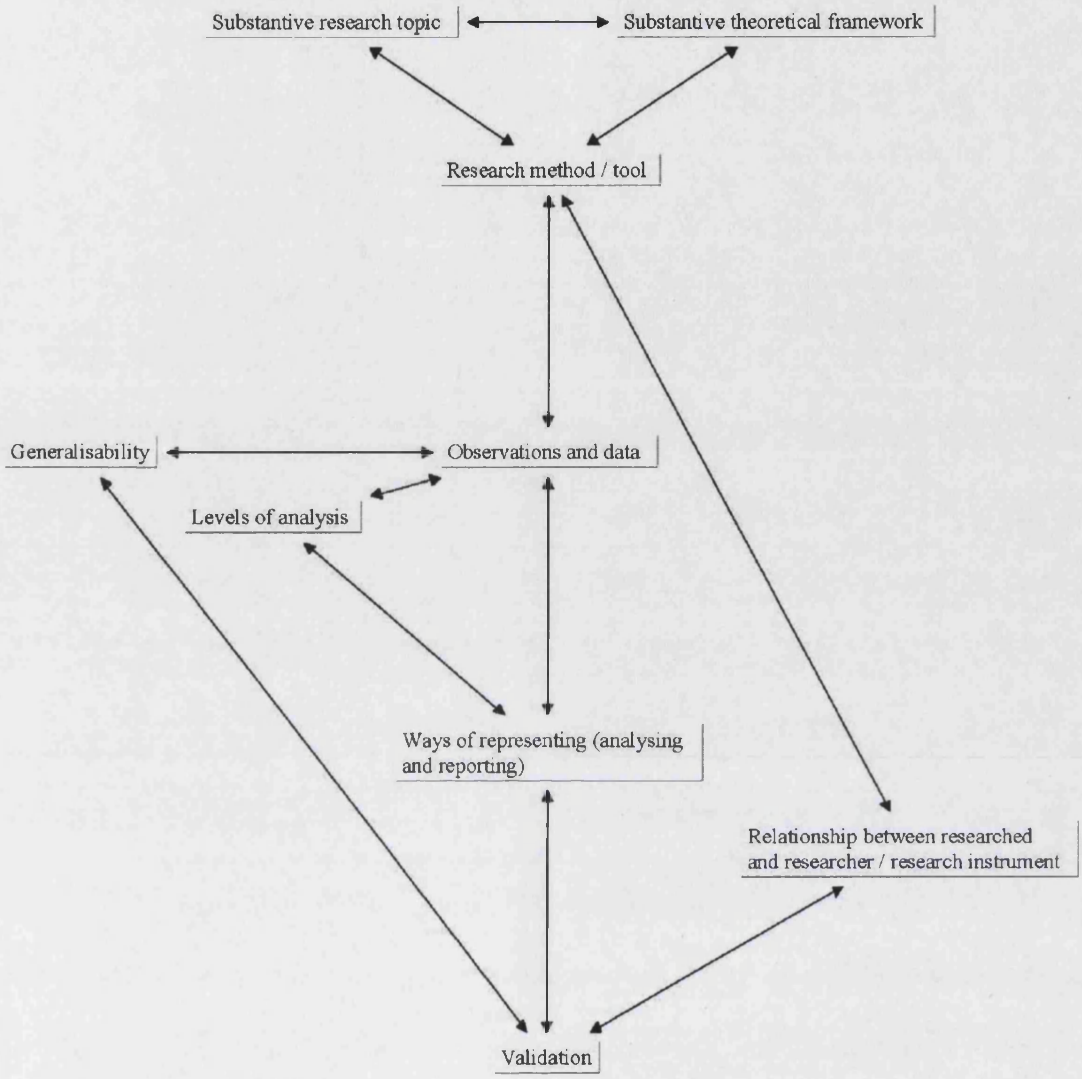
The taxonomy is presented in Figure 3.1. The elements of the scheme are headings representing sets of questions or choices, with links drawn between them where there seems to be a strong logical, theoretical or methodological connection. The intention is that from this framework, any piece of work could be described in terms of each of the themes. As Dawes and Smith (1985) note, researching attitudes is all about choices. In some cases, making these choices involves selecting from a number of clear, distinct and discrete possibilities. In others, it involves trying to untangle a number of interconnecting and ambiguous issues. In others still, practically speaking it involves very little, if any, reflection on the possible choices. Dawes and Smith (1985) further comment that often researchers are oblivious to the fact that choices exist, because certain phenomena have been studied or measured in the same way for so long that conventional wisdom has rendered the choices invisible. Gigerenzer (1991) goes further to suggest that even theories themselves may derive from the dominant research tools of the day. The more transparently we can understand our methods, then, the better. The intended purpose of the taxonomy is to make explicit the main choices or issues involved in attitudinal research – choices which represent both constraints and opportunities. Explicitly considering the constraints and opportunities implied by different approaches to empirical research could provide a valuable way of deciding on a research design for a substantive research question, or on the best way to continue researching from a particular point within an existing research project. The scheme will be used in this way in the discussions in Chapter 9.

The taxonomy, while having been developed initially with attitude research in mind, seems to have a broader application to social research than just to attitudinal surveys. In the description below I attempt to demonstrate this by giving a few examples from qualitatively as well as quantitatively oriented research.

The sub-sections provide explanations for the headings given in Figure 3.1. The order in which the headings appear in the diagram reflects in a very approximate way the logical order in which a survey researcher planning a project might encounter these questions. However, the location of the items is not intended to be strongly symbolic – the figure exists mainly to lend some narrative and ease of comprehension to the text. In the diagram, the three headings at the top – the substantive research question, theoretical framework and research method – are those that would usually define a research project. These three headings need very little explanation as they are largely self-evident to any researcher, and have been documented widely and in great detail in a host of research methods text books – so they are discussed only briefly.

The headings forming a vertical line underneath are slightly more involved. These are the main questions that would usually follow the specification of a research project, and define the research method(s) to be employed in it: that is, what data to collect, how to analyse them and how to validate the findings they suggest. The literature on measurement can add some insights to the common sense interpretation of these themes. Finally the three headings to either side of the main vertical line are additional considerations, the importance of which will vary from project to project. These are of particular interest for the empirical studies in the thesis, in the ways that they touch on measurement issues, and the challenges to survey research outlined in Chapter 2. They will be discussed in more technical detail in Chapter 4.

**Figure 3.1 A taxonomy of choices in researching attitudes**




### *3.4.1 Substantive research topic*

Most research begins with a concrete question of some sort. It may be fairly broad (for example, 'What is the nature of public opinion on biotechnology in the UK?') or very specific (for example, 'Do higher levels of scientific knowledge among UK females correspond to greater levels of concern about the risks entailed in therapeutic cloning?'). It may entail explanation, prediction, and/or description. It also sets out the parameters of the project, specifying some phenomenon and some population of interest. An example of the former might be attitudes or beliefs concerning a given topic, and an example of the latter might be the general public in one country, an interest group around an issue, female single mothers in full time work, etc. So there is a 'what' and a 'who'. These two elements may be distinct, and may define the boundaries of a project from the outset. In some studies, however, they may be theoretically and empirically intertwined. In such scenarios it may be a research objective per se to reach a description or understanding of these two elements. One polemical stream of PUS research, for example, seeks to deconstruct and problematise the notions of public, understanding, and science, claiming that existing objectifications of these concepts serve to hold the locus of control in the scientific sphere, at a distance from non-scientists (Felt, 2000).

### *3.4.2 Substantive theoretical framework*

Arguably, any social research project involves an element of substantive theory. In social psychology, for example, a study might be built specifically around Fishbein and Ajzen's (1972) model of attitudes and behaviour, or Tajfel's (1981) social identity theory, or Moscovici's (1984) theory of social representations. By contrast, and especially in social research conducted in a non-academic setting, there may be ostensibly no theoretical element to a research project. However, from a certain perspective it might be contended that even then, *some* substantive theoretical framework would be present, even if only tacitly. This point comes easily to social psychologists, in the sense that social researchers are at a fundamental level simply people studying people. As such, they draw on the knowledge they possess about the social world, and the resulting scripts, schemas and heuristics which make navigating a path through everyday life possible (Garfinkel, 1967; Schank & Abelson, 1977; Tversky & Kahneman, 1981). This applies as much to them and as it does to their research subjects – and to the interaction of both parties during the research process. A strong version of this argument would then be to claim that it is specifically *social psychological* theory which is omnipresent in social research. Pressing this argument is not important or necessary here – the point is made only to illustrate the idea that any research project could be positioned somewhere on a continuum of implicit to explicit adoption of one or more substantive theories, social psychological or otherwise.

Where a theory is actively employed, it may play a central or peripheral role in a project. In some scenarios, data may be collected specifically to test aspects of a theory, or to explore a phenomenon through the perspective of a theoretical orientation. In others, there may be only a loose fit between theory and data, where the theory chosen lends itself to interpretation of data but is not the *only* theory that could be used. The latter scenario applies to the empirical studies in this thesis, as to many similar studies in social psychology. My theoretical orientation to the interpretation of the data belongs to a social representations approach, but the data could alternatively be given an interpretation that fits with the classic model of the attitude, for example.

### *3.4.3 Research method or tool*

By a research tool I mean a survey, or set of interviews, or piece of participant observation, to name a few possibilities. In the context of the wealth of literature already existing on the subject of research design and research methods (see e.g. de Vaus, 2006), this theme needs little elaboration here – suffice it to say that the choice of research method should be closely linked to the substantive research question, and involves broadly two choices: how to select participants or research subjects, and what mode of data collection to use. In terms of selecting participants, the possibilities range from a complete enumeration, through the various types of probability and non-probability sampling (see e.g. Kish, 1965). These possibilities are cross-cut by the willingness of the people being studied; a census, in principle, requires universal participation whereas in ethnography and action research the researcher may have little control over who becomes a research subject.

In terms of data collection mode, the choices are numerous, and again, closely linked to the substantive aims of the research. To return to the example of distributions of opinions on a popular topic, the standard approach would be to administer a questionnaire. In exploring emerging currents of opinion on a little-known topic, focus groups would be well indicated. In ethnography and action research, multiple data collection modes would usually be needed, and a high degree of flexibility and spontaneity on the part of the researcher.

The great range of possible methods varies, for example, in terms of the extent to which each entails personal contact with subjects (internet surveys being remote, in-depth interviews often relying heavily on personal rapport), intervention (covert observation contrasted with action research), and formality of structure (experiments in contrast to participant observation). It is common and often commendable for more than one method to be drawn upon in a single research project. For example, this thesis concerns large-scale random-sampling survey research, but the data used form part of a multi-strand research project on *Life Sciences in European Society* (see e.g. Wagner & Hayes, 2005, p.347) including mass media monitoring and policy analysis. Exactly how to integrate data from 'mixed methods' is another question altogether, and the subject of a significant new concerted effort in social research (Tashakkori & Creswell, 2007).

### 3.4.4 *Observations and data*

In making a distinction between observations and data I draw on Coombs (1964). 'Observations' thus refers to information collected in a research project, in its rawest possible form: a tape recording and interviewer notes from an interview for example; unabridged survey responses; field notes and recollections of a participant observer – broadly, the immediate outcome of applying a research method. 'Data' comprise such information in the format in which they are analysed.

In his *Theory of Data* (1964), Coombs defines data as 'relations between points in space' (ibid., p.1). He goes on to propose a classification of types of data based on three binary distinctions. The first specifies whether the data comprise pairs of points or pairs of dyads. The former might result from asking questionnaire respondents to give answers to a set of items one by one. This is the norm in social surveys. Pairs of dyads are less commonly used in social surveys, and more commonly used in market research; these are produced by some comparison task, such as asking respondents to compare pairs of products. Coombs' second distinction specifies whether in the data there are one or two sets of points. To take again the examples mentioned, giving responses to items one by one implies two sets of points (respondents and items), while carrying out a paired comparisons task results in data from one set of points (items compared with items). These two distinctions result in a four-way classification of basic kinds of data. 'Single stimulus' data depict relations between pairs of points from different sets – this is the type of data most commonly produced from social surveys, as mentioned above. 'Preferential choice' data describe points from different sets, but in terms of relations between pairs of dyads. An example of this might be taken from market research, where respondents are asked to choose their favourite from pairs of stimuli. 'Stimulus comparison' data result from one set of points (e.g. items only) and the relations between pairs of points – an example of this might be the data derived from the development phase of a Likert scale. Lastly, 'similarities' data are derived from one set of points, and the relations between pairs of dyads, such as those produced from a paired comparisons exercise.

Each of these four classes of data may be further subdivided according to Coombs' third distinction: whether the relations between points or dyads are those of proximity or order. To take again the example of the single stimulus data analysed in this thesis, the attitudinal items appearing in Chapter 5, with Likert-type 'definitely agree' to 'definitely disagree' response categories in relation to statements such as 'GM food is useful for society', produce proximity relations. That is, respondents choose the category that most closely represents their opinion, with no information about the direction of the proximity contained in the response. Thus, for a person answering 'tend to agree', we have no way of knowing if his or her true position is a little more positive or a little more negative than the position implied by that answer. By contrast, the knowledge items analysed in Chapter 6 represent order relations, or dominance relations in the language of Guttman. A respondent theoretically needs to possess a certain amount of knowledge to answer a question correctly; answering correctly implies possessing knowledge at or above a critical threshold, which enables the respondent to 'dominate' the item.

Coombs' classification appears to have had limited uptake amongst quantitative researchers of attitudes and opinions. One reason for this may be that the generality of the scheme is lost on a survey research field that mainly generates single stimulus data. However, it is possible to translate data forms, as Coombs shows. Single stimulus data of the kind in the Eurobarometer can be transformed into other types of data, and indeed often are, as a part of standard analyses. For example, a correlation matrix of responses to a set of continuous items, or a contingency table of responses to a set of categorical items, are both types of similarities data. These might be reported as they are, or used as inputs in various multivariate analyses, notably including latent variable models. Generally, latent trait and class models used in this thesis begin from contingency tables, and factor analysis from a correlation or covariance matrix.

Coombs' point that one set of observations can be converted into different types of data not only increases its attractiveness for survey research but also raises the valuable insight that data are not straightforwardly *given*, as their name would imply, but involve an element of creativity on the part of the researcher. Whereas in some studies there is little or no appreciable distinction to be made between observations and data (for example where a set of survey responses are the very ones used as input in an analysis), in others, the data production process is more elaborate. For example in an interview, from the immediate observations of the actual dialogue between interviewer and interviewee, normally a tape recording is produced; from this a transcription is made; a coding frame is then applied to the transcription and the frequencies and patterns of codes are analysed to reach an interpretation of the conversation. In this process the data undergo several phases of transformation as part of the analysis itself: creation and analysis of data are combined, particularly in the coding stage.

### *3.4.5 Ways of representing (analysing and reporting)*

There are two senses to this theme: how to represent data to oneself as a researcher, i.e. how to analyse; and how to represent findings to an audience. Sometimes these will be synonymous, while sometimes they will constitute two distinct steps in a research project. Deciding how to present the results of a complex quantitative analysis to a lay audience, for example, can involve some difficult choices about which information is included and excluded, whether results can be summarised graphically, etc.

Depending on the piece of research, concerns under this heading may be technical or conceptual to varying degrees. On the conceptual level, the issue is straightforwardly, what is the nature of the representation we are creating? Perhaps it consists entirely of prose, with various narrative and rhetorical characteristics. Perhaps it involves images – such as illustrative photographs, or diagrams. In social anthropology there is a conventional way to depict kinship structures diagrammatically, for example. If it involves numbers, then it raises the variety of possibilities outlined in Sections 3.2 and 3.3 of this chapter, along with the technical issues and debates surrounding them.

Some examples of different possibilities illustrate the many conceptual choices involved in numerical representations of data. Guttman's scalogram model implies a depiction of the response process as an order relationship between respondent and stimuli such that respondents answering 'yes' to a certain stimulus are said to dominate that item. The resulting scale orders items along a line according to the degree to which they possess the attribute they are measuring. When the test is administered to a sample, the respondents are likewise ordered along this same line. The scale is only ordinal: distances between items or between respondents cannot be determined. The distances between elements may however have meaning; this is the aim in Thurstone's equal-appearing intervals, for example, and in conjoint analysis. By contrast, even the concept of relative location may have no place in the representation of the items, such as in the typology of logics of support and opposition for biotechnology, which is purely a classification of different types of opinions. So the spatial representation possible echoes S.S. Stevens' (1946) classification of levels of measurement. Elements of these may be combined in any one analysis. For example, hierarchical cluster analysis involves grouping items or respondents (a nominal summary variable), and combining sub-groups at different stages of aggregation (order) to form larger groups (new nominal summary variables), making for a dual process of both grouping and ordering.

The choice of which numerical representation to use to describe a set of data entails the issues and debates introduced briefly in Section 3.3. In most social science analyses the mapping between the ERS and NRS is weak, and pragmatically chosen, and invokes a range of means for assessing the degree to which the choice of NRS is reasonable, within the bounds of conventional social science thinking. The debate over the proper way to model categorical observed items in latent variable models (explained in detail in Chapter 4) also falls under this heading. The telling critiques by Pardo and Calvo (2002) cited in Chapter 2 are a good example of this issue as applied to quantitative research in PUS. These might all be reframed as issues of validation – see Section 3.4.9.

### *3.4.6 Generalisability*

There are two elements within this theme: generalisability of substantive *findings* – to a broader population; and generalisability of the research *instrument* employed – to other research subjects, perhaps even to those from a different conceptual population. To illustrate the first, we might for example build a statistical model from a sample of data, and from it make an inference of the findings to the broader population from which the sample was drawn. By contrast, the research may be a description solely of the people studied, and no generalisation made. So there is a formal distinction between studies that make use of inferential statistics and those that do not – but also between these are many shades of grey, which are the subject of a rich literature (e.g. Campbell & Stanley, 1963; Chambers & Skinner, 2003).

Within the second element, the generalisability of the research instrument, a distinction can be made between the dominant classical and contemporary approaches to attitudinal research. Attitude measurement literature in the classical vein, as described in Section 3.2, means developing an attitude scale by asking a set of judges to respond to a large pool of items, by selecting from this pool the collection of items that best represents the concept of interest in the way required, and for fine-tuning the scale according to particular theoretical and technical specifications. The tool is only administered to the sample of interest when it has been shown to possess certain characteristics – for example, equal-appearing intervals between items (Thurstone & Chave, 1929), or a strictly cumulative relationship between items (Stouffer et al., 1950). These carefully developed scales are then taken to be fit for re-use with other samples, that is they are taken to be transferable to other groups of subjects, even in other research projects.

This approach is still widely applied in certain fields of psychology, such as in clinical psychology, which employs a range of established psychiatric rating scales. In attitudinal survey research, by contrast, it is rare to invest so much time in scale preparation, often due to resource constraints. A more common procedure for constructing scales is to do so post hoc from the target sample, using statistical models such as latent trait or factor analysis models. The nearest to checking the generalisability of the research instrument is then the pilot test, which is undertaken with varying degrees of thoroughness. At a minimum, cognitive interviewing should be carried out for interview topic guides and questionnaires purely to check the intelligibility of questions used. More thorough pretesting involves, for example, piloting a questionnaire on a small sample and running some detailed analyses of questionnaire items to assess item functioning. Again, in attitudinal survey research this is not a common luxury, but it is hoped that the latent variable models applied to Eurobarometer data in this thesis will be seen as useful tools in scale development.

One of the key challenges in PUS surveys is the comparability of items and scales cross-nationally. The analyses in the empirical chapters of this thesis provide a contribution to assessing the transferability of composite indicators across countries. These analyses are limited to assessing whether items 'function' similarly in relation to the construct being modelled – as described in Chapter 2, this is a purely quantitative assessment of the 'generalisability' of the measures, and in itself an incomplete assessment of the transferability of the measures. More details of the quantitative strategy adopted will be given in the next chapter (Section 4.1.8).

### *3.4.7 Levels of analysis*

One or more levels of analysis may be involved in any research method. For example, experimental psychology studies will tend to be framed at the level of the individual, focusing on intra-individual processes. In more interpretative social psychology, by tradition the focus is on the individual in his or her social context; sometimes the individual-group relationship and sometimes the inter-individual relationship. For sociologists the group level of interest is society, so that the focus is on societal-level processes and the relationship between the individual and society.

In a conceptual sense the level or levels of analysis are not always clearly articulated in a research project, and sometimes there is disagreement between academics regarding the actual and preferable level of analysis. Survey research is often criticised by qualitatively inclined social psychologists for extricating respondents from their social contexts, and extracting information from them in an artificial setting, which cannot capture the complexities of their experiences with the substantive topic. Survey researchers might respond that this is not their intention at all, and that the survey research process should be considered as a number of communication acts between the researcher and subject (Dillman, 2000).

The connection between individual-level and group-level findings is particularly interesting and elusive in social psychological research. An example cited by Doise et al. (1993) illustrates the conundrum. In a study of political values, Kerlinger (1984) applies a factor analysis to regular between-subjects survey data, and derives from it dimensions of conservatism and liberalism. These dimensions undoubtedly describe patterns of associations at the group level. But he finds, further, that regardless of their own political convictions, when asked to recall items, respondents do this more successfully for those items strongly linked to the conservatism and liberalism factors than those items weakly linked to them. These items, which define the factor structure, resonate more strongly with respondents than the peripheral items. He comments, 'One cannot help but wonder … whether a certain attitude factor structure found from the responses of many individuals has some sort of representation in the cognitions of individuals' (ibid., p.229). He reasons that the process of taking one's particular position in relation to issues or objects involves taking stock of the positions of other people. That is, it must invoke processes of social identity (Tajfel, 1981), and effectively a 'symbolic interiorisation' of group-wide representations.

This seems a plausible theory in the case of a well-known topic such as political persuasion, but it would be heroic to assume that it would stand up to empirical scrutiny universally, and especially in the case of an unfamiliar topic (such as biotechnology). It implies that the individual has a broad knowledge of significant others, both out-group and in-group members. Where this is not present, and bearing in mind the systematically different ways in which people perceive in- versus out-group members, the chances of this process in aggregate leading to common representations seem slim. To clarify the distinction between representations at different levels of analysis, Gaskell (2001) invokes Harré's (1984) distinction between 'distributed' and 'collective' representations. Whereas in distributed representations every individual holds the same representation, in collective representations every person holds part of a representation, but as a whole it is only realised collectively[1]. This is a very useful conceptual distinction which is often lost or glossed over in attitudinal survey research. Some further comments on it will be made in the next chapter (Section 4.2.3).

### *3.4.8 Relationship between the researched and the researcher/research instrument*

This theme refers in a general sense to what happens during the research process, and the extent to which it is a prominent issue for the researcher. At one end of the continuum, for example, attitudinal survey research using Eurobarometer-type data relies on the assumption that respondents have a common understanding of the survey items and choose their answers using basically the same response process. In the language of Harré, this amounts to assuming distributed representations of meaning and processes of responding to questionnaires. Where this is not justified, it is a problem, and raises concerns about the quality of the data. Qualitative ways of investigating these concerns include cognitive interviewing. Quantitative approaches include survey experiments and quasi-experiments; for example, since 1991 the Eurobarometer on biotechnology has used a split ballot to ask about expectations of biotechnology. In one half of the survey, respondents are asked about 'biotechnology', and in the other half they are asked about the alternative term 'genetic engineering'. To varying degrees across countries, responses to the term 'biotechnology' remain more optimistic, on average, than to the traditionally negatively loaded term 'genetic engineering' (Gaskell et al., 2006). As I have mentioned already, varying meanings of words constitute one of

[1] Thus Harré's and Durkheim's collective representations are not quite the same phenomena.

the greatest obstacles in survey research, and this obstacle is only enlarged in comparative research, where different languages complicate matters further. In this sense, the theme of the relationship between the researched and the research instrument is often closely connected to the theme of generalisability. Qualitative ways of addressing these joint concerns include careful back-translation of questionnaires. A quantitative contribution which is not commonly used but which will be employed in the empirical chapters in this thesis, is to compare the statistical behaviour of sets of items between countries.

It is not just survey researchers who worry about possible biases induced in data by the presence and nature of the research instrument; this is a universal concern in social research (see e.g. Rosnow & Rosenthal, 1997). More obtrusive examples include the well known Hawthorne effect. But the relationship between the researched and the researcher is not always purely a technical nuisance, obscuring substantive findings. In action research, ethnography and other participatory research exercises, the relationship between researched and researcher is a defining feature of the project, and brings with it a number of different concerns with more strongly political and ethical emphases – issues such as accountability, privacy, and the avoidance of harm. Ethical concerns more broadly are a fundamentally important element of this theme.

### *3.4.9 Validation*

Essentially this comprises efforts to reassure the audience as well as the researcher of the credibility of the research project and findings. A great deal of social science literature is devoted to the concept of validity. Here I highlight just a few relevant points for this thesis. The first is to note that most of the existing literature concerns validation for quantitative studies, particularly in the sense of how confidently the results of analyses and models can be read – for example, does an experiment provide convincing evidence of a causal relationship between X and Y? A number of classifications of different types of validity have been proposed within this frame of reference (e.g. Shadish, Cook, & Campbell, 2002), alongside more general discussions of the broader concept of validity in the context of social research (Gaskell & Bauer, 2000).

In this thesis, validation is concerned with the narrowly focused issue of the validity of the measures created of the three PUS constructs. This relates to the idea of 'construct' validity, in the sense of Shadish et al. (2002): that is, the extent to which the observed items represent the more general construct or concept that they are intended to capture. This term 'construct validity' is itself used in a number of different ways in the literature, depending on the theoretical position taken on measurement; thus, the theme is closely tied to the theme of ways of representing. Dawes and Smith (1985) for example note that those devoted to representational measurement define a valid measure as one fulfilling certain axioms of quantity. In pragmatic measurement, validation is couched in terms of demonstrating convincing patterns of associations with external measures, of the same or of other constructs. Judd and McClelland (1998), for example, using the term 'construct validity' in a different sense from Shadish et al., define it as comprising convergent validity (specifically: is the measurement associated with other constructs according to theoretical expectations?), discriminant validity (does the measurement capture the construct of interest and *only* that? – that is, are the required patterns of independence with other variables observed?), and reliability (does the research instrument work in a consistent way between subjects?).

Contrasting with the multitude of possible techniques for assessing validity in quantitative research, validation in qualitatively-oriented research is a much less formally developed field. In fact a significant body of thought within qualitative research rejects the notion of formal validation procedures as redolent of the inherent positivism they dislike in quantitative research. Gaskell and Bauer (2000) contend that qualitative research requires its own quality criteria, functionally equivalent to but not straightforwardly lifted from quantitative methods. They see these quality criteria as necessary for public accountability, making an interesting connection to the theme of the relationship between the researcher and the researched.

In some studies validation is an ongoing issue, arising early in a research project via questions about the quality and properties of measurement instruments used. In others it is a more post-hoc exercise, undertaken as an appendix to the administration of the research method and analysis of data. Shadish et al. (2002) make the useful point that the concept of validity implies a dichotomy: valid versus invalid – although in practice, validity is not an attribute whose existence can be demonstrated indisputably.

## Conclusion

This chapter has given a brief overview of some relevant themes from the sizeable literatures in social psychology concerning attitudes and related constructs and their measurement. It began with an explanation for the catholic approach to terminology adopted in the thesis – the neutral spirit in which the terms attitudes, perceptions, opinions, etc. will be used, and the generality of the term 'attitude measurement'. It went on to convey, via a selection of examples, an impression of the wealth of literature on attitude measurement in social psychology, to stress that in addressing the methodological challenges of the thesis, I by no means start from a blank slate.

In an effort to organise the useful elements from this literature, the taxonomy of choices in attitudinal research was designed to draw out some key themes to help structure the discussions of the empirical findings from the case study data. The content of this chapter has remained general, abstract and conceptual. The next chapter introduces the particular statistical models that will be used in the empirical studies, beginning with technical, specific details, and followed by a conceptual discussion framed around the headings from the taxonomy in this chapter.

# 4 Latent trait and latent class models in attitudinal survey research

This chapter introduces the statistical models that are the methodological focus of the thesis. Latent variable models are many and diverse (cf. Bollen, 1989; Skrondal & Rabe-Hesketh, 2004; van der Linden & Hambleton, 1997), and the chapter thus covers only those actually employed in the empirical studies, although some references are made to related models where relevant. Section 4.1 contains the technical details of the models, taking as its point of departure the widely used normal linear factor model, and noting the difference between this and the models used in the thesis, which specify logistic models for the relationships between latent and manifest variables. The specific models applied in the empirical studies are then presented in turn: models for continuous latent traits, latent classes, and discrete latent traits. These are given for nominal observed variables, followed by a short description of the models for ordinal observed variables, which make a brief appearance in Chapter 7.

Having outlined the basic models, more precise details are given about the way in which the cross-national analyses are approached, clarifying the types of restrictions which may be applied to parameters of their measurement models. The essential features of the estimation method are outlined next, followed by an explanation of the diagnostic statistics employed for model selection purposes. The final two parts of Section 4.1 outline the calculation of two pieces of further output: posterior scores and class allocations derived from the models selected, and weighted estimates of the distribution(s) of latent variable(s), using sampling weights.

Section 4.2 is a conceptual discussion of the latent variable models applied in the empirical studies. It uses the taxonomy of the previous chapter to highlight a selection of features of the models which are of special interest. Some of these relate to attitudinal survey research generally, and some to the particular challenges of survey research in PUS which were outlined in Chapters 1 and 2. In the terminology of Chapter 3, latent variable models comprise a family of ways of representing data, so Section 4.2.1 begins by discussing the considerations involved in choosing between different types of latent variables to represent a set of items. Recalling the range of philosophical stances towards measurement that were sketched in Chapter 3, such a choice is quite open if one adopts a pragmatic approach to measurement, but restricted if one adopts a representational approach. The means of validation for a finally selected model are also quite different depending on whether one adheres to a representational or a pragmatic approach to measurement, though they rely heavily on the same or similar fit statistics. These issues are somewhat reflective and would not often be contemplated in the context of practical survey research, where the typical survey researcher will by default operate within a pragmatic frame of reference.

Section 4.2.2 takes a more practical turn, clarifying the ways in which latent trait and class models will be used in this thesis as tools to address the particular methodological challenges posed by PUS surveys. It is helpful to reframe these challenges in the terminology of the taxonomy. The problem of capturing complex constructs, and the question of how to analyse DK responses, can be thought of as questions regarding the relationship between the researched and the research instrument. The task of making valid comparisons between countries is a question of the generalisability of measures of constructs. The chapter concludes, in Section 4.2.3, by reflecting more broadly on the place of latent variable models within the field of social psychology, and on their current and potential use in attitudinal survey research.

## 4.1 Statistical specification for the latent trait and latent class models used in the thesis

Latent trait and latent class models may be thought of most simply as regression models with multiple observed response variables and a smaller number of unobserved explanatory variables. These models are variants on the General Linear Latent Variable Model (GLLVM) (Bartholomew & Knott, 1999), and as such are essentially generalised linear models (e.g. Dobson, 2002) but with latent rather than observed explanatory variables. The GLLVM, in turn, is part of the more general family of Generalised Linear Latent and Mixed Models (GLLAMMs) (Rabe-Hesketh, Skrondal, & Pickles, 2004). The connections between different types of latent variable models, and between latent variable and other models, are clarified in texts such as Bartholomew and Knott (1999) and Skrondal and Rabe-Hesketh (2004). I refer the reader to these sources for full details of the models introduced in this chapter; just the essential details of relevant models are presented in the sections below. This presentation relies heavily on the unified treatment of latent variable models given in Bartholomew and Knott (1999), and on the technical manual for the Latent GOLD software (Vermunt & Magidson, 2005), which is used to estimate most of the models in the thesis. I follow Bartholomew and Knott particularly closely in their style of notation. It should be emphasised that although this may look a little different from some treatments of the subject, there is nothing theoretically 'alternative' in the models used in this thesis – they are the standard latent trait and latent class models.

### *4.1.1 Variables*

The following conventional notation is used to denote the types of variables involved in the models:

$x_j$ (j=1,...,q) are $q$ latent variables (denoted alternatively by the vector $\boldsymbol{x}$),

$y_i$ (i=1,...,p) are $p$ observed or manifest variables (or vector $\boldsymbol{y}$), and

$z_k$ (k=1,…,r) are $r$ observed covariates; for our purposes group (country) variables.

Observed and latent variables may be categorical or continuous; the various kinds of latent variable model are defined according to whether latent and observed variables are the former or the latter. The basic classification of types of models is shown in Table 4.1. Here the term 'categorical' includes models for binary variables, and for nominal and ordinal polytomous variables, whether observed or latent. The earliest latent variable model, factor analysis, which is still the most widely used model in attitudinal survey research, was conceived for scenarios where both the latent and the manifest variables were continuous. Subsequently, however, models have been developed for all other possible combinations, including models for mixtures of different levels of measurement – for example, latent trait models where some observed variables are discrete and some are continuous (see e.g. ch.7 of Bartholomew & Knott, 1999).

**Table 4.1 Types of latent variable model**

| | | **Manifest variables $y_i$** | |
|---|---|---|---|
| | | **Continuous** | **Categorical** |
| **Latent variables** $x_j$ | **Continuous** | Factor analysis | Latent trait analysis |
| | **Categorical** | Latent profile analysis | Latent class analysis |

Adapted from Bartholomew and Knott (1999, p.3)

The focus of this thesis is on latent trait and class models, since all of the observed variables in the three empirical studies are categorical. Just as with regression models, when the response variables are categorical a linear link function between observed and latent variables is not appropriate, and an alternative must be used. We can, for example – as in this thesis – model the logarithm of the odds of responses falling in particular categories – equivalent to binary logistic regression for binary response variables, or multinomial logistic regression for nominal response variables. This key point is often paid scant attention in survey analysis, and one will often see factor analysis used for analysing categorical observed variables. The next section explains more fully why this is problematic, providing a point of departure from which to describe the trait and class models used in the thesis.

### *4.1.2 The normal linear factor model*

The normal linear factor model (see e.g. Bartholomew, Steele, Moustaki, & Galbraith, 2002; Bollen, 1989) is the best known of the latent variable models, and is easily and often implemented with standard software packages such as SPSS. Using the notation introduced above, i.e. for $y_i$ continuous response variables, and with up to $q$ latent variables denoted by $x_j$ and error terms by $e_i$ the model is given as:

$$y_i = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} x_j + e_i \qquad (1)$$

where $\alpha_{i0}$ is the intercept or constant, and each $\alpha_{ij}$ gives the loading for item $i$ on latent variable $j$. Usually (though this assumption can be relaxed) these latent variables or factors are assumed to be independent of each other. A core assumption of the model is that the observed items are conditionally independent, given the latent variable(s) – that is, that the explanatory variables account for all the association between them. The explanatory variables $x_1,..x_q$ are assumed to follow a multivariate normal distribution, as are the error terms. These imply that the response variables $y_1,..y_i$ are also multivariate normal.

The model presented here is known as the exploratory factor analysis model; here, loadings are free to vary between items, and between factors, if there is more than one factor. It is possible, in a simple modification to the model, to place restrictions on the loadings – in practice, this means most commonly that in a multiple factor model, certain items are allowed to load only on certain factors. This confirmatory factor analysis model may in turn be extended, so that particular relationships between the latent variables can be specified and tested: structural equation models (Bollen, 1989;

Jöreskog, 1973) typically involve modelling the relationships between several latent variables, alongside covariates.

Normal linear factor models, both exploratory and confirmatory, are often used for items with Likert-type responses, e.g. from 'strongly agree' to 'strongly disagree', and sometimes even for binary observed items. This approach is analogous to using a linear regression model to model categorical response variables, and is well known to be problematic. For example, with binary observed variables, modelling the expected values of $y_i$ (the left hand side of equation (1)) amounts to modelling the probability of responding in one category rather than another, given certain values on $x_j$. Any probability, by definition, must fall between 0 and 1. Yet if the relationship between the $y_i$ and $x_j$ is modelled as linear – if the right hand side of the equation is not restricted in any way – it is quite possible that values on the left hand side of the equation could fall outside these boundaries. Such values would be logically meaningless. This problem can easily be resolved, however, by using a non-linear link function between the $y_i$ and $x_j$, and abandoning the assumption of normality for $y_i$. This is the approach adopted in latent trait and class models. Since the observed variables analysed in this thesis are categorical, the normal linear factor model is not used at all in this thesis[2].

A variant on the normal linear factor model is available for ordinal and binary observed items. The Lisrel model (Jöreskog & Sorbom, 1996) employs the 'underlying variable approach', which begins by using a pre-processor programme to construct from the data a pseudo-correlation matrix. Hypothesising that the categorical variables correspond to categorised versions of continuous variables, it uses the proportions in the various categories to define thresholds along these hypothesised continua. Correlations between these hypothetical continuous variables are then calculated. This Lisrel model is not employed at all in this thesis for a number of reasons, chief amongst them being that most of the observed items in the analysis are nominal, and for those few items that are ostensibly ordinal, the ordering of categories is an issue to be tested empirically rather than assumed[3].

[2] It is not the focus of the thesis to consider the extent to which factor analysis may or may not return the same substantive interpretation of a set of items; such comparisons are arguably best achieved using simulation studies, and a formal statistical approach, in contrast to the applied spirit of this thesis.
[3] For comparisons of the Lisrel model with other models for ordinal observed items the reader is referred to Jöreskog & Moustaki (2001).

### *4.1.3 Latent trait models for binary and nominal observed variables*

In these models the problem of using a linear response function to link the observed and latent variables is solved by using a logit function, as given in the GLLVM:

$$logit\ \pi_i(x) = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} x_j \tag{2}$$

where $\pi_i(x)$ is the probability of a particular response for item $i$, given a value or set of values on the latent variable(s). Where these latent variables are continuous, they are usually referred to as latent traits. These latent variables are usually assumed to have standard normal distributions, as in (1).

The formula given in (2) implies binary manifest variables. In the empirical chapters in this thesis, however, models for binary items are only presented fleetingly; the main focus is on models for nominal polytomous variables. To rewrite the model for polytomous items, if we say that observed variable $i$ has $c_i>2$ categories, then denoting the categories by $s$ ($s$ takes the values $0,\ 1 \ldots c_i-1$, with $0$ as the reference category), the model for polytomous variables can be written as:

$$log\ \{\pi_{is}(x) / \pi_{i0}(x)\} = \alpha_{i0(s)} + \sum_{j=1}^{q} \alpha_{ij(s)} x_j \tag{3}$$

where $\pi_{is}(x) = P(y_i = s|x)$ is the probability of responding in category $s$ $(=1,2 \ldots c_i-1)$ to item $i$, given a value or set of values on the latent trait(s), $\pi_{i0}(x)$ is the probability of responding in the reference category to that item, and $x$ denotes the vector of values for the $q$ latent variables.

Latent trait models, in their various forms, are known collectively as Item Response Theory, or IRT, models (van der Linden & Hambleton, 1997). These have been developed in the field of educational testing, where the focus of interest is usually on modelling single traits representing knowledge, ability, and similar attributes. The well known model given in (2) is attributed to Birnbaum (1968), and this and the model in (3) can be referred to as two-parameter logistic (2-PL) models, in IRT terminology. In standard terminology, the model defined in (2) describes a binary logistic regression model, and in (3), a multinomial logistic regression model.

### *4.1.4 Interpretation of latent trait parameter estimates*

Latent trait models are interpreted by way of two key sets of parameter estimates that describe the relationship between the observed and latent variables. Firstly, the loadings $\alpha_{ij}$ give, for polytomous items, the effect of the latent variable on the log odds of responding in category $s$ rather than the reference category for an item. Higher values indicate greater discrimination power for category $s$ in relation to its reference category. The relative values of the discrimination parameters for different categories of an item can be informative; for example, we would expect the categories of a Likert item to be ordered, so that if 'strongly disagree' were the reference category, the coefficients would be steadily increasing (decreasing), with the highest (lowest) coefficient given to 'strongly agree'. If this is not the case, it would indicate that the item should be interpreted as nominal rather than ordinal, in relation to the latent trait. Some response categories may be more discriminating than others, in relation to a reference category; if some are not significantly different from each other, it may suggest that in future waves of the survey these categories may be collapsed, or that different response categories should be used. In the simpler case with binary items, there is only one discrimination parameter per item, so they can be used more straightforwardly as indicators of the discrimination power of *items* rather than categories of items.

Items exhibiting high discrimination power are particularly useful for the survey analyst as tools for differentiating between respondents in terms of the latent variable; those items with low discrimination power can be thought of as not doing much 'work' in this regard. As such, when using latent variables to model a psychological construct, and with an eye on future survey design, one can identify items with high discrimination power as those to repeat in future surveys, and those with low discrimination power as candidates for deletion or modification.

Alongside discrimination parameters, the model specifies constants or intercepts $\alpha_{i0}$. Whereas these are not of great interest in the linear factor model, in latent trait models they have heuristic value when expressed as difficulty parameters for the items. The difficulty of a particular response is defined as the probability of giving that response for the median individual on a trait. Fixing the values of all the latent traits to 0 (when the traits are normally distributed, this is the median individual), the probability for an individual of responding in category $s$ to item $i$ can be calculated from the following equation:

$$\hat{\pi}_{is}(0) = \exp[\hat{\alpha}_{i0}(s)] \;/\; \sum_{r=0}^{c_i-1} \exp[\hat{\alpha}_{i0}(r)] \;. \qquad (5)$$

Difficulty is an intuitive concept for items assessing knowledge of a topic, but can be equally interesting when analysing attitudinal items. In the latter case, the common sense meaning of 'difficulty' does not map so easily onto the interpretation of the difficulty parameter – but the following example illustrates the way in which it can be understood. On a scale of high to low engagement with biotechnology, a difficult item might be to agree that 'I would be prepared to take part in public hearings on the subject'; the average individual on the trait would be unlikely to agree to this. An easier item might be to agree that 'I have heard of biotechnology before'; the average individual may be very likely to agree to this. For clarity and convenience of presentation, and because of the general usage of the term 'difficulty' for this parameter in the IRT literature, I will use it throughout the thesis, both for items capturing knowledge as well as those capturing attitudes and other affective attributes. Identifying easy and difficult items can be a key part of forming a description of a latent variable. It can also inform future survey design – for example, if nearly everyone is likely to agree that they have heard of biotechnology before, whilst this may be an interesting and important finding in its own right, the utility of the item in a scale of engagement is rather limited. For purposes of scale construction, then, it may be dropped from future surveys, or at least, dropped from the scale in future.

The combined information from item loadings and intercepts can be represented graphically, by calculating a selection of fitted probabilities of item responses for a range of values for a latent trait $x_j$ (fixing the other traits at some values, if there is more than one trait). To do this for the polytomous latent trait model for example, from equation (3) we use:

$$\hat{\pi}_{is}(x) = exp\,(\hat{\alpha}_{i0}(s) + \sum_{j=1}^{q} \hat{\alpha}_{ij}(s)x_j\,) \,/\, \sum_{r=0}^{c_i-1} \exp(\hat{\alpha}_{i0}(r) + \sum_{j=1}^{q} \hat{\alpha}_{ij}(r)x_j) \qquad (6)$$

for selected values of $x_j$, where $\hat{\alpha}_{ij}$ denotes the parameter estimate (and $\hat{\alpha}_{i0}(0) = \hat{\alpha}_{ij}(0) = 0$). From such fitted probabilities we can draw Item Characteristic Curves (ICCs) or trace lines, which show at a glance the changing probabilities of choosing each of the response categories at any point along the latent trait. 'At any point' can be sensibly limited to values of approximately +/– 3 when the latent trait(s) are assumed to be standard normal. ICCs show for each item its discrimination power, via the steepness of the slopes of its response curves (the steeper the slope, the greater the discrimination), and its difficulty, by way of the location of the curve in the plot (the higher on the latent trait, the greater the difficulty).

There are two variants on the 2-PL model which it is useful to mention briefly, for comparison purposes – one with fewer parameters and one with more. The one-parameter logistic (1-PL) model is often used in educational testing. In this model, discrimination parameters are assumed to be the same for every item, and only the difficulty parameter is estimated separately for each. This is a random effects version of the well known Rasch model (Rasch, 1960). By contrast, sometimes an extra parameter is added (the 3-PL model: Birnbaum, 1968). This raises the height of the lower asymptote of the logistic response curve, and is commonly introduced as a 'guessing' parameter, in scenarios where it is deemed that the probability of a correct or positive response from those at the lowest end of the trait should be increased. Notably, a 3-PL model is used in Jon Miller's models of biotechnology knowledge items. I do not use a 3-PL model in this thesis – reasons for this will be outlined at the start of Chapter 6, where I give a brief discussion of Miller's analyses and motivation for the models used in the chapter.

A last key detail in the interpretation of latent trait models is the facility to rotate the solution, where there are two or more traits – as is routinely used in exploratory factor analyses (Bartholomew & Knott, 1999). Where the traits or factors are assumed to follow standard normal distributions, the likelihood of the model is not changed by rotation, so it simply provides a different angle of interpretation on the same model. For example in a two-trait model, if for each item we plot its loadings on two axes, it is easy to imagine rotating those axes clockwise or anticlockwise to reach a clearer picture of which items load on the different axes. The points in the plot do not move in relation to each other, just the directions of the axes. Rotation may be orthogonal, if the latent variables are constrained to be independent of each other, or 'oblique', if they are allowed to be correlated. In this thesis, where continuous multi-trait models are presented, the solutions are rotated to aid interpretation, with oblique rotation applied throughout, since there is no theoretical reason to enforce independence between the latent variables. Another way of aiding the interpretation of a trait model is to apply restrictions to the model specification, in the manner of a confirmatory factor analysis – for example, setting some item loadings to zero. This approach can be particularly useful in discrete trait models, to which rotation cannot be applied because the model entails no assumptions of a multivariate normal distribution for the latent variable(s) (see Section 4.1.6).

### *4.1.5 Latent class models for nominal observed variables*

The latent class model (Lazarsfeld & Henry, 1968) is basically the same as the latent trait model but with one categorical latent variable $x$ with, say, $j$ unordered categories. The model is of similar form to (2) but with explanatory variables $x_2, ..., x_j$ which are dummy variables for categories $2, ..., j$. The latent variable is thus nominal, unless further restrictions are imposed on the model. The model yields intercepts and slope estimates as with a trait model, but in interpreting a latent class model it is common to focus on the combined implications of these, that is on the estimates of the conditional probabilities $\pi_{is}(j)$ of giving particular responses to particular items, given membership of a particular class. For example, we might be interested in how the probability of 'definitely agreeing' that GM food is morally acceptable changes according to the latent class allocation of a respondent. Just as the 'meaning' of latent traits can be interpreted by reference to notably high or low loadings for the items, the content of the classes can be described by inspecting these conditional probabilities and looking for patterns – in simplest terms looking for high probabilities of giving particular responses in different classes. Having reached descriptions of the latent classes in a model, we might be interested in the overall proportions of respondents expected to belong to each of them. This information is contained in the estimated prior probabilities of belonging to class $j$, denoted by $\eta_j = P(x=j)$.

### *4.1.6 Discrete latent trait models*

Discrete latent trait models (Heinen, 1996) fall somewhere between latent class and continuous latent trait models. Those described here and presented in later chapters are implemented in Latent GOLD (Vermunt & Magidson, 2005). Discrete latent trait models can be thought of, in a rough way, as latent class models in which the latent variable is ordinal, or as latent trait models where the trait is categorical rather than continuous – though in each case, with an important qualification. Whereas in latent class models we model a single nominal categorical latent variable, in Latent GOLD we can model more than one ordinal latent variable. Specifically, each trait has two or more ordered levels, which are coded at equal intervals between 0 and 1; these codes are used as numerical scores in the measurement part of the model. No assumptions are made about the distribution(s) of the latent variable(s); rather, the probabilities of its levels are now parameters to be estimated as part of the model. In the case of two- or more trait models, the traits can be constrained to be independent of each other, or allowed to be correlated.

A discrete trait model can be thought of then as a class model with certain constraints on its parameters. For example, a model with three discrete traits, each with two levels, is straightforwardly a restricted version of an eight-class latent class model (Vermunt & Magidson, 2005). Thus a discrete trait model makes it possible to investigate the dimensionality of a set of items in a way that a class model does not. Restrictions along the lines of confirmatory factor analysis can also be tested with discrete trait models – such as fixing loadings of some items on some trait(s) to zero.

The more levels specified for the trait(s), the closer it becomes, conceptually, to a continuous trait model. A continuous trait model, in fact, is not quite continuous in practical computational terms in any case, since the problematic nature of the integration required to estimate the model necessitates the use of techniques such as numerical integration methods – in Latent GOLD, using the Gauss-Hermite quadrature. That is, when the model is estimated, the trait is approximated by a number of points along the continuum, weighted so that the trait approximates a normal distribution. In practice then, a ten-level discrete trait model is quite similar to a continuous trait model in which ten quadrature points are used for its estimation. The key difference between the two models in this case is that whereas in a continuous trait model the trait is assumed to be normally distributed, in a discrete trait model no particular form is imposed on the prior distribution of the latent variable. So a continuous trait model contains a number of quadrature points whose weights depend on the normal distribution, whereas in a discrete trait model, the comparable quantities are parameters to be estimated – that is, probabilities for each of the levels of the latent variable(s). Thus one particular benefit of a discrete latent trait model is that it allows us to explore, in a somewhat less restricted way, how a latent variable is distributed.

### *4.1.7 Latent trait and class models for ordinal observed variables*

In most models presented in this thesis the observed variables are treated as nominal, that is where the model in question returns for each item an intercept and slope parameter for each response category in relation to the reference category. If the response categories behave ordinally in relation to each other, then the item parameter estimates would be ordered correspondingly. We can *enforce* ordinality on an item's categories by using a model for ordinal responses. The most common of these is the proportional odds model. In Latent GOLD, however, the model employed for ordinal responses is the adjacent category logit model, or in IRT terminology, the partial credit model. This is written as

$$log\ \{\pi_{is}(x) / \pi_{i0}(x)\} = \alpha_{i0}(s) + \sum_{j=1}^{q} s\alpha_{ij}x_j \qquad (7)$$

where $s = 0,1,...c_i{-}1$ are used as equal-interval numerical scores for the response categories. This also implies

$$log\ \{\pi_{i,s+1}(x) / \pi_{is}(x)\} = [\alpha_{i0}(s+1) - \alpha_{i0}(s)] + \sum_{j=1}^{q} \alpha_{ij}x_j \qquad (8)$$

for all $s = 0,1,...c_i{-}2$. Note that this measurement model can be specified for discrete trait and class models, as well as for continuous trait models.

### *4.1.8 Measurement models, structural models and multiple group models*

Each of the models described above can be thought of as consisting of two parts (as outlined in Chapter 1): a measurement model, and a structural model. The measurement model comprises the intercept and slope parameters for the items, while the structural model comprises the (joint) distribution(s) of the latent variable(s). It is important to note that I adopt a broader definition for 'structural model' than is common in the literature on latent variable models. In the statistical literature, structural models are typically taken to refer to relationships between latent variables which are specified by means of structural equation models. I use the term 'structural model' to refer generally to the joint distribution of a set of latent variables; and for simplicity and convenience of presentation, I extend this definition to include the distribution of a single latent variable. It is unconventional to use the term in the contexts of exploratory and confirmatory latent variable models; I do so, however, in an explicit effort to make the account of latent variable models as clear and simple as possible for the non-statistical reader.

In the analyses used in the empirical chapters the measurement model is generally more interesting and important to the success of the analysis than the structural model, since the focus of the thesis is on deriving the best possible measures of three types of constructs. The distribution(s) of the latent variable(s) will be something to inspect with interest, but only after a good measurement model has been found.

The simplest kinds of restrictions on the measurement models are those in the spirit of a confirmatory factor analysis. For example, in a multiple-trait model, selected item loadings can be set to zero, so that certain items load only on certain traits. Some restrictions of this kind will be used with discrete trait models in Chapter 7. The slope and difficulty parameters can also be restricted to values other than 0; in a continuous single trait model, setting the slope coefficients to be equal for all items amounts to a random effects Rasch model (the 1-PL model introduced in Section 4.1.4). In latent class models, item parameters can be constrained in order to fit, for example, probabilistic Guttman-type models. In Guttman models (described in Chapter 3), defined for $j$ binary items, we have $j+1$ classes, with the items and classes ordered in a certain way. At one extreme, the typical response profile is to answer negatively to all items. In the adjacent class, the typical response profile is to answer only the easiest item correctly (or positively, if attitudinal items are being modelled); in the next class, respondents are predicted to answer the two easiest items correctly; in the next, the three easiest items – and so on, until the final class, in which every item prompts a positive response. Some models of this kind are presented in Chapter 5. They are fitted using the LEM programme, which allows item parameters for latent class models to be restricted to certain values, to equality, or to zero, making a range of Guttman-type models possible. For example, the basic probabilistic version of the Guttman model is attributed to Charles Proctor (1970), in which the conditional response probabilities, patterned in the appropriate way, are equal between both classes and items. The Equal Item Specific Error Rates model (or EISER model: Lazarsfeld & Henry, 1968), is less restricted, allowing these to vary between items but not classes. A number of further variants on these models can easily be specified by restricting parameter estimates in this way (Van den Wittenboer, Hox, & de Leeuw, 2000).

It is also possible to impose restrictions on the structural part(s) of a model. Structural models are explored only in a minor way in this thesis; the focus is on the measurement models. However, a few small-scale analyses of relationships between latent variables appear in the thesis. In Chapter 7 for example, there is some interest in testing whether associations exist between traits in a discrete trait model. The subject of Chapter 8 is in a sense about structural models, in that it comprises an analysis of the relationships between the three PUS constructs, and selected socio-demographic variables. However, this analysis does not follow a structural equation modelling approach; it is rather a loglinear model using the posterior scores and class allocations from the PUS constructs (see Section 4.1.11).

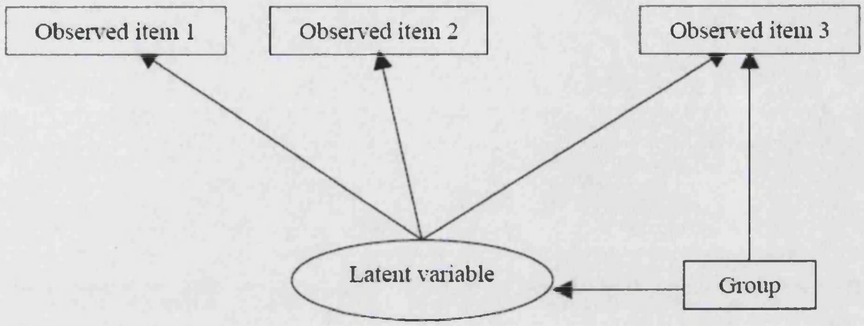
The most important use of model restrictions in the thesis is for multiple group analyses, where the aim is to derive latent variable models of constructs that are comparable, statistically speaking, between country samples. These involve, firstly, a structural model, for the latent variable(s) given country. It would be possible to restrict the distribution of the latent variable(s) to be the same for each country, but in cross-national analyses it is more usual that this is specifically a parameter of interest – for example, PUS researchers would be interested in how the distribution of knowledge varies from country to country. More elaborate specifications for structural models are the province of structural equation models, which are not used in this thesis.

It is the measurement part of the model – the part that defines the meaning of the latent variable – that is most usefully modified for the purposes of the cross-national analyses in my empirical studies, by allowing item response intercepts, or intercepts and slopes together, to vary by country. Figure 4.1 illustrates the two types of relaxations of parameters that are considered in this thesis. In latent trait models, freeing the intercept(s) or difficulty parameter(s) for an item amounts to including in the model direct effects of country on response probabilities for that item. This would allow us for example to specify that, all other things equal, a Danish respondent is more likely to answer a certain knowledge item correctly than is a Portuguese respondent. Freeing the slope or discrimination parameter(s) for an item as well as its intercept(s) would mean allowing an interaction between country, trait and response probabilities for that item. In the example of the knowledge items, this could mean for example that a knowledge item was highly discriminating in Portugal, but less so in Denmark. Freeing the slope parameter has the effect, then, of allowing the relationship between an item and its latent variable to vary by country – it implies that the item has a different interpretation for the attribute in question, between countries. In a model with more than one latent trait, these effects can be fixed and freed on one or more of the traits. This analysis broadly follows what in educational testing is termed 'differential item functioning' (DIF) analysis, employed to identify items which have significantly different slope and/or intercepts for different groups (Holland & Wainer, 1993).

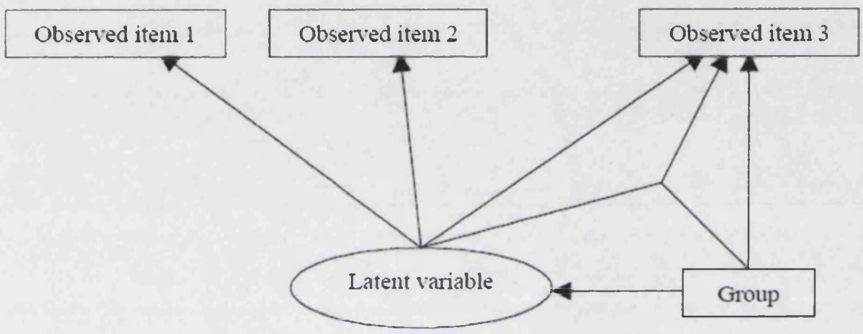
The same principle applies to latent class models, and makes most sense when these are expressed as logistic regression models. Allowing the intercept(s) for an item to vary by country means modelling direct effects between country and response probabilities for the item. So, for example, an Austrian respondent might be more likely than a Swedish respondent to agree that GM food is risky – regardless of latent class membership. Freeing both the intercept(s) and slope(s) for an item would mean allowing an interaction between country, latent variable and response probabilities. This would allow the effect of country on response probabilities to differ across classes. In a similar way to the latent trait model, the interaction term would allow the interpretation of the item in relation to the latent variable to vary from country to country.

**Figure 4.1 Two modifications to the measurement model of a latent variable with a covariate (group variable)**

**a. Direct effect between item 3 and group**




**b. Interaction between item 3, latent variable and group**




### *4.1.9 Estimation*

Estimation for all models is via Maximum Likelihood estimation (or more precisely, Bayesian estimation with vague prior distributions – see discussion below). In this thesis, almost all latent trait and class models are fitted using the programme Latent GOLD, version 4.0 (Vermunt & Magidson, 2005). The exceptions are Guttman-type class models presented in Chapter 5, and one type of discrete trait model in Chapter 6, which are fitted using the programme LEM (Vermunt, 1997). Other programmes which might be used to fit latent class and/or latent trait models include GLLAMM (Rabe-Hesketh, Pickles, & Skrondal, 2004), Mplus (Muthén & Muthén, 1998), SAS PROC LCA (Lanza, Collins, Lemmon, & Schafer, in press), BILOG-MG and MULTILOG-MG (du Toit, 2003).

Estimation for latent variable models is a fairly difficult computational task, and especially in complex models there is a possibility that iterations will converge to a local rather than a global maximum of the likelihood function. With older programmes this usually meant that a number of runs of the same model should be attempted from different starting values, to increase one's chances of finding the best model, with the greatest likelihood. Latent GOLD by default begins model estimation with ten sets of random starting values, and chooses the best of these from which to calculate model parameter estimates. An informal investigation undertaken early on in the empirical analyses for the thesis suggested that this default of ten sets seems to be sufficient for the algorithm to converge to the same solution on repeated runs. To be cautious, however, and since some of the models considered are relatively complex, 100 starting sets are used for all models estimated in Latent GOLD that are presented in the empirical studies.

Some additional features of the estimation settings used in Latent GOLD should be mentioned here (see Vermunt & Magidson, 2005 for further details of these). The first is that by default Latent GOLD specifies a prior distribution for the latent and conditional response probabilities – 'Bayes constants' in Latent GOLD terminology – to avoid boundary solutions, that is estimated probabilities of 0 or 1. The default values in Latent GOLD for these priors have been applied; these are quite weak priors, so the estimates from models using these differ little from Maximum Likelihood estimates.

It is possible in Latent GOLD to specify the number of quadrature points used in estimating continuous latent traits, and the number of levels in discrete traits. For continuous trait models, bearing in mind that too few quadrature points is likely to affect the results of the model (Lesaffre & Spiessens, 2001), twenty quadrature points are used for each continuous trait, unless otherwise stated. For discrete trait models, when using the models to approximate continuous latent variables and in the absence of any notable guidance in the literature, some preliminary analyses on the models in Chapter 6 indicated that beyond five levels, the model did not seem to change – parameter estimates would be very similar for a model with six levels compared with a model with seven levels, for example. Beyond eight levels, the estimation appeared to become more problematic, both in terms of failing to converge on a solution after a reasonable number of iterations, and especially with multiple traits, increasing estimation time considerably. The default number of levels used for discrete trait models in the thesis is seven, again, unless otherwise stated.

### *4.1.10 Diagnostics and model choice*

Model selection is a notoriously difficult topic, with no consensus in the literature on the best way to approach it. In this thesis I select models with reference to various fit statistics, and with due consideration to the interpretability of a model. I present a selection of the many diagnostics provided in Latent GOLD, alongside two newer and less widely used statistics. I pay greater attention to the latter, and will comment in Chapter 9 on how they appear to work in the range of models presented.

The statistics from Latent GOLD are standard and widely used, and I refer the reader to Vermunt and Magidson (2005) for details of their formulae. The following figures are presented in each table of fit statistics in the empirical chapters:

- The likelihood ratio chi-squared statistic (the deviance, or $L^2$)
- The number of degrees of freedom of the model (d.f.)[4]
- The bootstrap p-value for $L^2$ under the null hypothesis that the model generated the observed data. This is obtained by generating 500 samples of the same size from the fitted model, using the estimated values for its parameters. The p-value is the proportion of these samples whose deviance is greater than the deviance for the original sample.
- The Bayesian Information Criterion (BIC) = $L^2 - \log(N)$ d.f., where N denotes the sample size
- Akaike's Information Criterion (AIC) = $L^2 - 2$ d.f.
- Likelihood ratio comparison tests of nested models are sometimes used, in the standard way; the results of these are reported in the text where they are relevant.

The bootstrap p-value is recommended over the asymptotic p-value when there are large numbers of small expected frequencies in the data (i.e. frequencies less than 5), since under these conditions the $L^2$ may not follow a chi-squared distribution. Most of the models presented in this thesis are based on sparse contingency tables; the bootstrap p-value therefore provides a solution to this problem. However, it does not offer a solution to the fact that as a test of goodness of fit, the deviance is also sensitive to

[4] In Latent GOLD, where the number of cells in the full contingency table is smaller than N, the degrees of freedom are calculated in the usual way, as (number of cells in the table – number of parameters). Where this is not the case, degrees of freedom are given as (N – number of parameters).

sample size. Evidence from the analyses carried out (see especially Chapter 5 when this is encountered for the first time) is that samples of circa 1,000 cases per country are too large for $L^2$ to be useful for this purpose.

BIC and AIC are commonly used for large samples as alternative, 'penalised' model selection criteria (Kuha, 2004). For each, the lower the figure, the better the model. Although they seem to be increasingly relied on for choosing between alternative models, it is known that they are not unproblematic, even for non-latent variable models (e.g. Kuha, 2004; Weakliem, 1999). It is sometimes found that one appears too high, and the other too low, according to an external criterion, for example. Each is based on a different criterion of model selection and as such behaves slightly differently – for example, BIC tends to favour smaller models.

We are still lacking a means of assessing absolute model fit – so for this purpose some alternative fit statistics are provided[5]. The first is the statistic which will be most heavily used in the empirical chapters to make decisions on model selection. This is based on an approach suggested in Bartholomew et al. (2002) drawing on Bartholomew and Knott (1999) and Jöreskog and Moustaki (2001), that is to look for large standardised marginal residuals as indicators of poor fit. Specifically, for responses to each pair of items, we create a two-way marginal table, by collapsing over responses to the other variables, and then compare $O$, the observed frequency in a single cell of such a table, with $E$, the expected frequency for that same cell. The residual for each cell is calculated as $(O-E)^2/E$. If the residual is assumed to have a chi-squared distribution with one degree of freedom, then as a rule of thumb we can take standardised residuals greater than four to be a sign of poor fit (Bartholomew et al., 2002). The greater the number of large residuals, the worse the model. However, since the number of two-way margins varies from model to model, it is helpful to calibrate this criterion accordingly, and consider the *proportion* rather than the raw sum of all two-way margins in a model that are unacceptably large. The final diagnostic is therefore the percentage of two-way standardised marginal residuals that are greater than four. Some comments will be made in Chapter 9 on what seems like a plausible rule of thumb for a percentage that indicates a well fitting model. One of the benefits of the marginal residuals is that as well as the overall percentage figure, used to assess global model fit, high residuals can

[5] These and a number of additional calculations on the fitted models – including rotation of continuous trait solutions, and drawing ICCs for discrete and continuous trait models – are calculated using functions kindly written by Dr Jouni Kuha in S-Plus software.

be used to identify local fit, or lack of fit – for example, to identify for which items a model fits poorly. Three-way margins may also be calculated, along the same lines as above. In this thesis they are used to assess the fit of cross-national models: specifically, country by item by item tables, that is two-way margins as described above, conditional on country.

Finally, a variant on the calculation of residuals is presented, based on Jöreskog and Moustaki (2001). In this approach we sum the two-way marginal residuals for pairs of items, for all categories of those items. So, where $m$ denotes the number of response categories for an item, for items $i$ and $j$ we calculate the sum, $S_{ij}$, of all two-way standardised marginal residuals in the $(m_i*m_j)$ table. To take into account differing rates of $m$, we convert this into a common metric using $S_{ij}/(m_i*m_j)$. Then to reach a single figure to summarise the information for a model, we repeat this for all combinations of pairs of items, and take the mean of all the $S_{ij}/(m_i*m_j)$ as our final measure of goodness of fit. In the tables of fit statistics in the empirical chapters this is presented under the heading 'Jöreskog and Moustaki index'. For a single two-way table, values greater than four for $S_{ij}/(m_i*m_j)$ are suggested as denoting a poorly fitting model (Jöreskog & Moustaki, 2001); this seems to be too generous a rule of thumb for the mean of several of them. In the models presented, a value of 1 or above on the Jöreskog and Moustaki index tends to indicate a very poorly fitting model.

### *4.1.11 Posterior scores*

In the way of posterior analysis, we are interested in the properties of the distribution of $x_j$ given all of the $y_i$, and group variables (covariates) $z_k$, if included. For continuous traits, the posterior mean of each trait is informative. For discrete traits, as outlined earlier, the distribution(s) of the trait(s) are interesting. These can be represented by the posterior probabilities of the levels of the trait(s). They can also be represented by the posterior mean(s) of the trait(s), obtained from the sums of the scores (between 0 and 1) of the levels, weighted by their respective probabilities. Such scores are known as empirical Bayes (EB) or expected a posteriori (EAP) predictors (Skrondal & Rabe-Hesketh, 2004; Vermunt & Magidson, 2005). In trait models of all descriptions, the posterior scores are generically termed 'factor scores'. For the latent class model, we use the estimates of conditional and prior probabilities to obtain estimated probabilities for membership of each class, given response profiles. The analogous quantity to factor scores is then, for each response profile, the class to which it is most likely to belong, that is for which the posterior probability is highest. In this report I call these 'class allocations'; in statistical literature they are often termed empirical Bayes modal (EBM) or modal a posteriori (MAP) predictors (Skrondal & Rabe-Hesketh, 2004; Vermunt & Magidson, 2005). This posterior analysis provides a tool for further research questions – for example, factor scores or class allocations can then be used in regression analyses, as predictors of other variables, or we might test for differences in factor scores between different socio-demographic groups. In this thesis, the posterior scores and allocations from the three empirical studies will be used in an analysis in Chapter 8, to explore in a simple way the associations between them. Further comments on this approach, relative to alternative ways of analysing the associations between latent variables, are reserved for Chapter 8.

### *4.1.12 Sample weights*

As mentioned above, in the analyses in the empirical chapters the structural models are of secondary interest, relative to the measurement models. However, this is not to say that the distributions of the latent variable(s) capturing the PUS constructs are not of interest at all – they are in fact rather important in the context of PUS research. For the final joint cross-national latent class models presented at the ends of Chapters 5 and 7, the estimated prior probabilities of membership in each class is given for each country, applying the basic case-level weights provided in the original survey data set, and for the fifteen EU countries together, weighted according to their relative population sizes.

The weighting is carried out using the two-step procedure which is available in Latent GOLD and recommended by the authors of the programme (see Vermunt & Magidson, 2005, for details). This entails running the selected model initially without any weights, taking from this the measurement model parameter estimates and fixing them at these values in a second, weighted run of the model. The second run therefore uses the weights to adjust the country effects and the class sizes, but without affecting the parameter estimates in the measurement model. This allows us to retain the advantage that the unweighted analysis is likely to give more stable parameter estimates for the response variables.

The distribution of the final discrete trait model presented in Chapter 6 is *not* weighted in this way, due to computational difficulties. Although the overall EU distribution of the trait can be adjusted according to country population sizes, the case-level weights cannot be taken into account. However, on the basis of the very slight effects of the weights in Chapter 5 and 7, this is likely to make very little difference to the results presented, in this case.

Having outlined the essential features of the models to be used in the empirical studies, the next section takes a conceptual rather than a technical turn, using the taxonomy from Chapter 3 to highlight some key points which are of particular interest and importance to attitudinal survey research. Comments on other latent variable models are reserved for Chapter 9, at which point, in the light of the findings from the empirical studies, we can widen the discussion to mention some of the many more advanced latent variable models that might be used in future analyses.

## 4.2 Conceptual issues in latent trait and latent class models

Latent variable models are ways of representing data. As such, they are strongly linked to the theme of observations and data, since they require a particular kind of data as their input: a correlation or covariance matrix, or table of associations, according to the level of measurement of the observed variables, and for posterior scores or class allocations, a matrix of individual-by-variable data – that is, the format of a standard survey data set. Section 4.1 described only the models employed in this thesis, and therefore did not cover the many more elaborate data structures that can be analysed with recent innovations in latent variable modelling (see e.g. Skrondal & Rabe-Hesketh, 2004); they need not be the usual single stimulus data (in Coombs' terminology) from cross-sectional surveys. Moreover, there is no inevitable link between latent variable models and any particular type of research method or tool; they are often used for analysing survey data but any research tool could be used, in principle, as long as the appropriate type of data were obtained from it.

The following sections discuss some other themes of the taxonomy in a little more detail. First are two themes – ways of representing, and validation – which are issues for latent variable models – and more or less contentious, depending on one's viewpoint. Next are two themes which represent the challenges in survey research outlined in Chapter 2, which we can address directly using latent variable models – that is, the relationship between the research subject and the research instrument, and the question of generalisability of the measures we create, specifically their comparability between countries. Lastly, two themes are relevant for reflecting on the role of latent variable models in social psychology more broadly: the interpretation of the models, if framed carefully at the appropriate level of analysis, recommends them to a less individualistic and more social psychology. The current limited use of latent trait and class models, it is suggested, may be partly explained by the historical connections between certain models and certain branches of the social sciences – that is, certain theoretical frameworks.

### *4.2.1 Choosing a model: ways of representing data and validating models*

Latent variable models imply a particular representation of a set of survey data: a depiction of individuals in a space – with the nature of that space determined by the precise specification of the measurement and structural parts of the model. A number of choices are involved in model specification. For the measurement model, there is firstly the straightforward choice of link function between latent and observed variables, depending on the level of measurement adopted for the observed variables. The measurement model also includes the definition of any special relationships between observed and latent variables – for example, restricting certain item loadings to zero, in a confirmatory factor analysis, or constraining parameters to other values – for example, to equality in a Rasch model.

Decisions about the structural model concern the number of latent variables and their level of measurement, and if more than one, the relationship(s) between them. The choice of level of measurement of the latent variable(s) is worth a special mention specifically, for the reason that choices along these lines are statistically speaking rather more arbitrary than might be supposed. Besides the basic comments in Section 4.1.6 on the similarity between continuous and discrete trait models, Bartholomew and Knott (1999) demonstrate that latent trait and class models may be empirically nearly indistinguishable, and Heinen (1996) makes a similar case for latent class and discrete latent trait models. On an applied level, further, Bartholomew and Knott (1999) comment that the choice of the prior distribution for a latent variable makes little difference in practice to its parameter estimates.

In the light of this, the choice of the form of the latent variable may seem facile. Indeed, Borsboom, Mellenbergh and van Heerden (2003) contend that this may be true if one takes a constructivist stance towards the ontology of the latent variable. In such a case, the representation of the latent space may be chosen according to criteria of practical convenience. Such criteria may concern, for example, the intuitive clarity of interpretation offered by different models, or the political correctness of connotations of one model over another (cf. Miller's class-analogous model of scientific literacy, versus Durant's trait-type model of degrees of scientific knowledge). At their most pragmatic, the models may be used simply as a way of deriving scores to summarise a set of items. The survey researcher then has more license in such a choice than he or she might otherwise realise. By contrast, if one adopts a realist position with respect to the status of the latent variable, then the choice of the latent variable will be determined by a different set of external criteria. Borsboom et al. (2003) argue that this is the only right way to go about fitting a latent variable model; to hold a conviction that it *exists* as a social or psychological phenomenon, which is causally linked to its observed indicators.

The position taken regarding the ontology of the latent variable, and the relative emphasis given to a representational versus a pragmatic measurement methodology, have considerable consequences for how its validation is addressed. There are two senses to the theme of validation here. On one hand is the straightforward fact that latent variable models come equipped with a number of statistics for assessing model fit. As outlined in Section 4.1.10, assessing model fit in terms of these statistics is far from simple, since each works in a slightly different way. Nevertheless, assessing model fit can be a valuable part of investigating the structure of a set of items.

Statistics for assessing model fit operate *within* the framework of the model. They do not, by themselves, provide answers to questions about the meaningfulness or credibility of a model in relation to external criteria – that is, validity in its broader sense. However, they are often used as tools for assessing validity in this way, and sometimes from very different standpoints in relation to measurement theory. For example, proponents of a strict representational approach to measurement such as Michell (1990) and Barrett (2002) contend that validation in latent variable models should lie in a phase of instrument construction. Before latent variables are used in further analyses – be this in the form of structural equation models or using posterior scores in other analyses – latent traits should be modelled individually, using Rasch models to demonstrate that they possess 'quantitative' properties. The fit of a precisely specified measurement model is crucial to their definition of validity.

By contrast, Hayduk (see e.g. Ferguson, 2003), a notable author in the field of structural equation modelling, takes a cross-reference approach to validation. For him, the validity of any one latent variable is defined by the way it is connected to other variables, latent and/or manifest. The *structural* model is crucial to validity here. So for Hayduk validation is to be found in the success of a structural equation model, with success defined in terms of strict tests of global model fit, which support expected substantive relationships between elements of the model.

For the kinds of analyses demonstrated in this thesis, I would take a less rigid line than either of these perspectives and lean towards a pragmatic approach towards selecting models, not least because in terms of what is feasible with Eurobarometer-type social survey data, the requirements of Barrett on the one hand or Hayduk on the other will usually be impossibly difficult to satisfy. Global goodness of fit tests are sensitive to sample size (see Section 4.1.10), so difficult to satisfy with large samples. Hayduk's standards could be made reachable by applying more pragmatic criteria to these goodness of fit tests, or by applying additional, or alternative, criteria. However, it is worth noting the cautionary note that some statisticians would attach to structural equation models generally, however well fitting, since they involve estimating so many unknown quantities at once (Bartholomew & Knott, 1999).

At the other extreme, meeting Barrett's conditions would require more resources than are usually available for social surveys on public opinion. Although rigorous scale construction is regularly undertaken in educational testing, in some branches of market research and in clinical psychology, realistically in social surveys we are pleased to use the parameter estimates of imperfect scales to teach us about the substantive topic we are investigating (van de Vijver & Leung, 1997). For example, differential discrimination parameters in a set of knowledge items might be predominantly a nuisance to a test constructor in the field of education, but informative and intriguing to an attitude theorist. Taking an exploratory rather than confirmatory approach to finding a measurement model can be very useful in survey analyses – indeed, this approach is explicitly adopted in the empirical chapters.

### *4.2.2 Methodological challenges in PUS surveys: relationship between research subject and research instrument, and generalisability*

In Chapters 1 and 2 I noted that responses to single survey items are the outcome of a number of factors. Some of these are interesting – for example, the attitude we are trying to capture. Some, however, are a nuisance – for example, response effects caused by the survey instrument, response effects unique to the individual, and other factors which might be collectively termed 'measurement error', which manifest themselves, for example, in response instability. The consensus among attitude researchers is that the best way of capturing a construct of interest is to combine responses from several items on the topic, in order to try to distil from data the information really desired regarding the construct. In Chapter 3 I described some classic approaches to attitude scaling. Having now been introduced to latent variable models, it is fitting to point out that many latent variable models are actually probabilistic versions of classic attitude scaling techniques. They represent a fundamental improvement on the scaling models of Thurstone, Guttman and Likert because these were deterministic: theoretically speaking, they could not accommodate any response profiles which deviated from the theorised scale structure. In practice of course, deviations were tolerated, if they constituted a sufficiently low proportion of responses. But probabilistic attitude scaling models are vastly more flexible and arguably more theoretically credible since they explicitly incorporate the idea of measurement error. In this aspect alone they make a significant step in addressing one of the many issues we worry about in surveys regarding the relationship between the research subject and the research instrument.

Latent trait and class models can be used for directly investigating response effects, which are an important element of this theme. Response styles are perhaps most effectively studied via survey experiments, but even in purely observational data latent trait and class models can be used to reveal possible response styles. This may give a steer on the interpretation of a set of items. In Chapter 7, for example, latent variable models will be used to see whether Likert-type items whose response categories are logically ordinal actually behave as such, just by comparing the fit and interpretation of models for nominal observed items and models for ordinal observed items. A simple comparison of this type can be informative not only for other survey analysts using these items, but also for survey designers considering modifications to items and response categories for future surveys.

Latent trait and class models can also be used to begin to understand DK responses, which are a key concern in PUS surveys. As noted in Chapter 2, a full understanding of DK responses of course requires qualitative data from cognitive interviews and the like, but in the meantime we can gain valuable information from the survey data themselves with the use of latent variable models. For example, in a latent trait model, items containing a DK option can be treated as nominal, and the trace lines of DK responses compared with those of the other response categories – to see whether, for example, DK falls at the end of a continuum, whether it is closer to one response alternative than another, and in particular, whether it functions as a middle response category. On a more basic level, simply by including DK as a valid response category in a model rather than recoding or deleting it, such an analysis represents an improvement over much standard survey research practice.

In this thesis latent variable models will also be used explicitly to address the challenge of cross-national comparisons in PUS surveys. This is a question of generalisability, in Chapter 3's terms: to what extent does a particular latent variable representation of a construct hold in other country samples? Latent variable models are in fact often used in cross-national analyses, but the common approach to their use is to pool the data from all countries – that is, treat the data as if they were sampled from a common population – and simply run the analysis for the total data set. The comparison part of the analysis comes later, when posterior scores or class allocations are contrasted between country samples. A better approach, which will be adopted in the empirical studies, is to explicitly investigate the question of whether the same latent variable representation of a construct is found in different country samples. As outlined in Section 4.1.8, before making comparisons between countries on a construct such as 'knowledge', we need to find out whether 'knowledge' has the same interpretation in every country. This can be done by evaluating the comparability of the measurement models of the construct 'knowledge'. As noted in Chapter 2, statistical models cannot tell us anything definitive about the *meaning* of a construct, as such; qualitative research is needed to answer questions on the full interpretation of a construct derived in a latent variable model. Nevertheless, statistical analyses *can* tell us whether in different samples the items tend to behave in the same way and tend to be associated in the same way, which must be the first important step in sensitive comparative analyses.

The tools used for this kind of analysis come from and are used in a very sophisticated way in the field of educational testing. There, the problem is usually to find out whether binary items (correct versus incorrect) on a single trait of 'ability' are more difficult for one group of students than another, controlling for level of ability; say we want to compare male and female students, or students from two or three different ethnic backgrounds. Usually it is assumed that the slope parameters are equal for the groups to be compared, and that only the locations of the curves may vary. It is sometimes the case that multiple choice items are being analysed, and that the analyst is interested in whether different groups of students tend to systematically choose one particular incorrect answer (or 'distractor') over another. Thissen, Steinberg, and Wainer (1993) use the term Differential Alternative Functioning (DAF) for scenarios where we are interested in the trace lines of more than one category of polytomous items.

In attitudinal survey analysis we are usually faced with a scenario of investigating DAF rather than DIF. Indeed the scenario is more complex in a variety of ways for social survey analyses. Not only are the items often polytomous, but in comparative analyses there are usually many more than a few groups – in this thesis, for example, there are fifteen countries. And it is often the case that a single dimension is not sufficient to represent the variation in the data. So we have some potentially powerful tools, but a difficult setting in which to use them. In the empirical chapters, in fact, DAF analyses are put into effect more successfully with latent *class* models, than with trait models, representing quite a departure from DIF analyses in educational testing, where the latent variable of interest is typically a single continuous trait.

### *4.2.3 Latent variable models in social psychology: comments on levels of analysis and links to substantive theoretical frameworks*

The ways in which latent variable models allow us to interpret data have much to recommend them to social psychological research, especially research into attitudes and related constructs. Doise, Clemence, and Lorenzi-Cioldi (1993) point out that factor analysis (and equally, we might add, latent trait analysis) reconnects attitudinal research with Thomas and Znaniecki's focus on the individual–social relationship. For example, patterns of loadings given in factor or trait models are often described as structures of knowledge and affect, shared at the group level. Factor scores then represent the varying positions of individuals within those common structures, or in relation to common reference points. In this sense, latent variable models are well suited to analyses of social representations, which Doise et al. define as 'organising principles of variations in positions of different individuals' (ibid., p.5).

The precision with which Doise et al. distinguish between individual and group levels of analysis is refreshing. In both academic and non-academic attitudinal survey research one will often find reports of between-subjects models interpreted by recourse to within-subjects explanations, most commonly in that dimensions from factor analyses of cross-sectional survey data are often described as if they comprised mental representations within individuals. To take an example from PUS, in a study of attitudes towards science and technology, Miller and Pardo (2000) identify two factors from a set of confirmatory factor analyses: one representing the promise of science and the other representing reservations about science. These factors emerge from between-individuals analyses, and yet the authors make the common linguistic slip of claiming that the 'two schemas, the promise of science and technology and reservations about the impact of science and technology, operate simultaneously *in the minds of most individuals* in modern industrial societies' (ibid., p.125, my italics). As discussed in Section 3.4.7 in the previous chapter, such a claim does not follow directly from this model. Miller and Pardo interpret their model as if it describes a *distributed* representation, in Harré's (1984) terminology – that is, assuming that each individual holds this representation. In fact the analysis they have carried out works on the basis of between-subjects patterns of responses, and depicts something closer to a *collective* representation, that is, a picture of perceptions formed by combining information from a congregation of individual response patterns.

As such, the representation denoted by the model exists more at the *group* level than the individual level. It may well also exist as a distributed representation, but this remains an open, empirical question. As Borsboom, Mellenbergh, and van Heerden (2003) rightly say, it is not that psychological processes *cannot* be linked to latent variables, but this link needs to be tested rather than assumed. If a latent structure is found to 'work' at the both the individual and the group level, it can be said to be locally homogeneous, that is, the same for each individual. The misreporting of models along these lines actually does them a disservice, in a sense: it is precisely the individualistic tone of reporting that makes them unpalatable to the very interpretivist-minded social psychologists.

A more evidently problematic issue in the use of latent variable models in social psychology is the choice of particular form of latent variable model, given the level of measurement of the observed items. Common sense dictates that very few variables used in social psychological models are really continuous; many items using Likert response categories are logically speaking ordinal. This should lead researchers to choose latent trait and class models over factor analysis, for the reasons set out in Section 4.1.2 above – but factor analysis, and exploratory factor analysis in particular, remains the dominant tool, especially in survey analyses.

There are many plausible explanations for this. It might be attributable to institutional memory; to historical ties between certain models and certain substantive research questions, and particular research fields or theoretical frameworks. Latent trait models were developed not in social psychology but in the context of educational testing (e.g. Lord & Novick, 1968). Paul Lazarsfeld, who gave us latent class models (e.g. Lazarsfeld & Henry, 1968), was known as a sociologist, even though much of his research falls within the territory of social psychology. It is not surprising to learn, then, that factor analysis, the latent variable model of choice in social psychology, is largely attributable to Louis Thurstone (1931). The factor analysis model that we use today was in fact developed as a response to Spearman's (1904) work, in the context of intelligence testing. Thurstone's multiple factor analysis suggested that the construct of intelligence could be broken down into seven primary mental abilities or 'vectors of mind' (1935, 1947) rather than the one general 'g' which Spearman proposed. So factor analysis was conceived in a politically contentious context, where the implications of these models impacted directly on the life chances of the many people to whom IQ tests were administered (Gould, 1981/1996).

The slow uptake of latent trait and class models might also be partly attributable, more mundanely, to college and university training. In the same way that students are taught linear regression before logistic regression, factor analysis is simply more widely taught than latent trait and latent class analyses. Then, on a practical level, factor analysis is available in user-friendly general computer programmes such as SPSS. Perhaps the situation will change as more accessible programmes for latent trait and class models become available – in that sense, Latent GOLD follows in the footsteps of Lisrel and Amos, which made structural equation models more accessible for those with an aversion to programming language. Likewise, perhaps reducing reliance on bespoke programmes for these models will increase their availability for researchers who would use these models only occasionally – GLLAMM in Stata is one example of a package for latent variable models in general statistical software.

## Conclusion

This chapter has introduced the statistical models which are the methodological focus of the thesis. It outlined the technical details of the models to be employed in the following empirical chapters, noting why the commonly applied linear factor model is not appropriate for the data to be analysed, and how the latent trait and class models to be used can be fitted and interpreted, and evaluated for goodness of fit.

The second section of the chapter drew on the themes from the taxonomy of Chapter 3 to highlight some interesting features of the models in conceptual rather than technical terms. It raised the point that the choice of type of latent variable is statistically speaking quite arbitrary, and that the validation of any latent variable model is a contentious point, with different schools of thought insisting on fulfilling different criteria of validation. There is therefore certainly no gold standard for applying these models. The approach adopted in this thesis is decidedly pragmatic, both in principle and also simply because of the nature of the data and aims of the analyses.

This thesis, indeed, is very practically oriented. One of its purposes is to demonstrate how the models can be used to address the key challenges in PUS surveys: namely how to capture complex constructs in the best way, accounting for response variability and trying to identify response effects; how to include DK responses in models as 'valid' response categories; and how to explore the comparability of measures across countries. The models recommend themselves not only to these particular challenges, however, but also to the analysis of attitudinal data more broadly. I noted with regret that the results of these models are often interpreted in an unclear way with regard to their level of analysis. Reported more accurately, they resonate with a less individualistic and more social conception of attitudes, and provide a potential means of reconnecting current quantitative attitudinal research with its more socially oriented origins. Latent variable models are flexible and valuable tools which, in social psychology, are currently rarely used to their full potential. I hope that the empirical studies in the next chapters deliver convincing examples of how they might be employed well in analyses of PUS survey data.

# 5 Modelling choices: beyond positive and negative attitudes

This chapter begins the empirical work of the thesis with a study of perceptions of two applications of biotechnology: GM food and therapeutic cloning. It appeals to what Bauer and colleagues term the 'Public Understanding' stream in PUS research (Bauer et al., 2007), for which attitudes towards biotechnology are the central concern. The focus in this chapter is on evaluations of specific applications of biotechnology rather than on biotechnology in general. The former tend to be more interesting and useful in PUS research, not only because of the varying average levels of support and opposition observed for different biotechnologies, but also because of varying *types* of support and opposition observed.

The literature cited briefly in Chapter 2 (e.g. Gaskell et al., 2006) demonstrates that underlying the single pithy 'for' or 'against' verdict on each different application of biotechnology is a range of patterns of opinion. Besides the general sense that Europeans are on average supportive of medical biotechnologies and unenthusiastic about green biotechnologies, there is a good deal more to say about the contents of these judgements. For example, approval seems to be linked to a perception of low risk in agri-food and industrial biotechnologies, whilst there is a higher threshold for risk tolerance in medical applications. These patterns are set against the qualification that survey questions on these topics tend to return high rates of 'don't know' (DK) responses, perhaps because the topic is unfamiliar or difficult for many respondents. Nuances such as these serve as a particular warning against placing too much confidence in single indicator measures of opinions, such as, 'public opinion...is four to one against [GM maize]' (Deane, 2004), which are used often by the press, and with considerable impact.

The analyses in this chapter are therefore framed around the task of characterising and classifying types of opinion for an example of a red biotechnology (therapeutic cloning) and an example of a green biotechnology (GM food). The items analysed ask respondents for their views on four criteria of support and opposition for these technologies. The first part of the chapter concentrates on the British sample data from the 2002 Eurobarometer, excluding Northern Ireland, in which the survey was separately administered. The second part of the chapter broadens the models to the full European data set. Specifically, for the selected items, the following questions are addressed:

1. How many types of support and opposition can we sensibly speak of in the British sample? How are they characterised? Are there notable differences in these characterisations for GM food compared with therapeutic cloning? What do these analyses reveal in particular about the relationship of perceptions of risk to other criteria of support and opposition? What can we make of DK responses in these item sets?

2. To what extent do similar patterns hold in other country samples? What comparisons, if any, can be drawn between countries using these survey items?

3. With a view to future Eurobarometer surveys, can we make any recommendations, specific or general, for the design of items capturing evaluations of biotechnologies?

## 5.1 Data

As I have mentioned in previous chapters, the data in all empirical chapters in this report are taken from the Eurobarometer 58.0 administered in 2002. The set of items analysed here are part of the split ballot design of the survey. Half of the sample were asked for their opinions on GM foods and half on therapeutic cloning, defined respectively in the questionnaire as follows:

- Using modern biotechnology in the production of foods, for example to make them higher in protein, keep longer or improve the taste. *(Split ballot A)*

- Cloning human cells or tissues to replace a patient's diseased cells that are not functioning properly, for example, in Parkinson's disease or forms of diabetes or heart disease. *(Split ballot B)*

For each application respondents were asked the following questions[6]:

- Could you please tell me whether you definitely agree, tend to agree, tend to disagree or definitely disagree that this application is **useful** for society?
- And to what extent to you agree that this application is a risk [**risky**] for society?
- And to what extent do you agree that this application is **morally acceptable**?
- And to what extent do you agree that this application **should be encouraged**?

Responses are therefore on a four-point Likert scale, with the added possibility of giving a DK response. The phrases given in bold type will be used throughout the chapter to refer in abbreviated terms to these variables.

Table 5.1 presents the distributions of responses in the British data, and shows the much greater degree of overall support for therapeutic cloning than for GM food, following the well established pattern of greater support for medical than for agri-food applications of biotechnology, despite the similar proportions of judgements of riskiness for both applications. Note the high proportions of DK responses: between 14 and 20 per cent for these items.

[6] Each application was part of a set of three biotechnologies presented to respondents along with the same accompanying questions on support and opposition. The other applications were:
Split ballot A:

- Using genetic testing to detect diseases we might have inherited from our parents such as cystic fibrosis, mucoviscidosis, thalassaemia.
- Introducing human genes into animals to produce organs for human transplants, such as into pigs for human heart transplants (xenotransplantation).

Split ballot B:

- Taking genes from plant species and transferring them into crop plants, to make them more resistant to insect pests.
- Using genetically modified organisms to produce enzymes as additives to soaps and detergents that are less damaging to the environment.

**Table 5.1 Distribution of responses to questions regarding GM food and therapeutic cloning, British (GB) sample**

| | % responses[7] | | | |
|---|---|---|---|---|
| *Item/response category* | Useful | Risky | Morally acceptable | Should be encouraged |
| **GM food** | | | | |
| Definitely agree | 21 | 23 | 14 | 13 |
| Tend to agree | 31 | 28 | 29 | 24 |
| Tend to disagree | 18 | 19 | 22 | 21 |
| Definitely disagree | 15 | 12 | 16 | 22 |
| Don't know (DK) | 16 | 19 | 19 | 20 |
| Total | 100 | 100 | 100 | 100 |
| *n=508* | | | | |
| **Therapeutic cloning** | | | | |
| Definitely agree | 45 | 24 | 16 | 23 |
| Tend to agree | 28 | 32 | 39 | 34 |
| Tend to disagree | 8 | 18 | 19 | 15 |
| Definitely disagree | 5 | 11 | 10 | 11 |
| Don't know (DK) | 14 | 14 | 16 | 17 |
| Total | 100 | 100 | 100 | 100 |
| *n=506* | | | | |

## 5.2 Logics of support and opposition

The analyses in this chapter build on the idea of 'logics' of support and opposition, as outlined in Chapter 2 (Gaskell et al., 2003). Generally, across the range of applications asked about in the survey, three response patterns are most common, which can be labelled 'support', 'risk-tolerant support', and 'opposition'. Proportions in these groups are listed in the top half of Table 5.2. The classification discounts the difference between definitely and tending to agree and disagree (i.e. collapses the variables into yes/no/DK). The prevalence of these three groups seems fairly stable over time and between countries, and they carry a good deal of face and predictive validity (Gaskell et al., 2006), so they provide a sound starting point for the chapter. The table shows the relatively similar proportions of supporters for the two applications in 2002, with a far greater number of risk-tolerant supporters for therapeutic cloning, and smaller number of opponents, in contrast with GM food. Echoing similar findings from a range of PUS studies, this suggests that the element of risk in therapeutic cloning seems to be deemed acceptable, given its potential benefits, whereas GM food, when judged to be risky, is considered not worth the risk.

[7] Proportions calculated applying the basic weight in the data set which corrects for over- or under-sampling in particular sampling strata. Totals do not always sum to 100 per cent due to rounding.

**Table 5.2 Logics of support and opposition, GB**

| | Useful? | Risky? | Morally acceptable? | Should be encouraged? | % GB sample GM food | % GB sample cloning |
|---|---|---|---|---|---|---|
| Supporters | Yes | No | Yes | Yes | 12 | 18 |
| Risk-tolerant supporters | Yes | Yes | Yes | Yes | 14 | 29 |
| Opponents | No | Yes | No | No | 16 | 5 |
| Sub-total | | | | | 42 | 52 |
| Other logic (no DK) | | | | | 32 | 27 |
| DK (1 or more) | | | | | 25 | 22 |
| Total | | | | | 100 | 100 |
| *n* | | | | | *508* | *506* |

The difficulty with this classification is illustrated in the lower half of the table. Approximately half of the sample escape this typology for each application, which is a particular disadvantage if these group allocations are to be used in analyses with other variables. Around 30 per cent of respondents offer a different combination of responses from the three most popular groupings of substantive answers. Added to this, a substantial proportion of respondents answer DK to one or more items. In standard survey analyses these respondents would be deleted listwise from the classification of support and opposition. In the case of GM foods, 25 per cent of respondents give one or more DK answer, and approximately half of these are all-DK response profiles. For therapeutic cloning, 22 per cent give one or more DK response, with again approximately half of these being all-DK response profiles.

Overall in the British sample, for both applications 43 different response patterns are observed in terms of the collapsed variables with categories agree/disagree/DK. Acknowledging the intensifiers (the distinction between definitely and tending to agree and disagree) makes for 151 response patterns observed in the case of GM food and 132 for therapeutic cloning. So there is a considerable amount of variation in the data – although not nearly as much as there *could* be; given four items and five response categories for each, there are 625 different possible response profiles. With just over 500 respondents for each application, perhaps we should even be surprised at the relative homogeneity in the data. Either way, with many of those response profiles at present unclassified, a probabilistic approach to their analysis is a natural way to build on Gaskell et al.'s typology of logics, and latent class models constitute an ideal tool here.

## 5.3 Models considered in this chapter

The focus of this chapter is on latent class models, where the latent variable is nominal, and with all observed items treated as nominal, in order to accommodate DK responses. Section 5.4 focuses on the British sample. As a first step, unconstrained latent class models are fitted to enable us to decide, firstly, how many different groupings of responses are needed to adequately represent all the variation in the data, and secondly, how these groups might be characterised. It will be interesting, for example, to see if the three logics of support and opposition identified by previous research emerge cleanly from the data when analysed in their full complexity, including DK responses and intensifiers – and if any other notable patterns come to light.

The remainder of Section 5.4 is given to a selection of variants on these models, to investigate some particular questions about the structure of the data. It begins with a brief investigation into DK responses. This involves removing from the analyses those respondents who say DK to all items, and re-fitting the unconstrained latent class models. With the weight of the 'all-DK' responses removed, it becomes possible to suggest which type of sentiment tends to belong to those respondents who give partial-DK answers, who are hidden amongst the other classes in the previous models. This very simple analysis tells us a little more about these 'sometimes-DK' respondents. Another comparison exercise is used to investigate responses to the risk item, which is a further special point of interest for these data. In this case the composition of classes for GM food in the full British data in 2002 are compared with those fitted to the same questions fielded in the 2005 Eurobarometer. Lastly, thinking about the relationship of the four items to each other, a small selection of Guttman-type models are fitted to the 2002 data, for both GM food and therapeutic cloning (excluding DK responses), to investigate the notion of latent classes ordered from positive to negative attitude, and items ordered from easy to difficult.

Section 5.5 proceeds with a consideration of the second research question: how to expand the models to other country samples. Here the focus is on the original, unconstrained latent class models, since these provide the most general and inclusive representation of responses. The first concern is whether the measurement parts of the models (the conditional probabilities for responses given class membership) can be constrained to be equal among country samples. If the answer to this question is negative, it makes further comparisons between countries difficult or potentially meaningless. As a starting point, equality between measurement models from the separately sampled Britain and Northern Ireland, and East and West Germany are assessed. These studies enable us to judge whether we can sensibly take Germany and the UK to be country units in subsequent analyses, although they consist of two samples each, separately administered. The same principle of comparison is then applied to the group of fifteen EU countries. An initial qualitative assessment of the similarities and differences between their measurement models is followed by a formal statistical evaluation. The final models of the chapter enable comparisons to be made among countries while at the same time taking into account the key variations found between them. The allocations of respondents to classes derived from these models will be used as response variables in Chapter 8.

## 5.4 Results of latent class analyses of British data

### *5.4.1 Unconstrained latent class models, full data sets*

Using the full information in the sample, five-class solutions seem to be the most useful, judging on grounds of substantive interpretation and fit statistics; in particular, BIC and residuals suggest five classes as the most parsimonious well fitting model (see fit statistics in Table 5.3). Table 5.4 gives, for responses to questions on GM food, the probabilities of the various responses for each item, conditional on class membership. Notably high[8] probabilities are highlighted in grey. For example, conditional on membership in the first class in the table (looking at the first column of figures), a respondent has a 0.83 probability of definitely agreeing that GM food is *useful*, a 0.31 probability of agreeing that it is *risky*, a 0.86 chance of agreeing that it is *morally acceptable* and a 0.87 chance of agreeing that it should be *encouraged.* Given such a pattern of likely responses for people in this class we could characterise it as one of strong or definite support. This suggested label is included at the top of the column of figures, alongside other suggested labels for the remaining classes. The last row of the table gives the (unweighted) estimated prior probabilities of belonging in each class.

[8] In general, for all latent class models presented in the thesis, conditional probabilities of 0.4 or greater are highlighted in grey. This arbitrary rule of thumb derives from observations during analyses that where one conditional response for an item is greater than 0.4, other responses tend to have low probabilities of occurring. Generally though, conditional probabilities have been inspected quite closely, and in cases where there is no one response that is notably larger than all the others, but two or three with similar probabilities of occurring, these are outlined in black. 'Similar' is defined as within or up to 0.10 (to 2 d.p.) from the probability of the most likely response for an item, given a class.

For example, this model estimates that 13 per cent of the British public would belong to the definite support class[9].

**Table 5.3 Fit statistics, unconstrained latent class models, GB, full sample**

| Model | $L^2$ | d.f. | p (bootstrap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **GM food, full GB sample** | | | | | | | |
| 2 classes | 1,239 | 475 | <0.001 | 289 | -1,720 | 37.3 | 9.32 |
| 3 classes | 758 | 458 | <0.001 | -158 | -2,096 | 17.3 | 3.92 |
| 4 classes | 497 | 441 | <0.001 | -385 | -2,250 | 8.0 | 1.31 |
| 5 classes | 352 | 424 | <0.001 | -496 | -2,289 | 0.7 | 0.32 |
| 6 classes | 314 | 407 | <0.001 | -500 | -2,221 | 0.0 | 0.25 |
| 7 classes | 286 | 390 | <0.001 | -494 | -2,144 | 0.0 | 0.19 |
| **Therapeutic cloning, full GB sample** | | | | | | | |
| 2 classes | 1,038 | 473 | <0.001 | 92 | -1,908 | 39.3 | 6.57 |
| 3 classes | 654 | 456 | <0.001 | -258 | -2,185 | 20.0 | 2.86 |
| 4 classes | 457 | 439 | <0.001 | -421 | -2,276 | 7.3 | 1.28 |
| 5 classes | 316 | 422 | <0.001 | -528 | -2,312 | 0.7 | 0.33 |
| 6 classes | 272 | 405 | <0.001 | -538 | -2,250 | 0.7 | 0.29 |
| 7 classes | 238 | 388 | 0.004 | -538 | -2,178 | 0.0 | 0.24 |

Inspecting the patterns of conditional probabilities across classes and especially focusing on *utility*, *moral acceptability* and overall *encouragement*, we can identify groups from definite support, through moderate support and opposition to definite opposition, and a class where DK is the most likely response for each question. A striking feature of the results is the indeterminacy of responses to the *risk* item, for all but the strongest opposition and the DK groups. For the two classes of support and for moderate opposition, probabilities for the *risk* item are not clearly defined. So a clean distinction between support and risk-tolerant support does not emerge in an obvious way from the full data. Positions on risk in relation to other judgements carry too much variation to make responses to this item good indicators of more general attitudes; whereas we could quite confidently guess a respondent's class just from his answer to any one of the other three questions, we could not do the same with the *risk* item.

[9] These are unweighted probabilities, as indicated in the table. In the final models presented in the chapter, the models were refitted with these statistics adjusted, as described in Chapter 4 (Section 4.1.12). In practice these weights make very little difference to the estimated prior probabilities, and no difference to the measurement models, so the estimates are left unweighted for the interim models.

**Table 5.4 Conditional and prior probabilities, 5-class latent class model for GM food, GB, full sample**

| | | Definite support | Support | Opposition | Definite opposition | DK |
|---|---|---|---|---|---|---|
| Item | Response category | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(5)$ |
| Useful | DK | 0.00 | 0.00 | 0.01 | 0.03 | 0.79 |
| | Definitely disagree | 0.00 | 0.02 | 0.07 | 0.67 | 0.03 |
| | Tend to disagree | 0.02 | 0.07 | 0.61 | 0.13 | 0.02 |
| | Tend to agree | 0.14 | 0.69 | 0.25 | 0.09 | 0.14 |
| | Definitely agree | 0.83 | 0.22 | 0.04 | 0.08 | 0.02 |
| Risky | DK | 0.06 | 0.07 | 0.01 | 0.03 | 0.81 |
| | Definitely disagree | 0.25 | 0.10 | 0.07 | 0.25 | 0.02 |
| | Tend to disagree | 0.21 | 0.28 | 0.31 | 0.05 | 0.00 |
| | Tend to agree | 0.31 | 0.38 | 0.39 | 0.12 | 0.10 |
| | Definitely agree | 0.16 | 0.18 | 0.23 | 0.55 | 0.07 |
| Morally acceptable | DK | 0.00 | 0.00 | 0.03 | 0.00 | 0.95 |
| | Definitely disagree | 0.05 | 0.01 | 0.01 | 0.82 | 0.00 |
| | Tend to disagree | 0.00 | 0.13 | 0.79 | 0.11 | 0.03 |
| | Tend to agree | 0.09 | 0.80 | 0.14 | 0.06 | 0.00 |
| | Definitely agree | 0.86 | 0.05 | 0.03 | 0.00 | 0.02 |
| Should be encouraged | DK | 0.02 | 0.05 | 0.00 | 0.01 | 0.92 |
| | Definitely disagree | 0.00 | 0.04 | 0.14 | 0.99 | 0.02 |
| | Tend to disagree | 0.03 | 0.13 | 0.81 | 0.00 | 0.01 |
| | Tend to agree | 0.08 | 0.73 | 0.05 | 0.00 | 0.03 |
| | Definitely agree | 0.87 | 0.06 | 0.00 | 0.00 | 0.02 |
| $\hat{\eta}_j$ | (unweighted) | 0.13 | 0.29 | 0.21 | 0.19 | 0.19 |

Key

$\hat{\pi}_{is}(j)$ = estimated conditional probability of response in category $s$ for item $i$, given membership of class $j$

$\hat{\eta}_j$ = estimated prior probability of membership in class $j$

For therapeutic cloning, a five-class model also fits well. Conditional and prior probabilities are given in Table 5.5. As for the GM food solution, we see a clear DK class, and responses for *moral acceptability* and overall *encouragement* are again easily divisible into definite and moderate support and opposition. Again *risk* seems relatively independent of the other items, but with those in the moderate support and strong opposition groups likely to agree that therapeutic cloning is risky. A noteworthy point of difference from GM food is the greater tendency in general to consider therapeutic cloning useful. Those in the moderate support class are as likely to definitely agree as to tend to agree that it is useful, and those in the moderate opposition group are most likely to tend to agree that it is useful. Overall the proportions of respondents estimated to fall into the two opposition classes for therapeutic cloning are much smaller than for GM food: 16 per cent and 10 per cent for moderate and definite opposition groups for therapeutic cloning, compared to 21 per cent and 19 per cent for corresponding classes for GM food.

**Table 5.5 Conditional and prior probabilities, 5-class latent class model for therapeutic cloning, GB, full sample**

| Item | Response category | Definite support $\hat{\pi}_{is}(1)$ | Support $\hat{\pi}_{is}(2)$ | Opposition $\hat{\pi}_{is}(3)$ | Definite opposition $\hat{\pi}_{is}(4)$ | DK $\hat{\pi}_{is}(5)$ |
|---|---|---|---|---|---|---|
| Useful | DK | 0.00 | 0.01 | 0.00 | 0.10 | 0.86 |
| | Definitely disagree | 0.01 | 0.01 | 0.00 | 0.49 | 0.00 |
| | Tend to disagree | 0.01 | 0.04 | 0.34 | 0.12 | 0.00 |
| | Tend to agree | 0.00 | 0.47 | 0.49 | 0.14 | 0.12 |
| | Definitely agree | 0.98 | 0.47 | 0.17 | 0.15 | 0.02 |
| Risky | DK | 0.06 | 0.02 | 0.04 | 0.00 | 0.79 |
| | Definitely disagree | 0.20 | 0.06 | 0.14 | 0.29 | 0.00 |
| | Tend to disagree | 0.20 | 0.26 | 0.25 | 0.00 | 0.01 |
| | Tend to agree | 0.25 | 0.51 | 0.28 | 0.12 | 0.13 |
| | Definitely agree | 0.29 | 0.16 | 0.30 | 0.59 | 0.07 |
| Morally acceptable | DK | 0.01 | 0.03 | 0.05 | 0.04 | 0.98 |
| | Definitely disagree | 0.02 | 0.01 | 0.07 | 0.78 | 0.00 |
| | Tend to disagree | 0.06 | 0.10 | 0.77 | 0.14 | 0.00 |
| | Tend to agree | 0.25 | 0.83 | 0.09 | 0.00 | 0.01 |
| | Definitely agree | 0.66 | 0.03 | 0.03 | 0.03 | 0.01 |
| Should be encouraged | DK | 0.00 | 0.04 | 0.10 | 0.00 | 1.00 |
| | Definitely disagree | 0.00 | 0.00 | 0.07 | 0.91 | 0.00 |
| | Tend to disagree | 0.01 | 0.05 | 0.81 | 0.06 | 0.00 |
| | Tend to agree | 0.12 | 0.82 | 0.01 | 0.03 | 0.00 |
| | Definitely agree | 0.86 | 0.08 | 0.01 | 0.00 | 0.00 |
| $\hat{\eta}_j$ | (unweighted) | 0.22 | 0.38 | 0.16 | 0.10 | 0.14 |

### *5.4.2 Unconstrained models, investigating DK responses*

Earlier I noted that of those respondents giving one or more DK answer, about half gave a full set of four DKs, with the other half giving a mixture of DK and non-DK answers. Little more can be said in this analysis about the all-DK respondents, but it would be interesting to try to find out a little more about those who do not *always* say DK. In the next two models, the all-DK responses (63 respondents for GM food and 53 for cloning) are excluded from the analyses, and unconstrained latent class models re-run on the new data sets. Table 5.6 shows that once again, five classes are needed to adequately represent the variation in the data, both for GM food and for therapeutic cloning.

**Table 5.6 Fit statistics, unconstrained latent class models, GB, all-DK responses removed**

| Model | $L^2$ | d.f. | p (bootstrap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **GM food, all-DKs removed** | | | | | | | |
| 2 classes | 908 | 412 | <0.001 | 84 | -1,604 | 26.0 | 6.18 |
| 3 classes | 614 | 395 | <0.001 | -176 | -1,795 | 20.0 | 2.50 |
| 4 classes | 471 | 378 | <0.001 | -285 | -1,834 | 8.0 | 1.43 |
| 5 classes | 334 | 361 | <0.001 | -388 | -1,868 | 1.3 | 0.31 |
| 6 classes | 305 | 344 | <0.001 | -383 | -1,793 | 1.3 | 0.30 |
| 7 classes | 274 | 327 | 0.002 | -380 | -1,720 | 0.7 | 0.21 |
| **Therapeutic cloning, all-DKs removed** | | | | | | | |
| 2 classes | 740 | 420 | <0.001 | -100 | -1,828 | 29.3 | 4.02 |
| 3 classes | 545 | 403 | <0.001 | -261 | -1,920 | 16.0 | 2.49 |
| 4 classes | 390 | 386 | <0.001 | -382 | -1,971 | 5.3 | 1.06 |
| 5 classes | 294 | 369 | 0.002 | -444 | -1,962 | 0.0 | 0.27 |
| 6 classes | 254 | 352 | 0.006 | -450 | -1,899 | 0.0 | 0.30 |
| 7 classes | 231 | 335 | 0.160 | -439 | -1,818 | 0.7 | 0.19 |

Table 5.7 shows that with a five-class solution, approximately the same four classes are obtained as with GM food previously. It is just the last column, previously dominated by all-DK responses, which now looks a little different. Here it can be seen that alongside DK responses to *moral acceptability* and overall *encouragement*, responses for *risk* and *utility* are beginning to tend towards agreement. A remarkably similar pattern is found for therapeutic cloning (details presented in Table 5.8), but with a stronger leaning in the DK class towards agreeing that the application is useful and risky – these are the most likely responses, though they are perhaps not decisively high.

This distinction between two types of DK respondents seems to be an informative one – and, to anticipate Section 5.5, it will reappear in the multiple group analyses in the form of six-class models, in which two classes are given to two different types of DK response sets. These new DK classes suggest that for those who sometimes give a DK response, judgements of *moral acceptability* and overall *encouragement* are more difficult processes than judgements of *utility* and *risk*. The increased propensity for the unsure to consider applications useful is not surprising, given that this has already been observed as a general tendency among respondents. We would not necessarily have expected to see this change in the functioning of the *risk* item, however, given the variation it exhibits in the other classes – variation that we might suggest corresponds to uncertainty, at the group level.

**Table 5.7 Conditional and prior probabilities, 5-class latent class model for GM food, GB, all-DK responses excluded**

| Item | Response category | Definite support $\hat{\pi}_{is}(1)$ | Support $\hat{\pi}_{is}(2)$ | Opposition $\hat{\pi}_{is}(3)$ | Definite opposition $\hat{\pi}_{is}(4)$ | DK $\hat{\pi}_{is}(5)$ |
|---|---|---|---|---|---|---|
| Useful | DK | 0.00 | 0.00 | 0.01 | 0.03 | 0.33 |
| | Definitely disagree | 0.00 | 0.02 | 0.07 | 0.67 | 0.08 |
| | Tend to disagree | 0.02 | 0.07 | 0.60 | 0.13 | 0.12 |
| | Tend to agree | 0.14 | 0.69 | 0.27 | 0.09 | 0.38 |
| | Definitely agree | 0.84 | 0.22 | 0.05 | 0.08 | 0.08 |
| Risky | DK | 0.06 | 0.07 | 0.00 | 0.03 | 0.40 |
| | Definitely disagree | 0.25 | 0.10 | 0.07 | 0.25 | 0.06 |
| | Tend to disagree | 0.22 | 0.28 | 0.32 | 0.05 | 0.00 |
| | Tend to agree | 0.31 | 0.38 | 0.38 | 0.13 | 0.31 |
| | Definitely agree | 0.16 | 0.17 | 0.23 | 0.55 | 0.23 |
| Morally acceptable | DK | 0.00 | 0.00 | 0.01 | 0.00 | 0.78 |
| | Definitely disagree | 0.05 | 0.01 | 0.01 | 0.82 | 0.00 |
| | Tend to disagree | 0.00 | 0.12 | 0.80 | 0.12 | 0.13 |
| | Tend to agree | 0.09 | 0.81 | 0.15 | 0.06 | 0.00 |
| | Definitely agree | 0.86 | 0.05 | 0.03 | 0.00 | 0.08 |
| Should be encouraged | DK | 0.01 | 0.04 | 0.00 | 0.01 | 0.67 |
| | Definitely disagree | 0.00 | 0.04 | 0.11 | 0.99 | 0.12 |
| | Tend to disagree | 0.03 | 0.12 | 0.83 | 0.00 | 0.05 |
| | Tend to agree | 0.08 | 0.74 | 0.06 | 0.00 | 0.10 |
| | Definitely agree | 0.87 | 0.06 | 0.00 | 0.00 | 0.07 |
| $\hat{\eta}_j$ | (unweighted) | 0.15 | 0.32 | 0.23 | 0.21 | 0.08 |

**Table 5.8 Conditional and prior probabilities, 5-class latent class model for therapeutic cloning, GB, all-DK responses excluded**

| Item | Response category | Definite support $\hat{\pi}_{is}(1)$ | Support $\hat{\pi}_{is}(2)$ | Opposition $\hat{\pi}_{is}(3)$ | Definite opposition $\hat{\pi}_{is}(4)$ | DK $\hat{\pi}_{is}(5)$ |
|---|---|---|---|---|---|---|
| Useful | DK | 0.00 | 0.01 | 0.00 | 0.10 | 0.26 |
| | Definitely disagree | 0.01 | 0.01 | 0.00 | 0.49 | 0.03 |
| | Tend to disagree | 0.01 | 0.04 | 0.32 | 0.12 | 0.10 |
| | Tend to agree | 0.00 | 0.46 | 0.49 | 0.13 | 0.48 |
| | Definitely agree | 0.98 | 0.48 | 0.19 | 0.15 | 0.12 |
| Risky | DK | 0.06 | 0.02 | 0.03 | 0.00 | 0.16 |
| | Definitely disagree | 0.20 | 0.06 | 0.14 | 0.29 | 0.03 |
| | Tend to disagree | 0.20 | 0.26 | 0.24 | 0.00 | 0.10 |
| | Tend to agree | 0.25 | 0.51 | 0.27 | 0.12 | 0.46 |
| | Definitely agree | 0.29 | 0.15 | 0.33 | 0.59 | 0.24 |
| Morally acceptable | DK | 0.01 | 0.02 | 0.01 | 0.04 | 0.66 |
| | Definitely disagree | 0.02 | 0.01 | 0.05 | 0.79 | 0.06 |
| | Tend to disagree | 0.06 | 0.10 | 0.81 | 0.13 | 0.07 |
| | Tend to agree | 0.24 | 0.84 | 0.10 | 0.00 | 0.17 |
| | Definitely agree | 0.67 | 0.04 | 0.03 | 0.03 | 0.03 |
| Should be encouraged | DK | 0.00 | 0.01 | 0.02 | 0.00 | 0.97 |
| | Definitely disagree | 0.00 | 0.00 | 0.08 | 0.90 | 0.00 |
| | Tend to disagree | 0.01 | 0.05 | 0.88 | 0.06 | 0.01 |
| | Tend to agree | 0.12 | 0.85 | 0.01 | 0.04 | 0.02 |
| | Definitely agree | 0.86 | 0.09 | 0.01 | 0.00 | 0.00 |
| $\hat{\eta}_j$ | (unweighted) | 0.25 | 0.41 | 0.17 | 0.11 | 0.07 |

### *5.4.3 Investigating risk: comparing 2002 and 2005 survey waves*

Qualitative analyses would be needed to fully understand how respondents interpret and answer the question about risk, and what accounts for the varied patterns of responses to it. However, a little more information can be gleaned from the data with one last model and change of data. In the 2005 Eurobarometer the logics items were asked in relation to GM food (unfortunately not to therapeutic cloning), but with a different ordering of items. In the 2005 wave, the respondents were first asked whether GM food was *morally acceptable*, then whether it was *useful*, then whether it was *risky*, and finally whether it should be *encouraged*. Could it be that judgements on risk are more uniform within classes when it is preceded by two rather than one anchoring criteria?

Table 5.9 juxtaposes conditional probabilities for models run separately for samples from the 2002 and 2005 waves, with items ordered according to the 2002 wave. Note that these models are run for the full samples, i.e. with all-DK responses returned to the data set. A quick glance indicates that the overall patterns of probabilities remain consistent between the two time points, with the main differences found in the *risk* item. Respondents in the two support classes in 2005 are marginally *less* likely than their predecessors in the 2002 survey to say that GM food is risky, while those in the opposition classes in 2005 are notably *more* likely to judge GM food as risky than those belonging to these groups in 2002. It could be said that support and opposition are becoming more clearly defined by risk in 2005, or that *risk* has become more closely aligned with the other three items. However, whether this is a result of item ordering or of the passage of time is a matter that cannot be determined from these data alone.

**Table 5.9 Conditional probabilities, 5-class latent class models for GM food, GB, 2002 and 2005 waves**

| | | Definite support | | Support | | Opposition | | Definite opposition | | DK | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 2002 | 2005 | 2002 | 2005 | 2002 | 2005 | 2002 | 2005 | 2002 | 2005 |
| Item | Response category | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(5)$ | $\hat{\pi}_{is}(5)$ |
| Useful | DK | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.04 | 0.03 | 0.09 | 0.79 | 0.86 |
| | Definitely disagree | 0.00 | 0.00 | 0.02 | 0.00 | 0.07 | 0.02 | 0.67 | 0.71 | 0.03 | 0.00 |
| | Tend to disagree | 0.02 | 0.00 | 0.07 | 0.04 | 0.61 | 0.72 | 0.13 | 0.13 | 0.02 | 0.00 |
| | Tend to agree | 0.14 | 0.04 | 0.69 | 0.86 | 0.25 | 0.23 | 0.09 | 0.07 | 0.14 | 0.14 |
| | Definitely agree | 0.83 | 0.96 | 0.22 | 0.09 | 0.04 | 0.00 | 0.08 | 0.00 | 0.02 | 0.00 |
| Risky | DK | 0.06 | 0.16 | 0.07 | 0.13 | 0.01 | 0.03 | 0.03 | 0.04 | 0.81 | 0.82 |
| | Definitely disagree | 0.25 | 0.15 | 0.10 | 0.03 | 0.07 | 0.00 | 0.25 | 0.10 | 0.02 | 0.00 |
| | Tend to disagree | 0.21 | 0.29 | 0.28 | 0.43 | 0.31 | 0.08 | 0.05 | 0.01 | 0.00 | 0.00 |
| | Tend to agree | 0.31 | 0.21 | 0.38 | 0.40 | 0.39 | 0.73 | 0.12 | 0.06 | 0.10 | 0.17 |
| | Definitely agree | 0.16 | 0.19 | 0.18 | 0.02 | 0.23 | 0.17 | 0.55 | 0.78 | 0.07 | 0.01 |
| Morally acceptable | DK | 0.00 | 0.02 | 0.00 | 0.01 | 0.03 | 0.02 | 0.00 | 0.05 | 0.95 | 0.83 |
| | Definitely disagree | 0.05 | 0.03 | 0.01 | 0.03 | 0.01 | 0.06 | 0.82 | 0.72 | 0.00 | 0.05 |
| | Tend to disagree | 0.00 | 0.02 | 0.13 | 0.07 | 0.79 | 0.77 | 0.11 | 0.18 | 0.03 | 0.12 |
| | Tend to agree | 0.09 | 0.01 | 0.80 | 0.87 | 0.14 | 0.15 | 0.06 | 0.04 | 0.00 | 0.00 |
| | Definitely agree | 0.86 | 0.93 | 0.05 | 0.03 | 0.03 | 0.00 | 0.00 | 0.01 | 0.02 | 0.01 |
| Should be encouraged | DK | 0.02 | 0.02 | 0.05 | 0.12 | 0.00 | 0.02 | 0.01 | 0.00 | 0.92 | 0.89 |
| | Definitely disagree | 0.00 | 0.03 | 0.04 | 0.03 | 0.14 | 0.19 | 0.99 | 0.98 | 0.02 | 0.02 |
| | Tend to disagree | 0.03 | 0.14 | 0.13 | 0.25 | 0.81 | 0.79 | 0.00 | 0.02 | 0.01 | 0.06 |
| | Tend to agree | 0.08 | 0.26 | 0.73 | 0.56 | 0.05 | 0.00 | 0.00 | 0.00 | 0.03 | 0.03 |
| | Definitely agree | 0.87 | 0.54 | 0.06 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 |

**Table 5.10 Fit statistics, GB samples from 2002 and 2005 waves, GB, GM food**[10]

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | | | Jöreskog & Moustaki index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | All | 2002 | 2005 | All | 2002 | 2005 |
| Measurement model equal | 881 | 945 | <0.001 | -1,009 | -5,678 | 1.1 | 9.3 | 7.3 | 0.56 | 1.40 | 1.44 |
| *Use* free to differ between waves | 822 | 925 | <0.001 | -1,028 | -5,598 | 1.6 | 6.0 | 8.0 | 0.56 | 1.21 | 1.21 |
| *Risk* free to differ between waves | 778 | 925 | <0.001 | -1,072 | -5,641 | 0.0 | 4.0 | 0.7 | 0.35 | 0.82 | 0.79 |
| *Moral acceptability* free to differ between waves | 839 | 925 | <0.001 | -1,011 | -5,581 | 1.1 | 8.0 | 6.7 | 0.52 | 1.24 | 1.22 |
| *Encouragement* free to differ between waves | 835 | 925 | <0.001 | -1,015 | -5,585 | 1.6 | 11.3 | 10.7 | 0.53 | 1.29 | 1.33 |
| *Risk* and *use* free to differ between waves | 721 | 905 | <0.001 | -1,089 | -5,560 | 0.0 | 0.7 | 0.7 | 0.30 | 0.60 | 0.58 |
| *Risk*, *use* and *encouragement* free to differ between waves | 677 | 885 | <0.001 | -1,093 | -5,465 | 0.0 | 0.0 | 0.0 | 0.20 | 0.37 | 0.40 |
| Measurement model unconstrained | 650 | 865 | <0.001 | -1,080 | -5,353 | 0.5 | 0.7 | 0.0 | 0.23 | 0.32 | 0.33 |

[10] The residuals statistics in this table are given first for the 2002 and 2005 waves considered together ('All'), then separately, conditional on each wave in turn.

For a formal statistical comparison of the two five-class models, Table 5.10 gives fit statistics from a set of analyses where data from the two waves have been analysed together, with selected constraints applied to their measurement models. Likelihood ratio tests comparing models with single item-by-wave-by-class interactions with the fully constrained model are all significant at $p<0.05$, meaning that according to this test, every item functions differently in 2005 compared to 2002. Clearly, though, freeing the *risk* item results in the most dramatic improvement in fit, judging by all fit diagnostics in the table. So it seems that *risk* is the least stable item between waves.

### *5.4.4 Ordering the classes in subsamples without DK responses*

I have already mentioned that the relatively low conditional probabilities for the *risk* item, dispersed among response categories, might be taken to mean that it is weakly associated with the other three items. Another interpretation is to say that it represents a *difficult* item. *Utility* for example might be thought of as the easiest item, implying that agreeing that an application is useful would be a necessary but not a sufficient condition for agreeing that it should be encouraged. I explore this idea here with a brief consideration of some Guttman-type models.

In Guttman models items and classes are ordered simultaneously; at one end of the scale is the class where respondents give a negative answer to all items, and at the other end is the class where all items receive a positive response. In between, the items and classes are ordered so that in the second class in the scale respondents give a positive answer to only the easiest item, in the next they respond positively to the easiest and the next most difficult, and so on.

As noted in Chapter 4, the original Guttman model is deterministic: that is, probabilities for the response patterns defining the Guttman scale are constrained to be 1, and probabilities for responses that do not fit the pattern, 0. In practice, however, usually some responses will deviate from the Guttman pattern; these are counted as 'errors', and the proportion of these responses, for each item conditional on level of the scale, as an 'error rate'. Allowing response probabilities to take values other than 1 and 0, in other words allowing error rates to be different from 0, therefore amounts to modelling a probabilistic Guttman scale. As described in Chapter 4 (Section 4.1.8), these can be specified as restricted latent class models. For example, in the simplest of these, proposed by Proctor (1970), error rates are constrained to be equal across all items and all classes. Lazarsfeld and Henry (1968) proposed a more relaxed version of this, which allows error rates to vary between items.

Unfortunately Guttman-type models are only defined for dichotomous responses[11], so this means excluding all DK responses listwise from this set of analyses and collapsing responses over the intensifiers 'definitely' and 'tend to'. Despite this disadvantage the results are interesting enough to warrant this small diversion from the main path of the chapter.

The first step in the analysis is to determine the order of difficulty of the items. Judging from the frequencies of responses, collapsing over the intensifiers and excluding DK, *utility* is the most readily granted for both applications (61 per cent for GM food, 84 per cent for therapeutic cloning). For GM food the next easiest item is *moral acceptability*, with 52 per cent agreement, followed by 46 per cent agreeing to its general *encouragement*. With therapeutic cloning there is little to separate these two items, with 66 and 69 per cent agreement respectively. Lastly, regarding risk – considering *not-risky* rather than *risky*, so that the four items take the same direction on the scale – this item is the most difficult, with 38 per cent agreement for GM food and 35 per cent agreement for therapeutic cloning.

Table 5.11 gives fit statistics for a just a few illustrative, probabilistic versions of Guttman models. The Proctor model fits poorly for both applications, but freeing the error term for *not-risky* results in very well-fitting models. These are significantly better fitting according to likelihood ratio comparison tests ($p<0.01$), and clearly well-fitting in terms of marginal residuals and non-significant overall deviances ($L^2$). Any further relaxations of model constraints therefore seem unnecessary. For illustration, Lazarsfeld and Henry's (1968) model, allowing error rates to vary between items, is also presented. This returns a significant improvement in deviance for GM food ($p<0.01$) but not for therapeutic cloning ($p=0.58$).

[11] More elaborate unfolding models for polytomous items are not attempted since the spirit of this section is just to offer a brief example of different ways of modelling the latent space.

**Table 5.11 Fit statistics, Guttman-type models, GB**

| Model | $L^2$ | d.f. | p | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **GM food** | | | | | | | |
| Proctor model (equal error rates across items and classes) | 81 | 10 | <0.001 | 61 | 21 | 33.3 | 6.74 |
| Proctor model but with item-specific error for risk | 11 | 9 | 0.25 | -7 | -42 | 0.0 | 0.28 |
| Item-specific error rates for each item | 2 | 7 | 0.96 | -12 | -40 | 0.0 | 0.01 |
| **Therapeutic cloning** | | | | | | | |
| Proctor model (equal error rates across items and classes) | 77 | 10 | <0.001 | 57 | 17 | 37.5 | 5.23 |
| Proctor model but with item-specific error for risk | 8 | 9 | 0.51 | -10 | -46 | 0.0 | 0.18 |
| Item-specific error rates for each item | 7 | 7 | 0.42 | -7 | -35 | 0.0 | 0.12 |

Tables 5.12 and 5.13 show the conditional probabilities for the two modified Proctor models. Note that the ordering of item difficulty is different for therapeutic cloning than for GM food, although in practice, using the item ordering *useful*, *moral*, *encourage*, *not-risky* for therapeutic cloning makes only a slight difference to its fit. In both models the error rate for the risk item is quite high – around a third – mirroring its unpredictable behaviour in the unconstrained classes. Note also that the bulk of respondents, for both applications, are predicted to fall into the first two or last classes corresponding to Gaskell et al.'s classification of support, risk-tolerant support and opposition. Finally, echoing the findings from the first models of the chapter, for therapeutic cloning in particular a substantial proportion of respondents belong to the fourth class, where though respondents say that the application is useful they nevertheless answer negatively to the other three items.

**Table 5.12 Conditional probabilities, Proctor model with item-specific error for risk, GM food, GB**

| | Item | Category | Most positive $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | Most negative $\hat{\pi}_{is}(5)$ |
|---|---|---|---|---|---|---|---|
| 'Easiest' to agree | Useful | Agree | 0.92 | 0.92 | 0.92 | 0.92 | 0.08 |
| | | Disagree | 0.08 | 0.08 | 0.08 | 0.08 | 0.92 |
| | Morally acceptable | Agree | 0.92 | 0.92 | 0.92 | 0.08 | 0.08 |
| | | Disagree | 0.08 | 0.08 | 0.08 | 0.92 | 0.92 |
| | Should be encouraged | Agree | 0.92 | 0.92 | 0.08 | 0.08 | 0.08 |
| | | Disagree | 0.08 | 0.08 | 0.92 | 0.92 | 0.92 |
| Most 'difficult' to agree | Not risky | Agree | 0.67 | 0.33 | 0.33 | 0.33 | 0.33 |
| | | Disagree | 0.33 | 0.67 | 0.67 | 0.67 | 0.67 |
| | $\hat{\eta}_j$ | (unweighted) | 0.15 | 0.30 | 0.03 | 0.10 | 0.42 |

**Table 5.13 Conditional probabilities, Proctor model with item-specific error for risk, therapeutic cloning, GB**

| | Item | Category | Most positive $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | Most negative $\hat{\pi}_{is}(5)$ |
|---|---|---|---|---|---|---|---|
| 'Easiest' to agree | Useful | Agree | 0.95 | 0.95 | 0.95 | 0.95 | 0.05 |
| | | Disagree | 0.05 | 0.05 | 0.05 | 0.05 | 0.95 |
| | Should be encouraged | Agree | 0.95 | 0.95 | 0.95 | 0.05 | 0.05 |
| | | Disagree | 0.05 | 0.05 | 0.05 | 0.95 | 0.95 |
| | Morally acceptable | Agree | 0.95 | 0.95 | 0.05 | 0.05 | 0.05 |
| | | Disagree | 0.05 | 0.05 | 0.95 | 0.95 | 0.95 |
| Most 'difficult' to agree | Not risky | Agree | 0.66 | 0.34 | 0.34 | 0.34 | 0.34 |
| | | Disagree | 0.34 | 0.66 | 0.66 | 0.66 | 0.66 |
| | $\hat{\eta}_j$ | (unweighted) | 0.08 | 0.60 | 0.04 | 0.15 | 0.13 |

Before moving to the comparative analyses it would be useful to summarise the findings from the British data. Five-class models provide well fitting representations of responses to both GM food and therapeutic cloning. In both cases, these comprise classes that can be labelled from definite support, through moderate support and opposition to definite opposition, and a class where DK is the most likely response for each question. By and large, the classes are defined according to whether the biotechnologies are seen useful, morally acceptable and to be encouraged overall. Responses to these items mostly mirror each other, such that knowing a respondent's answer to one of these items, we could fairly confidently predict that he or she gave the same answer to the other two (we would be more confident of this for GM food than for therapeutic cloning). The risk item is an exception to this pattern, and for both GM food and therapeutic cloning behaves relatively independent of the other items. It is only in the definite opposition class that the most likely response for risk is clearly defined as definitely agree, and in the DK class, DK. So a distinction between risk-tolerant and risk-relaxed support does not emerge cleanly from the unconstrained models.

Although very similar, judgements of GM food and therapeutic cloning do not follow identical patterns in the British sample. For therapeutic cloning there is not such a close alignment of responses to utility with those for moral acceptability and overall encouragement. In all but the definite opposition and DK classes, respondents are most likely to view therapeutic cloning as useful. The proportions predicted to fall into the five classes are also a little different for GM food compared with therapeutic cloning, reflecting the well known trend of more support for red than for green applications of biotechnology.

Three short deviations from these main models add further shades of light to our interpretations of responses. Amongst those who give some but not all DK responses, a trend is evident for both biotechnologies towards agreeing that they are useful and risky. Whereas we might already expect to see this tendency towards acknowledging the utility of these applications, it is interesting that a trend towards a positive response can be observed for risk, given its indeterminacy across main support and opposition classes. Comparing the five class model for GM food with its equivalent from a more recent Eurobarometer survey, there is tentative evidence that amongst the population in general, responses to risk at the group level are becoming less unpredictable, and more closely aligned with the other three criteria of support. This inference comes with a caveat, however, that the changed pattern may be attributable to differences in question order in the two survey waves. An alternative way of thinking about the risk item is to say that it is not so much independent of the other items as just a more difficult item. Proctor models show that for both GM food and therapeutic cloning, the criteria of support can be ordered as if higher and higher hurdles on the scale of support, with the highest hurdle being to say that the application is not risky. That said, in order for the models to fit, the error rate attached to the risk item has to be freed, in which case it is in the order of one third. So although this ordering of criteria makes sense statistically, it only does so when the heterogeneity of responses to the risk item is part of the model.

## 5.5 Extending latent class analyses to other country samples

The second research question of the chapter is considered here: that is, how far can the summary variables derived in Section 5.4 be said to apply to other country samples, and how can fair comparisons be made between countries on the basis of them? Sample sizes for the countries in the 2002 Eurobarometer data set are given in Table 5.14. In this section, in order to reach the most inclusive possible representation of European opinions, I focus on the unconstrained latent class models for the full data sets, including DK responses and intensifiers.

Extending the analyses to other groups implies comparisons of the two different parts of these models, as set out in Chapter 4 (Section 4.1.8). We first need to study the measurement model, as denoted by the conditional probabilities – that is, the relationship between the items and the latent variable. If we can find a common measurement model across all groups (for us, countries) in the data, it gives us license to make comparisons between countries in terms of the structural model, here the prior probabilities – that is, the distribution of the latent variable.

**Table 5.14 Sample sizes for multiple group analyses**

| | Split ballot | | |
|---|---|---|---|
| | A (GM food) | B (therapeutic cloning) | Total |
| Austria | 529 | 479 | 1,008 |
| Belgium | 520 | 554 | 1,074 |
| Denmark | 505 | 495 | 1,000 |
| Finland | 506 | 494 | 1,000 |
| France | 484 | 520 | 1,004 |
| Germany | 1,028 | 1,017 | 2,045 |
| *(East* | *508* | *501* | *1,009)* |
| *(West* | *520* | *516* | *1,036)* |
| Greece | 498 | 503 | 1,001 |
| Ireland | 494 | 505 | 999 |
| Italy | 493 | 499 | 992 |
| Luxembourg | 295 | 304 | 599 |
| Netherlands | 482 | 516 | 998 |
| Portugal | 497 | 503 | 1,000 |
| Spain | 502 | 498 | 1,000 |
| Sweden | 500 | 500 | 1,000 |
| UK | 659 | 661 | 1,320 |
| *(Great Britain* | *508* | *506* | *1,014)* |
| *(Northern Ireland* | *151* | *155* | *306)* |
| Total | 7,992 | 8,048 | 16,040 |

### *5.5.1 Investigating the comparability of measurement models between subsamples of the UK and Germany*

As an introduction to the procedure used in the multiple group analyses, this section considers the preliminary question of whether it is statistically speaking valid to merge the two separately sampled regions of the UK (Britain and Northern Ireland) and Germany (East and West). Although there should be no serious concerns, in common sense terms, about merging them, since for each pair the questionnaire is in the same language and arguably the cultural settings are very similar, the fact of their being sampled separately suggests it would be prudent to check that their measurement models are equivalent rather than taking this for granted.

Table 5.15 gives fit statistics for models combining data from these two sets of samples. For Britain and Northern Ireland the most constrained models are the best fitting according to AIC and BIC, and are satisfactory in terms of numbers of marginal residuals. The model allowing conditional probabilities for risk to differ between samples is included for information, since informal inspection of the unconstrained models suggests that the greatest difference between them falls here. However, likelihood ratio comparison tests indicate no significant difference even between the unconstrained and completely constrained models (p=0.11 for GM food and p=0.10 for therapeutic cloning), so it is acceptable, according to this test, to treat the measurement models as equal and to pool the data from these two samples into one UK sample.

**Table 5.15 Fit statistics from testing measurement models between Great Britain and Northern Ireland, and between East and West Germany**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | | | Jöreskog & Moustaki index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **UK, GM food** | | | | | | All | GB | NI | All | GB | NI |
| Measurement model equal | 600 | 571 | <0.001 | -542 | -3,106 | 0.5 | 0.0 | 1.3 | 0.27 | 0.36 | 0.84 |
| *Risk* free to differ between samples | 572 | 551 | <0.001 | -530 | -3,004 | 0.5 | 0.0 | 0.7 | 0.23 | 0.33 | 0.56 |
| Measurement model unconstrained | 504 | 491 | <0.001 | -478 | -2,683 | 0.0 | 0.7 | 0.0 | 0.17 | 0.32 | 0.23 |
| **UK, therapeutic cloning** | | | | | | | | | | | |
| Measurement model equal | 569 | 573 | <0.001 | -577 | -3,152 | 0.0 | 0.7 | 2.0 | 0.30 | 0.52 | 0.79 |
| *Risk* free to differ between samples | 539 | 553 | <0.001 | -567 | -3,052 | 0.0 | 0.0 | 2.7 | 0.28 | 0.46 | 0.62 |
| Measurement model unconstrained | 472 | 493 | <0.001 | -514 | -2,730 | 0.0 | 0.7 | 0.0 | 0.19 | 0.33 | 0.28 |
| **Germany, GM food** | | | | | | All | West | East | All | West | East |
| Measurement model equal | 688 | 940 | <0.001 | -1,192 | -5,831 | 0.5 | 2.0 | 2.7 | 0.36 | 0.75 | 0.90 |
| *Use* free to differ between samples | 639 | 920 | <0.001 | -1,201 | -5,741 | 0.5 | 0.0 | 2.7 | 0.32 | 0.54 | 0.71 |
| *Risk* free to differ between samples | 657 | 920 | <0.001 | -1,183 | -5,723 | 0.5 | 2.7 | 1.3 | 0.32 | 0.61 | 0.71 |
| *Moral acceptability* free to differ between samples | 662 | 920 | <0.001 | -1,178 | -5,719 | 0.5 | 2.0 | 1.3 | 0.35 | 0.65 | 0.77 |
| *Encourage* free to differ between samples | 668 | 920 | <0.001 | -1,172 | -5,713 | 0.5 | 1.3 | 3.3 | 0.34 | 0.69 | 0.82 |
| Measurement model unconstrained | 562 | 860 | <0.001 | -1,158 | -5,403 | 0.0 | 0.0 | 0.0 | 0.23 | 0.28 | 0.31 |
| **Germany, therapeutic cloning** | | | | | | | | | | | |
| Measurement model equal | 709 | 929 | <0.001 | -1,149 | -5,724 | 2.1 | 2.0 | 6.0 | 0.50 | 0.82 | 0.94 |
| *Use* free to differ between samples | 686 | 909 | <0.001 | -1,132 | -5,609 | 2.1 | 2.0 | 2.7 | 0.43 | 0.65 | 0.82 |
| *Risk* free to differ between samples | 675 | 909 | <0.001 | -1,143 | -5,620 | 1.6 | 1.3 | 2.7 | 0.40 | 0.58 | 0.71 |
| *Moral acceptability* free to differ between samples | 689 | 909 | <0.001 | -1,129 | -5,605 | 2.6 | 1.3 | 3.3 | 0.49 | 0.75 | 0.95 |
| *Encourage* free to differ between samples | 674 | 909 | <0.001 | -1,144 | -5,620 | 1.6 | 0.7 | 5.3 | 0.48 | 0.71 | 0.86 |
| Measurement model unconstrained | 599 | 849 | <0.001 | -1,099 | -5,280 | 0.5 | 0.0 | 2.7 | 0.29 | 0.29 | 0.49 |

With Germany, the story is slightly more complicated. The greatest difference is found in responses regarding GM food and its usefulness. Although the likelihood ratio statistic is significant ($p<0.001$) for the difference between the unconstrained model and the model where use is free, Table 5.16 shows that the conditional probabilities for the latter are very similar between East and West Germany. AIC selects this model as the best fit, but BIC identifies the most constrained model as the best, and the marginal residuals suggest that the most constrained model fits well. It seems reasonable then to pool data from East and West Germany for GM food, and to argue that the likelihood ratio test is too sensitive for practical purposes here – perhaps due to sample size (with around 360 more German than UK respondents). On these grounds, results for therapeutic cloning suggest that combining samples from East and West Germany is satisfactory: p=0.015 for the difference in deviance between the most constrained and the unconstrained model; both AIC and BIC select the most constrained model as the best; and this choice is supported by low marginal residuals.

**Table 5.16 Conditional probabilities for East and West Germany GM food, with conditional probabilities free to vary for *utility* but fixed equal for other items**

| Item | Response category | Definite support $\hat{\pi}_{is}(1)$ | Support $\hat{\pi}_{is}(2)$ | Opposition $\hat{\pi}_{is}(3)$ | Definite opposition $\hat{\pi}_{is}(4)$ | DK $\hat{\pi}_{is}(5)$ |
|---|---|---|---|---|---|---|
| Useful (West) | DK | 0.00 | 0.02 | 0.04 | 0.01 | 0.74 |
| | Definitely disagree | 0.00 | 0.00 | 0.07 | 0.79 | 0.03 |
| | Tend to disagree | 0.02 | 0.12 | 0.72 | 0.19 | 0.00 |
| | Tend to agree | 0.17 | 0.73 | 0.17 | 0.01 | 0.18 |
| | Definitely agree | 0.81 | 0.13 | 0.00 | 0.00 | 0.05 |
| Useful (East) | DK | 0.00 | 0.00 | 0.04 | 0.09 | 0.87 |
| | Definitely disagree | 0.00 | 0.00 | 0.00 | 0.69 | 0.00 |
| | Tend to disagree | 0.08 | 0.02 | 0.76 | 0.18 | 0.03 |
| | Tend to agree | 0.04 | 0.81 | 0.16 | 0.01 | 0.08 |
| | Definitely agree | 0.88 | 0.17 | 0.04 | 0.02 | 0.03 |
| Risky | DK | 0.03 | 0.08 | 0.02 | 0.03 | 0.87 |
| | Definitely disagree | 0.18 | 0.03 | 0.01 | 0.10 | 0.01 |
| | Tend to disagree | 0.35 | 0.39 | 0.17 | 0.02 | 0.01 |
| | Tend to agree | 0.24 | 0.38 | 0.61 | 0.22 | 0.05 |
| | Definitely agree | 0.21 | 0.12 | 0.20 | 0.63 | 0.05 |
| Morally acceptable | DK | 0.00 | 0.03 | 0.09 | 0.00 | 0.90 |
| | Definitely disagree | 0.00 | 0.00 | 0.03 | 0.78 | 0.01 |
| | Tend to disagree | 0.01 | 0.06 | 0.76 | 0.20 | 0.00 |
| | Tend to agree | 0.23 | 0.81 | 0.11 | 0.01 | 0.08 |
| | Definitely agree | 0.76 | 0.10 | 0.00 | 0.00 | 0.00 |
| Should be encouraged | DK | 0.02 | 0.06 | 0.07 | 0.02 | 0.99 |
| | Definitely disagree | 0.01 | 0.01 | 0.14 | 0.96 | 0.00 |
| | Tend to disagree | 0.02 | 0.10 | 0.78 | 0.02 | 0.00 |
| | Tend to agree | 0.08 | 0.81 | 0.00 | 0.00 | 0.00 |
| | Definitely agree | 0.87 | 0.03 | 0.00 | 0.00 | 0.01 |
| (unweighted) $\hat{\eta}_j$ | West Germany | 0.13 | 0.31 | 0.19 | 0.19 | 0.19 |
| | East Germany | 0.09 | 0.38 | 0.23 | 0.18 | 0.11 |

### *5.5.2 Qualitative comparisons of measurement models in country-by-country analyses*

I now embark on the multiple group analysis proper. Here it would be far less surprising to see different combinations of latent classes between samples, given their correspondingly different cultural contexts, and perhaps most obviously, language differences. Before testing the comparability of measurement models for the full fifteen samples, an informal qualitative evaluation was carried out by running separate models for each country and each type of biotechnology. Table 5.17 summarises the patterns of conditional probabilities found across countries for GM food, and Table 5.18 for therapeutic cloning. Here '++' indicates that 'definitely agree' is the most likely response for a class, '+' corresponds to 'agree', etc., and '?' for DK. Where probabilities are close for two or more responses (e.g. '+/–'), they are listed in order from most to least likely[12].

For GM food, patterns of responses across classes are notably different in qualitative terms only for the risk item – particularly within the classes of definite support and moderate opposition. In Table 5.17 under *risky* the countries are ordered approximately from the smallest to the greatest likelihood, combined across classes, of agreeing that GM food is risky. Thus in the countries at the end of the list (Austria, Italy, Ireland and Luxembourg), respondents are likely to say that GM food is risky across all classes except the DK class, while at the upper end, in Greece, Sweden and Germany, those in the strong support class are most likely to disagree that it is risky. Only in Finland is the pattern of responses to risk is exactly aligned with the other items.

[12] Again, 'close' is defined by a difference of 0.10 (when rounded to two decimal places) or less from the most likely response.

**Table 5.17 Qualitative summaries of highest conditional probabilities from unconstrained 5-class models, GM food, 15 countries**

| | Definite support | Support | Opposition | Definite opposition | DK |
|---|---|---|---|---|---|
| **Useful** | | | | | |
| All countries | ++ | + | – | – – | ? |
| **Risky** | | | | | |
| Finland | – – | – | + | ++ | ? |
| Greece, Sweden | – – | + | + | ++ | ? |
| Denmark | – –/– | + | +/++ | ++ | ? |
| Germany | – | –/+ | + | ++ | ? |
| UK | – –/+ | + | + | ++ | ? |
| France | – –/+ | + | +/++ | ++ | ? |
| Belgium | –/++/+/– – | + | +/++/– | ++ | ? |
| Spain | +/++/–/– – | + | + | ++ | ? |
| Netherlands | ++/– –/+ | + | + | ++ | ? |
| Portugal | ++/–/– –/+ | + | + | ++ | ? |
| Austria | ++ | + | – | ++ | ? |
| Italy | ++/+ | + | + | ++ | ? |
| Ireland | ++ | + | ++ | ++ | ? |
| Luxembourg | ++ | +/++ | ++ | ++ | ? |
| **Morally acceptable** | | | | | |
| All except those listed below | ++ | + | – | – – | ? |
| Netherlands | ++ | + | +/– | – – | ? |
| Spain | ++ | + | –/+ | – – | ? |
| **Should be encouraged** | | | | | |
| All countries | ++ | + | – | – – | ? |

For therapeutic cloning, alongside differences on risk, a range of patterns are observed for *utility*. In Table 5.18 countries are ordered in this section from most to least positive. In the majority of countries all but the strongest opponents agree that it is useful (or say DK, if in a DK class). The rough comparisons suggest that in Europe-wide models for GM food and therapeutic cloning we might reasonably expect to be able to constrain the measurement models for *moral acceptability* and *encouragement* between countries, but need to be cautious with regard to *risk* and possibly to *utility*.

**Table 5.18 Qualitative summaries of highest conditional probabilities from unconstrained 5-class models, therapeutic cloning, 15 countries**

| | Definite support | Support | Opposition | Definite opposition | DK |
|---|---|---|---|---|---|
| **Useful** | | | | | |
| Denmark | ++ | ++ | ++/+ | – – | ? |
| Belgium, UK | ++ | +/++ | + | – – | ? |
| Spain | ++ | +/++ | +/- | + | ? |
| Luxembourg | ++ | + | ++ | – –/– | ? |
| Ireland | ++ | + | +/++ | – – | ? |
| Finland, France | ++ | + | + | – – | ? |
| Austria, Italy, Netherlands | ++ | + | +/– | – – | ? |
| Portugal | ++ | + | – | – –/+ | ? |
| Greece | ++ | ++ | – | – – | ? |
| Germany, Portugal, Sweden | ++ | + | – | – – | ? |
| **Risky** | | | | | |
| Finland | – – | – | + | ++ | ? |
| Germany | – | +/– | + | ++ | ? |
| Italy | –/++ | + | +/–/++ | – –/++ | ? |
| Greece | – –/–/++ | + | –/++ | ++ | ? |
| Austria | – –/+/++ | + | –/+ | ++ | ? |
| Sweden | – –/+ | + | + | ++ | ? |
| UK | ++/+/– –/– | + | +/++/– | ++ | ? |
| Spain | +/– –/+ | + | +/++ | ++ | ? |
| Denmark | +/– – | + | ++ | ++ | ? |
| Luxembourg | ++ | + | –/++ | ++ | ? |
| Belgium | +/++ | + | ++ | ++ | ? |
| Netherlands | ++/+ | + | + | ++ | ? |
| Portugal | ++ | + | + | ++ | ? |
| France, Ireland | ++ | + | ++ | ++ | ? |
| **Morally acceptable** | | | | | |
| All except those listed below | ++ | + | – | – – | ? |
| Spain | ++ | + | + | – – | ? |
| Ireland | ++ | + | –/? | – – | ? |
| Greece | ++ | + | –/– – | – – | ? |
| **Should be encouraged** | | | | | |
| All except those listed below | ++ | + | – | – – | ? |
| Luxembourg | ++ | + | + | – – | ? |
| Ireland | ++ | + | ?/– | – – | ? |

### *5.5.3 Joint models with fixed effects to describe attitudes across fifteen EU countries*

Table 5.19 gives fit statistics for a selection of models. The first specifies that not only is the measurement model the same for every country, but so is the distribution of the latent variable (i.e. the probabilities of the latent classes). Clearly it fits very poorly for both GM food and therapeutic cloning. A model where the measurement model remains fully constrained but the distribution of the latent variable is free between countries is also a poor fit – at least, certainly prohibitively poor for therapeutic cloning, though not so bad for GM food. Generally it is more difficult to achieve a well fitting model for therapeutic cloning than for GM food.

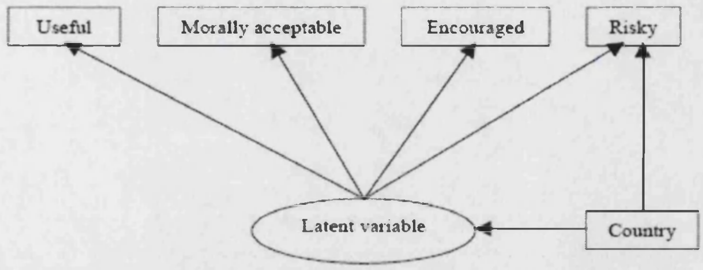
As a strategy for finding a satisfactory joint model for each biotechnology, the exploratory analyses of the previous section indicate that direct and interaction effects between country and item should be tested for *risk* and for *utility*. For clarity, Figure 5.1 illustrates the difference between the two models conceptually, with the example of *risk* as the item whose parameters are freed (see also Section 4.1.8). The direct effects model allows the probabilities of particular responses to *risk* to vary between countries, but in the same way for each class. The interaction model allows the country effects on response probabilities to vary between classes.

Besides considering direct and interaction effects for two items, freeing any more parameters for purposes of fit would really be undesirable. Although it would seem reasonable that a cross-country model would require some movement in one item, with any more relaxations of parameters it would be difficult to speak of a latent variable with a common interpretation between countries. An alternative modification to try to improve model fit is to increase the number of latent classes; six-class models are therefore also tested, and their fits presented below.

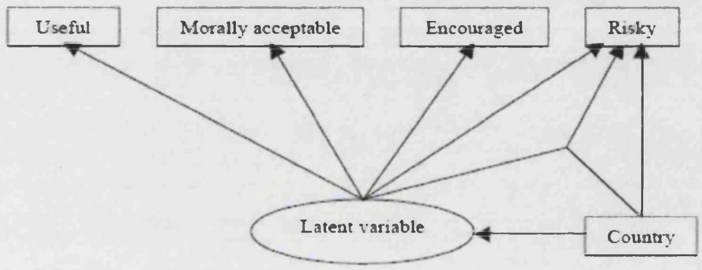
For most measurement models, AIC BIC and residuals favour six- over five-class models. The following discussion therefore focuses on the former. It is clear from the residuals that some sort of relaxation of the parameters for risk results in the best fitting models, for both GM food and therapeutic cloning. Relaxing parameters for *utility* does little to improve fit for therapeutic cloning, and for GM food although it results in some improvement in fit, the effect is by no means as marked as that resulting from freeing *risk*. To select the final models, then, it is simply a matter of choosing between direct and interaction effects for *risk*.

**Figure 5.1 Direct effects and interactions for the logics items**

a. Direct effect between country and risk




b. Interaction between country, risk and latent variable




In choosing between these models, we need to give the diagnostics for marginal residuals some close inspection. Although the overall figures indicate little difference between direct and interaction effects, when broken down by country (see Tables A.1-A.4 in the Appendix to the thesis), they reveal a great improvement in fit for most countries when moving from direct to interaction effects. For example, in the six-class model for therapeutic cloning with direct effects between country and *risk*, the percentage of two-way marginal residuals ranges from 4.0 in Spain to 28.7 in Germany, with a mean across countries of 11.1. With the interaction effect these are reduced to 2.7 for Spain and 18.0 for Germany, with an average across countries of 6.4. In the six-class model for GM food with direct effects between country and *risk*, the percentage of two-way marginal residuals ranges from 1.3 in the UK to 24.0 in Germany, with an average across countries of 8.2. With the interaction effect these are reduced to between 0.0 for Luxembourg and 15.3 for Germany, with an average across countries of 5.2. For GM food, the same six-class model with *risk* freed is deemed to be the best fit here. The sixth class is not absolutely crucial to model fit, but is retained for convenience of comparisons with therapeutic cloning, to anticipate the analyses to come in Chapter 8.

**Table 5.19 Fit statistics, testing measurement models between 15 countries**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **GM food, 5 classes** | | | | | | | |
| Country independent of latent variable and indicators | 8,364 | 7,908 | <0.001 | -7,452 | -62,699 | 39.3 | 5.22 |
| Measurement model equal | 7,396 | 7,852 | <0.001 | -8,308 | -63,164 | 7.6 | 1.13 |
| Direct country effect for *risk* | 7,042 | 7,796 | <0.001 | -8,550 | -63,014 | 2.7 | 0.77 |
| Interaction between country, latent variable and *risk* | 6,535 | 7,572 | <0.001 | -8,609 | -61,509 | 2.4 | 0.78 |
| Direct country effect for *utility* | 7,140 | 7,796 | <0.001 | -8,452 | -62,916 | 6.4 | 1.02 |
| Interaction between country, latent variable and *utility* | 6,815 | 7,572 | <0.001 | -8,329 | -61,228 | 6.4 | 1.05 |
| Measurement model free to differ between countries | 4,950 | 6,732 | <0.001 | -8,514 | -55,545 | 1.1 | 0.5 |
| **GM food, 6 classes** | | | | | | | |
| Country independent of latent variable and indicators | 8,013 | 7,891 | <0.001 | -7,769 | -62,898 | 38.7 | 5.03 |
| Measurement model equal | 6,924 | 7,821 | <0.001 | -8,718 | -63,357 | 6.0 | 0.91 |
| Direct country effect for *risk* | 6,578 | 7,765 | <0.001 | -8,952 | -63,200 | 1.3 | 0.57 |
| Interaction between country, latent variable and *risk* | 6,050 | 7,485 | <0.001 | -8,920 | -61,212 | 1.3 | 0.58 |
| Direct country effect for *utility* | 6,651 | 7,765 | <0.001 | -8,879 | -63,127 | 6.2 | 0.82 |
| Interaction between country, latent variable and *utility* | 6,330 | 7,485 | <0.001 | -8,640 | -60,931 | 5.3 | 0.82 |
| Measurement model free to differ between countries | 4,393 | 6,477 | <0.001 | -8,561 | -53,811 | 0.2 | 0.32 |
| **Therapeutic cloning, 5 classes** | | | | | | | |
| Country independent of latent variable and indicators | 8,589 | 7,964 | <0.001 | -7,339 | -63,032 | 34.9 | 5.39 |
| Measurement model equal | 7,806 | 7,908 | <0.001 | -8,010 | -63,312 | 14.9 | 2.03 |
| Direct country effect for *risk* | 7,200 | 7,852 | <0.001 | -8,504 | -63,415 | 7.8 | 1.40 |
| Interaction between country, latent variable and *risk* | 6,580 | 7,628 | <0.001 | -8,676 | -62,020 | 7.8 | 1.37 |
| Direct country effect for *utility* | 7,494 | 7,852 | <0.001 | -8,210 | -63,121 | 12.2 | 1.83 |
| Interaction between country, latent variable and *utility* | 7,110 | 7,628 | <0.001 | -8,146 | -61,490 | 13.1 | 1.81 |
| Measurement model free to differ between countries | 4,820 | 6,788 | <0.001 | -8,756 | -56,225 | 3.6 | 0.91 |
| **Therapeutic cloning, 6 classes** | | | | | | | |
| Country independent of latent variable and indicators | 8,089 | 7,947 | <0.001 | -7,805 | -63,380 | 33.1 | 4.98 |
| Measurement model equal | 7,140 | 7,877 | <0.001 | -8,614 | -63,699 | 12.7 | 1.57 |
| Direct country effect for *risk* | 6,533 | 7,821 | <0.001 | -9,109 | -63,803 | 5.6 | 0.88 |
| Interaction between country, latent variable and *risk* | 5,887 | 7,541 | <0.001 | -9,195 | -61,931 | 5.8 | 0.89 |
| Direct country effect for *utility* | 6,880 | 7,821 | <0.001 | -8,762 | -63,456 | 10.4 | 1.41 |
| Interaction between country, latent variable and *utility* | 6,477 | 7,541 | <0.001 | -8,605 | -61,340 | 10.7 | 1.40 |
| Measurement model free to differ between countries | 4,150 | 6,533 | <0.001 | -8,916 | -54,603 | 1.6 | 0.52 |

Tables 5.20 and 5.21 summarise the measurement model for GM food. The first table gives the conditional probabilities for *utility*, *moral acceptability* and *encouragement*, which are fixed to be the same for each country. The contribution of the final class is to allow for respondents who say DK to the last two items but tend to agree that GM food is useful, echoing the analysis of the British 'sometimes-DK' respondents earlier in the chapter. The second table summarises differences between countries for *risk*, just as in the qualitative analysis. There is little change here from the comparable table in the initial five-class qualitative analysis, aside from a new distinction between countries where the partly unsure say that GM food is risky (Austria, Denmark, France, Ireland, Portugal, Spain, and the UK) and those where the partly unsure say DK.

Tables 5.22 and 5.23 give the corresponding information for therapeutic cloning. Here, compared with responses to GM food, those in the second DK class are more likely, on balance, to agree or definitely agree to the *utility* item – reflecting this general tendency in the population. Once again the qualitative summary of responses to the *risk* item distinguish those countries where the propensity to agree that therapeutic cloning is risky extends to those in the partly-unsure class – in France, Germany, Ireland, Italy, Netherlands, Portugal, Spain, and the UK.

As a last note on the selection of these models over the direct effects models, it seems that the selection of one over the other has little effect in terms of the substantive interpretation of the classes. For example, the differences between the fitted probabilities in the six-class GM food models with direct and interaction effects for *risk* are mainly found in two classes. In the DK (2) class, people in some countries are more likely in one model than the other to agree that GM food is risky. There is also some difference in interpretation for *risk* in the definite support class, but not a clear difference – in both models this class represents the most mixed and least interpretable collection of responses to *risk*. So the differences between the two models are rather subtle. It is not the case that the direct effects model drastically misrepresents responses to *risk* in certain countries – rather the difference is in the detail of the second DK class, and in the definite support class, in which responses to *risk* are very indeterminate in any case. This highlights again that statistical fit and substantive interpretation do not always go hand in hand; it is not the case here that the slightly better-fitting interaction effect model offers a much clearer interpretation than the direct effects model – indeed in terms of substantive output, it is difficult to pinpoint exactly what value it adds compared to the direct effects model.

**Table 5.20 GM food: conditional probabilities for three items where measurement model is equal across countries, 15 countries**

| | | Definite support | Support | Opposition | Definite opposition | DK | DK (2) |
|---|---|---|---|---|---|---|---|
| Item | Response category | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(5)$ | $\hat{\pi}_{is}(6)$ |
| Useful | DK | 0.00 | 0.01 | 0.02 | 0.02 | 0.98 | 0.24 |
| | Definitely disagree | 0.01 | 0.01 | 0.09 | 0.82 | 0.01 | 0.08 |
| | Tend to disagree | 0.02 | 0.07 | 0.64 | 0.10 | 0.00 | 0.15 |
| | Tend to agree | 0.11 | 0.72 | 0.21 | 0.03 | 0.00 | 0.35 |
| | Definitely agree | 0.86 | 0.19 | 0.04 | 0.03 | 0.01 | 0.19 |
| Morally acceptable | DK | 0.01 | 0.01 | 0.03 | 0.03 | 0.99 | 0.53 |
| | Definitely disagree | 0.02 | 0.01 | 0.08 | 0.83 | 0.00 | 0.07 |
| | Tend to disagree | 0.03 | 0.09 | 0.71 | 0.08 | 0.00 | 0.11 |
| | Tend to agree | 0.14 | 0.81 | 0.16 | 0.04 | 0.00 | 0.20 |
| | Definitely agree | 0.80 | 0.09 | 0.02 | 0.02 | 0.00 | 0.10 |
| Should be encouraged | DK | 0.01 | 0.03 | 0.03 | 0.01 | 0.99 | 0.71 |
| | Definitely disagree | 0.01 | 0.02 | 0.20 | 0.96 | 0.00 | 0.05 |
| | Tend to disagree | 0.02 | 0.09 | 0.72 | 0.02 | 0.00 | 0.08 |
| | Tend to agree | 0.12 | 0.81 | 0.03 | 0.00 | 0.00 | 0.13 |
| | Definitely agree | 0.83 | 0.05 | 0.01 | 0.01 | 0.00 | 0.03 |

**Table 5.21 GM food: qualitative summary of highest conditional probabilities for the risk item, 15 countries**

| | Definite support | Support | Opposition | Definite opposition | DK | DK (2) |
|---|---|---|---|---|---|---|
| Finland | – – | – | + | ++ | ? | ? |
| Greece | – – | +/– | +/++ | ++ | ? | ? |
| Sweden | – – | + | + | ++ | ? | ? |
| Denmark | – –/- | + | + | ++ | ? | ++/? |
| Germany | - | –/+ | + | ++ | ? | ? |
| France | – –/+ | + | +/++ | ++ | ? | ++ |
| UK | – –/+ | + | + | ++ | ? | ?/+ |
| Netherlands | – –/++/+ | + | + | ++ | ? | ? |
| Belgium | –/++/+/– – | + | +/++/– | ++ | ? | ? |
| Portugal | ++/–/– –/+ | + | +/– | ++ | ? | +/++ |
| Spain | +/++/–/– – | + | ++/+ | ++ | ? | +/? |
| Italy | ++/+ | + | + | ++ | ? | ? |
| Austria | ++/+ | + | – | ++ | ? | ++ |
| Luxembourg | ++ | +/++ | ++ | ++ | ? | ? |
| Ireland | ++ | + | ++/+ | ++ | ? | ++/?/+ |

**Table 5.22 Therapeutic cloning: conditional probabilities for three items where measurement model is equal across countries, 15 countries**

| | | Definite support | Support | Opposition | Definite opposition | DK | DK (2) |
|---|---|---|---|---|---|---|---|
| Item | Response category | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(5)$ | $\hat{\pi}_{is}(6)$ |
| Useful | DK | 0.00 | 0.00 | 0.03 | 0.04 | 0.98 | 0.13 |
| | Definitely disagree | 0.01 | 0.00 | 0.04 | 0.62 | 0.00 | 0.04 |
| | Tend to disagree | 0.01 | 0.02 | 0.41 | 0.16 | 0.00 | 0.05 |
| | Tend to agree | 0.05 | 0.61 | 0.39 | 0.10 | 0.00 | 0.39 |
| | Definitely agree | 0.93 | 0.37 | 0.13 | 0.08 | 0.01 | 0.39 |
| Morally acceptable | DK | 0.01 | 0.02 | 0.02 | 0.02 | 1.00 | 0.51 |
| | Definitely disagree | 0.02 | 0.01 | 0.10 | 0.85 | 0.00 | 0.12 |
| | Tend to disagree | 0.03 | 0.11 | 0.69 | 0.09 | 0.00 | 0.15 |
| | Tend to agree | 0.19 | 0.83 | 0.17 | 0.01 | 0.00 | 0.15 |
| | Definitely agree | 0.75 | 0.04 | 0.01 | 0.04 | 0.00 | 0.06 |
| Should be encouraged | DK | 0.01 | 0.02 | 0.05 | 0.01 | 1.00 | 0.62 |
| | Definitely disagree | 0.00 | 0.01 | 0.10 | 0.91 | 0.00 | 0.02 |
| | Tend to disagree | 0.01 | 0.03 | 0.77 | 0.05 | 0.00 | 0.05 |
| | Tend to agree | 0.07 | 0.88 | 0.07 | 0.01 | 0.00 | 0.24 |
| | Definitely agree | 0.91 | 0.07 | 0.01 | 0.02 | 0.00 | 0.07 |

**Table 5.23 Therapeutic cloning: summary of highest conditional probabilities for the risk item, 15 countries**

| | Definite support | Support | Opposition | Definite opposition | DK | DK(2) |
|---|---|---|---|---|---|---|
| Finland | – – | – | + | ++ | ? | ? |
| Germany | – | –/+ | + | ++ | ? | + |
| Greece | – –/–/+/++ | + | –/++ | ++ | ? | ? |
| Italy | –/++/+ | + | –/+/++ | – – | ? | ++/? |
| Sweden | – –/+ | + | + | ++ | ? | ? |
| Austria | +/– –/++ | + | +/– | ++ | ? | ? |
| Spain | +/– – | + | +/– | ++ | ? | ++/+/? |
| Denmark | +/– – | + | ++ | ++ | ? | ? |
| UK | +/++/–/– – | + | ++/+ | ++ | ? | + |
| Belgium | +/++ | + | ++ | ++ | ? | ? |
| Netherlands | ++/+ | + | + | ++ | ? | +/? |
| Portugal | ++ | +/– | + | ++ | ? | ++/?/+ |
| Ireland | ++ | + | + | ++ | ? | ++ |
| Luxembourg | ++ | + | ++ | ++ | ? | ? |
| France | ++ | + | ++ | ++ | ? | ?/++/+ |

Finally, Table 5.24 gives for GM food and therapeutic cloning the estimated percentages of respondents belonging to each latent class within country. These are recalculated from the final models using the sampling weights to adjust for over- or under-sampling in different population strata (as described in Section 4.1.12). The last columns in the table give percentages of respondents in definite and moderate support together, likewise for the two opposition classes and for the two DK classes. Countries are listed in order of the proportions of respondents in the support classes, from most to least supportive overall.

**Table 5.24 Weighted percentages of respondents per class, GM food and therapeutic cloning, 15 countries**

| *% within country* | Def. support | Support | Oppos'n | Def. oppos'n | DK | DK (2) | TOTAL: SUPPORT | TOTAL: OPPOSITION | TOTAL: DK |
|---|---|---|---|---|---|---|---|---|---|
| **GM food** | | | | | | | | | |
| Spain | 18 | 30 | 23 | 13 | 12 | 5 | 47 | 36 | 17 |
| Germany | 20 | 25 | 11 | 12 | 18 | 14 | 45 | 23 | 32 |
| Portugal | 14 | 26 | 24 | 16 | 14 | 6 | 40 | 40 | 20 |
| Netherlands | 13 | 24 | 17 | 12 | 22 | 11 | 37 | 29 | 33 |
| Denmark | 11 | 27 | 30 | 14 | 11 | 7 | 37 | 44 | 18 |
| Sweden | 12 | 23 | 23 | 23 | 16 | 3 | 35 | 46 | 19 |
| Belgium | 16 | 19 | 19 | 34 | 4 | 8 | 35 | 53 | 12 |
| Luxembourg | 9 | 25 | 10 | 15 | 39 | 2 | 34 | 25 | 41 |
| Austria | 13 | 21 | 26 | 19 | 17 | 4 | 34 | 45 | 21 |
| Ireland | 16 | 16 | 8 | 14 | 37 | 9 | 31 | 22 | 46 |
| UK | 10 | 19 | 24 | 32 | 11 | 4 | 29 | 55 | 15 |
| France | 12 | 12 | 18 | 31 | 16 | 11 | 24 | 49 | 27 |
| Italy | 7 | 16 | 20 | 39 | 13 | 3 | 24 | 59 | 17 |
| Greece | 6 | 12 | 27 | 24 | 28 | 3 | 18 | 51 | 31 |
| Finland | 10 | 5 | 18 | 37 | 23 | 7 | 15 | 55 | 30 |
| Europe total (pop. weighted) | 12 | 21 | 22 | 20 | 18 | 7 | 33 | 42 | 25 |
| **Therapeutic cloning** | | | | | | | | | |
| Denmark | 39 | 26 | 13 | 9 | 5 | 8 | 65 | 22 | 13 |
| Sweden | 33 | 31 | 11 | 9 | 13 | 4 | 63 | 20 | 16 |
| Luxembourg | 33 | 29 | 6 | 6 | 20 | 6 | 62 | 13 | 26 |
| Belgium | 29 | 31 | 13 | 9 | 11 | 7 | 60 | 21 | 19 |
| Finland | 25 | 34 | 15 | 10 | 8 | 8 | 59 | 25 | 16 |
| Spain | 32 | 26 | 7 | 4 | 17 | 14 | 57 | 11 | 32 |
| UK | 22 | 33 | 17 | 8 | 12 | 8 | 56 | 25 | 19 |
| Italy | 23 | 32 | 16 | 6 | 11 | 11 | 56 | 23 | 22 |
| Portugal | 25 | 29 | 10 | 6 | 19 | 10 | 55 | 16 | 29 |
| France | 26 | 29 | 9 | 13 | 14 | 9 | 54 | 22 | 23 |
| Germany | 18 | 32 | 21 | 15 | 7 | 8 | 50 | 35 | 15 |
| Austria | 21 | 23 | 22 | 14 | 13 | 7 | 44 | 36 | 20 |
| Netherlands | 16 | 26 | 6 | 10 | 40 | 2 | 42 | 16 | 42 |
| Greece | 25 | 14 | 11 | 16 | 10 | 23 | 39 | 27 | 33 |
| Ireland | 20 | 17 | 5 | 9 | 37 | 11 | 37 | 14 | 48 |
| Europe total (pop. weighted) | 24 | 30 | 14 | 10 | 13 | 9 | 53 | 24 | 22 |

A few features of the table are worth pointing out. Firstly, it is easy to see that the figures in the total support and total opposition columns echo the well known pattern of more support for medical than for agri-food applications of biotechnology. In all countries except Germany, greater proportions of supporters are found for therapeutic cloning than GM food, and vice versa for proportions of opponents. For therapeutic cloning, in every country there are more supporters than opponents. For GM food the pattern is reversed for the majority of countries, and it is just in Germany, Ireland, Luxembourg, Spain and the Netherlands that we find more supporters than opponents. The overall range of total supporters across countries is roughly the same for GM food and therapeutic cloning (32 and 28 per cent ranges respectively). By contrast, negative perceptions are more widely spread between countries for GM food (37 per cent range) than for therapeutic cloning (25 per cent). The range of respondents falling into the DK classes is roughly the same for both applications, and notably large (35 and 36 per cent). Ireland, Luxembourg and the Netherlands are conspicuous in their high numbers of DK respondents. This has a considerable impact on the comparisons of opinions between countries. For example, for both biotechnologies Ireland has a high ratio of supporters to opponents, but its large numbers of DK respondents cause it to appear quite low down the list of supporters of GM food, and at the bottom of the table for therapeutic cloning.

## Summary

The models in this chapter have resulted in a number of substantive, methodological and practical findings. The substantive findings concern the distributions and compositions of the latent variables – that is, the content of the models. The methodological findings, broadly stated, relate to item functioning. The practical implications of the models include both findings concerning technical issues encountered in fitting those models, and practical recommendations for future survey design and administration. These three kinds of results are ideal types; most findings contain elements of all three themes. For example, the findings regarding the *risk* item could be thought of as a methodological matter of item functioning; or as a substantively interesting finding about the relationship of risk perceptions to other criteria of support for biotechnology; and they might suggest on a practical level that risk as a topic should be investigated in more detail in future surveys.

The summary of findings in this and in the following empirical chapters will focus on the results from the cross-national analyses, with the results from analysing the British sample adding further possible insights to these main patterns.

Across the fifteen European countries analysed:

- For both GM food and therapeutic cloning, three of the criteria of support – *utility*, *moral acceptability* and overall *encouragement* – are broadly positively associated with each other. The other criteria put to respondents – the *riskiness* of the biotechnologies – is rather weakly and in a broad sense negatively associated with the other items.
- For both GM food and therapeutic cloning, we can adequately represent the variation in responses to these items in six classes.
- For GM food, four of these classes form a common sense ordering, from definite opposition to definite support, defined in terms of responses to the three criteria of *utility*, *moral acceptability* and overall *encouragement*, where a particular response to one of these items (say 'tend to agree') will be a good predictor of the same response on the other two.
- Four very similar classes are found for therapeutic cloning. The main difference between the patterns of responses for therapeutic cloning and GM food is that in the former, those in the moderate opposition class are almost as likely to agree as to disagree that the biotechnology is useful.
- The other two classes are given to DK responses; one in which DK is the most likely answer to every question, and one in which the utility of the biotechnology (whether GM food or therapeutic cloning) is likely to be acknowledged.
- The behaviour of the *risk* item varies between countries – indeed, the fit of a Europe-wide model hinges on freeing this parameter between countries. However, this variation is not completely without pattern:
  - In most countries, those in the opposition classes are likely to agree (to varying degrees – usually most definitely in the definite opposition class) that the biotechnology in question is risky.
  - In most countries, those in the moderate support class are also likely to agree that the biotechnology in question is risky.
  - The greatest variation between countries in responses to *risk* are found in the definite support class. In some countries, agreement is most likely, in others disagreement, and in others still, there is no clearly identifiable most likely response.
  - In all countries, those in the first DK class are likely to give a DK response also to the *risk* item.

    - Responses to *risk* in the second DK class are mixed, with DK the most likely response in some countries and agreement the most likely response in others.
- Europe-wide, the balance of favour rests with therapeutic cloning rather than with GM food. Whereas a third of Europeans are estimated to support GM food (moderately or definitely), more than half are estimated to support therapeutic cloning. In all countries except Germany, there are greater proportions of supporters for therapeutic cloning than for GM food, and vice versa for proportions of opponents.
    - For therapeutic cloning, in every country there are more supporters than opponents.
    - For GM food the pattern is reversed for the majority of countries, and it is just in Germany, Ireland, Luxembourg, Spain and the Netherlands that we find more supporters than opponents.
- Considering Europe overall, approximately similar numbers of people are predicted to fall into a DK class for GM food and therapeutic cloning. We could perhaps interpret this as about equal levels of ambivalence for the two biotechnologies. However, this averages over considerable amounts of variation *between* countries in the proportions of respondents belonging to the DK classes. This must make a considerable impact on between-country comparisons of levels of support for biotechnology.

Some addition insights from the British sample are the following:

- There is tentative evidence that the relationship of the *risk* item to the other three might be evolving. Comparisons between British responses regarding GM food in 2002 and 2005 show that supporters in the more recent wave are less likely to disagree that GM food is risky, and opponents more likely to say it is risky. So responses to *risk* in 2005 seem to become more closely aligned with the other criteria of support. However, it remains to be discovered whether this represents a genuine shift in distribution of opinions an artefact of different question ordering in 2005.
- If DK responses are eliminated from the data set, the items can be ordered in Guttman-type models. This model implies that the four criteria follow an order of difficulty, in which saying that an application is not risky is the tallest order on the scale of support.

Lastly, two technical insights arising from the analyses in this chapter concern the difficulty in selecting one model from a range of possible models:

- In testing the equivalence of measurement models between just two countries, conditional probabilities that look very similar can nevertheless be significantly different from each other in likelihood ratio comparison tests. Comparing model deviances may often be too sensitive for practical purposes in comparative analyses with standard sample sizes of circa 1,000.
- By contrast, statistics of global model fit may often be too blunt for practical purposes in such analyses. In the models in focus here, diagnostics relating to marginal residuals appeared to be equivalent for direct effects and interaction models, but calculating them conditional on country revealed the latter models to be great improvements on the former.

The logics items seem to work very well together, on balance, and prompt perhaps only one immediate recommendation for future surveys. This regards the very different rates of DK responses between countries. The analyses in this chapter can only reveal and not explain such differences. It is very important to know whether they reflect genuinely differential rates of not-knowing amongst countries, or whether they reflect differences in fieldwork practices between the survey companies used to administer the questionnaire. If they are attributable to the former, this is a very interesting result, which could be contextualised with non-survey data. If they are due to the latter, this is at the least a nuisance and at the most a serious distortion effect on comparative results. In future survey waves, standardising fieldwork practices between countries should be a high priority.

A less pressing recommendation relates to the *risk* item. It will be interesting to see how the role of *risk* changes over time, and it would be useful in a future survey wave to use the split ballot to check whether the ordering of the items (*use*, *risk*, *moral acceptability* and *encouragement* in 2002; *use*, *moral acceptability*, *risk*, *encouragement* in 2005) has an effect on response profiles. It would of course also be interesting to add more criteria of support and opposition to the item sets, to add more detail to the interpretations offered in this chapter.

The next chapter focuses on a set of knowledge items, for which scaling rather than classifying is the task at hand. The ongoing questions concerning how to analyse and interpret DK responses and how to compare models between countries remain a central interest. In taking a scaling approach, however, attention turns from class models to trait models, and to the particular insights they offer into item difficulty and discrimination, and the dimensionality of the latent space used to represent responses.

# 6 Knowledge of biology and genetics: comparability in scaling

The measurement of scientific knowledge has attracted a great deal of attention in the area of PUS research. As outlined in Chapter 2, debates on this subject focus primarily around the content of knowledge or literacy scales – that is, what sorts of facts it might be reasonable to expect or wish members of the public to know about science and technology. Different facts, it is argued, may be more or less useful and difficult for scientists compared with laypeople. Such variations may apply, by a similar token, between different groups of survey respondents. This has implications both for survey designers, who prescribe the content of knowledge items, and survey analysts, who seek to make fair and useful comparisons between sets of respondents.

The analyses in this chapter focus on the insights that careful analyses of survey data can bring to this project. Survey analysts cannot advise survey designers on the content of survey items from a substantive theoretical or normative point of view. However, they *can* contribute to the construction of useful measures of knowledge, by describing the statistical behaviour of the items when they have been put to respondents. In this chapter an example of such a contribution is presented, in the form of the following research questions:

1. How can we characterise the behaviour of the set of 'knowledge' survey items analysed in relation to each other? Specifically, do they differ, and if so, how, in terms of their difficulty, and the degree to which they enable us to discriminate amongst respondents of different knowledge levels? Are any response effects found in the data? Is there any meaningful distinction to be made between substantively incorrect and 'don't know' (DK) responses?

2. How the variable 'knowledge' best be represented with the data? To what extent is a simple sum-score of correct responses an adequate or meaningful measurement of knowledge level? Do different models lead to very different knowledge scores for respondents?

3. How can fair and valid comparisons of levels of knowledge be made between respondents in different countries?

4. With a view to future Eurobarometer surveys, what recommendations can be made for the design of this item battery?

## 6.1 Data

The data, as in Chapter 5, are initially those referring to the British respondents from the 2002 Eurobarometer, later broadening the analyses to respondents from other European countries. The items for analysis are a set of ten statements about biology and genetics which respondents are asked to identify as true or false. A DK response is also permitted. This knowledge 'quiz' has been included in Eurobarometer survey waves dating back to 1991, and is a widely used source of information about biotechnology knowledge levels across Europe.

Table 6.1 gives frequencies for responses to the ten quiz items for British respondents. To address possible issues of response effects such as acquiescence bias, for some questions, 'true' is the correct answer, while for some questions 'false' is correct (DK is not correct for any item). In the table, grey highlighting indicates the correct response. The statements labelled *ktom3*, *kmod3* and *kbig3* are known by the survey designers as 'image' items, on the grounds that believing them to be true not only denotes incorrect knowledge, but also connotes susceptibility to menacing images of biotechnology. *Ktom3* contains ideas of difference, *kmod3* ideas of contagion, and *kbig3* conveys images of monsters.

The variation in responses to the items is clear from the table; some items appear easier than others, and some provoke more equivocation than others. For example, with one exception (*ktrgen3*) the 'true=correct' items yield more correct answers than those for which 'false=correct'. In all countries in the Eurobarometer 58.0, these knowledge items feature very high rates of DK responses. In the British data, in fact, DK rates exceed substantively incorrect responses for every item. This fact alone makes them worthy of some attention – quite apart from the theoretical points of interest in DK responses outlined in earlier chapters.

**Table 6.1 Distribution of responses to knowledge questions, British (GB) sample**

| *No.* | *Label* | *Statement* | *% responses*[13] TRUE | FALSE | DK |
|---|---|---|---|---|---|
| 1 | kbac3 | There are bacteria which live from waste water. | 88 | 3 | 9 |
| 2 | ktom3 | Ordinary tomatoes do not contain genes, while genetically modified tomatoes do. | 26 | 34 | 40 |
| 3 | kclo3 | The cloning of living things produces genetically identical copies. | 72 | 11 | 17 |
| 4 | kmod3 | By eating a genetically modified fruit, a person's genes could also become modified. | 15 | 54 | 31 |
| 5 | kmot3 | It is the mother's genes that determine whether a child is a girl. | 18 | 57 | 24 |
| 6 | kyea3 | Yeast for brewing beer consists of living organisms. | 80 | 6 | 14 |
| 7 | kprg3 | It is possible to find out in the first few months of pregnancy whether a child will have Down's Syndrome. | 84 | 6 | 9 |
| 8 | kbig3 | Genetically modified animals are always bigger than ordinary ones. | 26 | 37 | 37 |
| 9 | kchim3 | More than half of human genes are identical to those of a chimpanzee. | 56 | 13 | 31 |
| 10 | ktrgen3 | It is not possible to transfer animal genes into plants. | 28 | 29 | 43 |

*n=1014*

## 6.2 Scales of knowledge about biology and genetics

Recalling Chapter 2, this format of true/false/DK responses applied to set of statements is the predominant approach to measuring knowledge in PUS surveys. Usually the total number of correct responses is then taken as a respondent's knowledge score. In some instances this is converted into a dichotomy describing whether a respondent reaches a certain threshold, usually determined a priori – commonly the line is drawn where two-thirds or three-quarters of answers given are correct. Typologies of knowledge are never seen in the literature; where knowledge is specified as a categorical variable, it is this binary split of a previously continuous variable. So the reference point for this chapter is firmly in scales of knowledge, rather than in classifications.

The simple sum-score approach is potentially problematic in a number of ways. There is first of all the principled objection to the model's inability to allow for any

[13] Weighted frequencies, applying the basic sampling weight in the data set. Totals do not always sum to 100 per cent due to rounding.

measurement error. In contrast to the classification of logics, it may not be obvious to the layperson that this is a shortcoming, since with this scale we do not have the difficulty that the model leaves some response profiles unclassifiable. A problem which is more intuitively accessible for the non-statistician is that a simple sum-score means weighting every item equally in the scale, so that the same amount of credit is awarded for very difficult items as it is for very easy items. Of course, we could calculate a weighted sum-score – but we would need some prescriptive (i.e. normative, a priori) guidelines to use to decide how to assign weights.

Latent trait models both take into account measurement error and provide information about the relative weights that should be assigned to items in deriving scores from them. They also provide a means of exploring the data in more depth. In technical terms, we can inspect estimated difficulty and discrimination parameters with a view to future survey design. Those items that have low discrimination power, and those that are very easy, do little work for us in characterising respondents, and may be candidates for deletion from future survey waves. We can also use trait models to ascertain whether the items fit together to form a single scale, or whether they represent a multidimensional space. Latent variable models have indeed been used in the existing literature for this purpose – for example Miller (1998) uses factor analysis to identify a two-dimensional model of a set of knowledge items, with one factor representing 'construct knowledge' and the other denoting 'process knowledge'.

Miller's various analyses of these items – both the Oxford scale, and Eurobarometer-style questions focusing on biotechnology-related facts – constitute the most advanced approach in the PUS literature, to my knowledge. He takes binary items (where there are DK responses these are recoded as 'incorrect'), uses a preliminary factor analysis to identify items that form a unidimensional scale, then applies a three-parameter logistic model to these items, using the programme BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 1996). He uses this programme for both single and multiple group analyses – see for example Miller and Kimmel (2001) for analyses of literacy within the US, and Miller, Pardo and Niwa (1997) and Miller and Pardo (2000) for comparative studies of the EU, US, Japan and Canada. In the latter case he uses the 'non-equivalent groups equating' function in BILOG-MG to construct a common scale for groups which do not share a complete set of items. In this model, the item parameters are assumed to be equal across groups. A preliminary DIF analysis can be carried out with the programme, which tests whether the difficulty parameters can be assumed to be equal across groups (there is no facility for testing discrimination parameters in the same way). It is not clear if Miller carries out any DIF analysis, however.

The models used in this chapter differ from Miller's in two ways. Firstly, they use a two- rather than three-parameter logistic model. This is motivated by two concerns. The first is that a third, 'guessing' parameter is only weakly identified (Skrondal & Rabe-Hesketh, 2004; Thissen & Wainer, 1982). Thissen and Wainer (1982) demonstrate that the inclusion of a third parameter makes the estimated difficulty parameter unstable, and unfeasibly large sample sizes are required to obtain accurate estimates for it – the more so the easier the item. The second concern is a hunch that there is a more complex response style in the Eurobarometer items than guessing between two options. Recall from the previous section the high rates of DK responses, and the observation that the 'true=correct' items are answered correctly much more often than the 'false=correct' items. The data may therefore contain a mixture of acquiescence bias, propensity to guess, and propensity to profess ignorance – in addition to the knowledge we are trying to capture. As a full exploratory analysis of the data, we might gain some insights into possible response effects by retaining the distinction between DK and an incorrect response. It is unlikely then that a single dimension will be sufficient to represent the data – but it remains to be seen.

The second departure from Miller's approach concerns the multiple group analysis. Unfortunately it is not possible to conduct a full DIF analysis, as explained in Chapter 4, but we can add to the insights from Miller's comparative model by at least testing for measurement model equivalence between EU member states: in Miller's comparative analyses, Europe is treated as if one country. Other papers employing simpler analyses of the knowledge items between EU countries suggest that knowledge scales might be quite differently composed within Europe (e.g. Pardo & Calvo, 2004; Peters, 2000).

## 6.3 Models considered in this chapter

The analyses in this chapter are exclusively of trait rather than class models, in keeping with the theoretical perspectives outlined previously, and acknowledging the political objections sometimes raised against the concept of knowledge as a binary variable. The first section of the chapter focuses on the British sample. Analyses begin with the simplest models for binary responses: correct versus incorrect. Three models are briefly compared: a Rasch model (i.e. the latent trait model which is most similar to a simple sum-score), and one-trait and two-trait models with unconstrained loadings. Further comments on these are reserved for a later point in the chapter, where a fuller comparison of these and subsequent models is given.

Since DK responses are of particular interest, the remainder of the analyses are for nominal items, that is reinstating the distinction between incorrect and DK answers. Multidimensional trait models are used to investigate the idea that response effects may be captured in a trait, as an alternative to adding a guessing parameter to the model. A little time is spent on exploratory analyses using continuous trait models, and item characteristic curves (ICCs) for a two-trait model are shown to give an initial interpretation of differential item discrimination and difficulty. This addresses the first research question of the chapter. Having identified a trait that might reasonably be labelled 'knowledge', a further modification is made to the model by moving to *discrete* traits. As described in Chapter 4, in discrete trait models there is no normality restriction on the prior distribution of the latent variable. At this point, then, a comparison is made of the distributions of knowledge scores assigned to British respondents from a selection of models encountered in the earlier analyses. This addresses the second research question of the chapter.

In Section 6.5, attention turns to the third research question: comparative analyses between country samples. It proves difficult to find a well fitting joint model across the fifteen countries in the data set. Two sets of analyses are presented: the first focuses on trait models for polytomous items, and the second for binary items, abandoning the distinction between DK and substantively incorrect responses. The distributions of factor scores from these are compared, alongside simpler scales which might commonly be made from the data. Neither of the two sets of analyses produce very well fitting models, however. A third study is therefore included for comparison purposes, using a different data set, to investigate informally the extent to which altering the content of the items might improve the chances of finding a good model. In the 2005 Eurobarometer on PUS, a set of items were used from the 'Oxford' scale of knowledge, about science in general, rather than biology and genetics in particular. A small set of discrete trait models for polytomous items are used for these data, and a joint model successfully fitted for the same fifteen countries as in the 2002 data set.

## 6.4 Results of latent trait analyses of British data

### *6.4.1 Continuous latent trait models for binary items*

Table 6.2 presents fit statistics for a selection of trait models. Those for binary responses offer, as expected, limited enlightenment. It is notable that a Rasch model, which constrains all discrimination parameters to be equal, fits the data quite poorly. Improvements in fit are gained from allowing loadings to vary between items, and further, from increasing the number of traits to two, although this second trait does not seem to be needed, and is perhaps overfitting. A two-trait model serves to draw a distinction between those items for which 'true' is correct and those for which 'false' is correct; broadly speaking, on one trait we find steep ICCs for the former and virtually flat ICCs for the latter, and vice versa on the other trait. It is interesting that this difference in item functioning between these two types of items appears, and it provides extra motivation for reinstating the distinction between DK and incorrect responses, to further investigate possible responses styles.

**Table 6.2 Fit statistics: continuous latent trait models for binary items, GB**

| Model | $L^2$ | d.f. | p (bootstrap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| Rasch | 834 | 1,003 | <0.001 | -1,172 | -6,108 | 7.8 | 1.02 |
| 1 trait, unconstrained loadings | 785 | 994 | <0.001 | -1,203 | -6,095 | 2.8 | 0.65 |
| 2 independent traits, unconstrained loadings | 705 | 984 | 0.296 | -1,263 | -6,106 | 0.0 | 0.17 |

### *6.4.2 Continuous latent trait models for polytomous items, recognising DK responses*

The fit of the models for three-category nominal items improves steadily as more traits are added (see Table 6.3). The first trait to emerge contrasts DK responses at one end of the trait with correct responses at the other; probabilities of giving the incorrect substantive response remain low at all points on the trait. Two traits seem to allow the separation of knowledge from response effects – this is clearly seen when the solution is obliquely rotated. A third trait has no clear interpretation, however. Since fit statistics – particularly inspection of marginal residuals – indicate that two traits give an acceptable fit, this is the model presented in more detail below.

**Table 6.3 Fit statistics, continuous latent trait models for 3-category nominal items, GB**

| Model | $L^2$ | d.f. | p (bootstrap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| 1 trait | 3,382 | 974 | <0.001 | 1,434 | -3,360 | 6.4 | 0.96 |
| 2 traits | 3,263 | 954 | 0.002 | 1,355 | -3,340 | 2.7 | 0.61 |
| 3 traits[14] | 3,176 | 934 | 0.026 | 1,308 | -3,289 | 0.7 | 0.35 |

The two traits emerging from the second model are practically uncorrelated: with oblique rotation, their correlation is just -0.07. Figure 6.1 shows ICCs for each rotated trait, between values of –3 and +3 (since the latent variables are assumed in the model to be normally distributed, this interval covers a sensible range), in each case with the value of the other trait fixed at 0. In these, the five 'true=correct' items (items *kbac3*, *kclo3*, *kyea3*, *kprg3* and *kchim3*) are collected together in the left-hand column, and the 'false=correct' items (*ktom3*, *kmod3*, *kmot3*, *kbig3* and *ktrgen3*) on the right-hand side. Broadly speaking, we can see in them a 'DK' trait (Figure 6.1(a)) and a 'knowledge' trait (Figure 6.1(b)).

Taking Figure 6.1(a) first, we can see that for every item, as we move towards the higher end of the trait, the probability of giving a DK response increases, until at values of 2 or more, on every item the most likely response is DK. So respondents with the highest score on this trait are those that give a full set of DK responses. Moving in the

[14] Estimated with 10 nodes per trait, due to computational burden.

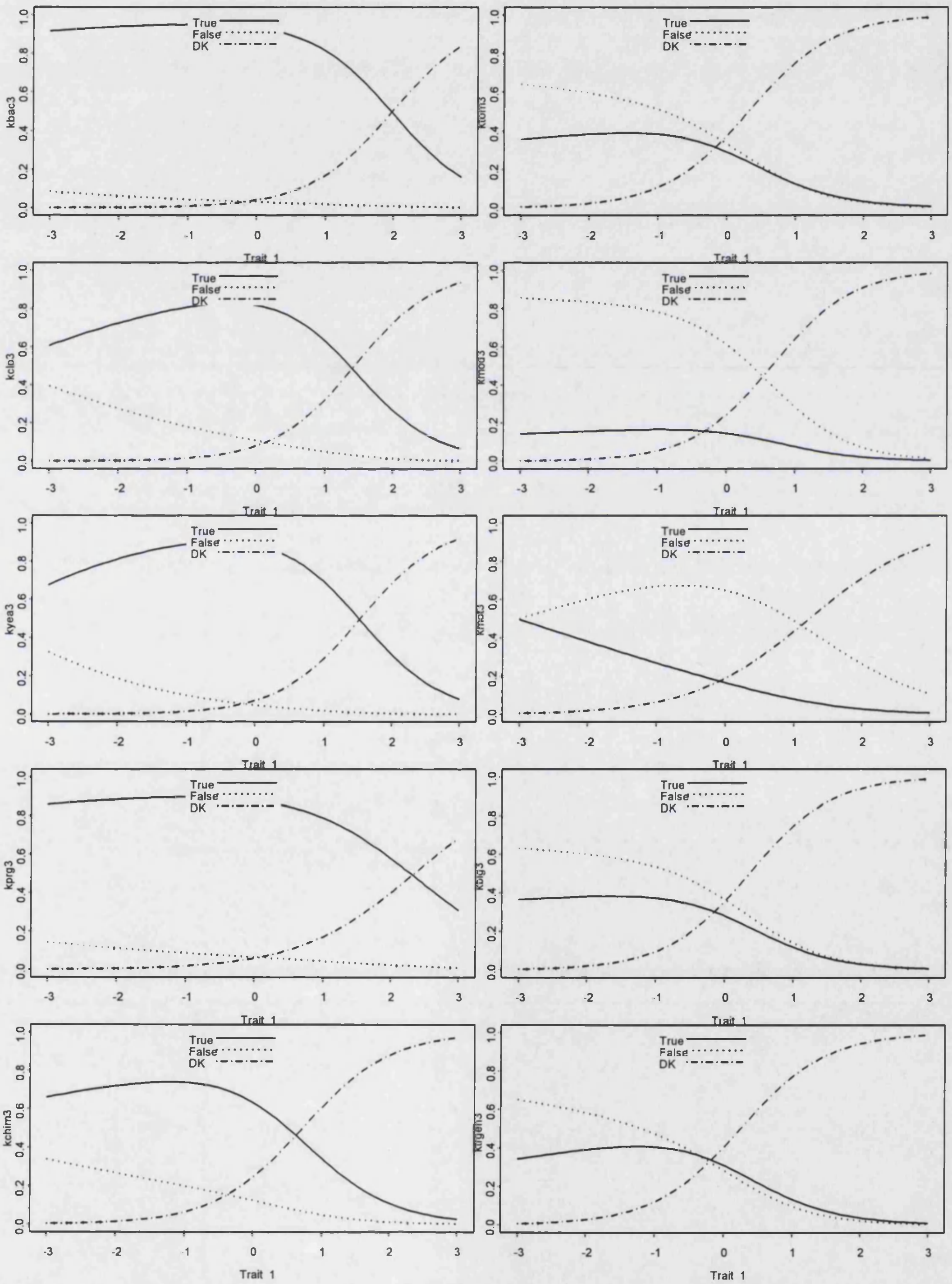
other direction, from the higher end towards the middle and lower section of the trait, correct responses become more likely – for all items, respondents at around the median and below, to scores of –3, are most likely to give the correct response, be it 'true' or 'false'. However, for the true=correct items and for *kmot3*, we can see the trace lines for the correct response begin to turn downwards at the lower end of the trait; this is most visible for kclo3 and kchim3, for which we can see that respondents with a score of –3 on the trait have a lower probability of giving a correct answer to these items than respondents with a score of –2. Extending the curves below –3, into the realms of very unusual scores, the *very* lowest scores theoretically possible on the trait correspond to the response 'false', for all but the item *kmot3*. So it seems reasonable to suggest that this trait captures information about confidence of response.

The knowledge trait is far more interesting, for our purposes. Fixing the value of the DK trait at 0, high scores on the knowledge trait are associated with giving correct substantive responses, 'true' or 'false' according to the item. Probabilities of correct responses are very high at two or three standard deviations above the mid-point of the trait for all but the two items *kchim3* and *ktrgen3*. For *ktrgen3* the slope for the correct response rises only very slowly, while for *kchim3* the probability of a correct response *never* rises above 0.7 at any point on the trait – in fact, the slope coefficient for the incorrect response 'false' is larger than for the correct response 'true', which means that in the theoretical model, in which the trait extends higher than +3, this item loads in the wrong direction. Within the expected range of the latent trait, it would not disrupt the scoring on the trait, but it is a good example of the effect of an item which loads the wrong way. According to the model, someone with a very high score on the trait would be more likely to answer this item incorrectly than correctly. This is clearly inconsistent with a trait meaningfully representing knowledge.

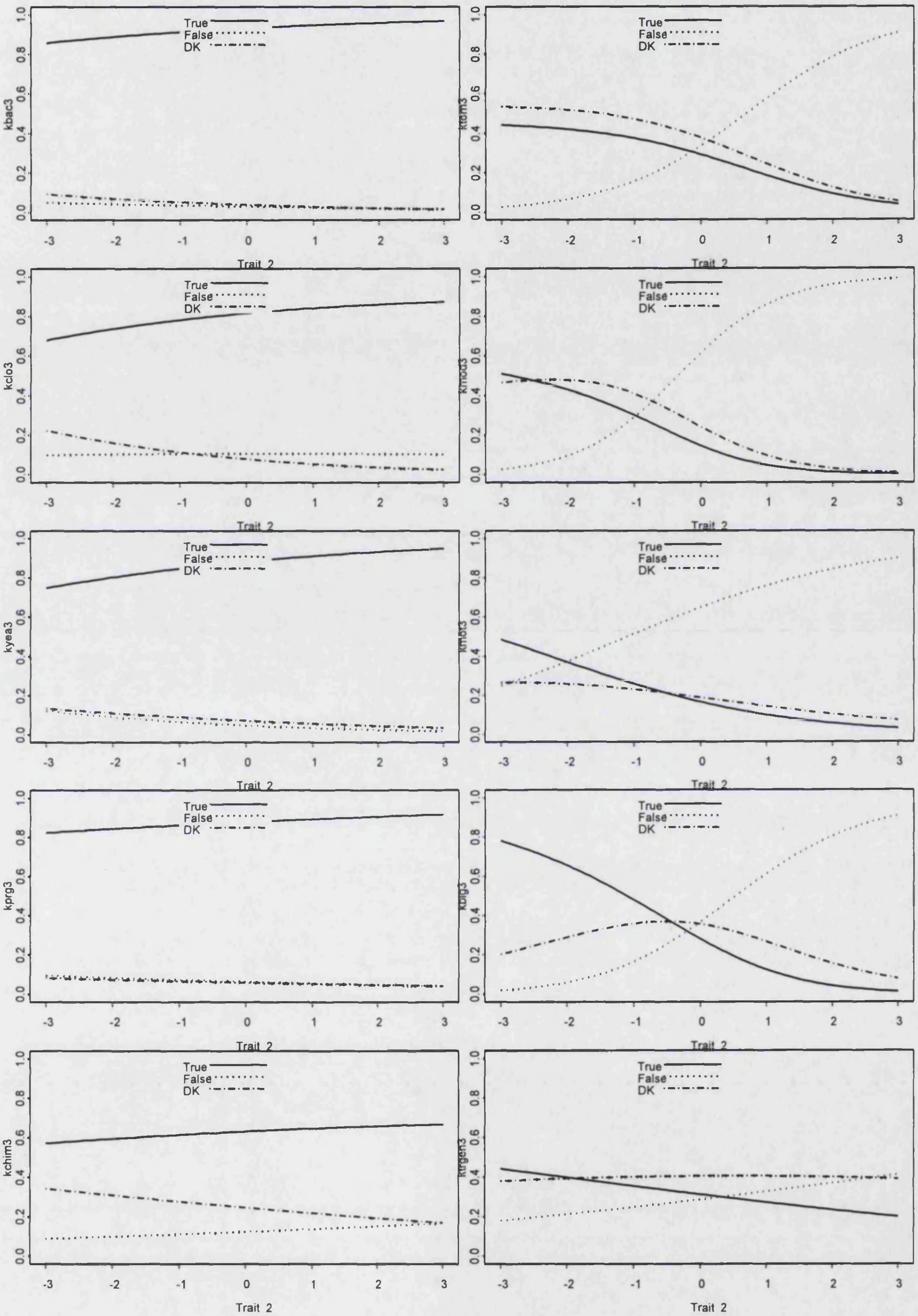
Within the expected range of the trait, *kchim3* causes no problems, but it is fair to say that it gives little information about a respondent's level of knowledge; respondents with scores of –3 and +3 are almost equally likely to give answer this correctly, with the odds slightly but only slightly in favour of a correct response. The other 'true=correct' items are equally ineffective for discriminating between respondents, with overwhelmingly high probabilities of responding correctly for most respondents. The very shallow ICCs for these items lie in stark contrast to the steep curves obtained for four of the 'false=correct' items (*ktom3*, *kmod3*, *kmot3* and *kbig3*). These are therefore the items which allow us to differentiate most sharply between respondents in terms of knowledge levels. This pattern is consistent with knowledge scales from other biotechnology surveys (e.g. Pardo & Calvo, 2004).

**Figure 6.1 Item characteristic curves from obliquely rotated, 2-trait continuous trait models for 3-category nominal items, GB**

**(a) ICCs for the 'DK' trait**



## (b) ICCs for the ‘knowledge’ trait



### *6.4.3 Discrete latent trait models for polytomous items*

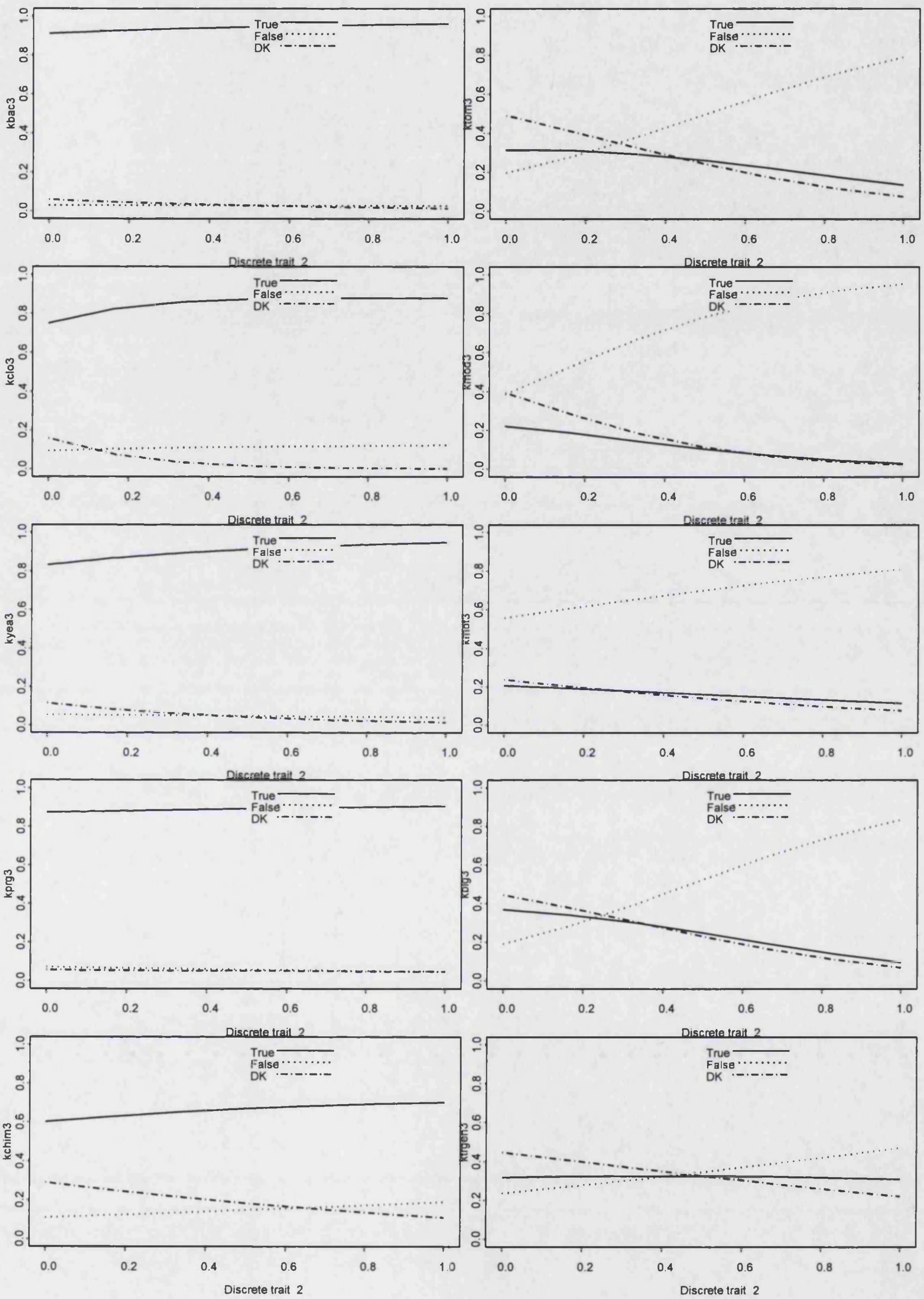
Having identified a trait that might be labelled 'knowledge', it is expedient (for the reasons outlined in Chapter 4) to move on to discrete trait models. Rotation of discrete traits is not possible; nevertheless, the substantive interpretation of traits from these models is practically identical to those from continuous trait models, as long as an adequate number of levels of discrete traits is specified (7 levels per trait are used throughout the thesis, as explained in Chapter 4, Section 4.1.6). A two-trait model for the ten knowledge items fits well ($L^2 = 3.224$; d.f. = 941; p(bootstrap) = 0.002; 2.7 per cent of standardised marginal residuals > 4; Jöreskog & Moustaki index = 0.58). Figure 6.2 presents the ICCs for the 'knowledge' trait for this model (the set of ICCs for the other trait is very similar to Figure 6.1(a), and is not presented here). The main differences between the curves for the continuous and discrete trait models lie with the strongly discriminating 'false=correct' items. The discrete trait model locates the points of intersection between curves where 'false' becomes the most likely response, further towards the lower end of the trait than is the case with the continuous trait model. This suggests that the posterior scores from the two models will follow somewhat different distributions.

A note of caution accompanies this model. Although the shapes of the trace lines for the 'knowledge' trait appear to define the scale logically – so that the highest level of the trait is occupied by those who answer every item correctly – the signs of the true and false slopes for *kclo3* and *kchim3* are in the wrong order: the probability of giving the incorrect response (for both items, 'false') *increases* as the level of the trait increases. This can be seen more easily in the parameter estimates, which are in the Appendix to this report (Table A.5). The slope estimates for 'true' and 'false' responses on these two items are actually not significantly different from each other (at $p<0.05$). However, the implication of the model for the calculation of posterior scores for the trait is that a person answering all other items correctly but incorrectly saying 'false' to these two items would be assigned a slightly higher score than a person answering all items correctly. Under this model, the former has a score of 0.943 and the latter, 0.917.

The next step in fitting a model to the British data would be to drop the problematic items from the scale, in order to find a model in which the slope coefficients are aligned in the directions which fit logically with a scale defining high knowledge at one end.

This exercise is in fact quite problematic. Removing items from the scale noticeably destabilises the items remaining – most notably, the 'true=correct' items, making it difficult to find a model in which the slope coefficients take the required signs. Within the scope of the several models that I attempted, I could not find one which represented any improvement over the ten item scale, especially in terms of producing factor scores that were logically ordered according to numbers of correct answers. So the ten item trait remains the final model for this section on the British sample, but with a caveat attached to it.

**Figure 6.2 Item characteristic curves for the 'knowledge' trait from a 2-trait discrete trait model for 3-category nominal items, GB**



### 6.4.4 Comparing scores from a selection of models

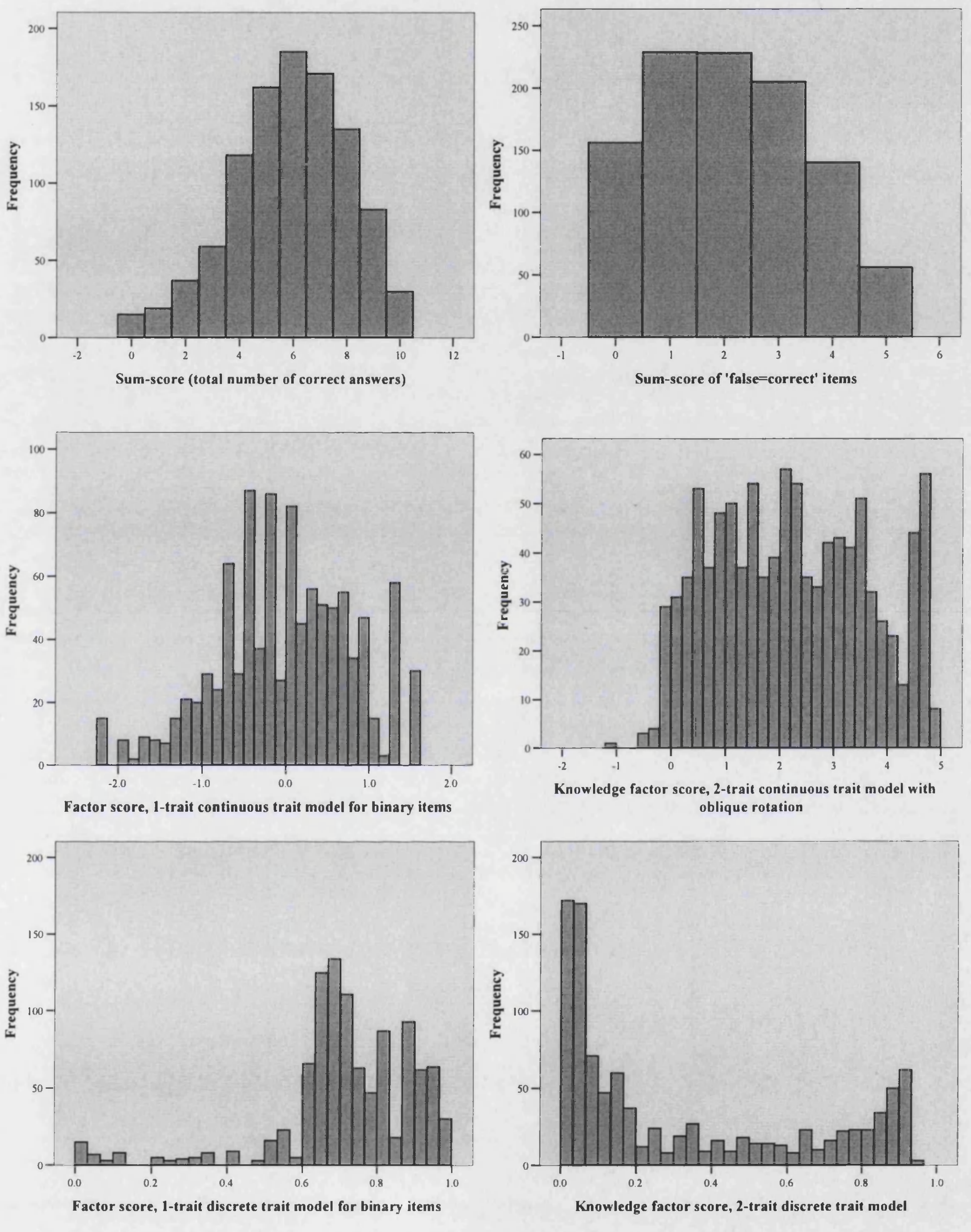
At this point it is useful to compare the distribution of scores from a selection of models considered so far in this chapter. Taking the simple sum-score as the point of departure, Table 6.4 shows that the single trait model for binary items produces, not surprisingly, very similar scores (r = 0.98). When DK and 'incorrect' are retained as separate response categories, the knowledge traits modelled are notably more highly correlated with a sum-score of the strongly discriminating 'false=correct' items than with the sum-score of all ten items (0.93 versus 0.85 for a continuous trait, and 0.87 versus 0.77 for a discrete trait). This is not surprising, given that in these models the ICCs for the 'true=correct' items are very shallow, and it is the 'false=correct' items that enable us to discriminate among respondents with different levels of knowledge.

**Table 6.4 Linear correlations between knowledge scores from a selection of models, GB**

| | Binary items (correct/incorrect), single scale | | | Scores from 'knowledge' trait in 2-trait models for polytomous items | |
|---|---|---|---|---|---|
| Pearson's correlation coefficient (r) | Sum-score of 10 items | Sum-score of 'false=correct' items | Factor score from 1-trait continuous trait model | Continuous trait model with oblique rotation | Discrete trait model |
| Sum-score of 10 items | 1.00 | | | | |
| Sum-score of 'false=correct' items | 0.85 | 1.00 | | | |
| Factor score from 1-trait continuous trait model | 0.98 | 0.87 | 1.00 | | |
| Continuous trait model with oblique rotation | 0.85 | 0.93 | 0.90 | 1.00 | |
| Discrete trait model | 0.77 | 0.87 | 0.83 | 0.89 | 1.00 |

The relatively high correlation of 0.77 between the simplest sum-score and the discrete knowledge trait score prompts the question of exactly how the two models differ, and whether there is any advantage or extra insight into the data gained by using one rather than the other. The set of histograms in Figure 6.3 illustrate the implications of each scoring system. In all but the discrete trait model, knowledge appears as approximately normally distributed among respondents. This is to be expected from the continuous trait models, due to the assumption of normality of the latent variable inherent in the models. To some extent it is not surprising for the sum-scores, since with guessing alone, a respondent would be likely to answer some of the items correctly. Strikingly however, in the discrete trait model, the distribution of knowledge scores is bimodal. Thus quite a different depiction of knowledge emerges from this model. Given the anomaly of its posterior scores, it would be unwise to make very bold claims about the distribution of the trait in this model. Nevertheless, with due caution, it is very interesting for researchers in PUS. Durant et al. (2000) propose a 'normalisation hypothesis' which states that as a society progresses along the path of industrialisation, knowledge of science will become more even in its distribution among the population. The inference is that in Britain, an advanced industrialised nation, we would expect to see the kind of distribution of knowledge scores produced by the continuous trait model or any of the sum-scores, rather than the distribution suggested by the discrete trait model.

**Figure 6.3 Histograms comparing distributions of knowledge scores from a selection of models, GB**




Before moving on to the multiple group analyses, a summary of the results from the British data might be useful. Very little time was spent on models of binary items – a one-trait continuous trait model fits these very well, on the condition that the item discrimination coefficients are free to vary between items. It is rather more interesting, however, to retain the distinction between DK and a substantively incorrect response, and for these polytomous items, two traits are needed to adequately represent the variation in the data. One trait seems to be defined by DK responses, while the other can be labelled as a measure of knowledge. In this knowledge trait, there is a broad distinction to be made between those items for which 'false' is the correct response, and those for which 'true' is correct. In the former, the trace lines for the correct response have steep slopes, whereas for the latter the curves are much flatter. So the 'false=correct' items are much more highly discriminating than the 'true-correct' items. This is found consistently, whether the model is for continuous or discrete traits. In the discrete trait model in fact, the instability of the slopes for the 'true=correct' items, wavering just above and below 0, creates some problems for calculating posterior scores. Overall, though, discrete and continuous traits return the same general story regarding the measurement model for knowledge. However, they return quite different stories regarding the structural model: posterior scores from a continuous trait model are approximately normally distributed, whereas scores from a discrete trait modal have a bimodal distribution.

## 6.5 Extending discrete latent trait analyses to other country samples

In this section I consider the third research question of the chapter: how to derive a measure of knowledge that enables fair comparisons between countries. The approach adopted is the same as that for the previous chapter: to take the final model from the last section (in this case, two discrete traits for polytomous items) as a starting point, and begin with brief two-group analyses of measurement models to assess whether it is feasible to combine regional samples for Germany and the UK. Following this, a qualitative comparison of country-by-country trait models is used to gauge informally whether it can reasonably be claimed that the same two traits are found in each sample. From this point, formal tests of the equality of measurement models between countries are conducted, in the same vein as Chapter 5. All models in this section are discrete rather than continuous trait models.

In order to manage the reader's expectations for the rest of the chapter I should state here that a joint model which is both well fitting and substantively meaningful has not been found, with the data at hand. The analyses in this section shed light on the particular difficulties entailed in finding such a model, and thus suggest other avenues to explore in future. Additional brief sets of analyses are offered as points of comparison for the polytomous biotechnology knowledge items. In the first, the items are modelled as binary rather than polytomous, reasoning that removing the distinction between the two types of incorrect responses may ease the task of modelling the items cross-nationally. This does not appear to be the case, however. The distributions of posterior scores from a selection of different models, and correlations between them are then presented as a further means of comparing the various possible measures of knowledge of biology and genetics. In the final section of the chapter, a different data set is used: the 2005 Eurobarometer on the Public Understanding of Science, which contains thirteen items from the so-called 'Oxford' scale described in Chapter 2 (Section 2.4.2). This has been more widely used in PUS, and is designed to assess knowledge about science in general, rather than specifically about biology and genetics. For this set of items, a satisfactory joint model *is* achieved.

### *6.5.1 Finding the best joint model for polytomous items*

Table 6.5 gives fit statistics for two-trait models for the UK and Germany, with measurement models fixed to equality between Great Britain and Northern Ireland, and between East and West Germany. The fits of these models are notably poorer than their equivalents in Chapter 5. Although for the models overall, the proportions of large standardised marginal residuals are not high, in Northern Ireland and in West Germany, they are more than 10 per cent. Inspecting the ICCs for the separate regions, informally, the high residuals do not seem to be attributable to any particular items in the case of Germany, whereas in the UK, the 'false=correct' items seem to be the site of the greatest differences between the two regions. A number of strategies might be attempted in order to improve the fit of these models: increasing the number of traits, dropping some items, or allowing some item parameters to vary between groups. These are strategies which need to be followed to find a joint model for the other countries in the data set. To anticipate: some items will be dropped from the scale in the course of finding a joint EU-wide model. When the number of items is reduced, marginal residuals suggest that combining Great Britain with Northern Ireland, and East with West Germany, is acceptable[15].

[15] In two-trait models for the five items retained in the final model in this section, percentages of large two-way marginal residuals are: UK = 0.8 (GB = 2.2, NI = 1.1); Germany = 0.8 (West = 4.4, East = 5.6).

**Table 6.5 Fit statistics from testing measurement models between Great Britain and Northern Ireland, and between East and West Germany**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | | | Jöreskog & Moustaki index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | All | GB | NI | All | GB | NI |
| UK, measurement model equal | 4,892 | 1,245 | <0.001 | 2,402 | -4,054 | 3.2 | 4.2 | 11.6 | 0.76 | 0.85 | 1.86 |
| | | | | | | All | West | East | All | West | East |
| Germany, measurement model equal | 8,604 | 1,970 | <0.001 | 4,664 | -6,414 | 6.0 | 10.4 | 7.2 | 1.11 | 1.38 | 1.27 |

Moving on to the explore a joint model of knowledge items for the fifteen countries, separate country-by-country analyses suggest that two-trait models are a feasible starting point. Using all ten items in the set, two traits fit well for all country samples: the percentage of large two-way standardised marginal residuals ranges from 0.2 in Finland to 7.2 in Germany, with an average of 2.4 per cent across the fifteen countries. In all countries, one trait can be reasonably labelled 'knowledge', while the interpretation of the other trait varies a little more between samples – of the range of interpretations, the most common is a response effect trait, with DK responses at one end, and 'false' at the other.

Focusing on the 'knowledge' trait, Table 6.6 gives a qualitative summary of the few items and few countries for which slope coefficients deviate from the pattern to be expected in a trait capturing knowledge. It reflects the model of the British data, in that many of the 'true=correct' items lack discrimination power in some countries, and in a number of cases, whilst the overall probability of a correct response is highest at the highest point of the trait, the slope for the incorrect response is increasing – implying the problems with factor scores encountered in the British data. However, these are not such a serious problem compared with the last item, *ktrgen3*, for which in five countries, at the 'high knowledge' end of the trait the probability of giving the incorrect response is greater than the probability of giving the correct response. In these cases it is very clear from the ICCs that the item does not fit logically with the others in the scale. From this point it is dropped from the item set. Repeating these exploratory analyses with nine items leaves the qualitative summary of them in Table 6.6 essentially unchanged.

**Table 6.6 Qualitative summaries of unusual ICCs on 'knowledge' traits, from 2-trait models, 15 countries**

| | 'True=correct' items | | | | | 'False=correct' items | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | kbac3 | kclo3 | kyea3 | kprg3 | kchim3 | ktom3 | kmod3 | kmot3 | kbig3 | ktrgen3 |
| Austria | | | | | | | | | | |
| Belgium | | | | | d | | | | | d, c-, i+ |
| Denmark | c-, i+ | | | | i+ | | | | | **I**, c-, i+ |
| Finland | | | | **c-, i+** | | | | i+ | | |
| France | c-, i+ | | **c-, i+** | **c-, i+** | **c-, i+** | | | | | |
| Germany | | c-, **i+** | c-, **i+** | c-, i+ | **i+** | | | | | **I, i+** |
| Greece | i+ | | | i+ | | | | | | |
| Ireland | i+ | | | | i+ | | | | | c-, i+ |
| Italy | c-, i+ | c-, i+ | c-, i+ | c-, i+ | | | | | | |
| Luxembourg | d | d | | d, i+ | | | | | | **I**, d |
| Netherlands | d | d | | d | d | | | | | **I**, d |
| Portugal | | | | i+ | | **I**, i+ | | | | |
| Spain | c-, i+ | c-, i+ | | c-, i+ | | | | | | **I, i+** |
| Sweden | d | | | d | | | | | | |
| UK | d | | | d | i+ | | | | | d |

<u>Key</u>

| | |
|---|---|
| d | Low discrimination: very flat ICCs |
| c- | Slope for correct response decreasing slightly with higher levels of 'knowledge' |
| i+ | Slope for incorrect response increasing slightly with higher levels of 'knowledge' |
| I | 'Incorrect' most likely response at top end of trait |
| regular font | Slight effect |
| **bold font** | Strong effect: more seriously problematic |

From the ICCs it is clear that some items have greater discrimination power than others. In a joint trait model, the relative discrimination of the items would be fixed to be the same between countries. As a brief preliminary analysis for such a joint model, an indicative analysis was carried out, focusing on item discrimination power, defined as the discrimination parameters of correct in comparison with incorrect responses – that is, ignoring the slope estimates for DK responses, for the moment. If the discrimination parameters of certain pairs of items were in a significantly different *order* in different countries – say, if *ktom3* were more highly discriminating than *kbac3* in some countries, but less highly discriminating than *kbac3* in others – it would be a clear sign that finding a well fitting joint model would be difficult. In fact, only two pairs of items (*kbac3* with *kyea3* and *ktom3*) appear to have significantly different relative discrimination powers, and then, only between Portugal and Spain, for the first pair, and Portugal versus Spain and Denmark, for the second.[16] This gives grounds for optimism that a cross-country

[16] This analysis was carried out with S-Plus software, using 95 per cent confidence intervals around the differences between slope estimates, and applying the Bonferroni correction to allow for multiple comparisons.

model of knowledge, with fixed measurement models between countries, might be feasible.

Unfortunately, it seems that the differences in ICCs between countries are nevertheless too large for a joint trait model, with the same measurement model for each country, to fit well. Table 6.7 gives fit statistics for a selection of models. In the joint version of the two-trait models for nine items, 42 per cent of standardised two-way marginal residuals are large. The number of large residuals is notably high for four items: *ktom3*, *kyea3*, *kbig3* and *kchim3*, in terms of two-way item-by-item margins and country-by-item margins, and three-way item-by-item margins. Dropping these from the scale almost halves the proportion of high residuals, but the rate is still 26.3 per cent. Increasing the number of traits also helps model fit: a three-trait model with these four problematic items removed reduces the proportion of high residuals to 15.9 per cent. By the standards of Chapter 5 however, this is still a poorly fitting model.

**Table 6.7 Fit statistics for 2- and 3-trait models, with measurement models constrained to be equal across 15 countries**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standard'd marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| 2 traits, 9 items (no k*trgen3*) | 45,188 | 15,945 | <0.001 | 13,298 | -109,205 | 41.8 | 8.14 |
| 2 traits, 5 items (*kbac3, kclo3, kprg3, kmod3, kmot3*) | 4,820 | 3,559 | <0.001 | -2,298 | -29,641 | 26.3 | 3.19 |
| 3 traits, 9 items (no *ktrgen3*) | 43,001 | 15,905 | -[17] | 11,191 | -111,005 | 24.8 | 2.84 |
| 3 traits, 5 items (*kbac3, kclo3, kprg3, kmod3, kmot3*) | 4,256 | 3,527 | | -2,798 | -29,896 | 15.9 | 1.64 |

Since the objective for this model is to derive a measurement of knowledge, there should be no compromise to the model by allowing a second trait, which fills the role of accounting for response styles, to differ between groups. That is, a feasible joint model might be one with a fixed trait representing 'knowledge', and a country-specific trait for response effects. Such a model can be fitted with LEM. However, for this set of items,

[17] The two 3 trait models were very burdensome in terms of time: more than three days for the 9 item model. Since the calculation of bootstrap p-values generally takes much longer than the estimation time for the model for which they are sought, it was practically speaking unfeasible to calculate them here.

it does not provide a solution. Both with nine items, and with a reduced set of five items (chosen by means of inspecting large two-way and three-way marginal residuals, as above), the fixed trait cannot be interpreted as 'knowledge'; it seems to be closer to a response effect, with DK and one end and 'false' at the other, regardless of whether this is the correct response. Albeit these models represent a great improvement in fit (13.2 per cent large two-way residuals for the nine-item model for example), they do not return a viable representation of 'knowledge'.

A number of useful observations from these analyses illustrate the instability of the measurement of knowledge using these items. In the models estimated in Latent GOLD, with equal measurement models between countries for all traits, it was found that different numbers and combinations of items produced quite different solutions – echoing the findings from the two-trait model of British data in Section 6.4.3. Some combinations of items failed to return a trait successfully representing correct 'knowledge' at one end, even when the model contained three traits. So although these items are intended to constitute a sample from a wide universe of knowledge items, the interpretation of the construct 'knowledge' seems to depend, more than is desirable, on the combination of items contributing to it.

This instability is also observed in the models fitted using LEM, in which the measurement model was free to vary between groups on one of the two traits. LEM does not include a facility to work through multiple sets of starting values, so to mimic this function, I ran each of the models mentioned above (with nine and five items) 50 times, selecting in each case the model with the lowest deviance as the final model. For both sets of items, nearly 50 different deviances were returned, indicating many local maxima for these models. For the initial model with nine items, I informally inspected the patterns of slope estimates for the fixed trait of these 50 models, to ascertain to what extent these different models returned traits with substantively different interpretations. Approximately a third of the fifty returned a trait that might be called 'knowledge'; the other two thirds returned something which might be called a response style trait; most often with DK response at one end and 'false' responses (regardless of the correct response) at the other.

From these analyses, then, it seems that finding a viable joint model for these items is a difficult task. The models attempted here either fit badly or do not identify a trait that could feasibly be interpreted as 'knowledge', and seem to be numerically unstable. Out of the models presented in this section, the three-trait model with five items is the best representation of 'knowledge', cross-nationally. The ICCs for the 'knowledge' trait from this model are presented in Figure 6.4 (ICCs for the other two traits are in the Appendix, Figure A.1). Note that for this trait low scores denote high levels of knowledge. I take the factor scores from this model for the analyses in Chapter 8 (with scores reversed, so that high scores denote high levels of knowledge).

**Figure 6.4 Item characteristic curves from 3-trait discrete trait models for 3-category nominal items, with measurement models equal for all traits, for 15 countries, 'knowledge'trait**

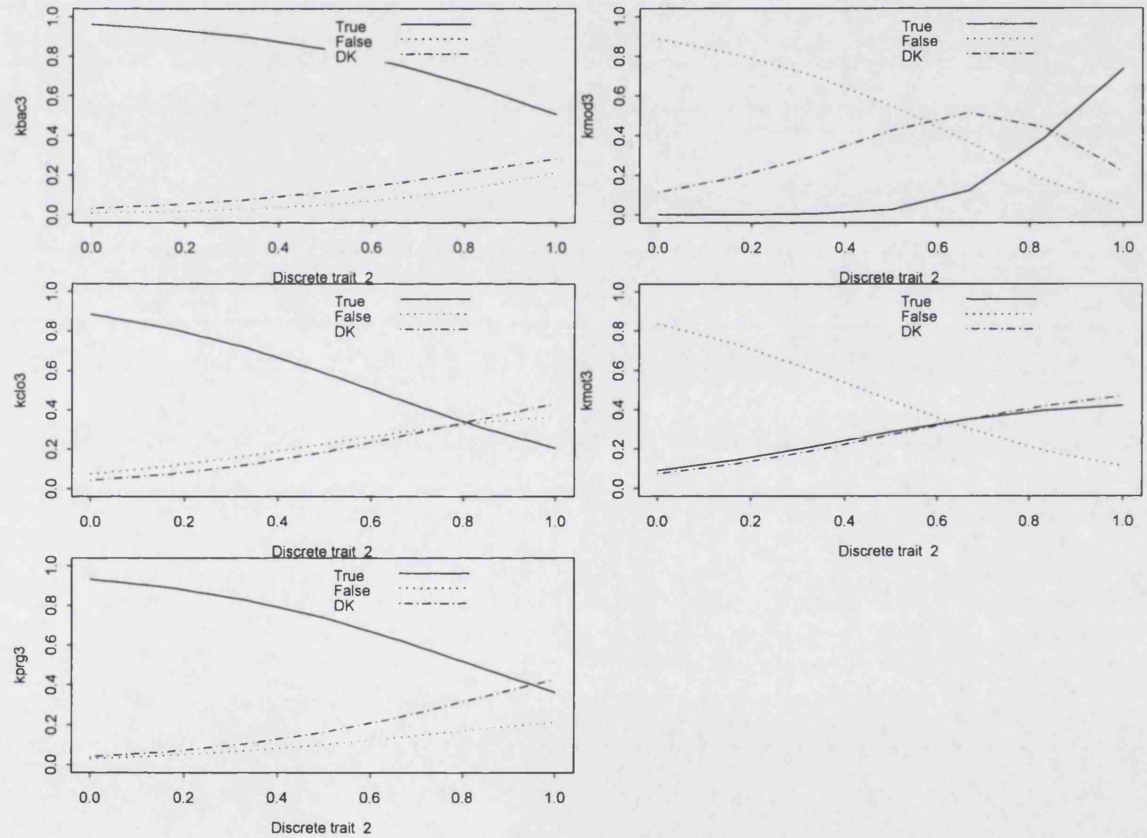



Table 6.8 shows the distribution of levels of knowledge according to this trait (reversed from the original model so that high levels of the trait denote high levels of knowledge). Specifically it shows the percentage of the population estimated to belong at each of the seven levels of the trait, by country, and for the fifteen countries together, weighted by their respective populations. Countries are ordered from highest to lowest mean knowledge score. The distribution of the trait among countries is consistent with expectations from the PUS literature: high levels of knowledge are found among the

Northern European countries, and with some exceptions, lower levels among those in the South. Overall, Europeans score quite highly on this scale, with very few people falling into the lower three levels of the trait, and with an EU wide average level of 0.68.

**Table 6.8 Percentages of respondents in each level of the final joint model of biotechnology knowledge items[18]**

| | Level | | | | | | | Mean |
|---|---|---|---|---|---|---|---|---|
| *% within country* | Low knowledge | | | | | | High knowledge | knowledge |
| Sweden | 0 | 0 | 0 | 7 | 10 | 47 | 36 | 0.85 |
| Denmark | 0 | 0 | 0 | 16 | 15 | 45 | 23 | 0.79 |
| Netherlands | 0 | 0 | 0 | 17 | 15 | 44 | 22 | 0.78 |
| UK | 1 | 0 | 1 | 22 | 17 | 42 | 18 | 0.75 |
| Finland | 1 | 0 | 1 | 30 | 19 | 36 | 12 | 0.70 |
| France | 2 | 0 | 1 | 30 | 19 | 36 | 12 | 0.70 |
| Luxembourg | 2 | 0 | 1 | 31 | 19 | 35 | 11 | 0.69 |
| Italy | 4 | 1 | 1 | 34 | 19 | 32 | 9 | 0.66 |
| Germany | 4 | 1 | 2 | 36 | 19 | 30 | 9 | 0.65 |
| Ireland | 5 | 1 | 2 | 36 | 18 | 30 | 9 | 0.65 |
| Spain | 5 | 1 | 2 | 37 | 19 | 29 | 8 | 0.64 |
| Belgium | 6 | 1 | 2 | 38 | 18 | 28 | 8 | 0.63 |
| Greece | 8 | 1 | 2 | 43 | 18 | 23 | 5 | 0.59 |
| Austria | 11 | 2 | 3 | 46 | 17 | 19 | 4 | 0.55 |
| Portugal | 18 | 2 | 3 | 45 | 15 | 15 | 3 | 0.49 |
| Europe total (pop. weighted) | 4 | 1 | 1 | 32 | 18 | 33 | 12 | 0.68 |

### *6.5.2 Finding the best joint model for binary items*

Following the lack of success in finding a well fitting joint model for the polytomous items, this section presents a short study modelling the binary versions of these items. This is of interest for two reasons: firstly, insofar as it might indicate whether the difficulties with the models above are reduced when the distinction between DK and substantively incorrect response is dropped; and secondly, because in the PUS literature it is far more usual to model the binary versions.

[18] Unweighted percentages, due to computation problems, as noted in Chapter 4, Section 4.1.12.

**Table 6.9 Fit statistics for 1- and 2-trait models of binary items**

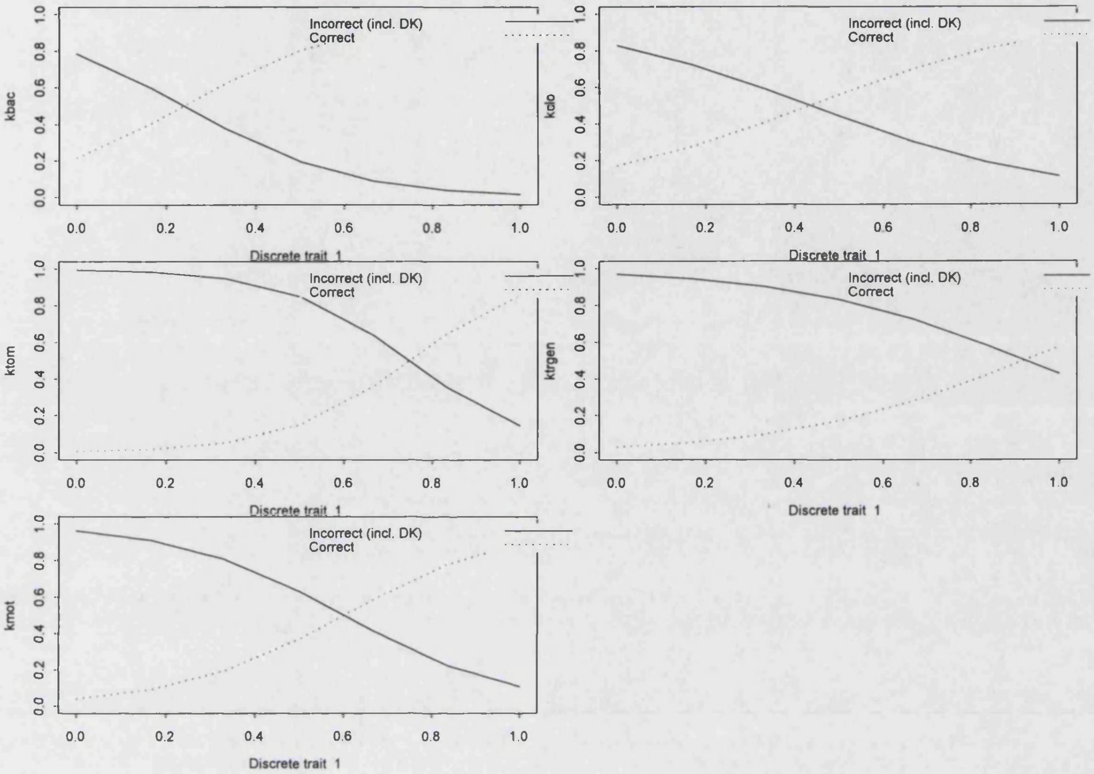
| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **Continuous trait, pooled data: no country covariate** | | | | | | | |
| 1 trait, 10 items[19] | 3,135 | 1003 | <0.001 | 1,129 | -6,579 | 27.8 | 276.58 |
| **Discrete traits, measurement model equal between countries** | | | | | | | |
| 1 trait, 10 items | 16,213 | 15,305 | <0.001 | -14,397 | -131,983 | 43.1 | 5.95 |
| 2 traits, 10 items | 14,449 | 15,274 | <0.001 | -16,099 | -133,447 | 25.4 | 2.83 |
| 1 trait, 5 items (*kbac, kclo, ktom, ktrgen, kmot*) | 1,459 | 435 | <0.001 | 589 | -2,753 | 26.3 | 2.92 |
| 2 traits, 5 items (*kbac, kclo, ktom, ktrgen, kmot*) | 1,021 | 409 | <0.001 | 203 | -2,939 | 14.2 | 1.09 |

Table 6.9 gives fit statistics for a selection of models. The first represents what would probably be the most common approach to modelling these cross-national data: that is, to run a one-trait continuous trait model on the pooled data from all countries, leaving the grouping variable out of the model altogether. Such a model fits quite poorly (27.8 per cent large two-way marginal residuals).

Moving on to discrete trait models, country-by-country analyses suggest that one trait is sufficient to represent the data: percentages of large two-way marginal residuals range from 0 in five countries, to 3.9 in Spain, and with an average of 2.10, and all items take loadings of the same sign. The country-by-country models may be qualitatively similar, but their parameters are different enough to make a joint model, fixing the measurement model between countries, fit very poorly (43.1 per cent large two-way marginal residuals). As above, both deleting problematic items from the scale, and increasing the number of traits, improves fit dramatically. With five items, 26.3 per cent of two-way marginal residuals are large. Notably, a somewhat different set of items are retained here in comparison with the model for polytomous items – in particular, *trgen* is included in the scale. A two-trait model for this set of items improves the fit further, but with flat response curves for three of the five items, it is of questionable value, and might be interpreted as a case of over-fitting. The one-trait model for five items is therefore arguably the preferred model from this section. ICCs for it are shown in Figure 6.5. It is interesting that correct responses to the item *trgen* are predicted to belong to only those at the very top of the scale.

[19] As this model was fitted to the pooled data set, without the country variable anywhere in the model, its fit statistics are not directly comparable to the other models described in the table.

**Figure 6.5 Item characteristic curves from a 1-trait discrete trait model for binary items, with measurement models equal for 15 countries**




### *6.5.3 Comparing scores from a selection of models*

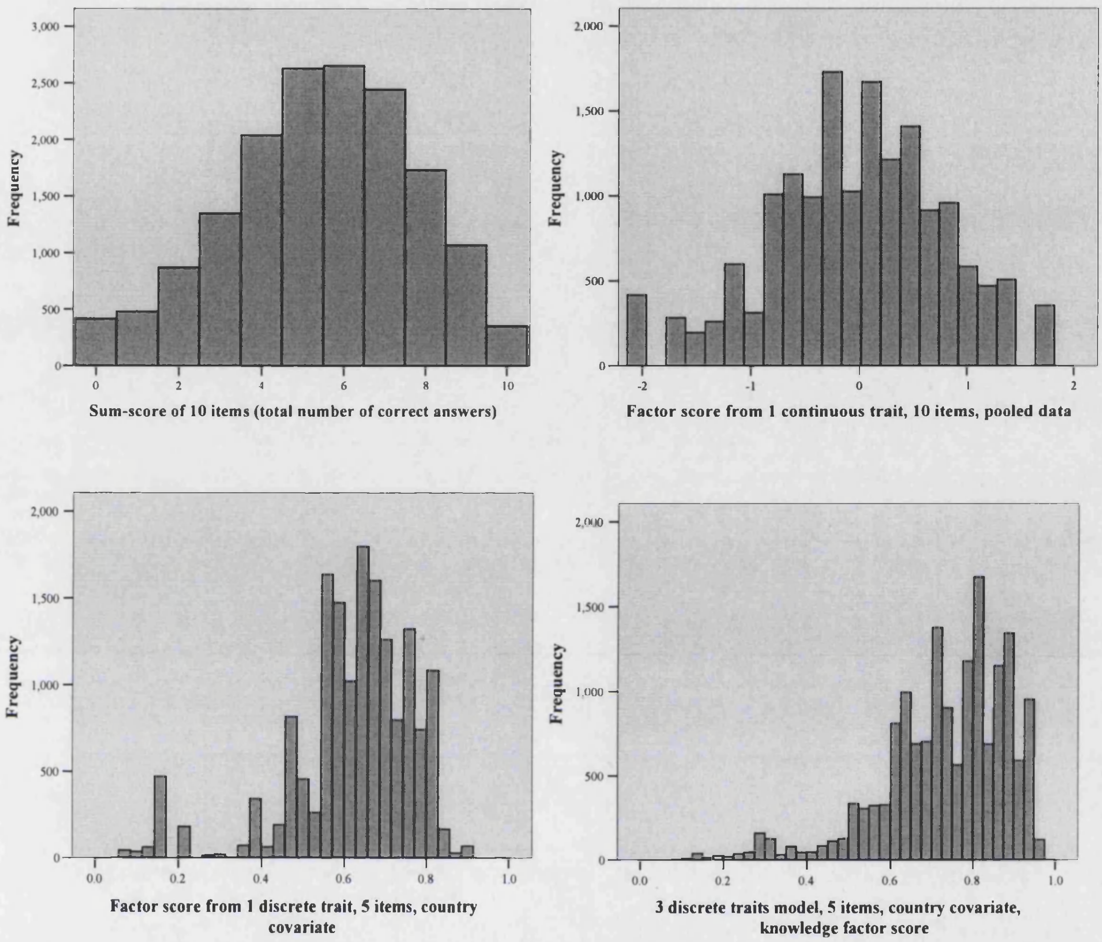
Despite the relatively poor fit of the two joint models of biotechnology items – or perhaps, because of this – it is instructive to review the scoring implications of the various different possible models of knowledge. Table 6.10 shows the correlations between a selection of models: firstly, a simple sum-score of the ten items, followed the factor score from a continuous trait model for the ten (binary) items, and the factor score from the best discrete trait model that could be found for a reduced set of five items. The last row of the table is given to the score from the final three-trait model for polytomous items; this is the most interesting, for our purposes. The final knowledge score is quite highly correlated with all of these alternative measures, with correlation coefficients approaching 0.8. The particular items included in each of these scales has some bearing on their correlations, of course. For example, the final discrete trait model score has a correlation of 0.79 with a sum-score of ten items, and of 0.93 with a sum-score of the same five items in the scale. Slightly stronger linear relationships are observed among the models for binary items, compared with the final model for polytomous items. Nevertheless, the relatively high correlations between all scores is encouraging.

**Table 6.10 Linear correlations between knowledge scores from a selection of models, data from 15 countries, weighted by respective populations**

| Pearson's correlation coefficient (r) | Binary items (correct/incorrect), single scale: Sum-score of 10 items | Factor score from 1 continuous trait, 10 items, pooled data | Factor score from 1 discrete trait, 5 items, country covariate | 3 discrete traits model, 5 items, country covariate, knowledge factor score |
|---|---|---|---|---|
| Sum-score of 10 items | 1 | | | |
| Factor score from 1 continuous trait, 10 items, pooled data | 0.99 | 1 | | |
| Factor score from 1 discrete trait, 5 items, country covariate | 0.83 | 0.84 | 1 | |
| 3 discrete traits model, 5 items, country covariate, knowledge factor score | 0.79 | 0.78 | 0.77 | 1 |

Lastly, the distributions of these scores are presented in Figure 6.6. The first two – the simple sum-score of ten items, and the factor score from a continuous trait model – return distributions that are approximately normal. The discrete trait models produced negatively skewed distributions of scores, whether the data are modelled as binary or polytomous. This is to some extent attributable to the items in the scales: sum-scores for the five items in each case are slightly negatively skewed.

**Figure 6.6 Histograms comparing distributions of knowledge scores from a selection of models, data from 15 countries, weighted by respective populations**




### *6.5.4 A point of reference: the Oxford scale of knowledge of science and technology*

Speculating on possible causes of the problems encountered in the sections above, the content of the items might be something to consider. Since in different models it was different items which caused the problems with fit, it seems not to be the case that odd question wording in a few places is to blame. A number of practical suggestions for the design of future knowledge items will be made at the end of this chapter, and in more depth in Chapter 9 (Section 9.4.2).

In the meantime, an informative comparison is to consider the same types of models for a different set of knowledge items. The Oxford scale, as described in Chapter 2, comprises a set of items of the same format as the biotechnology items, but relating to science in general, rather than focusing specifically on biotechnology. These items have been widely used in PUS research, and were posed in the 2005 Eurobarometer on biotechnology. Table 6.11 sets out the thirteen items from the 2005 Eurobarometer, with response distributions aggregated for the fifteen countries analysed in this thesis.

**Table 6.11 Distribution of responses to Oxford scale questions on knowledge of science: 2005 PUS Eurobarometer, 15 countries**

| *No.* | *Label* | *Statement* | *% responses*[20] | | |
|---|---|---|---|---|---|
| | | | TRUE | FALSE | DK |
| 1 | sun | The Sun goes around the Earth. | 31 | 65 | 4 |
| 2 | hot | The centre of the Earth is very hot. | 87 | 6 | 6 |
| 3 | oxygen | The oxygen we breathe comes from plants. | 80 | 15 | 4 |
| 4 | milk | Radioactive milk can be made safe by boiling it. | 10 | 75 | 16 |
| 5 | electrons | Electrons are smaller than atoms. | 45 | 30 | 25 |
| 6 | plates | The continents on which we live have been moving for millions of years and will continue to move in the future. | 88 | 5 | 7 |
| 7 | mother | It is the mother's genes that decide whether the baby is a boy or a girl. | 20 | 65 | 15 |
| 8 | dinosaurs | The earliest humans lived at the same time as the dinosaurs. | 22 | 67 | 11 |
| 9 | antibiotics | Antibiotics kill viruses as well as bacteria. | 40 | 49 | 11 |
| 10 | lasers | Lasers work by focusing sound waves. | 26 | 47 | 27 |
| 11 | radioactivity | All radioactivity is man-made. | 27 | 60 | 13 |
| 12 | animals | Human beings, as we know them today, developed from earlier species of animals. | 72 | 19 | 9 |
| 13 | month | It takes one month for the Earth to go around the Sun. | 19 | 65 | 16 |

*n=15518*

The analyses here are somewhat truncated, since they are, as in the previous section, intended to provide just a brief comparison with the biotechnology items. Modelling items as polytomous, and running two-trait discrete trait models, as in Section 6.5.1, country-by-country analyses identify a few items as problematic in terms of the signs of their loadings on a country's 'knowledge' trait. Deleting some items from the scale is

[20] Weighted frequencies, weighting each country's contribution to the total according to their respective population sizes. Totals do not always sum to 100 per cent due to rounding.

therefore a necessary step in finding a joint model. In order to make such a joint model as broadly comparable as possible with its biotechnology scale counterpart, nine items are retained.

**Table 6.12 Fit statistics for 1- and 2-trait models of polytomous Oxford items, with measurement models constrained to be equal across 15 countries**

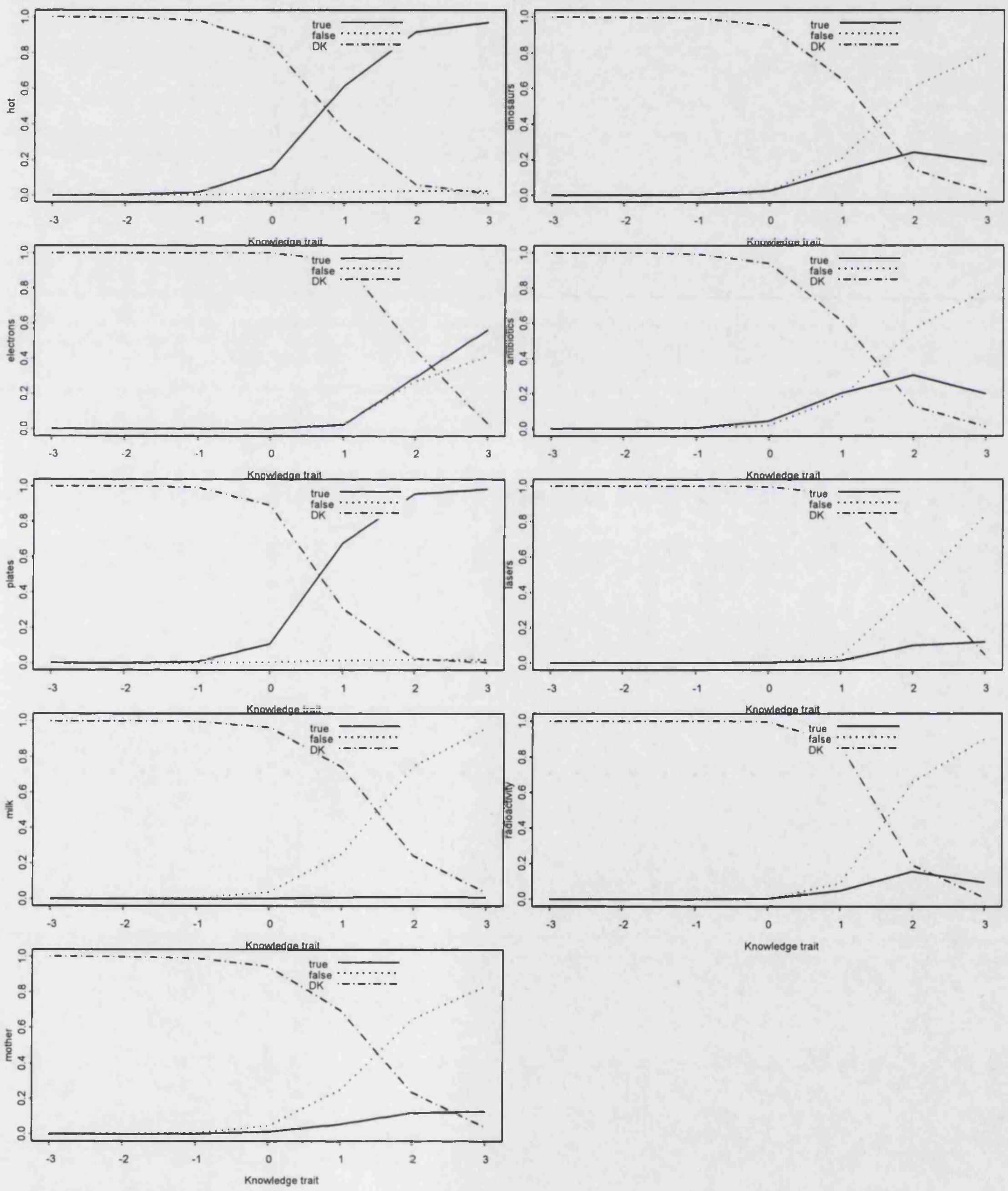
| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standard'd marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **2 traits; measurement models equal between countries for both traits** | | | | | | | |
| 9 items (no *sun, oxygen, animals, month*) | 35,391 | 15,423 | <0.001 | 4,545 | -113,437 | 28.1 | 2.91 |
| 6 items (no *sun, oxygen, animals, month, hot, electrons, plates*) | 11,295 | 10,843 | <0.001 | -10,391 | -93,337 | 27.4 | 2.72 |
| **2 traits; equal measurement model for 'knowledge' trait; for 'response style' trait, interaction between country, trait and all observed items** | | | | | | | |
| 9 items (no *sun, oxygen, animals, month*) | 31,295 | 14,904 | not available in LEM | 1,487 | -112,525 | 4.1 | 1.24 |

Table 6.12 shows that a two-trait model for these nine items, with equal measurement models between countries on both traits, fits poorly – although notably, not as poorly as the two-trait nine-item model for the biotechnology items (cf. 28.1 per cent large two-way marginal residuals for the former, versus 41.8 per cent for the latter). Reducing the number of items in the scale does very little to reduce the proportion of large two-way marginal residuals. However, relaxing the measurement model results in a great improvement in fit. The last model presented in the table is similar in form to that attempted for the biotechnology items: that is, allowing both slopes and intercepts to vary between countries on one trait, while constraining slopes to be equal across countries (with intercepts free to vary between countries) on the other. In this model only 4.1 per cent of two-way marginal residuals are large. Furthermore, the fixed trait can feasibly be interpreted as representing low to high knowledge. Figure 6.7 shows ICCs for this model, for UK respondents – that is, with the fixed, Europe-wide slopes but UK-specific intercepts.

A few observations are worth noting here. Firstly, the slopes for correct responses are relatively steep, for all items – compared with those in the biotechnology set (cf. Figure

6.4). Secondly, it is not the case that all of the 'false=correct' items are more difficult than the 'true=correct' items. The first three items in the diagram, in the left hand column, are those for which 'true' is the correct response. The second, 'Electrons are smaller than atoms', is a relatively difficult item, whose difficulty is not attributable to response style. These are two attractive features of the item set – features that would be very desirable in the biotechnology items.

**Figure 6.7 Item characteristic curves from final Oxford scale model, curves for UK**



## Summary

The analyses in this chapter have revealed a number of interesting findings, although these are coupled with a number of significant concerns. While the knowledge items in the Eurobarometer form relatively good within-country scales of knowledge, there are considerable obstacles to deriving a well fitting joint cross-national measure. To summarise the findings from the cross-national analyses of the biotechnology items:

- Treating the ten knowledge items as polytomous, that is retaining the distinction between DK and a substantively incorrect response, within each country sample in the data set a discrete trait model with two traits fits the ten knowledge items well.
- In all countries, one of the traits can be interpreted as capturing knowledge.
- In almost all countries, the other seems to capture the tendency to give DK responses, implying strongly that there *is* a meaningful distinction to be made between DK and substantively incorrect responses.
- On the knowledge traits, a few items are problematic in some countries:
  - For some of the 'true=correct' items, the slopes for the incorrect responses are increasing towards the high knowledge end of the scale, which could make factor scores inconsistent with a sum-score at the very end of the trait.
  - In five countries the substantively incorrect response for item *ktrgen3* is most likely at the high knowledge end of the trait.
- In some countries, those items for which 'false' is the correct response are more strongly discriminating than those for which 'true' is correct.
- These patterns of differential item functioning are not radically different from country to country. Between countries there appears to be a broadly similar ordering of items on the knowledge trait in terms of discrimination of the response 'false' in relation to the reference category 'true'. So we do not find significant difference in the relationship between correct and substantively incorrect responses, item to item.
- Despite the broad similarities in the interpretation of the trait model between countries, a joint model constraining all parts of the measurement model to be equal between them, fits very poorly.
- Some improvement to the model can be made by deleting particularly problematic items from the scale.

- However, doing so reveals the unstable nature of the solution: the successful identification of a 'knowledge' trait depends on the items included in the model.
- A further improvement to the model can be made by increasing the number of traits to three.
- In a three-trait model, using five items, we can identify one trait as representing 'knowledge', while the other two could be said to pick up different response effects.
- Allowing the measurement model to vary between countries for one of two traits does not result in a satisfactory model. Although freeing parameters improves the fit greatly, the fixed trait cannot be interpreted as representing high to low knowledge.
- Setting aside the distinction between DK and substantively incorrect responses, it is not noticeably easier to find a joint model for binary items. Again, deleting some items from the scale improves model fit.
- Comparing the scores from a selection of models of these items suggests that the different representations do not differ drastically. All are relatively highly linearly correlated with each other (0.77 or above). The distributions of these scores do vary in shape, however, according to the model used. Europe-wide, a simple sum-score of correct responses and a single continuous trait for binary items produce approximately normally distributed scores. By contrast, two discrete trait models (one for binary items and one for polytomous items) return factor scores with negatively skewed distributions.
- In this last model – the final model for the biotechnology items in this chapter, the distributions of scores vary somewhat between countries. With some exceptions, Northern European countries achieve high average scores, and those in the South, lower scores. For example, the mean knowledge level is 0.85 in Sweden, and 0.49 in Portugal.

Supporting findings from the analyses of British data are as follows:

- The differential discrimination powers of the items should not be ignored:
    - A Rasch model, treating the ten items as binary (correct/incorrect) and constraining their slope parameters to be equal, does not fit the data well.
    - Since the Rasch model is the closest probabilistic model to a simple sumscore, we can infer from it that the latter does not provide the best possible representation of the data.

- For these items, a two parameter logistic model (allowing the slope parameters to differ by item) does fit the data.
- However, estimated slope coefficients for some of the 'true=correct' items are unstable, since they are essentially flat. This presents problems for calculating posterior scores, as described above.
- The distribution of posterior factor scores on the knowledge trait is markedly affected by the choice of trait model. With a continuous trait model, the distribution of scores is approximately normal – similar to the distribution of a simple sum-score of correct responses. However, with a discrete trait model, the distribution of scores is bimodal, with strikingly large proportions of people at the very highest and very lowest ends of the trait.

Having not found a very well fitting joint model for the biotechnology items, an additional, brief set of analyses were carried out on a set of items from the Oxford scale of knowledge about science in general, fielded in the 2005 PUS Eurobarometer. Considering a joint model of nine items, for the fifteen European countries:

- Two-trait models, with slope parameters fixed to equality between countries, do not fit the data well. Reducing the number of items in the scale to six only marginally improves the fit (in terms of proportion of large two-way marginal residuals).
- Allowing the measurement model to vary between countries on one trait results in a very well fitting model, in which the trait that is fixed between countries can be interpreted as defining low to high knowledge.

In terms of ways to proceed with the biotechnology items, further exploratory analyses might be informative. For example, Pardo and Calvo's (2004) analyses of the Oxford items suggested that the properties of a sum-score of them varied between different socio-demographic groups. Drawing on this, we might expect the scale to work differently amongst those with, say, higher and lower levels of education. If a well fitting cross-national model could be found within those strata, it would be informative as to where and how the joint model does not work. A similar approach could be taken using clusters of countries; perhaps, for example, a well fitting model could be found within North Europe and South Europe.

A more dramatic recommendation would be to modify the items in future surveys. Item characteristics suggest that the 'true=correct' items in particular give cause for concern.

In some cases they are not very strongly discriminating, and therefore not of great value to the scale, whilst in others their flat curves actually cause problems in terms of posterior scores. An experimental approach would be required to determine with certainty whether these items are problematic because of their easy content (many people know the correct answer), or because they are hostage to 'acquiescence bias' (many people in doubt will default to a positive response). Such an experiment would seem a valuable first step in improving the scale, if these items are indeed obscured by response effects, since nuisance factors only make the task of finding a good cross-national scale of knowledge more challenging. A fuller discussion of this point, and some suggestions for future surveys, are made in Chapter 9.

The next chapter explores quite a different set of items. Whereas knowledge is a concept that has been of central interest in PUS for a number of years, and whose measurement has sparked a great deal of debate and engendered a strong convention in survey research, engagement with science is relatively speaking a more recent interest. The measurement of engagement with science has received much less formal and direct theoretical attention – it is not even clear whether a typology of engagement or a scale of engagement is indicated. So the next chapter adopts a more exploratory approach to examining the relationships between a more diverse range of items.

# 7 Engagement with science and biotechnology: a matter of degree or of kind?

The theme of this chapter is one of central importance to the current Science and Society turn in PUS research (Bauer et al., 2007): engagement with science and with biotechnology. In this scheme, members of the non-expert audience are freed of accusations of deficit of knowledge or deficit of positive attitudes, and the onus is placed on the science community to build relations of trust and cooperation with them. Within this perspective, public engagement with science and technology is a key indicator of the fertility of the climate for consultation. By many accounts, the publics of surveyed countries are not well furnished with knowledge of science (Jon D. Miller, 2004), and biotechnology in particular is an unfamiliar topic for them (Gaskell et al., 2003). As such, engagement with science and technology is a lifeline for the success of the Science and Society project. Without enthusiasm for reading, hearing and talking about the subject matter, be it in formal or informal settings, the public's capacity for meaningful participation in science-making can only remain limited.

In contrast with studies of attitudes towards and knowledge about science and technology, there has been relatively little discussion, and virtually no debate, about the best way of capturing the concept of engagement using survey data. This is no accident – Science and Society is associated with the use of action research and the rejection of survey methods. However, many researchers involved practically in public consultation exercises such as the UK's *GM Nation?* explicitly state the need for good survey data, both as components of these projects (Pidgeon et al., 2005) and as quality indicators in evaluations of them (Rowe et al., 2005). Surveys are therefore not obsolete in this new research wave, far from it – paradoxically they have a role to play in assessing the success of the shift away from outmoded science literacy and PUS approaches. However, given that these prior approaches have tended to be denounced via criticisms of the survey methods typically employed in them, Bauer et al. (2007) point out that there is a real risk of duplicating work in future research by ignoring the existing literature on science indicators in surveys – a 'reinvention of the wheel, but this time for a different car' (ibid., p.86).

In developing measurements of engagement with science, then, we do have recourse to the general lessons learned in the literature on other science indicators. In many ways, the same types of obstacles might be expected in this case as in capturing positive and negative attitudes, for example. So the task in this chapter is not very different from those already encountered. An appreciable difference, however, lies in the lack of a dominant baseline or conventional model from which to begin. Existing literature showcases a variety of possible approaches, employing different collections of variables to model concepts that are arguably either components of engagement or synonyms for it. Some of these were described in Chapter 2 (Section 2.4.3). They include 'interest' (Evans & Durant, 1995), 'informedness' (Pardo et al., 2002) and 'attentiveness' (Jon D. Miller & Pardo, 2000). Gaskell et al. (2006) use the broadest range of items to construct a typology of 'modes of engagement', including knowledge of biotechnology as a constituent feature of this model.

So the general research question of this chapter is how to capture the concept of engagement using the 2002 Eurobarometer survey data. In other words, the task is to investigate the associations between the items that in common sense terms relate to engagement, and more specifically, to find the best latent variable model to represent the data. Judgements of the 'best' representation will be made on the grounds of fit statistics and the interpretation of parameter estimates, with a little more emphasis on the former, given the methodological focus of the thesis. The following working research questions can be derived from this objective:

1. How are the various items relating to engagement associated with each other? Do any important response effects or styles emerge when they are analysed together? How can summary measures of responses to these items be created, and how might they be characterised? Is a classification of types of engagement meaningful? Can engagement be represented using one or more latent traits?

2. As a corollary to the findings from these explorations, how can fair and valid comparisons of engagement be made between respondents in different countries?

3. With a view to future Eurobarometer surveys, what recommendations can be made for the design of items capturing the concept of engagement with science and technology, and/or with biotechnology?

## 7.1 Data

Again the data are taken from the 2002 Eurobarometer, using the British sample for initial analyses. In contrast to the previous two chapters, here there are three sets of relevant items, taken from different parts of the questionnaire. Table 7.1 gives frequencies for eight questions designed to capture elements of engagement. The first four correspond to cognitive and affective aspects of engagement with science and technology in a broad sense: how interested and how knowledgeable respondents feel about the topic (hereafter referred to as the 'science' items). The remaining items are different in two respects: they focus specifically on biotechnology, and ask about behavioural elements of engagement rather than on cognition or affect (hereafter they will be referred to as the 'biotechnology' items). Two items ask only about hypothetical behaviours: would respondents, in principle, be willing to participate in a public forum or use the media to find out about biotechnology? The last two items ask for reports of actual behaviours – whether respondents have ever talked about biotechnology, and whether they have been exposed to coverage of biotechnology in various media forms.

The frequency distributions for these items show a relatively even spread of responses among the first four questions. In contrast with the data considered in previous chapters, these contain very few 'don't know' (DK) responses; at most, 2.5 per cent (only four per cent of response profiles for these four items include one or more DK response). Because of the very low rates of DK responses for these four items, in the analyses in this chapter they will be treated as missing, and Full Information Maximum Likelihood estimation used, which avoids listwise deletion of response profiles containing DK responses for these items[21].

The next two questions provoke much more equivocation, especially on taking part in public hearings and discussions, where 13 per cent will not be drawn, and 59 per cent of respondents would tend to decline to participate. Such a lack of enthusiasm for discussing biotechnology in a formal setting echoes low levels of experience of discussing it in any setting; nearly two thirds of respondents have never talked about biotechnology with anyone. Vocal engagement may be a tall order, then. However, nearly two thirds of respondents say they would be happy to engage with biotechnology

[21] In all models containing these items, the number of large marginal residuals, and the Jöreskog & Moustaki index, are calculated only for full, non-DK response profiles. That is, when calculating these statistics for the two-way tables in the data, those containing a DK response (e.g. a DK response for one item and a non-DK response for the other item) are not included in the final summary fit statistic.

in a more passive way, by reading articles or watching television programmes on the topic. This is not already a widespread habit, however; half of the sample have not, in the last three months at least, heard or read about biotechnology in the mass media. Where they have done so, it is most commonly on television or in newspapers.

**Table 7.1 Distribution of responses to engagement questions, British (GB) sample**

| *Label* | *Statement* | *% responses*[22] | | |
|---|---|---|---|---|
| ***Science items*** | | | | |
| | | Most of the time | Some of the time | Hardly any of the time |
| scint | I am interested in science and technology. | 26 | 36 | 37 |
| scinf | I feel well informed about science and technology. | 17 | 37 | 46 |
| scund | I understand science stories in the news. | 28 | 40 | 31 |
| scconf | I become confused when I hear conflicting views on science and technology. | 28 | 42 | 30 |
| *n=1002;994;1001;991* | | | | |

| | | | | |
|---|---|---|---|---|
| ***Biotechnology items*** | | | | |
| | | Tend to agree | Tend to disagree | Don't know |
| discuss | I would be prepared to take part in public discussions or hearings about biotechnology. | 28 | 59 | 13 |
| readtv | I would take time to read articles or watch television programmes on the advantages and disadvantages of biotechnology. | 67 | 25 | 8 |
| *n=1014* | | | | |

| | | | | | |
|---|---|---|---|---|---|
| | | Yes, frequently | Yes, occasionally | Yes, only once or twice | No, never |
| talkbr | Before today have you ever talked about modern biotechnology with anyone? | 4 | 19 | 13 | 65 |
| *n=1012* | | | | | |

| | | |
|---|---|---|
| | Before this interview, over the last three months, have you heard or read anything about issues involving modern biotechnology? | |
| heardbio | No. | 50 |
| npaper | Yes, in newspapers. | 27 |
| radio | Yes, on the radio. | 9 |
| mags | Yes, in magazines. | 9 |
| televis | Yes, on television. | 35 |
| www | Yes, on the internet. | 3 |
| forgot | Yes, does not remember where [*spontaneous*]. | 2 |
| *n=1014* | | |

[22] Weighted frequencies, applying the basic sampling weight in the data set. Totals do not always sum to 100 per cent due to rounding.

## 7.2 Types and levels of engagement

The choices and difficulties involved in analysing these data are a combination of those noted for the attitude items and the knowledge items. Here, however, the task of modelling the items might be expected to be more challenging. In technical terms there are two primary potential complications. Firstly, there are many items, similarly to the knowledge quiz. Finding a joint cross-national model generally becomes more difficult as the number of items in the set increases. And whereas the knowledge questions are taken as a sample of items from a broader universe of possible questions, making it feasible to drop certain items from a scale if they are shown to function irregularly, here the set of items more obviously represent different elements of engagement – how one feels, how one behaves, what one would in principle be willing to do. Excluding items from the set here carries greater consequences for the substantive representation of engagement that can be achieved with the data. Secondly, the collection of items are not from a single battery, but dispersed among other items in the questionnaire, and with quite an assortment of response formats. Given the response effects found in the previous two studies of sets of very similar items, greater problems might be anticipated in reconciling patterns of associations blurred by a variety of response effects.

In substantive terms, it is not obvious in this case which type of model would be preferable. Common sense suggests that is reasonable to think of some people as more engaged than others, implying a single dimension of engagement. Constructing a scale from these items, however, is not a simple business. For example, how should the media items be treated? Assigning scores (equal or otherwise) for every type of media in which a respondent has heard about technology results not in a scale of degree of exposure to biotechnology in the media, but in some sort of scale of multi-media-ness. It is not clear what substantive interpretation such a scale would have; moreover it is highly likely that it would represent a conflation of general media consumption levels and those specific to biotechnology – and we know that the former are strongly related to socio-demographic factors (Sturgis & Allum, 2006). If responses are combined into a dichotomised version (any versus none) however, the more finely grained information provided in them is lost. Whilst some forms of media are not commonly used by some respondent types (e.g. internet use is negatively associated with age), some forms of media are widely consumed across all demographic profiles (for example, television) –

so it might be useful to retain *some* form of distinction between common and less common media types.

A scale can be created in many ways. It entails questions of the relative weights to assign to different items, and in this case, to categories within items – thus the problems with weighting discussed in Chapter 6 are amplified here. Following a conventional sum-score approach, typically, scores of 0, 1 and 2 might be given to the three response categories in the science items, scores of 0–3 for the item *talkbr*, 0 and 1 for *discuss* and *readtv* (but how then to score DK?) and a point for every media item. Or the scores could be rescaled, to each have a range of 0–1. Without some careful item analysis, such as with latent trait models, deciding how to weight items remains a problematic exercise.

In a scaling approach, higher scores on the scale could signify a number of combinations of responses; relatively high scores could be allocated both to those who are emotionally engaged (interested, informed and the like) but not widely exposed to biotechnology in the media, as well as to those who are emotionally detached but who have heard about biotechnology before. If it is useful to make a distinction between these two cases, then a two-trait model, separating emotional engagement from media exposure, might be indicated. Or a typology of responses might give a useful representation of the data. We might begin with something analogous to that used for the attitude items, identifying and labelling the most frequent response profiles. And again, the problems with then including less common response profiles into the classification arise, as with the logics items.

Any researcher may have good reasons, developed from substantive theories, for choosing to model degrees of engagement or kinds of engagement – indeed, examples of both formats are seen in the existing literature for the engagement-related constructs cited earlier in this chapter (and in Chapter 2). It is not within the remit of this thesis to make a judgement between the two on substantive grounds, but to provide evidence regarding the statistical behaviour of the items, which might contribute to the choice of model. Both class and trait models will be used in this chapter, and their results compared.

## 7.3 Models considered in this chapter

Taking a cautious approach, the analyses relating to the first research question begin by separating the science items from the biotechnology items. In each set, and treating all variables as nominal, latent class models – the least restrictive of the models used in the thesis – will be employed to identify basic patterns in response profiles. From this point, tighter structures can be tested on the data with discrete latent trait models. These will be used to investigate dimensionality and any response effects in the items, and to ascertain whether those items with ostensibly ordinal response categories behave as such when modelled together. A brief comparison of the best-fitting trait versus class models will then be given, including a comparison of the posterior scores and class allocations derived from them – these are, in a sense, the practical implications of the models for the survey analyst wishing to derive a composite indicator of engagement.

Having taken these sets separately, the question is whether and (if so) in what way they fit together. Does a distinction between engagement with science and technology versus engagement with biotechnology – and/or affective versus behavioural engagement – suggest itself? Trait and class models will be used to investigate this possibility. Scores and class allocations from a trait and a class model will be presented, once again, as part of a comparison of the two final representations of these combined data.

The second research question, focusing on country comparisons, is approached in the same vein as in Chapters 5 and 6, though more similarly to the former than the latter. It is surprisingly difficult to find a well fitting joint model for the engagement items. As such, trait models are not presented; rather, the focus is on class models. Likewise, combining the science and biotechnology items increases the difficulty of finding a well fitting joint model, so these two aspects of engagement are taken separately. For each set of items, then, I return to the class models found for the British data, and run these country-by-country, making qualitative comparisons of their measurement models. Building on these, joint models are fitted, where the measurement models can be constrained to be equal across country samples, enabling fair comparisons to be made between them.

## 7.4 Results of latent trait and latent class analyses of British data

### *7.4.1 Affective and cognitive elements of engagement with science and technology*

Table 7.2 gives fit statistics for a selection of class and trait models. Beginning by considering the least constrained models, that is treating the observed and latent variables as nominal, just three latent classes are required to return an acceptable fit to the data.

**Table 7.2 Fit statistics for models of science items, GB**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **Class models for nominal items** | | | | | | | |
| 2 classes | 558 | 195 | <0.001 | 168 | -792 | 48.1 | 8.7 |
| 3 classes | 202 | 186 | <0.001 | -170 | -1,086 | 1.9 | 0.4 |
| 4 classes | 151 | 177 | 0.012 | -203 | -1,074 | 0.0 | 0.1 |
| **Class models for ordinal items** | | | | | | | |
| 2 classes | 588 | 199 | <0.001 | 190 | -789 | 57.4 | 9.7 |
| 3 classes | 305 | 194 | <0.001 | -83 | -1,038 | 24.1 | 3.6 |
| 4 classes | 246 | 189 | <0.001 | -132 | -1,063 | 14.8 | 1.8 |
| 5 classes | 214 | 184 | <0.001 | -154 | -1,059 | 11.1 | 1.5 |
| **Discrete trait models for ordinal items** | | | | | | | |
| 1 discrete trait | 303 | 194 | <0.001 | -85 | -1,040 | 24.1 | 3.6 |
| 2 correlated discrete traits | 204 | 183 | <0.001 | -162 | -1,063 | 1.9 | 0.8 |
| **Discrete trait models for nominal items** | | | | | | | |
| 1 discrete trait (7 levels) | 273 | 190 | <0.001 | -107 | -1,042 | 27.8 | 3.0 |
| 1 discrete trait (20 levels) | 269 | 177 | <0.001 | -85 | -956 | 25.9 | 2.9 |
| 2 correlated discrete traits (2 levels each) | 163 | 185 | 0.024 | -207 | -1,117 | 0.0 | 0.2 |
| 2 correlated discrete traits (7 levels each) | 144 | 175 | 0.104 | -206 | -1,068 | 0.0 | 0.1 |

Conditional and prior probabilities for this three-class model are presented in Table 7.3. It classifies responses cleanly into a class for low engagement (with respondents here most likely to be interested, informed and understand science stories in the news 'hardly any of the time', and become confused by conflicting views on science 'most of the time'), a class where the middle, 'some of the time' response, is the most likely for every question, and a class defined by high levels of engagement (being interested, informed and understanding science stories 'most of the time', and becoming confused 'hardly any of the time'). A four-class model provides a further distinction, though not a very clear one, between the middle responses and low engagement classes: in this extra class, respondents are likely to answer 'hardly any of the time' to the first two questions, and 'some of the time' to the second two – although responses for these two items are not very clearly defined.

**Table 7.3 Conditional and prior probabilities, 3-class latent class model for science items, GB**

| Item | Response category | High engagement $\hat{\pi}_{is}(1)$ | Middle responses $\hat{\pi}_{is}(2)$ | Low engagement $\hat{\pi}_{is}(3)$ |
|---|---|---|---|---|
| I am interested in science and technology | Hardly any of the time | 0.01 | 0.14 | 0.86 |
| | Some of the time | 0.11 | 0.72 | 0.13 |
| | Most of the time | 0.88 | 0.14 | 0.02 |
| I feel well informed about science and technology | Hardly any of the time | 0.04 | 0.24 | 0.95 |
| | Some of the time | 0.29 | 0.71 | 0.04 |
| | Most of the time | 0.66 | 0.05 | 0.01 |
| I understand science stories in the news | Hardly any of the time | 0.00 | 0.12 | 0.73 |
| | Some of the time | 0.13 | 0.72 | 0.22 |
| | Most of the time | 0.86 | 0.16 | 0.05 |
| I become confused when I hear conflicting views on science and technology | Hardly any of the time | 0.57 | 0.21 | 0.24 |
| | Some of the time | 0.29 | 0.63 | 0.26 |
| | Most of the time | 0.14 | 0.16 | 0.51 |
| $\hat{\eta}_j$ | (unweighted) | 0.22 | 0.41 | 0.37 |

Key

$\hat{\pi}_{is}(j)$ = estimated conditional probability of response in category *s* for item *i*, given membership of class *j*

$\hat{\eta}_j$ = estimated prior probability of membership in class *j*

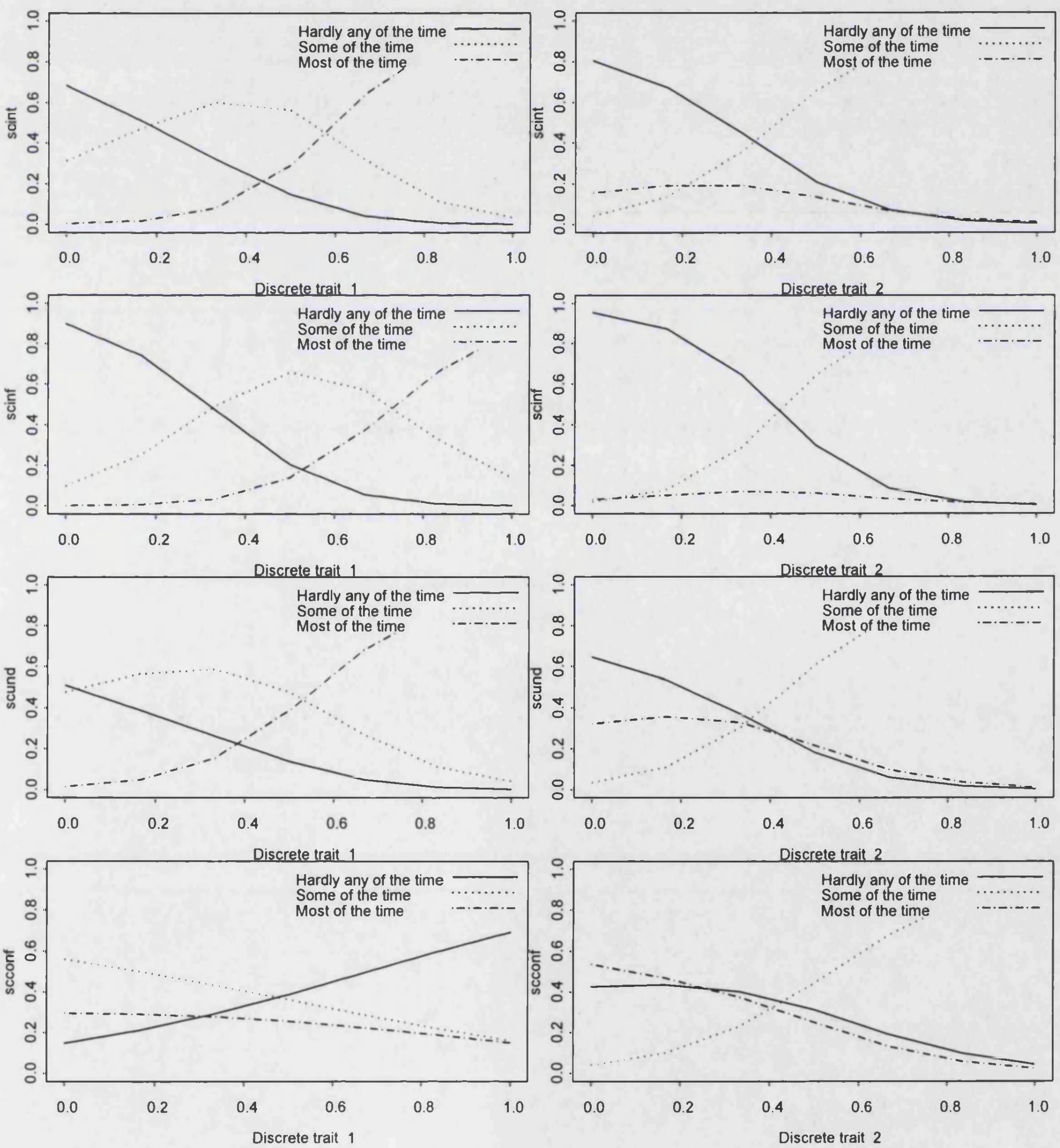
It seems therefore that interest in science and technology tends to go hand in hand with confidence in one's grasp of the subject. However, two further observations should be noted. Firstly, the pattern is stronger for the first three items in the set, and weaker in relation to the last item. The more irregular functioning of the last item might be attributed to a variety of factors. In terms of the mechanics of the survey response process, it may be slightly more cognitively challenging simply by virtue of having a negative connotation, in contrast with the other items. In terms of substantive content, it may be logically linked to levels of engagement in a number of ways, making for some degree of heterogeneity in its meaning among respondents. For example, the statement is a non sequitur for respondents who are unexposed to conflicting views on science (making responses for this group error-prone), while exposed-but-detached respondents might hear conflicting views on science but remain nonchalant regarding their incompatibility, i.e. some of these unengaged respondents might answer in a way which we would take to denote high levels of engagement with the topic. By the same token, some highly engaged respondents may be more apt to become confused by conflicting views on science and technology; it may even be this confusion that motivates them to become better informed on the subject.

The second observation from this model is that the data do not seem to behave ordinally. The latent variable is nominal, and although in common sense terms the middle class might be thought of as falling between the other two, no ordering of classes is imposed in the statistical model – the middle class is not *located* between the other two, statistically speaking. Using discrete trait models as a means of testing the ordinality of the latent variable, it is clear that response profiles cannot be represented accurately on a single scale, regardless of the number of categories given to the scale. Increasing the number of classes (see the example of 20 levels on a discrete latent trait, in Table 7.2) does little to improve fit if only one ordinal latent variable is used. By contrast, a two-trait discrete trait model with only two levels per trait, yielding a cross-classification of four classes, fits the data very well. (In this model the four classes have the same interpretation as those in a four-class unrestricted latent class model, with high, middle and low classes, and one describing low–mid engagement.) Likewise, it is difficult to find a well fitting class model when the *observed* items are treated as ordinal. A three-class model for nominal observed items fits well, but when they are treated as ordinal, even five classes fail to provide a well fitting representation of the data.

This hint at the non-ordinality of the observed items is more directly elucidated with latent trait models. Using discrete latent trait models (with the default of seven levels per trait as an approximation to a continuous variable), regardless of whether the items are modelled as ordinal or nominal, single-trait models fit poorly, but two-trait models fit well. In the model for ordinal items, one trait could be interpreted as representing 'engagement', with large positive loadings for the first three items and a negative loading for the last. On the other trait all items take positive loadings (decreasing in size from the first to the fourth); this might be interpreted as a response effect trait, with those at the top end of the trait answering 'most of the time' to all items.

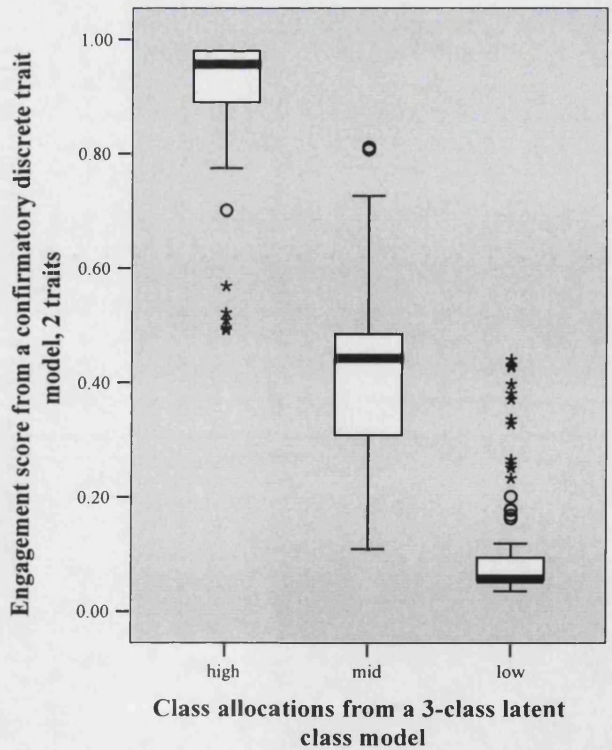
The model for nominal variables offers a different interpretation for a response effect trait, illustrated in the item characteristic curves (ICCs) given in Figure 7.1. The first trait (items in the column on the left hand side of the figure) represents low, running through medium, to high engagement. The second could be taken as depicting a response style, but this time with 'some of the time' responses at the top end of the trait, and a mixture of the responses broadly corresponding to low engagement at the lower end. It may be that this second trait captures some kind of undecided element in responses; we might speculate that it fills a similar function to a DK response. So trait models both for nominal and for ordinal observed items suggest a response effect in the data. In both, two traits are needed to obtain satisfactory model fit, and in both, the second trait could be interpreted as representing a response style; however, the nature of the response style differs according to the model.

**Figure 7.1 Item characteristic curves from an exploratory 2-trait discrete trait model for science items, GB**



It seems then that there is a multidimensional structure in the data which cannot be ignored, but which can speak equally well via a two-trait model (two ordered latent variables) or a three-class model (one nominal latent variable) – though they each carry different emphases. Comparing scores from the trait model for nominal observed items with class allocations from the class model (see the box plot in Figure 7.2), those in the high and low engagement classes take positions at the high and low ends of the engagement trait, with little variation. Those falling into the 'middle responses' latent class are distributed along the remaining range of the variable, with a median score around the centre of the trait. So the trait model provides finer distinctions along a dimension of engagement among those giving one or more 'some of the time' responses – and explicitly models the tendency to give this answer, as a separate dimension of response patterns. In choosing between class and trait models we must decide whether it is important to tease this tendency apart from a purer scale of engagement. It might be contended that we should not create a scale of engagement at all – that we should instead model types. In such a representation, 'some of the time means' simply some of the time, and no judgement need be made about the location of a set of such answers relative to other response profiles.

**Figure 7.2 Comparison of engagement scores from trait model with class allocations from a 3-class model; science items, GB**



### *7.4.2 Behavioural elements of engagement with biotechnology*

As a result of initial analyses, some items in this set are simplified. Amongst the types of media where respondents may have heard about biotechnology, those with very low frequencies (*radio, magazines, the internet, respondent does not remember where*) are combined into an *other media* variable. A binary version of the question 'Before today, have you ever talked about biotechnology with anyone?' is used ('no' versus 'yes') rather than its full four-category version. All variables in this section are nominal: mostly binary, but with the two items on willingness to engage containing the third, DK, category. Fit statistics for a selection of models are given in Table 7.4.

**Table 7.4 Fit statistics for models of biotechnology items, GB**

| Model | $L^2$ | d.f. | p (bootstrap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **Class models for nominal items** | | | | | | | |
| 2 classes | 247 | 126 | <0.001 | -5 | -625 | 7.4 | 1.10 |
| 3 classes | 183 | 117 | <0.001 | -51 | -626 | 3.7 | 0.49 |
| 4 classes | 136 | 108 | <0.001 | -80 | -612 | 1.2 | 0.30 |
| 5 classes | 100 | 99 | 0.144 | -98 | -586 | 0.0 | 0.15 |
| **Discrete trait models for nominal items** | | | | | | | |
| 1 discrete trait | 221 | 121 | <0.001 | -21 | -616 | 7.4 | 0.96 |
| 2 discrete traits, confirmatory model *(model shown in Figure 7.3)* | 134 | 110 | 0.026 | -86 | -627 | 1.2 | 0.28 |

Beginning again with latent class models, although a three-class solution fits well, a four-class model returns a clearer interpretation. The conditional probabilities for it, presented in Table 7.5, identify a high engagement class, in which respondents are likely to have talked about biotechnology before, and likely to agree, in principle, to take part in public hearings on the topic and to take time to read articles or watch television programmes about it. By way of exposure to biotechnology via the media, they are likely to have heard about it on television, with even chances of having read about it in newspapers, but slightly less than even chances of having heard about it in other media. Next is a moderately engaged class, similar to the first but in which respondents are unlikely to want to take part in public discussions on the subject. Those in the low engagement class are likely to answer negatively to all questions, except regarding willingness to read an article or watch a programme about biotechnology, which they are marginally likely to be willing to do. Finally a DK class represents those with profiles representing low engagement in terms of talking and hearing about biotechnology, and DK responses for the hypothetical participation questions.

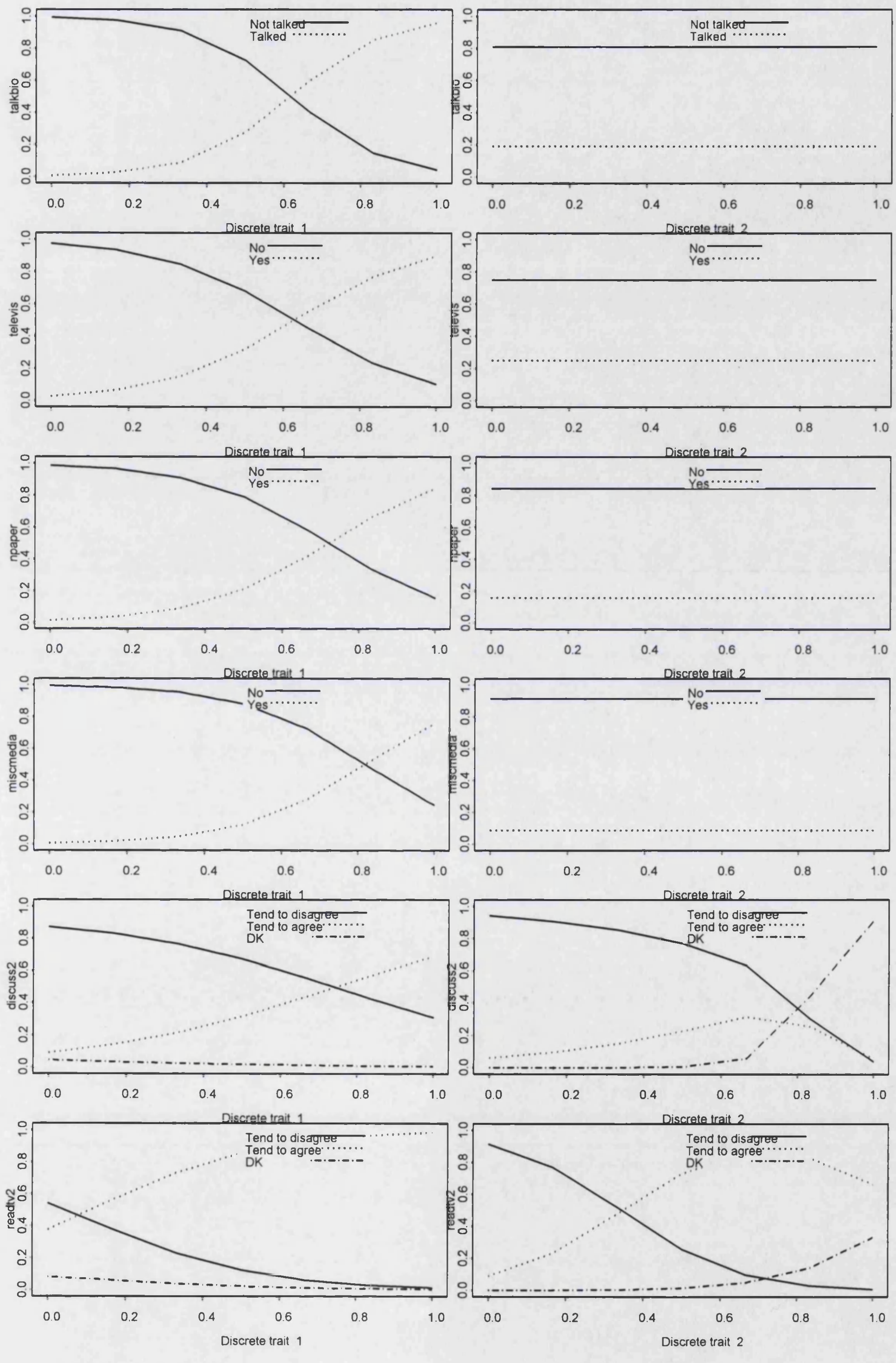
**Table 7.5 Conditional and prior probabilities, 4-class latent class model for biotechnology items, GB**

| Item | Response category | High $\hat{\pi}_{is}(1)$ | Mid $\hat{\pi}_{is}(2)$ | Low $\hat{\pi}_{is}(3)$ | Low, plus DK $\hat{\pi}_{is}(4)$ |
|---|---|---|---|---|---|
| Before today, ever talked about modern biotech? | No | 0.32 | 0.34 | 0.97 | 0.92 |
| | Yes | 0.68 | 0.66 | 0.03 | 0.08 |
| Would take part in discussions or hearings. | DK | 0.15 | 0.01 | 0.06 | 0.83 |
| | Tend to disagree | 0.06 | 0.99 | 0.78 | 0.15 |
| | Tend to agree | 0.79 | 0.01 | 0.16 | 0.02 |
| Would watch TV programme or read articles. | DK | 0.00 | 0.05 | 0.06 | 0.53 |
| | Tend to disagree | 0.04 | 0.21 | 0.43 | 0.01 |
| | Tend to agree | 0.96 | 0.74 | 0.50 | 0.46 |
| Over last 3 months, have heard or read about biotech via.. | | | | | |
| Television | No | 0.37 | 0.38 | 0.94 | 0.82 |
| | Yes | 0.63 | 0.62 | 0.07 | 0.18 |
| Newspapers | No | 0.50 | 0.51 | 0.95 | 0.94 |
| | Yes | 0.50 | 0.49 | 0.05 | 0.06 |
| Other media (radio, magazines, internet, forget where) | No | 0.57 | 0.75 | 0.98 | 0.94 |
| | Yes | 0.43 | 0.25 | 0.02 | 0.06 |
| $\hat{\eta}_j$ | (unweighted) | 0.26 | 0.22 | 0.44 | 0.08 |

As with the science items, latent trait models can be used to look for an ordered latent space. A one-trait model does not fit prohibitively poorly – but two traits improve the fit considerably. The contribution of the two-dimensional model is to provide a response effect trait for the two items with DK responses (parameter estimates for other items on this trait are not significant at the 5 per cent level). Figure 7.3 shows ICCs for a more parsimonious, confirmatory version of this model, in which just these two items load on the second trait, teasing the response effect apart from the main engagement trait. On the main engagement trait we can see that the media items can be ordered in terms of difficulty, though the difference between them is only slight. The trace lines for positive and negative responses intersect at the point where both are equally likely; at higher positions on the trait a positive response is more likely than a negative response. This intersection is at a slightly higher point on the trait for *miscmedia* compared with *npaper*, and *npaper* compared with *televis*. So some types of media exposure to biotechnology are associated with slightly higher levels of engagement, but only slightly higher levels. This is reflected in the conditional probabilities in the class model – for example, overall every respondent is more likely to give a negative than a positive response to *miscmedia*, but relatively, those in the high engagement group have a greater chance of saying that they have heard or read about biotechnology in these other media forms. The distinction between the types of media is retained in these models to illustrate the small differences, but combining the media types together results in no great loss of information. For example replacing the three media items with one, a four-class latent class model fits well[23], and records whether respondents have heard or read about biotechnology in *any* media. A very similar pattern of response is then found to that in Table 7.5, with those in the high and mid classes likely to say they have heard about biotechnology from some source, and those in the low and low-plus-DK classes likely to say they have not heard or read about it from any source.

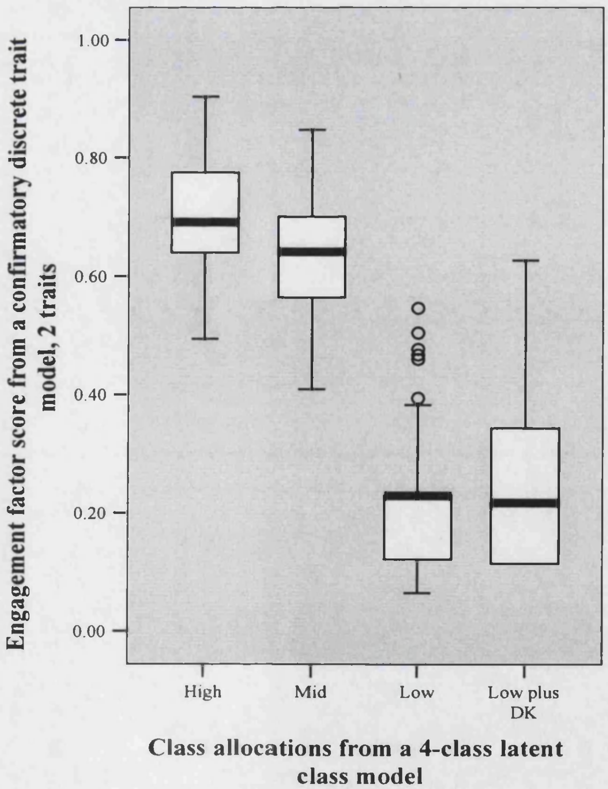
[23] $L^2 = 28$, d.f. = 8, p (bootstrap) $< 0.001$, AIC = 12, BIC = -27; no two-way standardised marginal residuals greater than 4; Jöreskog & Moustaki index = 0.13. Conditional probabilities are described in the text but not presented here since this example is given just as an illustration.

**Figure 7.3 Item characteristic curves from a confirmatory 2-trait discrete trait model for biotechnology items, GB**



Comparing the posterior class allocations and scores from the class and trait models respectively, Figure 7.4 suggests that the class model distinguishes between certain response profiles in a way that the trait model does not. For example, the distinction between the high and mid engagement groups in the class model is based solely on responses to the item *I would be prepared to take part in discussions or public hearings*, and as such (and noting from Figure 7.3 that the trace lines for this item are not very steep), the distributions of trait scores in the high and mid classes of engagement with biotechnology are very similar. Likewise, median trait scores for those in the two low engagement classes are approximately equal, though in the DK class there is a greater variance of scores, with relatively more respondents positioned at higher levels of the trait. By contrast, around the centre of the trait, at scores of 0.5 or so, we find whiskers and outlying scores for all the classes, and boxes for none of them – that is, the class allocation for a person with a score of 0.5 on the trait is not clearly defined. So the class and trait models entail slightly different emphases in scoring; the class model distinguish more finely between respondents at the ends of the continuum but not in the middle, and vice versa with the trait model.

**Figure 7.4 Comparison of engagement scores from trait model with class allocations from a 4-class model; biotechnology items, GB**



### *7.4.3 A combined indicator of engagement with science and technology and with biotechnology*

Beginning with class models, taking forward the three- and four-class models from Sections 7.4.1 and 7.4.2, up to twelve classes might be needed to achieve adequate fit. In fact, anything from four classes upwards would be acceptable in terms of fit statistics (see Table 7.6). Being able to combine the items together into such a small number of classes suggests a relatively strong association between the two sets of items.

**Table 7.6 Fit statistics for selected models for all engagement items, GB**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **Unrestricted class models** | | | | | | | |
| 2 classes | 3,374 | 981 | <0.001 | 1,412 | -3,416 | 21.5 | 2.95 |
| 3 classes | 2,984 | 964 | <0.001 | 1,056 | -3,689 | 9.9 | 1.74 |
| 4 classes | 2,788 | 947 | <0.001 | 894 | -3,767 | 2.6 | 0.71 |
| 5 classes | 2,670 | 930 | <0.001 | 810 | -3,768 | 2.0 | 0.59 |
| 6 classes | 2,579 | 913 | 0.016 | 753 | -3,741 | 2.0 | 0.45 |
| 7 classes | 2,519 | 896 | 0.006 | 727 | -3,683 | 1.7 | 0.31 |
| **Discrete trait models: initial basic models tested** | | | | | | | |
| 1 discrete trait | 3,022 | 976 | <0.001 | 1,070 | -3,734 | 14.2 | 1.96 |
| 2 discrete traits, exploratory: all items loading on each trait; correlated traits | 2,719 | 953 | 0.004 | 813 | -3,877 | 6.3 | 0.84 |
| 2 discrete traits, confirmatory: science items loading on one trait, biotech loading on second trait; correlated traits | 2,773 | 969 | <0.001 | 835 | -3,934 | 7.3 | 1.0 |
| 2 discrete traits, confirmatory: all items loading on one trait, *discuss* and *readtv* loading on second trait; correlated traits | 2,960 | 963 | <0.001 | 1,034 | -3,706 | 12.3 | 1.68 |
| **3 discrete traits confirmatory: science items loading on first and second traits, biotech items loading on third trait** | | | | | | | |
| All factors uncorrelated | 2,915 | 956 | <0.001 | 1,003 | -3,702 | 34.0 | 5.35 |
| Correlation between science-substantive and science-response-style traits | 2,914 | 955 | <0.001 | 1,004 | -3,696 | 34.0 | 5.33 |
| Correlation between two substantive traits | 2,648 | 955 | 0.018 | 738 | -3,963 | 2.3 | 0.63 |
| Correlation between two substantive traits, and between science-substantive and science-response-style traits | 2,635 | 954 | 0.028 | 727 | -3,968 | 2.3 | 0.64 |
| All factors correlated | 2,631 | 953 | 0.036 | 725 | -3,965 | 2.3 | 0.61 |

In order to explore the nature of this association fully, I present below not the four-class model but the seven-class model. Although seven classes are not necessary to achieve statistically good fit, they return the clearest interpretation of classes. Moreover, a seven class solution includes a class where the DK responses for *discuss* and *readtv* are well defined. This is particularly useful, given the ongoing concern with DK responses in the thesis.

The conditional probabilities for this model are given in Table 7.7. It identifies three sets of classes based on responses to the science items: high, mid and low engagement with science, and within those sets, classes of high and low engagement with biotechnology. This suggests that there is a two-dimensional structure to the data. Moreover, there seems to be an association between these two dimensions. For example, in the two classes of high science engagement, a distinction can be made between those who would and would not take part in public discussions on biotechnology (corresponding to high and low engagement with biotechnology). In the low science set however, respondents in *neither* high nor low biotechnology engagement classes would be willing to take part in such discussions. In the third low science class, this item is most likely to elicit a DK response. This last class, accounting for only a small group of respondents, contains the clearest profile of the unengaged, where respondents are almost certain to answer 'hardly any of the time' to the first three science items. Taking another illustrative contrast, respondents in *both* high science classes would be most likely willing to watch a television programme or read an article about biotechnology, as would those in both mid science engagement classes – but in the low science engagement set, likely replies are variable, with 'yes', 'no' and DK corresponding to high, low and DK groups respectively.

**Table 7.7 Conditional and prior probabilities, 7-class latent class model, GB**

| Item | Response category | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(5)$ | $\hat{\pi}_{is}(6)$ | $\hat{\pi}_{is}(7)$ |
|---|---|---|---|---|---|---|---|---|
| | Science class type: | High | High – | Mid | | Low + | Low | Low – |
| | Biotech class type: | High | Low | High | Low | High | Low | DK |
| I am interested in science and technology | Hardly any of the time | 0.00 | 0.04 | 0.03 | 0.21 | 0.55 | 0.89 | 1.00 |
| | Some of the time | 0.06 | 0.23 | 0.69 | 0.74 | 0.41 | 0.08 | 0.00 |
| | Most of the time | 0.94 | 0.73 | 0.28 | 0.05 | 0.04 | 0.03 | 0.00 |
| I feel well informed about science and technology | Hardly any of the time | 0.04 | 0.04 | 0.10 | 0.33 | 0.82 | 0.91 | 0.99 |
| | Some of the time | 0.19 | 0.40 | 0.84 | 0.67 | 0.14 | 0.08 | 0.01 |
| | Most of the time | 0.77 | 0.56 | 0.06 | 0.00 | 0.04 | 0.01 | 0.00 |
| I understand science stories in the news | Hardly any of the time | 0.00 | 0.02 | 0.00 | 0.19 | 0.44 | 0.75 | 1.00 |
| | Some of the time | 0.02 | 0.33 | 0.70 | 0.72 | 0.50 | 0.16 | 0.00 |
| | Most of the time | 0.98 | 0.65 | 0.29 | 0.09 | 0.05 | 0.08 | 0.00 |
| I become confused when I hear conflicting views on science and technology | Hardly any of the time | 0.72 | 0.41 | 0.28 | 0.10 | 0.24 | 0.23 | 0.30 |
| | Some of the time | 0.17 | 0.42 | 0.60 | 0.71 | 0.53 | 0.12 | 0.21 |
| | Most of the time | 0.10 | 0.17 | 0.12 | 0.19 | 0.23 | 0.65 | 0.49 |
| Before today, ever talked about modern biotech? | No | 0.13 | 0.75 | 0.26 | 0.98 | 0.47 | 0.94 | 0.98 |
| | Yes | 0.87 | 0.25 | 0.74 | 0.02 | 0.53 | 0.06 | 0.02 |
| Would take part in discussions or hearings | DK | 0.06 | 0.09 | 0.11 | 0.16 | 0.14 | 0.00 | 0.60 |
| | Tend to disagree | 0.33 | 0.62 | 0.38 | 0.67 | 0.61 | 0.88 | 0.36 |
| | Tend to agree | 0.61 | 0.29 | 0.51 | 0.17 | 0.25 | 0.12 | 0.04 |
| Would watch TV programme or read articles | DK | 0.00 | 0.05 | 0.03 | 0.12 | 0.03 | 0.04 | 0.47 |
| | Tend to disagree | 0.10 | 0.15 | 0.03 | 0.32 | 0.20 | 0.54 | 0.28 |
| | Tend to agree | 0.90 | 0.80 | 0.94 | 0.56 | 0.77 | 0.42 | 0.24 |
| Over last 3 months, have heard or read about biotech via... | | | | | | | | |
| Television | No | 0.33 | 0.74 | 0.35 | 0.93 | 0.39 | 0.88 | 1.00 |
| | Yes | 0.67 | 0.26 | 0.65 | 0.07 | 0.61 | 0.12 | 0.00 |
| Newspapers | No | 0.37 | 0.77 | 0.47 | 0.94 | 0.61 | 0.97 | 0.98 |
| | Yes | 0.63 | 0.23 | 0.53 | 0.06 | 0.40 | 0.03 | 0.02 |
| Other media (radio, internet, mags, forget) | No | 0.37 | 0.92 | 0.66 | 0.97 | 0.76 | 0.98 | 0.95 |
| | Yes | 0.63 | 0.08 | 0.34 | 0.03 | 0.24 | 0.02 | 0.05 |
| $\hat{\eta}_j$ | (unweighted) | 0.10 | 0.12 | 0.17 | 0.18 | 0.16 | 0.20 | 0.07 |

Once again, the relationship between the science and biotechnology dimensions of engagement can be explored more directly with latent trait models. Fit statistics in Table 7.6 give striking support to the story emerging from response patterns in the latent class model: the data do not fit a unidimensional scale, and the two dimensions needed to represent them are correlated with one another.

Looking first at the initial two-trait models, in an exploratory model, both traits seem to represent low to high engagement, with loadings clearly separating science from biotechnology items (the former have high loadings in the expected direction on one trait and low loadings on the other, and vice versa). Fixing these two sets of items to load only on separate traits increases the proportion of high two-way marginal residuals only slightly. However, forcing them to load only on a single dimension (the last model in this set, with the other trait reserved to represent DK responses to *discuss* and *readtv*) reduces model fit markedly. The data do not therefore fit a single dimension well.

The best model, judged by a combination of fit, parsimony and substantive interpretation, includes one trait each for the science and biotechnology items, alongside an extra trait to take account of the 'middle response' effect of the science items. The response effect trait for the two biotechnology items with DK responses can be dropped here, since it makes little contribution to model fit. Fit statistics in the last section of Table 7.6 demonstrate clearly that the crucially important correlation in the model is that between the substantive science and biotechnology traits. Including this correlation reduces the percentage of high two-way marginal residuals from 34 to 2.3. The correlation between the substantive and response effect science traits does not reduce the proportion of high residuals, and as such is not vital to the model. However, it results in a significant reduction in $L^2$ ($p<0.001$), whereas the remaining correlation, between the biotechnology trait and the response effect trait, is clearly not needed ($p=0.048$ for the comparison of the last two models in the table).

The final trait model for this section is therefore the one which allows the two dimensions of engagement with science to be correlated, and the two substantive engagement traits to be correlated. The measurement model for this reflects very closely the separate models illustrated in Figures 7.1 and 7.3, so further details of them are not presented here. Considering briefly the distributions of factor scores for the science and biotechnology engagement traits, each is not far from normally distributed, though for engagement with science there are three slight peaks, at the ends and in the middle of the scale, reflecting the large proportion of respondents who give the same answer to the first three items. The scores from these traits are highly correlated with one another, as to be expected from the model fitting exercise (Pearson's r = 0.81). It is useful to note that the scores from this combined model are highly correlated (approximately 0.95 in both cases) with their counterpart scores from the separate models of science and biotechnology, implying that these are quite stable solutions, not dramatically affected by their correlation with each other.

The analysis of data from the British sample therefore suggests that both kinds (classes) and degrees (traits) can be used to represent the data in a statistically satisfactory way. Firstly, there is a reasonably clear distinction between affective engagement with science and technology generally, and behavioural engagement with biotechnology in particular. These kinds of engagement are not independent of one another, however, but positively associated. Within the type of science engagement we can model degrees of engagement, but in doing so we need to take into account the two-dimensional structure of the data; the item categories are not straightforwardly ordinal in relation to each other. Within the type of biotechnology engagement, items more or less straightforwardly form a unidimensional scale from low to high engagement. In terms of statistical fit, these two types of engagement can be modelled together equally well either with traits or with classes.

## 7.5 Extending latent class analyses to other country samples

As noted in Section 7.3, it is difficult to find a well fitting joint model for these data. Initial exploratory analyses suggested that a latent trait model is not feasible for the cross-national data, and that combining science and biotechnology items into one model is likewise problematic for the full data set. In this section I therefore present a selection of latent class models, fitted separately to the science and biotechnology items. So the starting point for Section 7.5.1 is the model reached for science engagement items with British data; that is a three-class model. Following a brief comparison of measurement models for Great Britain and Northern Ireland, and East and West Germany, to check that it is acceptable to combine these into the UK and Germany, the analyses proper begin with qualitative comparisons of country-by-country three-class models. From these a number of steps are taken to reach a joint model. Section 7.5.2 follows a similar process for the biotechnology items. So the outcomes of the chapter are two cross-national models, one for affective engagement with science and technology, and one for behavioural engagement with biotechnology.

### *7.5.1 Affective and cognitive elements of engagement with science and technology*

Beginning with the three-class model reached for the British data, a first simple analysis suggests that it is acceptable to combine Great Britain and Northern Ireland into one UK sample, and likewise to combine the two separately sampled regions of Germany into one group. The very low numbers of large marginal residuals in the models with fully constrained measurement models (Table 7.8) obviate the need for formal comparisons of fit with less restricted models. A note to accompany Table 7.8 which is useful to keep in mind for the joint model is that this chapter's models contain, on the whole, far fewer two-way margins than those in Chapters 5 and 6. We may thus need to recalibrate the idea of what counts as 'many' large residuals. For example, conditional on region there are 63 two-way margins for these science items, compared to 150 for the logics items in Chapter 5 – so the stated 4.8 per cent large residuals in Great Britain and Northern Ireland correspond to just three residuals in each case. These involve a variety of questions and responses, that is they do not seem to identify any one particularly problematic item. For example, the joint model underestimates the number of Britons and overestimates the number of Northern Irish who say they are interested in science and technology 'most of the time' but feel informed only 'some of the time', and does likewise for those who say they understand science stories in the news 'most of the time' but feel informed only 'some of the time'.

**Table 7.8 Fit statistics from testing measurement models between Great Britain and Northern Ireland, and between East and West Germany**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | | | Jöreskog & Moustaki index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | All | GB | NI | All | GB | NI |
| UK, measurement model equal | 397 | 348 | <0.001 | -299 | -2,103 | 4.5 | 4.8 | 4.8 | 0.42 | 0.74 | 1.42 |
| | | | | | | All | West | East | All | West | East |
| Germany, measurement model equal | 522 | 465 | <0.001 | -408 | -3,022 | 2.2 | 0.0 | 1.6 | 0.50 | 0.61 | 0.67 |

Three-class models fit well in most countries in the data set, echoing the analyses of the British data[24]. Inspecting the measurement models in these country-by-country analyses suggests that the last item, which for the British sample notably fitted less well with the other items, is problematic generally. Table 7.9 presents a qualitative summary of the most likely responses in each class, for each country. This shows clearly that the first three items mirror each other consistently across countries, with a very few exceptions, whereas the last item brings with it considerable variation. It is not just between countries that this item produces such heterogeneity of responses: within countries, conditional probabilities are generally much lower than for the other items. Indeed, even the most likely responses listed in the table are not very clearly defined. In the light of the comments from Section 7.4.1 on the multiple possible interpretations of this

[24] The percentage of standardised marginal residuals > 4 is on average 4.1 across countries, with a range of 0.0 in Portugal to 11.1 in Finland.

question, it seems sensible to discard it at this stage. With the remaining three items, three-class solutions fit very well country-by-country[25], and qualitative inspections of most likely responses reveal exactly the same patterns as in the top half of Table 7.9.

**Table 7.9 Qualitative summaries of highest conditional probabilities from unconstrained 3-class models, 15 countries**

| *Items and countries* | *Classes and responses* | | |
|---|---|---|---|
| | High engagement | Middle responses | Low engagement |
| **Interested** | | | |
| All countries except... | Most | Some | Hardly |
| Greece | Most | Some/Most | Hardly |
| **Informed** | | | |
| All countries except... | Most | Some | Hardly |
| Finland | Most/Some | Some | Hardly |
| **Understand** | | | |
| All countries except... | Most | Some | Hardly |
| Denmark | Most | Some | Hardly/Some |
| Sweden | Most | Some | Some/Hardly |
| **Become confused** | | | |
| Ireland, Netherlands, UK | Hardly | Some | Most |
| Denmark, Sweden | Hardly | Some | Most/Some |
| Germany | Hardly | Some | Some/Most |
| Finland | Hardly | Some | Some/Hardly |
| Luxembourg | Hardly | Some | Hardly/Some |
| Austria | Hardly/Some | Some | Some/Most/Hardly |
| Greece, Portugal | Some/Hardly | Some | Most |
| Italy | Some | Some | Most |
| Spain | Some | Some | Most/Some/Hardly |
| Belgium | Some | Some | Hardly/Most |
| France | Some/Most | Some | Hardly/Most |

Key

| | |
|---|---|
| Most | Most of the time |
| Some | Some of the time |
| Hardly | Hardly any of the time |

From this analysis, it is surprising that a joint three-class model, with measurement model constrained to be equal across countries, fits so poorly (Table 7.10). It seems that there is no single culprit item responsible for this, more a matter of the differences between countries in the relative magnitudes of the conditional probabilities for the three classes. The model does fit notably better in some countries than others, which

[25] For every country, no standardised marginal residuals > 4.

might suggest that some clusters of countries share more similar measurement models in this regard. However, an informal inspection does not reveal any clear groupings.

As an alternative way of investigating this idea, I ran some class models country-by-country, for just those respondents who do *not* give one of the three common sets of answers, to try to identify any patterns in these uncommon response profiles. These comprise approximately half of the sample in each country (ranging from 48 per cent in Ireland to 64 per cent in Greece and Finland). The analyses do not help, however, in suggesting groups of countries with similar sets of response profiles. With three-class models, amongst the fifteen countries seven different types of class emerge, and with four-class models, ten different types. Amongst these classes, some countries seem to share a similar set of classes, but these tend to be only pairs or trios of countries. Moreover these apparent groupings are quite unstable, and alter in composition when the models are changed from three-class to four-class. These analyses suggest that it is not a systematic divergence in patterns in the data that accounts for the lack of fit of a three-class joint model. So there is no motivation to attempt to divide the data set into smaller sets of countries within which to fit models. An alternative, and for comparison purposes better strategy is to continue with the full fifteen country data set, and simply increase the number of classes.

A seven-class solution is selected as a final model for these items, on the basis of fit statistics and interpretability. A six-class model, though apparently well fitting according to Table 7.10, does not return a very clear interpretation, and in fact gives cause for some concern in terms of numbers of large two-way marginal residuals. Although *overall* only 1.9 per cent of two-way residuals for this model are large, conditional on country, rates are still very high in some instances; ranging from 3.7 in Austria to 29.6 per cent in Finland, with an average of 11.3 per cent among the fifteen countries. In a seven-class model, by contrast, they range from 0 in France, Portugal and Sweden, to 14.8 per cent in Belgium and Italy, but with an average across countries of just 5.9 per cent (see Tables A.6 and A.7 in the Appendix for full details). The patterns of conditional response probabilities are quite clear from this model, whereas an eighth class only serves to duplicate one of the classes from it. Seven classes are therefore retained.

**Table 7.10 Fit statistics for models of science items, with measurement models constrained to be equal across 15 countries**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| 3 classes | 2,240 | 756 | <0.001 | 728 | -5,080 | 22.2 | 2.20 |
| 4 classes | 1,844 | 735 | <0.001 | 374 | -5,273 | 16.0 | 1.50 |
| 5 classes | 1,464 | 714 | <0.001 | 36 | -5,450 | 8.0 | 1.00 |
| 6 classes | 1,210 | 693 | <0.001 | -176 | -5,500 | 1.9 | 0.56 |
| 7 classes | 1,041 | 672 | <0.001 | -303 | -5,466 | 0.0 | 0.28 |

A seven-class model returns an intuitively appealing set of classes. In between the primary classes of high, mid and low engagement (the first, middle and last columns in the table, labelled accordingly) there are two sets of two extra classes. So amongst those who say they are interested in science and technology 'most of the time', we can identify those who say that they however feel informed only 'some of the time' (High–), and those who further say that they understand science stories in the news only 'some of the time' (Mid +). From the opposite end of the table, amongst those who say they feel informed about science and technology 'hardly any of the time, there are those who say they nevertheless understand science stories in the news 'some of the time' (Low +) and those who say they are interested only 'some of the time' (Mid –). So the classes can be thought of as grouped into three sets (Mostly, Sometimes and Hardly) on the basis of the most usual response. For example, those classes under the heading 'Sometimes' all imply a 'some of the time' response for two out of three items (albeit that this is only marginally true for the group Mid –, and a different researcher might wish to classify it as a 'Hardly' class).

**Table 7.11 Conditional probabilities, final 7-class model for science engagement**

| | | Mostly | | Sometimes | | | Hardly | |
|---|---|---|---|---|---|---|---|---|
| | | High | High – | Mid + | Mid | Mid – | Low + | Low |
| Item | Response category | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(5)$ | $\hat{\pi}_{is}(6)$ | $\hat{\pi}_{is}(7)$ |
| I am interested in science and technology | Hardly any of the time | 0.01 | 0.02 | 0.00 | 0.10 | 0.00 | 0.58 | 0.91 |
| | Some of the time | 0.09 | 0.44 | 0.13 | 0.85 | 0.80 | 0.42 | 0.09 |
| | Most of the time | 0.90 | 0.54 | 0.87 | 0.06 | 0.19 | 0.00 | 0.00 |
| I feel well informed about science and technology | Hardly any of the time | 0.02 | 0.05 | 0.06 | 0.00 | 0.71 | 0.98 | 0.95 |
| | Some of the time | 0.07 | 0.95 | 0.81 | 0.96 | 0.29 | 0.01 | 0.04 |
| | Most of the time | 0.90 | 0.00 | 0.14 | 0.04 | 0.00 | 0.01 | 0.01 |
| I understand science stories in the news | Hardly any of the time | 0.01 | 0.02 | 0.03 | 0.06 | 0.47 | 0.02 | 0.90 |
| | Some of the time | 0.15 | 0.38 | 0.66 | 0.80 | 0.49 | 0.84 | 0.08 |
| | Most of the time | 0.84 | 0.60 | 0.31 | 0.14 | 0.04 | 0.14 | 0.02 |

Table 7.12 reports the percentages of respondents in each of the classes, by country and overall (recalculated from the final models using the sampling weights, as described in Section 4.1.12). The last three columns combine proportions into the three aggregated groups Mostly, Sometimes and Hardly, and countries are ordered according to the total proportions in the first of these, i.e. the two high engagement classes. This shows that rates of engagement with science tend to be higher in Scandinavian countries, and lower in southern European countries. The distribution amongst the three levels is interesting. Whereas with the logics items we found between countries roughly the same patterns of relative proportions in different classes (for example, consistently more opponents than supporters for GM food), with science engagement there is much more heterogeneity across countries. Whereas across the fifteen countries overall approximately a third of the population are located at each level of engagement, these proportions vary markedly from country to country. For example, more than half of Swedes are predicted to be highly engaged, and only a quarter low engaged, whereas only 9 per cent of Portuguese are predicted to fall into the high engagement class, and the rest divided evenly between mid and low engagement. Looking a little more closely at these proportions, the detailed seven-column part of the table possibly suggests one of the reasons for the difficulty in achieving a well-fitting joint model with a smaller number of classes. For some countries, no people are predicted to belong to certain classes – for example, no one in Luxembourg or Portugal is predicted to fall into the class High –, and likewise no one in the Netherlands and Germany is predicted to belong to the class Mid +.

**Table 7.12 Weighted percentages of respondents in science engagement classes**

| *% within* | Mostly | | Sometimes | | | Hardly | | TOTAL: | TOTAL: | TOTAL: |
|---|---|---|---|---|---|---|---|---|---|---|
| *country* | High | High – | Mid + | Mid | Mid – | Low + | Low | MOSTLY | S'TIMES | HARDLY |
| Sweden | 25 | 33 | 3 | 8 | 5 | 16 | 10 | 58 | 16 | 26 |
| Netherlands | 19 | 30 | 0 | 7 | 8 | 16 | 20 | 49 | 15 | 36 |
| Italy | 17 | 24 | 6 | 25 | 5 | 9 | 14 | 41 | 36 | 23 |
| Denmark | 29 | 11 | 2 | 29 | 0 | 14 | 15 | 41 | 31 | 29 |
| Germany | 22 | 17 | 0 | 26 | 2 | 16 | 17 | 39 | 28 | 33 |
| Austria | 21 | 7 | 5 | 26 | 5 | 13 | 23 | 28 | 36 | 36 |
| Luxembourg | 25 | 0 | 21 | 20 | 7 | 8 | 19 | 25 | 48 | 26 |
| UK | 18 | 7 | 3 | 23 | 6 | 13 | 30 | 25 | 32 | 43 |
| Finland | 11 | 10 | 14 | 13 | 19 | 15 | 18 | 20 | 46 | 34 |
| France | 15 | 2 | 17 | 23 | 17 | 5 | 21 | 17 | 57 | 26 |
| Belgium | 12 | 5 | 9 | 27 | 5 | 10 | 32 | 16 | 41 | 42 |
| Spain | 11 | 4 | 5 | 29 | 12 | 9 | 30 | 16 | 46 | 38 |
| Greece | 11 | 1 | 36 | 18 | 20 | 0 | 14 | 13 | 73 | 14 |
| Ireland | 9 | 1 | 7 | 25 | 6 | 14 | 39 | 10 | 38 | 52 |
| Portugal | 9 | 0 | 12 | 19 | 14 | 5 | 41 | 9 | 45 | 46 |
| Europe total (pop. weighted) | 17 | 12 | 7 | 24 | 8 | 11 | 22 | 29 | 38 | 33 |

### *7.5.2 Behavioural elements of engagement with biotechnology*

Recalling the potential difficulties in interpreting the different media items, and in anticipation of difficulties in cross-national model fit, this section uses the combined media item: having heard about biotechnology from any media source, versus not having heard about it from any source. So there are four nominal variables to analyse. Beginning with the four classes suggested by Section 7.4.2, Table 7.13 shows the results of a brief analysis which suggests that for these four class models it is acceptable to combine Great Britain and Northern Ireland into one UK sample, and likewise to combine the two separately sampled regions of Germany into one group. In this analysis there are just 37 two-way margins conditional on country, meaning that only two residuals are large for Northern Ireland and one each for West and East Germany.

**Table 7.13 Fit statistics from 4-class models, testing measurement models between Great Britain and Northern Ireland, and between East and West Germany**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | | | Jöreskog & Moustaki index | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | All | GB | NI | All | GB | NI |
| UK, measurement model equal | 85 | 40 | <0.001 | 5 | -202 | 0.0 | 0.0 | 5.4 | 0.24 | 0.34 | 0.95 |
| | | | | | | All | West | East | All | West | East |
| Germany, measurement model equal | 112 | 40 | <0.001 | 32 | -193 | 1.8 | 2.7 | 2.7 | 0.50 | 0.66 | 0.87 |

Proceeding with the exploratory analyses, four-class models fit well within each of the fifteen countries[26]. However, the composition of these classes varies somewhat. The qualitative summary of them given in Table 7.14 implies that fitting a cross-national model to these data will not be a straightforward matter. Although for each type of engagement there is a core of at least six countries which share broadly the same pattern of likely responses, there is a good deal of variation around these cores – moreover, the core group of countries changes in composition from class to class. Not every class group is found in every country, and in certain countries some types of classes are found twice. For example, there are two high engagement classes and no mid/mixed engagement class in Belgium, Finland and the UK. Likewise there are two low engagement classes and no mid/mixed engagement class in Ireland and Italy, and there are no DK classes in France or Spain. Although admittedly these claims rest on the

[26] The percentage of standardised marginal residuals > 4 is on average 1.1 across countries, with 0 in ten countries, up to 5.4 in Spain.

judgement of the researcher in grouping responses patterns qualitatively, even a few changes to the classification would not change the overall verdict of considerable heterogeneity in measurement models between countries. Looking across the rows of the table, and looking down the columns of the table, it is not easy to pinpoint any particular source of this heterogeneity – it does not seem to be the case that one particular item or some particular countries are notably different than the others. So the analysis does not indicate that it would necessarily be helpful to test any particular interaction of item and latent variable; neither does it suggest any clusters of countries. It does clearly suggest, however, that a joint four-class model with measurement models constrained to be equal across countries will fit poorly.

**Table 7.14 Qualitative summaries of conditional probabilities from unconstrained 4-class models, 15 countries**

| *Classes and countries* | *Items and responses* | | | |
|---|---|---|---|---|
| | Have talked about biotech (ever) | Have heard of biotech in media (in last three months) | Would take part in a discussion or hearing | Would read an article / watch a programme |
| **High engagement** | | | | |
| Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Luxembourg, Netherlands, UK | Yes | Yes | Yes | Yes |
| Belgium, Italy, Portugal, Spain, Sweden, UK | Yes | Yes | No | Yes |
| Finland | No | Yes | Yes | Yes |
| **Low engagement** | | | | |
| Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, Portugal, Spain | No | No | No | No |
| Denmark, Netherlands, UK | No | No | No | Yes/No |
| France, Ireland, Sweden | No | No | No | Yes |
| Luxembourg | No | No | Yes/No | Yes |
| Luxembourg | No/Yes | Yes | No | No |
| **Mid/mixed engagement** | | | | |
| Austria, Denmark, Germany, Greece, Portugal, Sweden | No | No | Yes | Yes |
| Italy | No | Yes | No | Yes |
| France | Yes | Yes | No | No |
| Netherlands | Yes/No | Yes | No | Yes/No |
| **DK** | | | | |
| Austria, Belgium, Finland, Ireland, Italy, Portugal, Spain, UK | No | No | DK | DK |
| Germany, Netherlands | No | Yes | DK | DK |
| Spain | Yes | Yes | DK | DK |
| Denmark | Yes | No/Yes | DK/No | DK |
| Greece | No | No | DK | Yes |
| Luxembourg | No | Yes/No | DK | Yes |
| Sweden | Yes | Yes/No | No | Yes/DK |

Table 7.15 demonstrates that this is indeed the case. In the absence of any evidence from the qualitative analysis to suggest relaxing particular item parameters, increasing the number of classes is taken as a first step towards improving model fit. A six-class model returns a clearly interpretable measurement model, and as such is to be preferred over a seven-class model which, though better fitting statistically, contains two classes which are hard to define. In the six-class model, no item-by-item two-way marginal residuals are large, but conditional on country the average percentage of large residuals is 19.6, ranging from 0 in the Netherlands to 40.5 in Sweden (more information is given in the Appendix, Tables A.8 and A.9). From this point, since model fit is still quite poor, but increasing the number of classes does not seem to be fruitful, it is worth revisiting the idea of testing for any notable improvements in model fit gained by freeing item parameters. In the absence of a steer from the qualitative analyses, each item is tested in turn. The greatest gain is achieved by allowing an interaction between the latent variable and one of the hypothetical behaviour questions, and all fit statistics suggest that freeing *readtv* brings a slight improvement in model fit over freeing *discuss*. In this final model, two-way marginal residuals are low overall, and conditional on country they range from 0 in Finland, Ireland, the Netherlands and Portugal, to 21.6 per cent in Denmark, with an average across countries of 5.8 per cent.

**Table 7.15 Fit statistics for models of biotechnology items, with measurement models constrained to be equal across 15 countries**

| Model | $L^2$ | d.f. | p (b'strap) | AIC | BIC | % 2-way standardised marginal residuals >4 | Jöreskog & Moustaki index |
|---|---|---|---|---|---|---|---|
| **Measurement model equal between countries** | | | | | | | |
| 4 classes | 2,455 | 456 | <0.001 | 1,543 | -1,960 | 27.3 | 3.47 |
| 5 classes | 1,649 | 435 | <0.001 | 779 | -2,563 | 18.2 | 1.40 |
| 6 classes | 1,288 | 414 | <0.001 | 460 | -2,721 | 13.9 | 1.00 |
| 7 classes | 1,032 | 393 | <0.001 | 246 | -2,773 | 7.5 | 0.62 |
| **6 classes, investigating interactions** | | | | | | | |
| Interaction between *talkbio* and latent variable | 866 | 330 | <0.001 | 206 | -2,329 | 6.4 | 0.44 |
| Interaction between *heardbio* and latent variable | 899 | 330 | <0.001 | 239 | -2,296 | 6.4 | 0.48 |
| Interaction between *discuss* and latent variable | 635 | 246 | <0.001 | 143 | -1,747 | 4.8 | 0.57 |
| Interaction between *readtv* and latent variable | 584 | 246 | <0.001 | 92 | -1,798 | 4.3 | 0.41 |

Table 7.16 gives patterns of conditional probabilities for the three items where the measurement model is constrained to be equal between countries. This shows that the two binary items asking respondents whether they have heard and talked about biotechnology mirror each other very closely. Within the two levels of these items we can clearly identify groups in each of the three possible response categories for the third item, expressing willingness to take part in a public discussion on the topic; agreeing and disagreeing, and responding DK.

**Table 7.16 Final 6-class model for engagement with biotechnology, conditional probabilities for three items where measurement model is equal across countries**

| | | High report, high willingness | High report, low willingness | High report, DK willingness | Low report, high willingness | Low report, low willingness | Low report, DK willingness |
|---|---|---|---|---|---|---|---|
| Item | Response | $\hat{\pi}_{is}(1)$ | $\hat{\pi}_{is}(2)$ | $\hat{\pi}_{is}(3)$ | $\hat{\pi}_{is}(4)$ | $\hat{\pi}_{is}(5)$ | $\hat{\pi}_{is}(6)$ |
| Ever talked about biotech? | No | 0.18 | 0.40 | 0.28 | 0.71 | 0.98 | 0.93 |
| | Yes | 0.82 | 0.60 | 0.72 | 0.29 | 0.02 | 0.07 |
| Heard about biotech in last 3 months? | No | 0.03 | 0.23 | 0.12 | 0.89 | 0.94 | 0.92 |
| | Yes | 0.97 | 0.77 | 0.88 | 0.11 | 0.06 | 0.08 |
| Would take part in discussions or hearings. | DK | 0.01 | 0.00 | 0.94 | 0.06 | 0.03 | 0.96 |
| | Disagree | 0.26 | 0.96 | 0.06 | 0.33 | 0.93 | 0.04 |
| | Agree | 0.73 | 0.04 | 0.00 | 0.61 | 0.04 | 0.01 |

Table 7.17 presents a qualitative summary to show how responses to the other ‘willingness question’, that is willingness to read articles or watch television programmes on biotechnology, varies between countries. In the table, following the format used in Chapter 5, ‘+’ indicates the response ‘tend to agree’, ‘–’ denotes ‘tend to disagree’ and ‘?’ is used where DK is the most likely response. Countries are ordered approximately according to the numbers of classes in which positive responses are expected, from the greatest number of positive classes to the least. Following a few unusually positive countries at the top of the table we can see a set – Belgium, Denmark, Ireland, Italy, the Netherlands and the UK – which follow the same pattern (this in fact is the profile that emerges from the six-class model where the measurement model is fixed to be equal between countries). According to this pattern, those in the ‘low report’ classes tend to make the same judgement on reading articles as on taking part in public discussions, while those in the ‘high report’ classes respond positively to this item, regardless of whether they would be willing to take part in discussions. It seems then that agreeing to take part in discussions on biotechnology is a more demanding item, or represents a higher bar in terms of levels of engagement, than

reading articles and watching programmes on the topic. In the three countries at the top of the table, even those who have not heard or talked about the subject before and who would be unwilling to participate in discussion on it would still be willing to read about it, in principle. By contrast, in a few countries – those towards the bottom of the table – even in the high report classes, low willingness to discuss biotechnology goes hand in hand with low willingness to consume media on the subject.

**Table 7.17 Final 6-class model for engagement with biotechnology, qualitative summary of highest conditional probabilities for fifteen countries, for the item 'I would be prepared to read an article or watch a television programme about biotechnology'**

| | High report, high willingness | High report, low willingness | High report, DK willingness | Low report, high willingness | Low report, low willingness | Low report, DK willingness |
|---|---|---|---|---|---|---|
| Sweden | + | + | + | + | +/– | ? |
| Luxembourg | + | + | + | + | +/– | ?/– |
| France | + | +/– | + | + | –/+ | ?/+ |
| Belgium | + | + | + | + | – | ? |
| Denmark | + | + | + | + | – | ? |
| Ireland | + | + | + | + | – | ? |
| Italy | + | + | + | + | – | ? |
| Netherlands | + | + | + | + | – | ? |
| UK | + | + | + | + | – | ? |
| Greece | + | – | + | + | –/+ | +/? |
| Austria | + | +/– | + | + | – | ? |
| Finland | + | –/+ | + | + | – | ? |
| Germany | + | – | + | + | – | ? |
| Portugal | + | –/+ | ?/+ | + | – | ? |
| Spain | + | +/– | ? | + | – | ? |

Finally, Table 7.18 presents the proportions predicted to belong to each class (again, recalculated using sample weights, as outlined in Section 4.1.12). Countries are ordered according to proportions in the class representing the highest level of engagement. It is perhaps heartening for those working in the field of science communication that overall in the fifteen countries listed, more than half of the population is predicted to have heard and talked about biotechnology before, with nearly a third also willing in principle to take part in discussions on biotechnology and read articles or watch programmes about it. This enthusiasm varies by country, however. Whereas nearly half of the French belong to this keen group, only 13 per cent of Spaniards could be identified in the same way. In Spain, more than a third of the population is predicted to give a full negative set of responses, reporting not to have been exposed to biotechnology before and being unwilling to participate in learning or talking about it.

The ordering of countries approximately reflects that for engagement with science, though with a few exceptions – for example, Sweden appears somewhere in the middle of the list, on account of the fact that a high proportion of otherwise engaged respondents would prefer not to take part in public discussions on biotechnology (42 per cent belong to this high report, low willingness class, and just 23 per cent to the highest engagement class). The two DK classes are much more sparsely populated in this set of items than in the logics of Chapter 5. A few countries, however, are notable in the much higher proportions of respondents found in them – for example, Italy, the Netherlands and Germany in the 'high report' class, and Ireland and Portugal in the 'low report' class. It might be recalled that Ireland and the Netherlands had notably high rates of DK responses in the logics analyses too, begging the question of whether these high rates represent genuinely different levels of certainty in these countries, or fieldwork company styles.

**Table 7.18 Weighted percentages of respondents in biotechnology engagement classes**

| *% within country* | High report, high willingness | High report, low willingness | High report, DK willingness | Low report, high willingness | Low report, low willingness | Low report, DK willingness |
|---|---|---|---|---|---|---|
| France | 48 | 18 | 6 | 11 | 14 | 3 |
| Luxembourg | 42 | 22 | 4 | 20 | 10 | 2 |
| Germany | 41 | 16 | 12 | 19 | 8 | 3 |
| Finland | 40 | 26 | 5 | 15 | 11 | 3 |
| Denmark | 40 | 19 | 6 | 24 | 10 | 1 |
| Netherlands | 26 | 35 | 10 | 7 | 17 | 3 |
| Austria | 24 | 15 | 7 | 38 | 9 | 7 |
| UK | 23 | 20 | 5 | 13 | 32 | 7 |
| Sweden | 23 | 42 | 7 | 14 | 12 | 2 |
| Ireland | 23 | 15 | 6 | 18 | 26 | 12 |
| Belgium | 22 | 26 | 6 | 9 | 28 | 9 |
| Italy | 22 | 36 | 16 | 7 | 15 | 4 |
| Greece | 21 | 2 | 1 | 40 | 29 | 5 |
| Portugal | 16 | 14 | 4 | 31 | 24 | 12 |
| Spain | 13 | 30 | 8 | 5 | 36 | 7 |
| Europe total (pop. weighted) | 30 | 23 | 9 | 14 | 19 | 5 |

## Summary

The analyses in this chapter raise a number of findings and questions about the data. Once again, it seems prohibitively difficult to fit a joint model to all items and all countries. This should not be cause for such great concern as in the previous chapter, since the items available for the analyses are broad in content as well as varied in format – it would perhaps be a little over-optimistic to expect to find a joint model for them. However, retaining the distinction between affective and cognitive engagement with science in general, and behavioural engagement with biotechnology in particular, it *is* possible to find well fitting class models, though not well fitting trait models (within the constraints of the trait models we are able to fit with the available software).

Considering first of all the items asking respondents about their affective engagement with science and technology in general, across the fifteen European countries analysed, the following has been found:

- Country-by-country, three classes adequately summarise the various response profiles in answer to three questions: whether one is interested in science and technology, feels informed about it, and understands science stories in the news.
- A fourth item – whether one becomes confused on hearing conflicting views on science – behaves irregularly in all countries, statistically speaking. A number of interpretations could be attributed to this item, which itself is not substantively important (in the way that the risk item was important to the logics analyses) – so for the purposes of developing a summary indicator of engagement with the other items, it is expedient to drop the item.
- The three classes can be described as denoting low, mid and high levels of engagement, corresponding to responses 'hardly any of the time', 'some of the time', and 'most of the time'. In the majority of cases, a respondent will give the same answer to each of these three questions.
- Although within each country, three-class models fit well, for a joint cross-national model with measurement models constrained to be equal between countries, seven classes are required to obtain a meaningful, well fitting model. Recalling that I added just one class to the logics items to ensure a really well fitting model, the addition of *four* classes to the science items seems to indicate that there is more variation in the detail of the response probabilities with these items compared with the logics items.

- The seven-class cross-national solution identifies the three dominant classes (low, mid and high engagement), plus two which could be positioned in common sense terms between low and mid engagement, and two which could be positioned between mid and high engagement.
- In the fifteen countries as a whole, just under a third of the population might be described as highly engaged, just over a third as mid-level engaged, and a third as not very engaged. These proportions vary considerably from country to country, though with an element of expected patterns – for example, in many of the northern European countries we would predict seeing more highly engaged people than unengaged people; the opposite is the case in many southern European countries.

For the items assessing behavioural engagement with biotechnology, across the fifteen European countries analysed:

- A six-class model of engagement fits well across the fifteen countries, if parameters for the item *I would be prepared to read articles or watch television programmes about biotechnology* are allowed to vary between countries.
- In these six classes, the criteria of having talked about biotechnology before, and having heard or seen or read about it in some form of media in the three months preceding the survey, mirror each other closely. Three classes are given to those who say 'yes' to both, and three classes to those who say 'no' to both.
- Within the two levels of having heard of and talked about biotechnology, there are three levels of response to the question *I would be prepared to take part in public discussions or hearings* on the topic: positive, negative and DK responses. So the classes can be labelled sensibly in terms of high or low reported engagement (having talked or heard), and high, low or DK willingness to engage (to take part in discussions).
- The other question on hypothetical engagement, that is willingness to read an article or watch a programme about biotechnology, is connected to the model in a slightly more complex way. The dominant pattern among the fifteen countries is that across all the high report engagement classes, people are likely to agree to this item. Across the low report engagement classes, responses to *I would read or watch* mostly mirror responses to *I would discuss*. This trend is seen in six of the fifteen countries. In some other countries (for example, Sweden, France), there is a slightly higher chance of even those in the low report classes giving a positive response to this item. By contrast, in some countries (e.g. Portugal, Spain), responses in the

high report classes are more closely aligned with responses to the question of willingness to participate in public discussions.

- More than half of the population of the fifteen EU countries is predicted to have high reported levels of engagement with biotechnology, in terms of having heard of and spoken about it before. 30 per cent have high reported levels of engagement and high levels of willingness to engage (that is, to participate in public discussions and read articles or watch programmes on the topic). Only 19 per cent are predicted to have low levels of reported engagement and low levels of willingness.
- Proportions of highly and less engaged people vary, country-by-country. For example, in France, we would expect to see 48 per cent of the population in the highest engagement class and 14 per cent in the lowest class. In Spain we would find almost the reverse: 13 per cent in the highest engagement class and 36 per cent in the lowest class.

Some additional findings arose from the more detailed analyses of the British sample:

- For models of the science items, the observed variables do not behave ordinally. It seems that there is a response effect, which manifests itself as an extra dimension, associated with the response 'some of the time'.
- A single ordered latent variable is not adequate for representing the science items. The non-ordinality of the latent variable 'engagement with science' can be addressed adequately (statistically speaking) by either a nominal class model, or a two-trait model.
- Turning to engagement with biotechnology, a similar allowance needs to be made for the 'willingness' items, which contain a number of DK responses. We either need two traits to represent these engagement items (one of which is a DK trait for the willingness items), or a class model in which a class is given to DK responses.
- Different types of media exposure to biotechnology are associated with different degrees of engagement. For example, we would expect only those at the top end of an engagement trait to have read about biotechnology on the internet or heard about it on the radio. Informally inspecting the ICCs of these items, however, indicates that these differences are fairly small.
- For engagement with science and/or biotechnology – that is, whether modelled separately or together – it is possible to find equally well fitting class and trait models. While differing slightly in the emphasis they give to distinctions between

certain response patterns, the posterior class allocations and trait scores from broadly agree.

- Class and trait models that combine the science and biotechnology engagement items both suggest that affective engagement with science, and behavioural engagement with biotechnology, can be thought of as distinct though closely related constructs.

In answer to the question, 'a matter of degree or of kind?', we might still wish to say 'both'. However, *kinds* seem to win over *degrees* in these analyses. Although it makes intuitive sense to speak of levels of engagement, a number of different types emerge in the data, making it a difficult task to model degrees, especially cross-nationally. We have, for example, a number of kinds of response formats, a number of possible response effects, and we are attempting to find models for items with a number of types of content: affective and cognitive, reported behaviours and hypothetical willingness, sometimes in relation to science and technology in general, and sometimes to biotechnology in particular. It is not surprising then that achievable joint models are those that define kinds rather than degrees. Although the classes from these models can be ordered in common sense terms, they cannot be ordered statistically, via a one-trait latent trait model, without a considerable cost to model fit.

To offer a contribution towards the last research question, that is making recommendations for future survey design, two simple points could be made. The first is simply to stress that the item *I become confused when I hear conflicting views on science and technology* does not seem to be an efficient use of a survey item, and should be a candidate for deletion from future survey waves. The second suggestion comes from the point made above that it is difficult to derive a measure of engagement when there is so much 'going on' in the data. A useful initial way to take these items forward into the next wave might therefore be to write a battery of ten or more questions, with the same or similar question and response formats. In PUS the distinction between generalised and specific attitudes and knowledge is a matter of ongoing interest (see e.g. Allum et al., forthcoming), but with the 2002 data set it is impossible to say whether there is a genuine separation between engagement with science and engagement with biotechnology, because the difference in item content is accompanied by a difference in item format.

These three chapters have demonstrated that when focusing on a single country sample it is easy to find relatively well fitting and clearly interpretable models of the constructs of logics of support and opposition for GM food and therapeutic cloning, knowledge of biology and genetics, and engagement with science and with biotechnology. These models not only provide useful information about the content of the constructs, but also provide valuable information about item functioning, which can feed into future survey design. When moving to cross-national models, however, it is not at all an easy task to find models that fit well when the measurement models are constrained to be equal between countries. I reached acceptable models for logics of support and opposition, and for engagement, but was not so successful in modelling knowledge cross-nationally. Before discussing the findings and methodological implications of these studies, the next brief chapter draws them together. Taking the scores and class allocations derived from the final models in these chapters, it presents in four simple loglinear models to assess the relationships between them.

# 8 Associations between attitudes, knowledge and engagement

In this chapter I draw together the final measures of attitudes, knowledge and engagement produced in Chapters 5, 6, and 7, in a brief analysis of the associations between them. Specifically, I take the factor scores and class allocations from these models, and treat them as observed variables in regression modelling. The first part of the chapter is given to the results from the analyses of the British data, and the second part to the Europe-wide measures. Recalling that the attitude items for GM food and therapeutic cloning were part of the split ballot design of the survey and each posed to half of the sample, this makes for four analyses. Each contains the measures for attitudes, knowledge and engagement, plus three socio-demographic variables of central interest in PUS: gender, age and level of education.

The focus of the empirical analyses in this thesis has been on the measurement models for the three PUS constructs, rather than on the structural models of relationships between the constructs. The purpose of this chapter is to give just a brief and simple illustration of possible joint analyses of these measures and their associations with socio-demographic variables. A more comprehensive analysis, and more comprehensive treatment of this task, is a topic for future studies. A short, general discussion of the ways in which cross-national analyses might be carried forward will follow in the next chapter (see Section 9.6).

For the present, it is important to note the relative advantages and disadvantages of analysing the associations between posterior scores in order to investigate the relationships between latent variables. This method is known as the 'three-step approach' (Bolck, Croon, & Hagenaars, 2004), with the steps comprising: firstly, estimating separate measurement models for the latent constructs; secondly, deriving factor scores or class allocations from these measurement models; and thirdly, treating these scores as observed variables in regression or other analyses, with other explanatory variables, as appropriate. This is sometimes termed the 'two-step approach' in fact, combining the first two steps. The alternative, 'one-step' approach involves structural equation models (e.g. Bollen, 1989; Jöreskog, 1973), that is, estimating both measurement models for the constructs, and the relationships between the constructs (plus other explanatory variables), simultaneously. Structural equation models (as described in Chapter 4, Section 4.1.2) involve linear models; the type of structural equation model required for this chapter would involve categorical latent variables, and logistic models for the measurement parts of the model.

The three-step approach is not without shortcomings (cf. Croon, 2002; Skrondal & Laake, 2001). There are many methods for computing scores, and thus as many different resulting representations of latent variables. The particular method used may therefore produce biased scores, effectively measuring the values of the latent variables with error. Analyses of associations between such measures may then themselves be biased. Typically, for example, such analyses may suggest weaker associations between the constructs than actually exist (Bolck et al., 2004). For a discussion of some of the issues associated with the empirical Bayes scores used in this thesis (cf. Chapter 4, Section 4.1.11), the reader is referred to Skrondal and Rabe-Hesketh (2004).

Bearing in mind these caveats, the three-step method is arguably more suitable than the one-step approach for the analyses in this chapter. Although the one-step approach has the advantage of producing unbiased estimates of the structural model, this is true only when the model is correctly specified. This method therefore requires a good deal of faith in the truth of the model; some would say, faith of heroic proportions (Bartholomew & Knott, 1999). For exploratory analyses of the kinds presented here, a three-step approach is more suitable simply because it is more cautious (Bolck et al., 2004). Practically speaking, indeed, although software for linear structural equation models is accessible, it is not currently available for structural models with logistic measurement models.

For these analyses, then, I use the scores from the relevant models in Chapters 5–7: one continuous and three or four categorical measures (for the British and the cross-national data sets, respectively). The joint distribution of these measures can be specified as conditional Gaussian (Edwards, 1995); that is, as the joint multinomial distribution of the categorical variables, combined with a normal distribution for the continuous variable, given the categorical variables. Specifying a model for these constructs, given the three socio-demographic explanatory variables (all of which are categorical), involves estimating the model in two parts, using standard regression models. A loglinear model is used for the relationships among the categorical variables, with the margin between the three socio-demographic variables saturated. A regular linear regression is used for the continuous variable (knowledge), given the categorical PUS measures and the socio-demographic variables.

In the tables below I show, for the loglinear parts of the models, odds ratios for pairs of variables, between pairs of adjacent categories on these variables. Recalling that the latent class models involve nominal, not ordinal latent variables, it should be stressed that for the purposes of calculating the odds ratios I have arranged the categories into pairs purely on the basis of a common sense interpretation of the classes, not on any formal ordering between them. For the linear part, the usual coefficients from the linear model are presented, that is the effects of the explanatory variables on the predicted knowledge score.

## 8.1 British data

The analyses of the British data, undertaken in the first part of each empirical chapter, have resulted in the creation of one measure for each of the constructs of interest, using the methods described in Chapter 4 (Section 4.1.11). From Chapter 5, there are latent class models to summarise types of attitudes towards GM food and therapeutic cloning. For each application, five classes can be defined as: definite opposition, opposition, support, definite support and DK. From Chapter 6, one of the two traits from a discrete trait model can be used to represent knowledge of biology and genetics. Finally, from Chapter 7 there is a six-class model to represent affective engagement with science and behavioural engagement with biotechnology. The six classes for this model range from 'low science, low biotechnology' to 'high science, high biotechnology'.

The models presented below were arrived at through a simple process of model selection. For the loglinear models, I began from all two-way interactions, plus the necessary three-way interaction between the three socio-demographic variables (age, education and gender), using likelihood ratio comparison tests for backward elimination. In each case, adding the three-way interaction implied by the remaining two-way terms did not result in any significant improvement in the model fit. For the linear models, standard tests were used to assess the significance of main effects only.

### *8.1.1 GM food*

For GM food, the final model, in extended Goodman notation (Agresti & Finlay, 1997)[27], is:

[Age*Ed*Gen][Ed***Eng**][Gen***Eng**][Ed***Know**][Age***Know**][**Eng*Know**][**Eng*Att**].

That is, the model contains the following two-way associations involving the PUS variables:

- level of education and engagement
- gender and engagement
- education level and knowledge
- age and knowledge
- engagement and knowledge
- engagement and attitude.

It thus excludes a number of two-way associations, that is:

- engagement and age
- knowledge and gender
- attitude and knowledge
- attitude and engagement
- attitude and all socio-demographic variables in the model.

Tables 8.1 and 8.2 summarise the implications of this model. Table 8.1 gives odds ratios for adjacent categories for the three categorical response variables obtained from the loglinear part of the model. Moving from left to right in the table, the ratios compare being in a 'higher' engagement class with being in a 'lower' one. So it can be seen, for example, that men tend to be more engaged with science and biotechnology than women: odds ratios in most cells for gender are greater than 1.

The relationship between education level and engagement is little more complex, with a suggestion that it is behavioural engagement with biotechnology rather than affective engagement with science that is more strongly affected by education. For example, the odds on a person with high (versus low) education belonging in the Mid science/High

[27] In addition to Goodman notation for the loglinear part of the model, this shows explanatory variables included in the linear model for knowledge as two-way interactions involving knowledge.

biotechnology group rather than the Mid science/Low biotechnology group are 2.84. By contrast, the odds on a person with high (versus low) education belonging in the Mid science/Low biotechnology group rather than the Low+ science/High biotechnology group are 0.68. Similarly, the odds on a current student versus someone with high education belonging in the Mid science/High biotechnology group rather than the Mid science/Low biotechnology group are 0.28, and of belonging in the Mid science/Low biotechnology group rather than the Low+ science/High biotechnology group, 2.70. This general patterns of odds ratios alternating above and below 1 suggests that education seems to be positively associated with behavioural engagement with biotechnology. Unfortunately it is not possible to say whether it is the element of biotechnology rather than science, or of behaviour rather than affect, which is responsible for this pattern. The increased awareness of biotechnology found in those with higher levels of education has been found in other studies (e.g. Jon. D. Miller & Kimmel, 2001); this is often theorised to be part and parcel of a more general awareness of public issues (e.g. Gaskell et al., 2006).

Patterns of association between engagement and attitudes towards GM food are not as clearly defined. A few prominent ratios from this part of the table are worth noting, however. Firstly, the odds ratios of belonging in the DK attitude class versus the Definite opposition class alternate above and below 1 with high and low biotechnology classes, respectively. So higher levels of behavioural engagement with biotechnology are more likely to go together with a strongly opposing than an ambivalent position towards GM food. Secondly, it is not straightforwardly the case that higher levels of engagement are associated with more positive attitudes towards GM food. A trend towards definite rather than moderate support is seen in the upper classes of engagement, but elsewhere in this section of the table, the pattern is mixed.

**Table 8.1 Odds ratios involving engagement with science and biotechnology with significantly associated variables, GB, GM food subsample**

| | | Engagement with science and biotechnology | | | | | |
|---|---|---|---|---|---|---|---|
| Item | Category | L–s/DKb vs. Ls/Lb | L+s/Hb vs. L–s/DKb | Ms/Lb vs. L+s/Hb | Ms/Hb vs. Ms/Lb | H–s/Lb vs. Ms/Hb | Hs/Hb vs. H–s/Lb |
| Gender | Male vs. female | 1.04 | 1.33 | 1.18 | 0.86 | 1.93 | 1.33 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 0.00 | * | 0.68 | 2.84 | 1.13 | 4.00 |
| | Still studying (young) vs. 20+ years old | * | * | 2.70 | 0.28 | 2.00 | 0.25 |
| Attitude towards GM food | Def. opposition vs. DK | 0.43 | 6.68 | 0.29 | 1.78 | 0.78 | 1.50 |
| | Opposition vs. Def. Oppos'n | 0.57 | 1.82 | 1.45 | 1.08 | 1.78 | 0.92 |
| | Support vs. Opposition | 1.49 | 0.47 | 1.43 | 0.61 | 1.29 | 0.84 |
| | Def. support vs. Support | 0.50 | 1.55 | 0.67 | 1.28 | 1.33 | 3.21 |

* Expected frequency of 0 for L–s/DKb for high education group.

Key

| | |
|---|---|
| Ls/Lb | Low science, low biotechnology |
| L–s/DKb | Low– science, DK biotechnology |
| L+s/Hb | Low+ science, high biotechnology |
| Ms/Lb | Mid science, low biotechnology |
| Ms/Hb | Mid science, high biotechnology |
| H–s/Lb | High– science, low biotechnology |
| Hs/Hb | High science, high biotechnology |

Table 8.2 presents the parameter estimates from the linear part of the model, relating to the associations between knowledge and the other variables. Just those variables that are significantly related to knowledge have been retained in the model. The table shows that higher levels of knowledge are associated with belonging in higher engagement groups. Education level is also associated with level of knowledge: specifically, those who studied until they were 20 years of age, or older, are predicted to have significantly higher levels of knowledge about biology and genetics than those with lower levels of education – including those who are still at school or college. Higher levels of knowledge are also associated with younger respondents: those in the 55+ age group have significantly lower levels of knowledge than the youngest respondents.

**Table 8.2 Parameter estimates from linear regression of knowledge on significantly associated variables, GB, GM food subsample**

| Item | Category | Knowledge of biology and genetics Estimate | Std. error | p |
|---|---|---|---|---|
| Intercept | | 0.278 | 0.036 | <0.001 |
| Engagement with science & biotechnology | Low science, low biotech (ref.) | | | |
| | Low- science, DK biotech | -0.078 | 0.071 | 0.271 |
| | Low+ science, high biotech | -0.002 | 0.045 | 0.958 |
| | Mid science, low biotech | 0.033 | 0.042 | 0.439 |
| | Mid science, high biotech | 0.151 | 0.044 | 0.001 |
| | High- science, low biotech | 0.175 | 0.050 | 0.001 |
| | High science, high biotech | 0.346 | 0.054 | <0.001 |
| Education (age at which completed) | Up to 19 years old (ref.) | | | |
| | 20+ years old | 0.091 | 0.043 | 0.032 |
| | Still studying (15-19 years old) | 0.018 | 0.063 | 0.779 |
| Age | 15-34 (ref.) | | | |
| | 35-54 | -0.064 | 0.035 | 0.065 |
| | 55+ | -0.074 | 0.035 | 0.036 |

### *8.1.2 Therapeutic cloning*

For therapeutic cloning, the final model is:

[Age*Ed*Gen][Ed***Eng**][Age***Eng**][Age***Know**][**Eng*Att**][**Eng*Know**][**Know*Att**].

That is, it contains two-way associations between:

- level of education and engagement
- age and engagement
- age and knowledge
- engagement and attitude
- engagement and knowledge
- attitude and knowledge.

It therefore excludes the following two-way associations:

- engagement and gender
- knowledge and gender
- knowledge and level of education
- attitude and gender
- attitude and education
- attitude and age.

Noticeably, the associations included in the models for therapeutic cloning and GM food are not quite the same. Both models include associations between age and knowledge, education and engagement, and engagement and attitude. Both exclude associations between knowledge and gender, and between attitude and age. However, there is some variation around this common core – and not only in terms of associations involving attitudes towards applications of biotechnology. For example, in the GM food subsample the association between engagement and gender is significant, but it is excluded from the model for the therapeutic cloning subsample. These differences could be due to random variation between the two subsamples, but might alternatively be explained by substantive considerations, given the presence of essentially different attitude variables in the two models. It is beyond the remit of this chapter to explore these differences further, but they should be noted for future comparisons of attitudes towards different applications of biotechnology.

Table 8.3 shows odds ratios for variables involved in significant two-way associations in the loglinear part of the model for therapeutic cloning. Interestingly, a similar pattern of odds ratios, alternating above and below 1, is found for the middle versus younger age group; the implication is that those in the 35-54 years of age group are more likely to have high levels of behavioural engagement with biotechnology. This pattern is also found with level of education – similarly to GM food, but more clearly in this case. Finally, again, associations between engagement and attitude are more complex. A similar pattern of odds ratios is found for strong opposition versus DK classes, but otherwise, the patterns are somewhat more mixed.

**Table 8.3 Odds ratios involving engagement with science and biotechnology with significantly associated variables, GB, therapeutic cloning subsample**

| Item | Category | L–s/DKb vs. Ls/Lb | L+s/Hb vs. L–s/DKb | Ms/Lb vs. L+s/Hb | Ms/Hb vs. Ms/Lb | H–s/Lb vs. Ms/Hb | Hs/Hb vs. H–s/Lb |
|---|---|---|---|---|---|---|---|
| | | Engagement with science and biotechnology | | | | | |
| Age | 35-54 vs. 15-34 | 0.46 | 4.06 | 0.64 | 1.44 | 0.91 | 1.10 |
| | 55+ vs. 35-54 | 2.55 | 0.38 | 0.85 | 1.36 | 1.01 | 0.53 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 0.52 | 5.00 | 0.52 | 4.31 | 0.24 | 11.00 |
| | Still studying (young) vs. 20+ years old | 6.00 | 0.25 | 1.00 | 0.67 | 4.50 | 0.00 |
| Attitude towards therapeutic cloning | Def. opposition vs. DK | 0.47 | 2.78 | 0.63 | 1.44 | 0.73 | 7.33 |
| | Opposition vs. Def. Oppos'n | 1.67 | 1.18 | 0.99 | 0.88 | 1.27 | 0.54 |
| | Support vs. Opposition | 0.56 | 0.91 | 1.96 | 1.49 | 0.31 | 1.56 |
| | Def. support vs. Support | 0.50 | 3.67 | 0.80 | 0.71 | 4.44 | 0.60 |

Key

| | |
|---|---|
| Ls/Lb | Low science, low biotechnology |
| L–s/DKb | Low– science, DK biotechnology |
| L+s/Hb | Low+ science, high biotechnology |
| Ms/Lb | Mid science, low biotechnology |
| Ms/Hb | Mid science, high biotechnology |
| H–s/Lb | High– science, low biotechnology |
| Hs/Hb | High science, high biotechnology |

Table 8.4 summarises the relationships between knowledge and significantly associated variables. For therapeutic cloning, in contrast to GM food, attitude is significantly associated with knowledge, although rather weakly. Those in the two extreme classes, 'definite opposition' and 'definite support', have significantly different levels of knowledge, but there is very little difference to speak of with regard to the other classes. It is noteworthy, however, that the direction of the relationship is in line with conventional expectations: that is, the strong supporters are predicted to have higher levels of knowledge than the strong opponents of therapeutic cloning. Engagement is broadly speaking positively associated with knowledge, with those in the higher classes predicted to have significantly higher levels of knowledge that those in the lowest classes. And lastly, age is roughly negatively associated with knowledge, as before, with those in the 55+ age group significantly likely to achieve a lower score on the knowledge scale than those in the 15-34 age group.

**Table 8.4 Parameter estimates from linear regression of knowledge on significantly associated variables, GB, therapeutic cloning subsample**

| | | Knowledge of biology and genetics | | |
|---|---|---|---|---|
| Item | Category | Estimate | Std. error | p |
| | Intercept | 0.254 | 0.051 | <0.001 |
| Attitude towards therapeutic cloning | Definite opposition (ref.) | | | |
| | DK | 0.085 | 0.056 | 0.128 |
| | Opposition | 0.042 | 0.054 | 0.444 |
| | Support | 0.031 | 0.048 | 0.521 |
| | Definite support | 0.146 | 0.051 | 0.005 |
| Engagement with science & biotechnology | Low science, low biotech (ref.) | | | |
| | Low- science, DK biotech | -0.140 | 0.057 | 0.015 |
| | Low+ science, high biotech | 0.046 | 0.045 | 0.316 |
| | Mid science, low biotech | -0.011 | 0.043 | 0.794 |
| | Mid science, high biotech | 0.197 | 0.044 | <0.001 |
| | High- science, low biotech | 0.081 | 0.049 | 0.100 |
| | High science, high biotech | 0.316 | 0.050 | <0.001 |
| Age | 15-34 (ref.) | | | |
| | 35-54 | -0.049 | 0.033 | 0.136 |
| | 55+ | -0.150 | 0.032 | <0.001 |

## 8.2 European data

The joint cross-national models from Chapters 5, 6 and 7 give us two measures of engagement from latent class analyses (a seven-class model for biotechnology and a six-class model for science), along with factor scores for knowledge of biology and genetics, and class allocations from six-class models for attitudes towards GM food and therapeutic cloning.

The classes for GM food and therapeutic cloning are as before, but with the addition of a second class for DK responses, named 'DK (2)'. As before, the classes defining engagement are slightly more complex. Affective engagement with science is defined by seven classes which can be ordered from 'low' to 'high'. Behavioural engagement with biotechnology is defined in six classes, ranging from 'low report, low willingness' to 'high report, high willingness'. Once again a key is provided with the relevant tables to explain the shortened labels used.

The models were approached in the same way as those for the British data. The loglinear parts of the model began with backward elimination from all two-way interactions (plus the necessary three-way interaction between the socio-demographic variables). For both GM food and therapeutic cloning, just one two-way interaction was found not to be significant: the association between gender and engagement with biotechnology. Since the purpose of this chapter is to give an overall impression of the patterns in the data rather than a detailed account of all the associations between the variables, the loglinear models presented below are these simple versions: with all but one two-way interaction included. Further analysis suggested that a number of three-way interactions were also significant[28], but to report them here would raise the level of complexity of this chapter beyond its intended spirit.

### *8.2.1 GM food*

Table 8.5 shows, firstly, that in the fifteen European countries as a whole, higher levels of engagement with science are mostly found in men rather than in women. The relationship between age and engagement is less clear, with even odds in three of the twelve comparisons in the table. Those in the middle age group are on the whole more likely to be found in the middle engagement classes ('mid' and 'mid+') compared with the young and the old. And engagement appears to be broadly positively associated with level of education.

**Table 8.5 Odds ratios involving engagement with science with significantly associated variables, 15 countries, GM food subsample**

| | | Engagement with science | | | | | |
|---|---|---|---|---|---|---|---|
| Item | Category | Low+ vs. Low | Mid– vs. Low+ | Mid vs. Mid– | Mid+ vs. Mid | High– vs. Mid+ | High vs. High– |
| Gender | Male vs. female | 0.81 | 1.00 | 2.00 | 2.00 | 0.50 | 2.00 |
| Age | 35-54 vs. 15-34 | 1.00 | 0.75 | 1.22 | 1.64 | 0.50 | 1.00 |
| | 55+ vs. 35-54 | 0.75 | 1.33 | 0.59 | 0.85 | 1.00 | 1.00 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 1.17 | 1.50 | 1.33 | 3.00 | 0.31 | 1.35 |
| | Still studying (young) vs. 20+ years old | 1.50 | 0.33 | 1.33 | 0.75 | 1.11 | 2.19 |

In Table 8.6 we can see that those in the middle, 35-54 years age group, generally speaking belong to classes of higher engagement with biotechnology than younger and

[28] Specifically, where Bio = biotechnology engagement, Sci = science engagement and Att = attitude (towards GM food or therapeutic cloning, as appropriate):

- For GM food, the following model was reached from adding significant three-way way interactions (four-way interactions implied were not tested): [Age*Ed*Gen] [Bio*Att*Ed] [Att*Ed*Sci] [Att*Gen*Sci] [Ed*Gen*Sci] [Age*Att*Sci] [Bio*Ed*Sci] [Bio*Age].
- For therapeutic cloning, the same approach results in the following model: [Age*Ed*Gen] [Ed*Gen*Sci] [Age*Bio*Sci] [Bio*Att*Sci] [Bio*Ed*Sci] [Age*Att] [Ed*Att] [Gen*Att]

older people. Higher levels of education are clearly associated with higher levels of engagement with biotechnology. There also seems to be, roughly speaking, a positive relationship between engagement with biotechnology and with science. This holds amongst most comparisons of pairs of adjacent categories, but with notable exceptions for comparisons between the highest three categories of science engagement.

**Table 8.6 Odds ratios involving engagement with biotechnology with significantly associated variables, 15 countries, GM food subsample**

| | | Engagement with biotechnology | | | | |
|---|---|---|---|---|---|---|
| Item | Category | Lr/DKw vs. Lr/Lw | Lr/Hw vs. Lr/DKw | Hr/DKw vs. Lr/Hw | Hr/Lw vs. Hr/DKw | Hr/Hw vs. Hr/Lw |
| Age | 35-54 vs. 15-34 | 0.91 | 1.17 | 1.22 | 0.81 | 1.24 |
| | 55+ vs. 35-54 | 0.99 | 0.84 | 0.88 | 1.27 | 0.84 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 1.15 | 1.61 | 1.04 | 1.12 | 1.49 |
| | Still studying (young) vs. 20+ years old | 0.75 | 0.53 | 0.83 | 1.07 | 0.84 |
| Engagement with science | Low+ vs. Low | 1.55 | 1.78 | 1.38 | 1.04 | 1.29 |
| | Mid– vs. Low+ | 0.44 | 1.13 | 0.50 | 1.62 | 0.93 |
| | Mid vs. Mid– | 1.56 | 1.14 | 2.13 | 0.44 | 1.82 |
| | Mid+ vs. Mid | 1.29 | 1.75 | 0.24 | 2.62 | 1.44 |
| | High– vs. Mid+ | 0.50 | 0.80 | 10.00 | 0.50 | 0.42 |
| | High vs. High– | 1.00 | 1.50 | 0.33 | 1.00 | 2.00 |

Key

| | |
|---|---|
| Lr/Lw | Low report, low willingness |
| Lr/DKw | Low report, DK willingness |
| Lr/Hw | Low report, high willingness |
| Hr/DKw | High report, DK willingness |
| Hr/Lw | High report, low willingness |
| Hr/Hw | High report, high willingness |

Table 8.7 presents odds ratios involving attitude towards GM food. It supports, first of all, the conventional wisdom that greater support is found among men than women. Women are, in their turn, more likely to be found in one of the DK classes than are men. Another general finding supported by the data is the negative association between age and attitude – that is, with the younger expressing more enthusiasm for GM food than the older. In terms of associations with education level, those who are still studying are more likely than the highly educated to give a statement of strong (versus moderate) support for GM food. However, students and those with relatively low levels of education are much more likely to belong to a DK class than to give a statement of strong opposition to GM food.

Looking to engagement with science and engagement with biotechnology, two points are worth noting. Firstly, and not surprisingly, those with lower levels of engagement are more likely than their more engaged peers to have ambivalent attitudes towards GM food. Secondly, the association between engagement with biotechnology and attitude towards GM food is notably weaker (with many odds ratios in the last three columns of the table close to 1) than between science engagement and attitude. In the latter case, however, the nature of the association is unclear.

**Table 8.7 Odds ratios involving attitude towards GM food with significantly associated variables, 15 countries**

| | | Attitude towards GM food | | | | |
|---|---|---|---|---|---|---|
| Item | Category | DK vs. DK(2) | Def.oppos'n vs. DK | Opposition vs. Def.oppos'n | Support vs. Opposition | Def.support vs. Support |
| Gender | Male vs. female | 0.98 | 0.83 | 1.13 | 1.16 | 1.02 |
| Age | 35-54 vs. 15-34 | 1.08 | 1.14 | 0.87 | 0.87 | 1.07 |
| | 55+ vs. 35-54 | 0.96 | 0.89 | 0.88 | 1.12 | 0.88 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 0.96 | 1.30 | 1.12 | 0.94 | 1.07 |
| | Still studying (young) vs. 20+ years old | 1.29 | 0.39 | 1.50 | 0.89 | 1.50 |
| Engagement with science | Low+ vs. Low | 0.68 | 1.48 | 1.17 | 1.03 | 0.78 |
| | Mid- vs. Low+ | 1.50 | 1.00 | 0.70 | 1.17 | 1.22 |
| | Mid vs. Mid- | 0.63 | 1.06 | 1.52 | 0.83 | 1.06 |
| | Mid+ vs. Mid | 1.59 | 1.42 | 0.75 | 0.94 | 1.42 |
| | High- vs. Mid+ | 0.67 | 0.50 | 2.00 | 1.00 | 0.75 |
| | High vs. High- | 0.75 | 2.67 | 0.50 | 1.50 | 1.33 |
| Engagement with biotechnology | Lr/DKw vs. Lr/Lw | 0.75 | 2.43 | 0.80 | 0.92 | 1.02 |
| | Lr/Hw vs. Lr/DKw | 0.79 | 1.79 | 1.08 | 0.80 | 1.05 |
| | Hr/DKw vs. Lr/Hw | 0.72 | 0.65 | 1.02 | 1.20 | 1.00 |
| | Hr/Lw vs. Hr/DKw | 1.16 | 1.46 | 1.04 | 0.94 | 1.11 |
| | Hr/Hw vs. Hr/Lw | 0.89 | 1.43 | 0.94 | 1.09 | 0.88 |

Table 8.8 presents the significant parameter estimates for the linear part of the model, describing the relationship between knowledge and other variables. Broadly speaking, a positive view on GM food is associated with higher levels of knowledge, a negative view with lower levels of knowledge, and an ambivalent attitude with the lowest levels. Engagement is related to knowledge in a less clear way. Those in the higher science engagement classes would be predicted to have higher levels of knowledge than those in the lowest class, but there is no monotonic relationship between engagement class and knowledge level. For engagement with biotechnology, the notable distinction is between high and low reports of engagement: those who have heard of biotechnology

before are likely to have higher levels of knowledge. Unusually, the model predicts men to take slightly lower scores on the knowledge scale than women. This pattern is counter to standard findings, and would require careful investigation in future studies. Those with higher levels of education are predicted to score more highly on the knowledge scale than both those who left education at a younger age, and those who are still studying. This suggests that it is not proximity to the text book which favours respondents in this measure, as has sometimes been implied in the PUS literature. Finally, the older respondents are less likely than the other two groups to take high scores on the knowledge scale.

**Table 8.8 Parameter estimates from linear regression of knowledge on significantly associated variables, 15 countries, GM food subsample**

| | Knowledge of biology and genetics | | | |
|---|---|---|---|---|
| Item | Category | Estimate | Std. error | p |
| Intercept | | 0.658 | 0.006 | <0.001 |
| Gender | Female (ref.) | | | |
| | Male | -0.014 | 0.003 | <0.001 |
| Education (age at which completed) | Up to 19 years old (ref.) | | | |
| | 20+ years old | 0.048 | 0.004 | <0.001 |
| | Still studying (15-19 years old) | -0.006 | 0.008 | 0.464 |
| Age | 15-34 (ref.) | | | |
| | 35-54 | -0.001 | 0.004 | 0.818 |
| | 55+ | -0.046 | 0.004 | <0.001 |
| Attitude towards GM food | Definite opposition (ref.) | | | |
| | DK(2) | -0.034 | 0.015 | 0.020 |
| | DK | -0.022 | 0.005 | <0.001 |
| | Opposition | 0.001 | 0.005 | 0.867 |
| | Support | 0.016 | 0.005 | 0.001 |
| | Definite support | 0.014 | 0.006 | 0.017 |
| Engagement with science | Low (ref.) | | | |
| | Low + | 0.090 | 0.006 | <0.001 |
| | Mid - | 0.040 | 0.007 | <0.001 |
| | Mid | 0.056 | 0.005 | <0.001 |
| | Mid + | 0.043 | 0.007 | <0.001 |
| | High - | 0.125 | 0.006 | <0.001 |
| | High + | 0.094 | 0.006 | <0.001 |
| Engagement with biotechnology | Low report, low willingness (ref.) | | | |
| | Low report, DK willingness | -0.069 | 0.008 | <0.001 |
| | Low report, high willingness | 0.007 | 0.006 | 0.246 |
| | High report, DK willingness | 0.032 | 0.007 | <0.001 |
| | High report, low willingness | 0.054 | 0.005 | <0.001 |
| | High report, high willingness | 0.038 | 0.005 | <0.001 |

### *8.2.2 Therapeutic cloning*

Table 8.9 shows again a broad trend of higher engagement with science among men than among women, and a weak pattern of association between age and engagement. Once again, those who completed their education after the age of nineteen are generally more likely to be in higher engagement classes than those who finished their education before this age.

**Table 8.9 Odds ratios involving engagement with science with significantly associated variables, 15 countries, therapeutic cloning subsample**

| | | Engagement with science | | | | | |
|---|---|---|---|---|---|---|---|
| Item | Category | Low+ vs. Low | Mid- vs. Low+ | Mid vs. Mid- | Mid+ vs. Mid | High- vs. Mid+ | High vs. High- |
| Gender | Male vs. female | 1.08 | 1.67 | 0.92 | 1.09 | 1.00 | 2.00 |
| Age | 35-54 vs. 15-34 | 0.92 | 1.19 | 0.85 | 0.71 | 1.00 | 1.00 |
| | 55+ vs. 35-54 | 0.84 | 0.93 | 0.83 | 1.50 | 0.67 | 2.00 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 2.16 | 0.83 | 1.25 | 1.20 | 2.00 | 1.00 |
| | Still studying (young) vs. 20+ years old | 0.63 | 1.00 | 1.20 | 0.83 | 1.36 | 0.73 |

Table 8.10 shows again, but in a less clear way than before, the pattern of those in the middle, 35-54 years age group tending to belong to classes of higher engagement with biotechnology than younger and older people. The trend of higher levels of education associated with higher levels of engagement is much more clearly visible. And once again the broadly positive relationship between engagement with biotechnology and with science can be seen, but again not in a clear way.

**Table 8.10 Odds ratios involving engagement with biotechnology with significantly associated variables, 15 countries, therapeutic cloning subsample**

| | | Engagement with biotechnology | | | | |
|---|---|---|---|---|---|---|
| Item | Category | Lr/DKw vs. Lr/Lw | Lr/Hw vs. Lr/DKw | Hr/DKw vs. Lr/Hw | Hr/Lw vs. Hr/DKw | Hr/Hw vs. Hr/Lw |
| Age | 35-54 vs. 15-34 | 0.96 | 1.23 | 0.93 | 1.13 | 1.10 |
| | 55+ vs. 35-54 | 1.09 | 0.76 | 0.90 | 1.06 | 0.94 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 1.12 | 1.55 | 1.30 | 1.07 | 1.29 |
| | Still studying (young) vs. 20+ years old | 0.74 | 0.75 | 0.90 | 0.83 | 0.94 |
| Engagement with science | Low+ vs. Low | 1.58 | 1.71 | 1.17 | 1.27 | 1.16 |
| | Mid- vs. Low+ | 1.08 | 1.22 | 0.55 | 1.05 | 1.30 |
| | Mid vs. Mid- | 0.49 | 1.21 | 2.26 | 0.48 | 1.57 |
| | Mid+ vs. Mid | 1.54 | 1.50 | 0.65 | 0.87 | 2.14 |
| | High- vs. Mid+ | 0.80 | 0.75 | 3.33 | 1.40 | 0.27 |
| | High vs. High- | 1.00 | 2.00 | 0.50 | 0.86 | 2.80 |

Key

| | |
|---|---|
| Lr/Lw | Low report, low willingness |
| Lr/DKw | Low report, DK willingness |
| Lr/Hw | Low report, high willingness |
| Hr/DKw | High report, DK willingness |
| Hr/Lw | High report, low willingness |
| Hr/Hw | High report, high willingness |

Table 8.11 shows, in a similar way to the comparable table for GM food, that men are more likely than women to favour this biotechnology, and less likely to respond in an ambivalent way to these items. There is also some evidence of a negative association between age and attitude. There is weak evidence of a positive association between education level and attitude, again with the greater propensity among those with lower levels of achieved education (students, and those who finished their studies before the age of twenty) to be ambivalent rather than definitely opposed to therapeutic cloning.

In terms of engagement, for therapeutic cloning the weaker relationship seems to be with *science* engagement, rather than with biotechnology engagement (in contrast to the pattern for GM food); many more ratios close to 1 are found in the former section of the table than the latter. Those with higher levels of engagement with biotechnology tend to – on the whole – adopt more positive views of therapeutic cloning.

**Table 8.11 Odds ratios involving attitude towards therapeutic cloning with significantly associated variables, 15 countries**

| | | Attitude towards therapeutic cloning | | | | |
|---|---|---|---|---|---|---|
| Item | Category | DK vs. DK(2) | Def.oppos'n vs. DK | Opposition vs. Def.oppos'n | Support vs. Opposition | Def.support vs. Support |
| Gender | Male vs. female | 1.07 | 0.88 | 1.04 | 1.12 | 1.06 |
| Age | 35-54 vs. 15-34 | 0.83 | 1.13 | 0.93 | 1.08 | 0.97 |
| | 55+ vs. 35-54 | 1.31 | 1.00 | 0.81 | 0.87 | 1.15 |
| Education (age at which completed) | 20+ years old vs. Up to 19 years old | 0.84 | 1.11 | 0.96 | 1.13 | 1.14 |
| | Still studying (young) vs. 20+ years old | 1.25 | 0.50 | 1.25 | 1.16 | 0.92 |
| Engagement with science | Low+ vs. Low | 1.01 | 1.39 | 1.12 | 1.00 | 0.98 |
| | Mid- vs. Low+ | 0.76 | 0.73 | 1.29 | 1.17 | 1.00 |
| | Mid vs. Mid- | 0.82 | 1.38 | 0.96 | 0.89 | 1.04 |
| | Mid+ vs. Mid | 1.00 | 0.67 | 1.38 | 0.72 | 1.69 |
| | High- vs. Mid+ | 1.67 | 1.20 | 1.00 | 1.00 | 0.67 |
| | High vs. High- | 0.80 | 1.25 | 1.00 | 1.00 | 1.50 |
| Engagement with biotechnology | Lr/DKw vs. Lr/Lw | 0.44 | 2.08 | 1.11 | 0.89 | 1.45 |
| | Lr/Hw vs. Lr/DKw | 0.82 | 1.84 | 1.10 | 0.83 | 1.05 |
| | Hr/DKw vs. Lr/Hw | 0.88 | 0.65 | 1.28 | 1.15 | 0.72 |
| | Hr/Lw vs. Hr/DKw | 1.58 | 1.27 | 0.67 | 1.21 | 1.12 |
| | Hr/Hw vs. Hr/Lw | 0.48 | 2.22 | 0.88 | 1.03 | 1.19 |

Table 8.12 echoes the comparable table for GM food. Generally, the highest levels of knowledge are found among the most supportive, and the lowest levels of knowledge among the ambivalent. Higher engagement with science broadly corresponds to higher knowledge, and those who report having heard about biotechnology before are also more likely to score highly on the knowledge quiz. Once again the significant difference in knowledge for different levels of education lies between those who completed their studies before and after the age of twenty, and in terms of age, the significant difference is between the 55+ year olds, and the youngest respondents.

**Table 8.12 Parameter estimates from linear regression of knowledge on significantly associated variables, 15 countries, therapeutic cloning subsample**

| | Knowledge of biology and genetics | | | |
|---|---|---|---|---|
| Item | Category | Estimate | Std. error | p |
| Intercept | | 0.665 | 0.008 | <0.001 |
| Gender | Female (ref.) | | | |
| | Male | -0.017 | 0.003 | <0.001 |
| Education (age at which completed) | Up to 19 years old (ref.) | | | |
| | 20+ years old | 0.048 | 0.004 | <0.001 |
| | Still studying (15-19 years old) | 0.009 | 0.008 | 0.276 |
| Age | 15-34 (ref.) | | | |
| | 35-54 | -0.004 | 0.004 | 0.306 |
| | 55+ | -0.040 | 0.004 | <0.001 |
| Attitude towards therapeutic cloning | Definite opposition (ref.) | | | |
| | DK(2) | -0.057 | 0.019 | 0.002 |
| | DK | -0.015 | 0.007 | 0.029 |
| | Opposition | -0.003 | 0.007 | 0.720 |
| | Support | 0.026 | 0.006 | <0.001 |
| | Definite support | 0.036 | 0.006 | <0.001 |
| Engagement with science | Low (ref.) | | | |
| | Low + | 0.078 | 0.006 | <0.001 |
| | Mid - | 0.007 | 0.008 | 0.379 |
| | Mid | 0.040 | 0.005 | <0.001 |
| | Mid + | 0.040 | 0.007 | <0.001 |
| | High - | 0.116 | 0.007 | <0.001 |
| | High + | 0.073 | 0.006 | <0.001 |
| Engagement with biotechnology | Low report, low willingness (ref.) | | | |
| | Low report, DK willingness | -0.100 | 0.008 | <0.001 |
| | Low report, high willingness | -0.014 | 0.006 | 0.013 |
| | High report, DK willingness | 0.014 | 0.007 | 0.071 |
| | High report, low willingness | 0.041 | 0.005 | <0.001 |
| | High report, high willingness | 0.032 | 0.005 | <0.001 |

## Summary

The models presented in this chapter are intended to give just an impression of the ways in which the results from analyses in Chapters 5, 6 and 7 could be carried forward to explore the relationships between the PUS constructs and other background variables. With the caveat that the models are highly simplified versions of a complex reality, a number of interesting two-way associations have been found.

Notably, in the British data, for both GM food and therapeutic cloning subsamples:

- Positive associations are seen between education and behavioural engagement with biotechnology, and between engagement and knowledge.

- A negative association is found between age and level of knowledge.
- Although there are significant associations between engagement and attitude, the nature of the associations is rather complex – it is not possible to straightforwardly say that the two variables are positively associated.

In addition, for the GM food subsample:

- Men are significantly more engaged with science and biotechnology than women.
- Education level is positively related to knowledge level.

And for the therapeutic cloning subsample:

- Those aged 35-54 years old are likely to be more engaged with science and biotechnology than those aged 15-34 years.
- Those with higher levels of knowledge are marginally more likely than those with low levels of knowledge to adopt positive attitudes towards therapeutic cloning.

In the full fifteen country data sets, all two-way associations were found to be significant, except the association between gender and engagement with biotechnology. Among these associations, very similar patterns can be observed for both the GM food subsample and the therapeutic cloning subsample. These are as follows:

- Men are more likely than women to be found in higher science engagement classes. They are also more likely to have positive attitudes towards biotechnology (be it GM food or therapeutic cloning). Counter to conventional wisdom, the models also predict that men are likely to have lower levels of knowledge than women regarding biology and genetics.
- Education level is positively associated with engagement, both with science and with biotechnology. Those with higher levels of education are also likely to score more highly in terms of knowledge than those who left education before the age of twenty, and than those who are still studying. There is tentative evidence of a weak positive relationship between education and attitude towards therapeutic cloning. The relationship between education and attitude towards GM food is less clear, however.
- Those in the 55+ age group are likely to have lower levels of knowledge than those in the two younger age groups. The relationship between age and engagement is a little complex, but there is a suggestion that those in the middle, 35-54 year old age group, are likely to be more highly engaged with biotechnology. Age is negatively

related to attitudes: that is, younger people are more likely to have positive views on GM food and therapeutic cloning.

- There is a weak positive relationship between knowledge and attitude towards biotechnology (either GM food or therapeutic cloning). Generally, those with the highest levels of knowledge are likely to take the most positive views, those with lower levels negative views, and those with the lowest levels of knowledge are most likely to be ambivalent.
- The relationship between the two kinds of engagement is not clear. Broadly speaking it appears to be positive, though not straightforwardly so.
- Engagement is broadly speaking positively associated with knowledge. This relationship is not monotonic – so although this holds quite clearly for the comparison between very high and very low levels of engagement, for the categories in between the pattern is less clear. Engagement and attitudes towards biotechnology are also somewhat complex, and differ a little for the two applications of biotechnology. Attitudes towards GM food are only weakly associated with biotechnology engagement; they are more strongly associated with science engagement, although the nature of the association is not clear. By contrast, attitudes towards therapeutic cloning are only weakly associated with science engagement, and more strongly with biotechnology engagement – here, in a fairly clear positive way.

Having explored briefly the associations between the measures created in Chapters 5, 6 and 7, the final chapter comprises a summary of the substantive and methodological findings from those studies, and their practical implications for the design of future biotechnology surveys. This is set within a general discussion of the value of latent trait and latent class models for attitudinal survey research in social psychology.

# 9 Summary and discussion

This chapter provides a summary of the main findings from the empirical studies, and a discussion of their implications, in the light of the aims of the thesis. It begins with a reminder of the rationale for illustrating the use of latent trait and latent class models, describing briefly the challenges in survey analysis which they are well suited to meet, and recalling the research questions around which the empirical studies were framed. The results of the empirical studies are then presented thematically. First, the substantive findings are summarised, that is, in terminology of the taxonomy of Chapter 3, the representations of the constructs given by the models used. Next the methodological findings are presented, that is, the particularities of item functioning, elucidating two key themes from the taxonomy: generalisability, and the relationship between the research subjects and the research instrument. Thirdly, the implications of the analyses for future Eurobarometer survey design are described, in the form of practical recommendations and suggestions for new or modified items, or new observations and data. Some technical results follow, in the form of notes on the performance of the different fit statistics used for model selection. This relates to the theme of validation from the taxonomy.

Having thus looked back on the results of the empirical studies, the last two sections look forward, to possible future analyses. It is outside the remit of the thesis to make broad recommendations regarding which other variables or constructs need to be taken into consideration in PUS studies, but it is appropriate to comment methodologically on cross-national analyses more broadly, which have been a significant theme in the thesis. These comments relate not only to the issue of generalisability, but also touch briefly on the possibility of varying the levels of analysis used. The closing section reflects more generally on the added value that latent trait and class models can bring to analyses of attitudinal survey data, connecting the models to the theoretical framework of social psychology, and survey research methods.

## 9.1 Recalling the rationale and aims of the thesis

Chapter 1 began with the longstanding problem in survey research that when we ask respondents a question, their answers are a reflection of a host of factors, some of which are interesting to us, and some of which are a nuisance, possibly distorting the information we really seek. To derive an indicator of a construct of interest, such as attitudes towards GM food, it is better not to rely on a single item as fully informative, but to combine the information from a set of items which relate to that construct, and create a composite indicator. There are many ways in which this might be done; indeed, attitude measurement was one of the primary pursuits in social psychology during the mid-twentieth century, producing many sophisticated methods for this purpose. The approaches of the classic attitude measurement theorists Thurstone, Guttman and Likert were described in Chapter 3 for illustration, alongside a brief outline of relevant contributions to this topic from measurement theory in social psychology more broadly. Latent variable models are one approach we might take. These models are well known to social psychologists in the form of factor analysis, but less well known in the form of latent trait and latent class models.

As a set of statistical models, a latent variable approach has no strong or necessary connection to any single social psychological theory, nor to any single conceptual or theoretical specification of the attitude. This is important to note, since it is the echo of the dominant positivist and individualistic model of the attitude that makes the idea of attitude measurement via survey methodology so unattractive to more interpretatively minded psychologists. Survey research undoubtedly has limitations as a means of capturing complex, context-dependent and socially constructed concepts such as attitudes and opinions. Nevertheless, I hope to have shown that within their specified remit, that is to provide a broad and shallow map of the basic distributions of these constructs, they are not as completely individualistic in nature as one might suppose. Unfortunately, it is common to find the results of latent variable models reported in a way in which the level of analysis is unclear, or misleading; where a set of dimensions or classes which arise from between-subjects analyses are described as if they represent within-subjects representations. But latent variable models do not necessarily reflect a single representation of a concept, shared identically in the minds of all respondents. Instead they are something closer to a collective representation (Harré, 1984) – that is, a representation that results from the combined input data from all respondents. Although that representation is restricted to include the variables and responses that are preset by the survey designers, I would nevertheless contend that latent variable models afford us a depiction of perceptions which is closer to a social representations approach (Moscovici, 1961) than many of the classic attitude measurement models allow.

If this is agreed, a useful consequence follows. Whereas the classic models of the attitude made sharp conceptual distinctions between affect, cognition and behaviour, often depicting the relations between them in diagrammatic form, with social representations there is a far less rigid distinction between these elements of a representation. Social representations are systems of values, ideas and practices (Moscovici, 1973); representations are formed through behaviours; knowledge and affect are inextricably linked. This implies that in analyses we need not demarcate affect, cognition and behaviour items too distinctly. However, we do need to bear in mind that survey methodologists identify particular measurement problems associated with particular types of question, such as errors in recalling past behaviours, and errors of social desirability bias in committing to future behaviours. In surveys requiring very accurate reports of behaviours, this might be a particular concern – for example, in a political poll it might be important to gauge accurately whether respondents voted in the past, and will vote in the future. However, in an opinion survey on a more abstract topic such as perceptions of biotechnology, it might be said that the accuracy of responses is less at issue, and that questions on behaviours are intended more as general reflections on people's dispositions towards an object or issue. By this token, in the empirical studies in this thesis there was no conceptual need to distinguish rigidly between questions of cognition, affect and behaviour – they could all be included together in a latent variable model.

Chapter 4 described the primary technical details of latent variable models, and their particular value in the analysis of the Eurobarometer case study data. As probabilistic models, they immediately represent an improvement over the classic deterministic attitude measurement scales, addressing the first general feature of survey data – that is, response variability which can be termed 'measurement error'. As logit models, the latent trait and latent class models considered here represent an improvement over factor analysis when we are analysing nominal observed items, as is mostly the case with Eurobarometer surveys. By virtue of being suitable for nominal data, they again immediately make the inclusion of DK responses a simple matter. There is no need to ignore or recode DK responses, as would be necessary in order to use factor analysis. This is a particular benefit for Eurobarometer data, where DK answers are returned in high numbers. Finally, latent variable models can be used for sensitive cross-national analyses, allowing the comparison of measurement models between country samples, to gauge the extent to which the constructs analysed can be given the same interpretation between countries, rather than simply assuming such and treating the data as if Europe were a single population. The models thus address the key challenges in the data which, in the language of the taxonomy, can be described as issues of the relationship between the subject and the instrument, the form of data to be analysed, and the problem of determining the generalisability of the measures created from the data.

The empirical studies were designed so as to highlight a range of types of models, while also addressing some key concerns in PUS. Thus, scales of attitudes have been criticised in the PUS literature, and the suggestion made that looking for types of attitudes might be fruitful. In the first empirical chapter, therefore, I used latent class models to analyse a set of items for which an alternative means of classifying types of attitudes has already been found to work well. For knowledge, the PUS community (particularly this side of the Atlantic) favours continuous measures over typologies. I therefore focused on latent trait models for the second empirical chapter, but applying these models to items on biotechnology rather than the more heavily used and analysed items on science broadly. For engagement, the literature offers no clear steer on the best way to proceed. Both types and scales have been developed by other analysts, containing various combinations of items. So in the third empirical chapter I adopted an exploratory approach, using trait and class models in parallel for the British data, and on the basis of feasibility studies, I proceeded with class models in the multiple group analyses.

In all studies I have made a concerted effort to include and interpret DK responses. This contrasts with most previous analyses of these items, in which DK responses have tended to be ignored; typically unclassified in the attitudinal items, and recoded as 'incorrect' in analyses of the knowledge items. I did this simply by treating items that had any more than a negligible number of DK responses in them as nominal. I have also made a concerted effort to choose the best level of measurement for the observed variables. In the study of the construct 'engagement', one set of items (affective engagement with science and technology) had ostensibly ordinal response categories. I used the models to investigate the ordinality of the variables, rather than assuming it. Lastly, the studies of the three PUS constructs each involved the challenge of cross-national comparisons. Initial two-group studies were used to ascertain whether it was feasible to combine the separately sampled Great Britain and Northern Ireland into one UK unit, and likewise the two separately sampled regions of Germany. The task was then to find a well fitting representation whose measurement model could be sufficiently constrained between countries, in order to produce a comparable cross-national measure of the construct being studied.

Although each empirical study addressed a slightly different set of specific concerns (traits or classes, DK responses or ordinal responses, etc.), three general research questions or tasks were common to all three studies. Each began with a set of items chosen to represent a construct. The ultimate objective of each study was to find a well fitting model of those items, where the measurement model could be constrained to be equal, or as near as possible to equal, between the fifteen countries in the data set. This process was informed by a set of detailed preliminary analyses of the British sample. The analyses were informative, substantively, as to how the construct could be depicted, and methodologically, how the items functioned in relation to one another. The latter is essentially a rephrasing of the former, since the picture built of a construct depends on the ways in which the constituent items function together in a model. But conceptually, it helps to make a distinction between substantive and methodological interpretations of the models. So the first general research task could be described as concerning the substantive description of the content and structure of the models reached to represent attitudes, knowledge and engagement, within the constraints of the representations allowed by latent trait and class models.

The second general research question concerned the ways in which the constituent items function in relation to each other. This involved, for example, possible response styles (a facet of the relationship between the research subject and the research instrument), the ordinality or otherwise of variables, and the extra challenge of cross-national comparability (i.e. generalisability).

Arising from this, the third research question was practical in nature, asking what implications could be drawn from the analyses to inform future survey design for these sets of items. This involved, for example, the identification of items that did not fit well with the others in a set, and suggestions for ways to try to reduce response effects – that is, suggestions for new observations and data to reduce the impact of the relationship between the research subject and research instrument on the representations of the constructs sought. So the analyses have resulted in findings that could be located at three levels: the substantive and methodological interpretations arising from the analyses, and their practical survey design implications. The first relates to the theme of ways of representing; the second to generalisability and the relationship between the research subjects and survey instrument; and the third, as a result of these findings, to suggestions for new observations and data.

The empirical studies in this thesis have been presented not only as a vehicle through which to demonstrate the value of latent trait and class models in survey research, but also as freestanding studies, as a contribution to PUS research. The substantive and practical findings are most relevant to PUS researchers: the former in a general sense, and the latter specifically for those writing future PUS surveys, particularly of the Eurobarometer type. The methodological results should be of interest more broadly to those analysing attitude surveys. So there are three ideal-type audiences for these three sets of results. The highlights from these results are given in the next section. Within each of these sets they are presented first for attitudes, then knowledge, then engagement. The substantive results section also includes a summary of the patterns found in the loglinear analyses in Chapter 8. There is some overlap in the contents of the different sections, since many of the results have substantive, methodological and practical implications. Hopefully, however, the different terminology and emphasis given in each section are illuminating.

## 9.2 Substantive results: representations of the three constructs

The results reported below are shortened and simplified summaries of those described in the chapters to which they refer. They focus on the final, cross-national models of the constructs, rather than on the detailed analyses of the British data from the first halves of the studies. The results from this section are freestanding, and the intention is that they can be read and easily understood by a reader unfamiliar with statistics. This is done with two motivations: firstly, to highlight the substantive findings per se, since the focus of the thesis so far has been much more on the methodological implications of the analyses conducted; and secondly, to demonstrate and hopefully convince the reader that the models used need not be unpalatable to a lay audience.

### *9.2.1 Attitudes*

In this set of items respondents were asked for their opinions on a number of specific applications of biotechnology. Results are presented here for two applications, defined in the questionnaire as follows:

- **Genetically modified (GM) food**: using modern biotechnology in the production of foods, for example to make them higher in protein, keep longer or improve the taste.
- **Therapeutic cloning**: cloning human cells or tissues to replace a patient's diseased cells that are not functioning properly, for example, in Parkinson's disease or forms of diabetes or heart disease.

Half of the respondents were asked about GM food, and half about therapeutic cloning. In each case, they were asked specifically to consider to what extent they thought it was *useful for society*, *risky for society*, *morally acceptable*, and to what extent it *should be encouraged.* They were asked to choose from one of the following response categories: 'definitely agree', 'tend to agree', 'tend to disagree', 'definitely disagree', and 'don't know' (DK).

Latent class models can be used to identify patterns in people's responses to these items. Specifically, they can be used to find clusters of typical sets of answers, which can be taken as exemplars of the different types of reactions to GM food and therapeutic cloning amongst Europeans. For both GM food and therapeutic cloning, six general types of attitudes can be identified across the fifteen EU countries. These types are summarised in Table 9.1. Each type, or class, is represented by a column, with a label at the top of the column to describe it. Under the label, the most likely response or responses are given, for each of the criteria: *utility*, *moral acceptability*, *risk* and *encouragement.* The figure at the bottom of each column gives the percentage of Europeans estimated to belong to each class.

For both GM food and therapeutic cloning, the model is such that patterns of responses to the criteria *utility, moral acceptability* and *encouragement* are the same, country by country. To take the first application, GM food, for example: those in the 'definite support' class for GM food are likely to definitely agree that it is useful, morally acceptable and that it should be encouraged. However, responses to the *risk* item vary from country to country. In fact, in this most positive class, they vary so much that it is difficult to say what the trend is, across Europe. In the next, 'support' class, the typical response to this item is to 'tend to agree'. That is, the 'support' class consists of those who 'tend to agree' to every item. These would be termed 'risk-tolerant' supporters by some (e.g. Gaskell et al., 2003). Those in the opposition classes are also likely to agree that GM food is risky, while contending that it is not useful, nor morally acceptable, nor that it should be encouraged. Beyond these four classes which range from the most to the least supportive, there are two classes of equivocators. In the first of these, respondents are likely to give a DK response to all items, while in the second class, respondents are likely to say DK to questions of moral acceptability and overall encouragement, but agree that GM food is useful, while in some countries also agreeing that it is risky.

The patterns of responses defining the classes are very similar for GM food and therapeutic cloning. There are two key differences between the two halves of Table 9.1, which are highlighted in bold type for clarity. The first is in the patterns themselves. In the 'opposition' classes, respondents tend to disagree that GM food is useful, whereas in the case of therapeutic cloning they are split between agreement and disagreement. In the second DK class also, there is a tendency towards stronger agreement that therapeutic cloning is useful. So overall, there is a greater acknowledgement of the utility of therapeutic cloning, even among its opponents, compared to GM food. The second key difference shown in the table is in the proportions of supporters and opponents for the two technologies, across Europe. Whilst there are similar proportions of 'DK' respondents for each application, there is notably more support for therapeutic cloning than for GM food. 24 per cent of Europeans are classed as strong supporters of therapeutic cloning, whereas only 12 per cent are classed in the same way with reference to GM food. By contrast, 20 per cent are definite opponents of GM food, but only 10 per cent with regard to therapeutic cloning.

There are some interesting country-by-country similarities and differences. Europe-wide, opinion tends to favour therapeutic cloning more than GM food. Comparing the absolute percentages, in all countries except Germany, there are more supporters for therapeutic cloning than for GM food, and more opponents for GM food than therapeutic cloning. This pattern holds, despite the fact that numbers in the DK classes vary quite considerably between countries. Taking each application separately, for therapeutic cloning, in every country there are more supporters than opponents. For GM food, in the majority of countries we find more opponents than supporters – except in Germany, Ireland, Luxembourg, Spain and the Netherlands, where we find more supporters than opponents.

**Table 9.1 Classifications of attitudes towards GM food and therapeutic cloning**

| | Definite support | Support | Opposition | Definite opposition | DK | DK (2) |
|---|---|---|---|---|---|---|
| **GM food** | | | | | | |
| Useful | Definitely agree | Tend to agree | **Tend to disagree** | Definitely disagree | DK | **Tend to agree** |
| Morally acceptable | Definitely agree | Tend to agree | Tend to disagree | Definitely disagree | DK | DK |
| Should be encouraged | Definitely agree | Tend to agree | Tend to disagree | Definitely disagree | DK | DK |
| Risky | mixed | mostly Tend to agree | mostly tend to/definitely agree | Definitely agree | DK | mostly DK, tend to /definitely agree |
| % Europeans (pop. weighted) | **12** | 21 | 22 | **20** | 18 | 7 |
| **Therapeutic cloning** | | | | | | |
| Useful | Definitely agree | Tend to agree | **Tend to disagree/ tend to agree** | Definitely disagree | DK | **Definitely/tend to agree** |
| Morally acceptable | Definitely agree | Tend to agree | Tend to disagree | Definitely disagree | DK | DK |
| Should be encouraged | Definitely agree | Tend to agree | Tend to disagree | Definitely disagree | DK | DK |
| Risky | mixed | mostly Tend to agree | mostly tend to/definitely agree | Definitely agree | DK | mostly DK, tend to/definitely agree |
| % Europeans (pop. weighted) | **24** | 30 | 14 | **10** | 13 | 9 |

### *9.2.2 Knowledge*

A longstanding question for those interested in public perceptions of biotechnology is the relationship between knowledge about the topic and enthusiasm for its applications. Is it the case that to know biotechnology is to love it? Surveys asking people for their opinions on biotechnology therefore often include a number of questions aiming to assess their level of knowledge on the subject. The Eurobarometer has, for many years, posed a set of ten questions to respondents. These are, more specifically, a set of statements relating to biology and genetics. Respondents are asked to say whether they think each one is true or false, or whether they don't know. For some of the statements, 'true' is the correct answer, while for some 'false' is correct. The focus for this section is on responses to five of the ten statements. These are presented in Table 9.2, which shows the statements, and the percentages of respondents across the fifteen European countries who gave each of the three possible responses, to each one. For each item, the percentage highlighted in grey indicates those people who gave the correct answer. So, for example, 84 per cent of respondents correctly said it is true that 'there are bacteria which live from waste water'.

**Table 9.2 Europe-wide distributions of responses to knowledge items used in final scale**

| *% responses* | True | False | DK |
|---|---|---|---|
| There are bacteria which live from waste water. | 84 | 3 | 12 |
| The cloning of living things produces genetically identical copies. | 66 | 16 | 18 |
| It is possible to find out in the first few months of pregnancy whether a child will have Down's Syndrome. | 79 | 7 | 14 |
| By eating a genetically modified fruit, a person's genes could also become modified. | 20 | 49 | 31 |
| It is the mother's genes that determine whether a child is a girl. | 23 | 53 | 24 |

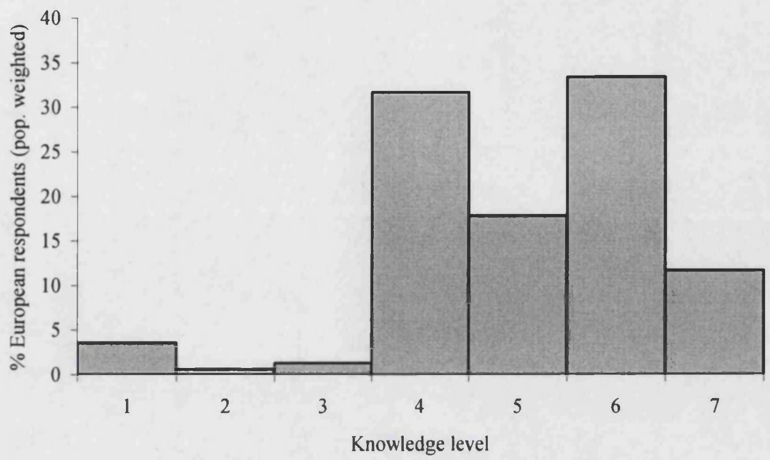
On average, people answered these questions well. Those statements for which 'true' is the correct response were answered correctly more often than those for which 'false' is correct. This raises some worries about a possible response style or bias in the data, specifically the tendency to give a positive response to a survey item (in this case, to say 'true' rather than 'false'), especially when in doubt. Indeed, the table suggests that self-doubt may be a common sentiment among Europeans: for all items, high proportions of respondents give DK answers – as much as 31 per cent for the statement on genetically modified fruit.

We can combine responses to these items to make a scale of knowledge of biology and genetics. A simple way of doing this would be to add up for each respondent the number of items answered correctly. There are two disadvantages to this, however. Firstly, it usually means assigning equal weights to all items in the scale, even though some may be easily answered by many people, and some may require a higher level of knowledge. Secondly, it usually does not allow the possibility of distinguishing between a substantively incorrect response and a DK response. It may sometimes be useful to do this, especially if we wish to investigate possible 'response effects' in the data: two other potential response styles are a propensity to guess, and the opposite – the tendency to say DK if one is at all unsure.

Latent trait models can be used to create a scale of an attribute such as knowledge, while taking these points into consideration. The model best representing these data consists of three dimensions, or traits. Two of these take account of the response styles mentioned above: the tendency to say 'true' to every question, regardless of the content, or to say 'DK' to every question. One dimension can be labelled 'knowledge', for which people at the top end of the scale answer every item correctly, and those at the lowest point answer incorrectly, with either a DK or a substantively incorrect response. For our purposes we are interested just in this dimension.

The scale of knowledge defined by this dimension is like a continuous scale, but due to technical details of the model, split into seven sections or levels, where 1 is the lowest level and 7 is the highest. The distribution of knowledge across Europe according to this scale is depicted in Figure 9.1. The histogram shows the percentages of respondents at each of these levels. Most Europeans, it appears, have quite high levels of knowledge according to the five items modelled, with the vast majority of people placed at the mid point or higher on the scale.

**Figure 9.1 Distribution of cross-national knowledge scale**




### *9.2.3 Engagement*

The public's engagement with biology is a topic of increasing interest for scientists and science policy-makers. Involving the public in decision-making on the future of biotechnology is predicated upon a public that is interested in and aware of the topic. But 'engagement' is an abstract concept, and there is no consensus as to how best to capture it with survey questions. The analyses reported in this section take the same form as in Section 9.2.1. That is, latent class models are used to try to reach a classification of types of engagement.

The data for this section are relatively diverse, consisting of a number of questions that respondents were asked at different points in the questionnaire. Some relate specifically to biotechnology, while others ask about science and technology more broadly; some ask respondents how they feel (that is, affective elements of engagement), while some ask about their behaviours – what they have done in the past and what they would be willing to do. These questions are split into two groups for the purpose of analysis. So two models are presented: one for affective engagement with science and technology in general, and one for behavioural engagement with biotechnology in particular.

Table 9.3 shows the groups identified for affective engagement with science and technology across Europe. They can be arranged in a logical order (though it is important to note that in terms of the statistical model used, the classes are not ordered). Again, the groups – seven groups this time – are listed in columns. The three questions used to form this typology are listed in the left hand column, with the most likely response to each question for each class in the corresponding cell. So, for example, the most highly engaged respondents will tend to say that they are interested in science and technology, feel well informed on the topic and understand science stories in the news, most of the time. 17 per cent of Europeans can be describe in this way. In the 'Mid' class, people will say 'some of the time' to each of these questions. 24 per cent of Europeans are placed in this middle category. And in the lowest engagement class, occupied by 22 per cent of Europeans, feelings of interest, informedness, and understanding of media science reports are likely to be experienced 'hardly any of the time'. These are the main types of engagement, which account for more than half of respondents. The remaining respondents are distributed amongst the other classes, giving different combinations of responses, as described in the table.

**Table 9.3 Classifications of types of affective engagement with science and technology**

| | High | High – | Mid + | Mid | Mid – | Low + | Low |
|---|---|---|---|---|---|---|---|
| I am interested in science and technology | Most of the time | Most of the time | Most of the time | Some of the time | Some of the time | Hardly any of the time | Hardly any of the time |
| I feel well informed about science and technology | Most of the time | Some of the time | Some of the time | Some of the time | Hardly any of the time | Hardly any of the time | Hardly any of the time |
| I understand science stories in the news | Most of the time | Most of the time | Some of the time | Some of the time | Hardly any/some of the time | Some of the time | Hardly any of the time |
| % Europeans (pop. weighted) | 17 | 12 | 7 | 24 | 8 | 11 | 22 |

Table 9.4 shows the results of a similar analysis for items capturing behavioural engagement with biotechnology. Here, six classes are identified across Europe. The first two, which ask if respondents have ever talked about biotechnology, or if they have heard or read about it (on television, in newspapers and magazines, on the radio, or the internet, or via some other media source), mirror each other very closely. So if a respondent has heard about biotechnology in the last three months, he or she is very likely to have talked about the topic before. We define people as 'high report' when they say 'yes' to both of these questions, and 'low report' when they say 'no'. Respondents were also asked whether they would be prepared to take part in discussions or public hearings on biotechnology. Their responses can be defined as high, low or

DK willingness depending on whether they answered 'yes', 'no', or 'DK', respectively, and all three responses are observed both for those who report having heard of and talked about biotechnology before and those who report not having done so.

So for example, those in the first class, in the first column in the table, are likely to have both heard of and talked about biotechnology already, and are likely to say they would be prepared to participate in public hearings. This class is labelled 'high report, high willingness'. There is a final question, however, for which responses vary a little between countries. This asks whether respondents would be willing to watch a television programme or read an article on the subject of biotechnology. Broadly, this seems to be a less demanding request than the idea of attending a public hearing. So in all of the 'high report' classes, the response to this question is likely to be 'yes', in most countries. In the 'low report' classes, people are likely to feel similarly about both passive and active forms of engagement with the topic.

Across Europe, nearly a third (30 per cent) are predicted to fall into the highest engagement class, with 23 per cent giving positive responses on all counts except the idea of taking part in a public discussion. 19 per cent of Europeans are expected to be broadly negative on the subject of engagement. The indication of a highly engaged Europe may be encouraging, or concerning, depending on one's perspective. Scientists and science policy makers would welcome the idea that people are aware of biotechnology and are in principle willing to engage further with it. But survey methodologists would warn against the possibility that 'social desirability bias' might creep into survey responses – that is, the tendency for respondents to sometimes give the answers that they think they *should* give, rather than the answers that accurately reflect their position. This is another example of a response effect, which is always a concern in survey research. On a positive note, however, it may be that where respondents were previously unfamiliar with the topic, the process of taking part in the survey genuinely increases their interest in it.

**Table 9.4 Classifications of types of behavioural engagement with biotechnology**

| | High report, high willingness | High report, low willingness | High report, DK willingness | Low report, high willingness | Low report, low willingness | Low report, DK willingness |
|---|---|---|---|---|---|---|
| Ever talked about biotech? | Yes | Yes | Yes | No | No | No |
| Heard about biotech in last 3 months? | Yes | Yes | Yes | No | No | No |
| Would take part in discussions or hearings? | Yes | No | DK | Yes | No | DK |
| Would watch programme/read article? | Yes | mostly Yes | mostly Yes | Yes | mostly No | mostly DK |
| % Europeans (pop. weighted) | 30 | 23 | 9 | 14 | 19 | 5 |

### *9.2.4 Relationships between the three constructs*

Having explored the three concepts of attitudes, knowledge and engagement, the question arises, are they related with one another? And if so, in what way? Some simple analyses were conducted to explore the relationships between these variables, and between these and three socio-demographic variables: gender, age, and level of education. These analyses were carried out for all fifteen European countries at once, so the results below should be taken to be Europe-wide.

The results of these analyses can be summarised as follows:

- Men are more likely than women to be found in higher science engagement classes, although there is no appreciable difference between the sexes in terms of engagement with biotechnology. Men are more likely to have positive attitudes towards biotechnology (be it GM food or therapeutic cloning). The analyses also suggest men tend to have lower levels of knowledge than women regarding biology and genetics. This is a surprising finding, and runs counter to the patterns that are usually found in analyses of public perceptions of biotechnology. As such, it should be taken with a degree of caution: further research should be undertaken to clarify it.
- Those with higher levels of education are likely to be more highly engaged, both with science and with biotechnology, than those with low levels of education. They are also more likely to have higher levels of knowledge about biology and genetics than those with less education (including those who are still studying). There is

tentative evidence that high levels of education are linked with more positive attitudes towards biotechnology, but the relationship is not very clear.

- Younger people are likely to have higher levels of knowledge about biology and genetics than older people: those above the age of 55 years, in particular, are likely to have lower levels of knowledge than the young and middle aged. Middle aged people (between 35 and 54 years old) are likely to be more engaged with biotechnology than both older and younger people. The young stand out as being notably more positive towards biotechnology, be it GM food or therapeutic cloning.
- As people's knowledge of biology and genetics increases, so their attitude towards biotechnology becomes slightly more positive. It should be noted, though, that we cannot tell from the analysis whether it is knowledge that leads to positive attitudes, or a positive disposition which motivates one to learn more about biotechnology. Probably, a little of both occur. Those with higher levels of knowledge also tend to be more engaged with science and with biotechnology.
- The relationship between engagement and attitudes towards biotechnology is somewhat complex. There is reasonable evidence that higher levels of engagement with biotechnology go along with positive attitudes towards therapeutic cloning. But the link between engagement and attitudes towards GM food is less clear.

## 9.3 Methodological findings: patterns reflecting the relationship between the respondent and the survey instrument, and the generalisability of representations

Just the principal findings from the empirical studies are recapitulated in this section. Greater emphasis is given to the joint models than to the results from the British data, and to the sets of items which for which it was more challenging to find a joint cross-national model. The summaries below involve some repetition of the results described in Section 9.2, in that it is informative to describe substantive relationships between variables in terms of item functioning. Some characteristics of item functioning might be explained as response effects, or other potentially problematic relationships between the respondent and the survey instrument. These make the task of finding good and generalisable models of constructs more difficult, and as such, they are taken forward into Section 9.4, as points of attention for those designing future waves of the Eurobarometer.

### *9.3.1 Attitudes*

Latent class analyses show, for both GM food and therapeutic cloning, strong positive associations between three criteria of support and opposition: *utility*, *moral acceptability* and *encouragement*. The fourth criterion, *risk*, is weakly and broadly speaking negatively associated with the other items.

Using class rather than trait models seems to be a helpful feature of this analysis. In a contrasting approach, Pardo et al. (2002) ran a factor analysis of sets of logics items from the 1996 Eurobarometer, and found a two-factor model of attitudes towards biotechnologies, with risk items loading on one dimension, and all other criteria on the first factor. Whilst one interpretation of the latent class models from Chapter 5 might be simply to say that *risk* is independent of the other items, this seems an unnecessarily simplistic line to adopt. A latent class model shows clearly, in a way that a trait model cannot, where risk does and does not align with the other items. For example, responses to the risk item are very unpredictable for those in the 'definite support' class, but well defined in other classes. Responses are only well defined up to a point, however: there is generally speaking a good deal of heterogeneity in responses to this item, both within and between countries. For each application of biotechnology, in order to achieve a well fitting joint model of attitudes, parameters for this item must be freed between countries. The implication is that the relationship of perceptions of risk to the classification of types of attitude varies between countries.

Another helpful feature of the latent class models for the logics items is the possibility of including DK responses, rather than excluding them from the classification of types of attitudes. While a notable proportion of respondents give a full set of DK answers to the four attitude questions, an equally notable proportion give a mixture of DK and substantive answers. It is interesting to see in which direction these latter respondents lean, and on which questions they tend to reserve judgement. Where these people give DK responses, it tends to be more often on the criteria of moral acceptability and overall support than on utility and risk. The final joint models for the fifteen countries each contain a 'DK (2)' class, in which the typical respondent is unsure whether the biotechnology in question is morally acceptable, unsure whether it should be encouraged overall, but likely to agree that it is useful, and in about half of the countries surveyed, also agree that it is risky.

### *9.3.2 Knowledge*

The task of creating a scale of knowledge has resulted only in partial success. There seems to be too much variation between countries to entertain a model in which difficulty and discrimination parameters of items can be constrained to be equal across Europe. The analyses have produced some useful insights, however, and may be a valuable stepping-stone towards a more satisfactory future model.

Firstly, it appears that it is useful to retain the distinction between DK and substantively incorrect responses. Recalling that it is inadvisable to add guessing parameters to trait models, particularly when some observed items are easy (Thissen & Wainer, 1982), allowing a second trait seems to be a good way of accounting for response effects in the data. Within each country, a two-trait model for the ten items treated as polytomous variables fits well. It returns one trait which can be interpreted as representing 'knowledge' (more or less well, depending on the country), while the other varies a little more between countries, but which can usually be interpreted as representing response effects.

The representations of knowledge in these trait models are not perfect. For a number of items, particularly those for which 'true' is the correct response, the trace line for the *incorrect* response curves slightly upwards towards the high end of the scale. This creates problems when factor scores are derived from the trait, giving the counterintuitive result that those respondents who answer all questions correctly may not be positioned at the very highest point on the scale. This instability in slope parameters is often due to the fact that the 'true' and 'false' curves in these items are very shallow, that is, they have low discrimination power.

A further notable characteristic of the 'true=correct' items is that they are, broadly speaking, easier to answer correctly than the 'false=correct' items. It is rather important to ascertain the reason for this. It could be that they simply happen to be easier in the common sense meaning of the term – that is, that they test facts which are genuinely widely known. It could also, however, be that the high rates of correct answers are a result of acquiescence bias, i.e. in this case a response tendency towards saying 'true'. My analyses suggest tentatively that the latter may be the case. In many of the trait models, within countries and in joint models, the non-knowledge trait(s) could often be interpreted as representing response style, often with DK at one end and 'false' at the other, i.e. the least confident to the most confident responses.

The magnitude of discrimination of 'true' versus 'false' responses varies between items and between countries. This variation is not so strong that pairs of items are in a significantly different order of discrimination, from country to country. However, clearly the magnitudes of these parameters are too diverse to yield a well fitting joint model where these are constrained to be equal between groups. Dropping some items from the set of ten, and increasing the number of traits, improves fit, but only when undertaken in quite dramatic proportions. The final joint model for the knowledge items uses three traits to represent five items. With a small number of items, a three-trait solution is quite unstable; deriving a model in which one trait can be interpreted as representing 'knowledge' depends more than is desirable on the particular combination of items in the model. The results of the models estimated in LEM add further evidence to this point. The model fitted specified one trait with equal discrimination parameters between countries and one allowing them to differ between groups. In fifty runs of this model, nearly fifty different deviances were returned. In the best fitting model, the fixed trait could be interpreted as response style rather than 'knowledge', with all DK responses most likely at one end of the trait, and all 'false' at the other end.

Analyses of the British data suggest that the distributional assumptions on the latent variable make a substantial difference to the allocation of posterior scores. In the continuous trait model fitted to the ten items in the set, factor scores form a nearly normal distribution, following the assumption of a normally distributed latent variable inherent in the model. However, using a discrete trait model, factor scores are distributed bimodally. This carries important implications for PUS, which has witnessed a number of theoretical predictions regarding the distribution of knowledge in more and less advanced industrial nations. However, the instability of the models of knowledge makes it premature to begin noting the shapes of the distribution of knowledge in different countries. In the joint cross-national model, indeed, the distribution of knowledge is negatively skewed.

Some practical suggestions for the scale will be made in Section 9.4.2. Here, however, it may be useful to reflect on the methodological objectives for a measurement of knowledge of biology and genetics. Sturgis and Allum (2006) pragmatically note that the success or otherwise of a measure must be decided according to the purpose for which it is needed. For these items, the within-country trait models of knowledge provide well fitting models of knowledge, with intuitive interpretations. Some of the slope estimates would create small anomalies in the factor scores derived from them, however – so they are not unproblematic. Nevertheless, these problems are slight, and the chances for resolving them are great, in comparison to the much more difficult task of finding a well fitting and meaningful cross-national model for the data.

It is worth pointing out specifically the challenge for a cross-national model. Borrowing from item response theory (IRT) used in the context of educational testing, a well fitting comparable model means fixing both discrimination and difficulty parameters to be equal across samples of the populations of all fifteen countries, for ten or fewer items. In the usual scenario of IRT, the challenge is to find a unidimensional scale from a much larger pool of items, to measure knowledge of a subject defined by a preset syllabus, administered to students of the same age who have been taught that syllabus in the preceding year. With the biotechnology items, there are only ten questions, administered to random samples covering all ages and all educational backgrounds of entire national populations. On these terms, it might be fair to say that we should more surprised to find a model for the data at all, than dismayed that the final model for these items fits poorly. A joint latent trait model of these items may be too much to ask of such a diverse population and broad field of knowledge.

This is far from a call to abandon the measurement of scientific knowledge. For within-country analyses, the existing two-trait models already deliver a good deal of useful information about the items. For cross-country comparisons, given the importance of knowledge to the PUS research field, more efforts to improve the scale would be valuable. Methodological critiques of biotechnology and science knowledge scales emphasise this. Pardo and Calvo (2004) point out that the weak association often found between measures of knowledge and attitudes might be partly attributable to the quality of the scales used. In their meta-analysis of the relationship between knowledge and attitudes, Allum et al. (forthcoming) find the greatest variance in their model attributable to the measures used, and very little to cross-national differences. This in particular provides motivation to work towards a better fitting cross-national measure.

### 9.3.3 *Engagement*

This summary should be prefaced with a reminder of the heterogeneity of the items in the study of engagement. Some items refer to science and technology in general, and some to biotechnology in particular; some ask respondents for affective judgements; some for reports of past behaviours; some for statements of willingness regarding potential future behaviours. Moreover, the items are dispersed throughout the questionnaire, rather than posed in a single battery, as the case with the logics and knowledge items. And many of them refer to less abstract concepts than opinions and knowledge. For example, the question asking if the respondent has heard about biotechnology on the internet is rather difficult to compare between countries where internet access is itself unevenly distributed. The culture for or against public meetings also varies markedly between European countries, making comparisons of the question *Would you attend a public hearing on biotechnology?* difficult too. A cross-national model could not be found to fit all of these items together, but well fitting joint models were found separately for those that refer to science and technology, and for those that refer to biotechnology.

Considering first the items relating to affective engagement with science and technology, the analyses began with four items, but the fourth, *I become confused when I hear conflicting views on science and technology*, was quickly dropped. It was found to be broadly negatively associated with the other items in the set, but in quite different ways from country to country. The other three items, however, are quite positively strongly associated. In fact, country-by-country, three-class models fit well, and for each country, are characterised by the same three typical patterns of responses: these are simply response profiles where the same reply is given to each item. However, the magnitudes of the conditional probabilities are clearly sufficiently different between countries to make a three-class model too restrictive. Constraining the measurement model to be equal across countries, at least five classes are needed to achieve good model fit, but even here, the interpretation of the extra classes is not clear. The addition of more classes sharpens the interpretation of the solution; seven classes give the clearest patterns of response types. So the final model contains approximately half of respondents in the three main classes (giving the same response to every question), with two classes in between the extremes and the middle class; in this latter class respondents answer 'some of the time' to every item.

It is important to note that although the classes can be ordered in common sense terms, this does not mean that they are ordered statistically. The latent variable is nominal here. Indeed, imposing any ordinality on the latent variable makes it much more difficult to achieve model fit. A discrete trait model does not fit the data. Likewise, although the response categories of the observed items are, in common sense terms, ordinal ('hardly any of the time', 'some of the time', 'most of the time'), analyses of the British sample data suggest that they do not behave ordinally, statistically speaking. For the British sample, trait and class models for ordinal observed items do not fit the data well. Modelling the items as nominal, a two-trait model fits the British data, where on the first trait the parameters follow a logical order ('some of the time' in between 'hardly any' and 'most of the time'), but on the second trait, the middle response 'some of the time' has the highest loading. So the second trait could be interpreted as reflecting a tendency to give a middle response. This interpretation of a second trait does not hold for all countries, however; in preparation for the multiple group analyses, two trait models were run country-by-country (but not reported in Chapter 7, since they did not lead to a good joint trait model), revealing that the interpretation of the second trait varies a good deal.

In the items regarding behavioural engagement with biotechnology, different methodological concerns arise. For the cross-national analysis there are two pairs of items. The first pair ask if respondents have talked about or heard of biotechnology previously. There is a strong positive association between responses to these items. The second pair of items ask if respondents would be prepared to take part in a public discussion on biotechnology, or to take the time to watch a programme or read an article about it. Responses to these items are positively associated with each other, but with an interaction effect: conditioning on negative responses to the talked/heard items, they mirror each other closely, but conditioning on positive responses to the talked/heard items, they are not as strongly associated. This is a function of a general tendency for people to be more willing to engage in a passive way (reading an article) than an active way (taking part in a discussion). An additional qualification to make to the connection of the passive-participation item with the other three is that patterns of responses vary significantly from country to country; it is necessary to free the parameters for this item in order to achieve a well fitting model.

Once again, a well fitting joint model requires more classes than well fitting country-specific models. Within countries, four classes are sufficient to represent the variation in the data. But the composition of these classes vary widely – unlike the very similar patterns found country-by-country for the science items. It is not at all surprising, then, that it is necessary to expand the number of classes to accommodate such variation.

## 9.4 Practical comments, recommendations and suggestions for future Eurobarometer surveys on biotechnology: new observations and data

The recommendations and suggestions given below, based on the objective of creating robust measures of constructs from the Eurobarometer data, are tentative ideas. It is worth repeating here the praise that Durant and colleagues (2000) give to the Eurobarometers, noting that they are without rival in their coverage over time and between countries. Finding good measures which could be generalised to previous data sets could be an asset to analyses of these data. It is also worth recalling that in their meta-analysis of the relationship between knowledge and attitudes, Allum et al. (forthcoming) find that a large proportion of variance in their model is explained by the different measures used. Deriving standardised measures for these key constructs would therefore seem to be a valuable enterprise.

### *9.4.1 Attitudes*

The logics are perhaps the most successful of all the items analysed in this report. Since they have been included in the survey since 1996, it would be valuable to retain them in future survey waves, in order to track aggregate changes in attitudes over time. Of the four items, *risk* is notably unusual in its behaviour; responses to the other three items are strongly associated, but with *risk*, patterns of responses are much more varied, within as well as between countries. No specific recommendations for particular questions to pose in future follow immediately from the analyses in this report. But I would make a more general recommendation: simply to ask more questions about risk in future surveys, in order to try to understand the variation in responses more fully. The content of such questions is, however, a matter for theorists in PUS and in public perceptions of risk.

The ordering of the criteria might be given some attention: an investigation of 2005 British data for GM food suggested that perhaps responses to *risk* are becoming more closely aligned with the other items. However, this cannot be confidently ascertained, since the question order was different in 2005: in this more recent wave, respondents were asked in order whether they thought the biotechnologies were *morally acceptable*, *useful*, *risky*, and *to be encouraged*. In 2002 by contrast, respondents were asked first if they thought the biotechnology in question was *useful*, then *risky*, then *morally acceptable*, and *to be encouraged*. If resources allowed, it would be helpful to use the split ballot to test for question order effect.

A matter of greater concern in these items is the different aggregate DK response rates between countries. For therapeutic cloning, for example, proportions of respondents in the two DK classes range from 13 per cent in Denmark to 48 per cent in the Netherlands. It would be very useful to know whether this reflects genuine differences in rates of not-knowing between countries, or whether it is a function of survey company fieldwork procedures. Between the 1999 and 2002 waves of this survey some notable changes in DK rates can be observed in some countries. For example, in answer to the question, *To what extent do you think GM food should be encouraged?*, between 1999 and 2002 DK rates fell from 44 to 16 per cent in Luxembourg, and rose from 7 to 42 per cent in the Netherlands. If fieldwork procedures are responsible for these trends, it invokes an urgent call for efforts to standardise administration procedures between countries and over time.

### *9.4.2 Knowledge*

Considering first of all the internal workings of the scale within countries, the most striking feature of the knowledge items is that many of them are relatively easy, and therefore not diagnostic, as the majority of respondents answer them correctly. However, it may not be that the items straightforwardly test facts that are very widely known – the high rates of correct answers may be attributable to response effects such as guessing, or acquiescence bias. Indeed, those items for which 'true' is the correct response are easier (that is, more people answer them correctly) than those for which 'false' is correct. The biotechnology items and the Oxford items are very similar in this regard. In their methodological analysis of the latter, Pardo and Calvo (2004) suggest that the scale could be improved by adding or substituting more difficult items in the set. They specifically recommend using more 'false=correct' items to increase the difficulty level of the test. They also suggest offering a four-point Likert answer scale, to allow respondents to differentiate between whether they think each statement is 'definitely' or 'probably' true or false, to alleviate guessing or other response effects. However, I would take a different approach. Response styles such as acquiescence bias and guessing are known to be more likely among certain demographics, including cultural groups (Smith, 2003). Increasing the number of 'false=correct' items might therefore lead the scale to favour a particular type of respondent, making it even more open to charges of bias.

It would undoubtedly be useful to try to increase the number of difficult items, but I would suggest this would be more effective if it could be ensured that it was the content of the item, not the required response, which was difficult. So it would be advantageous if these more difficult questions were mixed, with some requiring 'true' and some 'false' as correct responses. The four-point Likert response scale may to some extent reduce the possible effects of guessing and acquiescence bias, but I think a more likely successful strategy might be to remove the true versus false dichotomy from the exercise altogether, instead asking respondents to choose between the two. Many of the original Oxford and biotechnology knowledge scale items could easily be reformulated in this style. For example, *It is the mother's genes that determine the sex of the child* would become a task of choosing between the statements *It is the mother's genes..* and *It is the father's genes...* Multiple choice items might also be considered. For example, the statement, *Ordinary tomatoes do not contain genes, while genetically modified ones do*, could be reformulated as a question, such as:

*Which of the following contains genes?*

A. Human beings
B. Fruits and vegetables
C. GM fruits and vegetables
D. A and B
E. B and C
F. All of the above.

Ideally the order of the first three response options would also be rotated. Multiple choice questions might be more complex to analyse, and more costly to field, but should be seriously considered as a possible way of alleviating the response effects in the data.

Changing the item format altogether is a relatively drastic move: new items may always work less well than established ones, and new items prevent the analysis of trends over time. A more moderate strategy would be to add more difficult items to the existing set for the next survey, and evaluate the effectiveness of this before considering changing the question format altogether in future waves. Adding items seems to carry a smaller risk of failure, although due consideration needs to be given to respondent fatigue and the relative importance of this construct versus the other topics that need to be covered in the questionnaire. The matter requires a strategic decision by the survey designers. If a model could be found to fit items in the current format, then this may be the best solution to creating a measure of knowledge. However, such a model would be likely to contain more than one trait, to account for response effects, and might not be as intuitive to a lay audience as a unidimensional knowledge scale.

### *9.4.3 Engagement*

Considering first the items relating to affective engagement with science and technology, the fourth, *I become confused when I hear conflicting views on science and technology* should be a candidate for deletion from future survey waves. Country-by-country, it is found to function in different ways relative to the other items in the set, and it generally has low discrimination power. In Chapter 7 I commented that this item could be interpreted in a variety of ways. For example, the response 'hardly any of the time' would be applicable to those who were highly engaged with science and technology and familiar with the topic, as well as to those who were very much unengaged, so that they never heard such conflicting views to begin with. So interpreting responses to this item seems a difficult exercise.

Another idea for the science items is to consider increasing the number of response categories, from three to four or five. This would be a slightly risky strategy, jeopardising the possibility of tracking trends over time. But it may be that the lack of fit of the joint three-class model relates to measurement error among those respondents who do not give one of the three main types of response; perhaps for these people, the available response categories divide up the response continuum too coarsely. Before acting on this, it may be useful to conduct a preliminary analysis of the 2005 Eurobarometer. In this survey some similar items are posed, but with four response categories: 'never', 'rarely', 'sometimes' and 'often'. If a joint model is easier to fit with these items, it may be an indication that a change in response scale should be considered. A thorough future study could make use of the split ballot to investigate three- versus four-point response scales. This would be useful, certainly in the British data, for investigating the apparent response effect of answering in the middle response category. If such answers are in some cases satisficing responses, then four points may remove or attenuate this effect. It may also affect DK response rates of course, which in the current data set are negligible.

There is no obvious or particular problem with the biotechnology items. The only suggestion I have to improve them is the more general suggestion for all of the engagement items; that is to reduce the number of 'things going on' in the data: the different elements of engagement covered, in different parts of the questionnaire. Increasing the number of questions posed, and putting them together in one set, with similar response formats, might make it easier to begin untangling the different elements in the data. It might be useful, for example, to explore further whether there is really a meaningful distinction to be made between engagement with science generally, and biotechnology specifically – simply by asking the same questions of each of these, directly.

## 9.5 The performance of selected fit statistics: validation

Validation has been addressed almost exclusively in terms of model fit in this thesis (barring a cursory consideration of construct validity, in its narrow sense, in Chapter 8). With this specific focus, a number of observations can be made regarding the success of statistics used for model selection.

Taking each in turn, the bootstrap p-values for model deviances are on the whole very small (<0.001) for the models presented in the empirical chapters. Although the bootstrap p-value overcomes the problems with calculating asymptotic p-values from sparse tables, it cannot help with the sensitivity of the statistic to large sample sizes. The Eurobarometer sample sizes are not very large, but it seems that they are sufficiently so to make tests of statistical significance too sensitive for practical purposes. Models which would achieve large p-values would probably be too complex to be useful, or even interpretable. Likelihood ratio comparison tests of nested models are also sensitive, often indicating significant differences between models which seem remarkably similar in interpretation.

The penalised statistics AIC and BIC have produced no new insights. Characteristically, BIC tends to favour smaller models, and AIC larger models. It is rare in the empirical studies that they indicate the same best model from a set, and they seem to bear fairly little relation to the statistics summarising marginal residuals.

For model selection I have relied most heavily on the two new statistics summarising proportions of large standardised marginal residuals from two-way tables. These appear to have worked very well. Three qualifications might be suggested for their use in future. The first is simply to point out that in models with very few parameters, one might wish to set a slightly lower standard to denote good fit in the 'percentage of two-way standardised marginal residuals >4', since the conversion to percentages upweights the result (this was commented on in Chapter 7). The second is that the Jöreskog & Moustaki index is effectively calibrated on a slightly different scale to the 'percentage of large marginal residuals'. Jöreskog and Moustaki (2001) suggest a critical value of 4 for indicating problematic model fit, for single two-way tables. For the mean of many of these values, I generally took values less than 1 to denote good fit. So as a rule of thumb, 1 might be taken as an approximate criterion for this statistic. The third point relates specifically to the cross-national analyses, and the use of the 'percentage of large marginal residuals' statistic. I often found that this figure could be quite small for the model overall, while concealing large values conditional on country. So in multiple group analyses, it would be wise to consider fit statistics both overall, and conditional on group. For the latter, a useful way of comparing models could be to take the mean of the percentages across countries.

## 9.6 Cross-national comparisons: comments on generalisability and levels of analysis

Cross-national comparisons are of key interest in PUS. In this thesis I have focused on how to derive measures that enable comparisons to be drawn between countries, without any particular reference to the use to which they might be put to test substantive theories in PUS. In fact, societal-level theories appear relatively often in the literature. Indeed, on a fundamental level, historical epochs are defined by the status of science, and its relationship with the general population. In contemporary PUS studies, societal-level theories mostly refer to the effects of industrialisation on the general climate of support for science and technology. Industrialisation, so the theory goes, at first brings with it increased enthusiasm for the promise of science and its contribution to improved quality of life. In stages of late industrialisation, or post-industrialisation, the negative consequences of science and technology begin to be recognised, and science may be viewed with more scepticism overall, or the public may be more discerning regarding different manifestations of science and technology. So we expect to see broadly positive attitudes towards all biotechnologies metamorphosing into more polarised attitudes, specific to certain technologies, as countries develop economically. In terms of changes in knowledge, the 'normalisation hypothesis' would lead us to suppose that as a country develops economically, the concomitant increases in overall quality of life means that higher levels of education are more widely available, so that knowledge of science becomes more evenly diffused. This implies that the shape of the distribution of knowledge changes over time, from bimodal (specialised) to unimodal (more evenly distributed).

We are not yet in a position to test either of these theories thoroughly. This is not only because the final model reached for knowledge in Chapter 6 is poorly fitting. These theories, about change over time, require a new effort to find models to cover several survey waves, fixing measurement models to be equal across time points, as well as between countries. It may well be a tall order to find such a stable model for a new and dynamic topic such as biotechnology. Nevertheless, such an approach should be preferred over the currently common method of using single cross-sectional data sets to test these theories, and taking less advanced, or more recently industrialised countries, as a window on history. Such an approach is methodologically questionable. The problem with inferring that less industrialised countries are equivalent to more industrialised countries in former years is analogous to cross-tabulating age with attitudes at the individual level, and neglecting to test for cohort effects. It is, moreover, normatively problematic, and politically sensitive. For both reasons, it would be very useful to develop robust measures that can be carried through several survey waves. This alone is motivation for trying to first find a way of fitting a good model to the knowledge items in their current format, before resorting to wholesale changes of item content and structure.

This approach is itself not without shortcomings, however. The cross-national measures presented in the empirical chapters are fixed effects models: in the structural part of the model, parameters are estimated individually for each country. They thus do not yield themselves to the multilevel models which would be ideal for testing the industrialisation hypothesis, or similar societal-level theories. The fixed effects of countries with regard to attitudinal constructs explain all of the country-level variance, preventing us from entering other country-level variables such as national income, dominant religion, etc., into such a model. The measures I have created in the empirical studies are therefore not a good solution for testing societal-level theories.

Random effects models might therefore be considered for future analyses. These would involve additional considerations. For example, the data requirements for such a model would be greater: fifteen units may be too few for a multilevel model, but perhaps more recent Eurobarometers, with 25 or more country samples, might make this approach feasible. A random effects model would also involve imposing some additional distributional assumptions on the structural parts of the models. This may not necessarily be disadvantageous – in fact, it may be the logical next step in the development of these analyses, following the fixed effects models presented in this report. Where the accuracy and fit of the measurement models between countries is a particular worry, or a direct research question, a fixed effects model is clearly indicated. Having explored this research question, a random effects model might be attempted next. Software already exists for this purpose (Rabe-Hesketh, Pickles et al., 2004).

Having suggested using more involved statistical models, it is important to reiterate the caveat on cross-national analyses stated in Chapters 1 and 2. In the empirical studies I hope to have improved on standard cross-national analyses of survey data which imagine that respondents are from the same population, and which assume that constructs such as attitudes and knowledge have the same interpretation in every country. The empirical studies have aimed to answer the question of whether this is a fair assumption, rather than take it for granted. But they have only addressed this question on a statistical level, and it would be stretching their credentials to claim that they are fully sensitive cross-national analyses. For such purposes qualitative data should be considered, and perhaps more formalised mixed methods approaches. The purely quantitative studies in this thesis have yet to be validated in this way. Having established measures of constructs which are statistically speaking comparable between countries, it is the task of PUS researchers to comment on the extent to which their substantive interpretation is comparable.

Countries are not the only units between which comparisons might be made. Indeed, the analyses in this thesis are open to the charge of methodological nationalism (Beck, 2003) – that is, the tendency in the social sciences to unquestioningly adopt the nation state as a natural unit of analysis. There are other notable points of comparison in the literature, both below and above the national level. Some are socio-demographic – for example, in their methodological analysis of the Oxford scale of scientific knowledge, Pardo and Calvo (2004) analyse its metric properties not only between countries but also for groups of different ages and different levels of education. Some are attitudinal – Hviid Nielsen, Jelsoe, and Ohman (2002) compare logics of resistance towards biotechnology, characterising 'blue' resistance based on traditional arguments and values, moral and ethical concerns, and 'green' resistance based on a cost-benefit analysis of uses versus risks. A common comparison is also to compare regions: southern versus northern European countries, for example; or as in Miller's analyses, Europe versus North America, versus Japan.

There are a number of reasons for explicitly exploring cross-national comparisons within Europe. One is substantive, and simply relates to the fact that for many science actors it is important to know how the climate of opinion varies from one polity to another. PUS research is strongly tied to a political context in which the nation state is an unignorable unit of analysis. Methodologically, motivations for cross-national comparisons include concerns over systematically varying meanings corresponding to the different languages in which the surveys are administered, as well as systematically varying measurement errors due to survey administration, which is organised on a country-by-country basis. But country comparisons are not the only contrasts that could be made. Theory might suggest others, and empirical results might suggest still others. Had I divided the EU sample into age bands rather than countries in Chapter 5, for example, I may not have found similar measurement models for logics of support and opposition between groups. Indeed, in their meta-analysis of the relationship between knowledge and attitudes, Allum et al. (forthcoming) find very little variance of the model attributable to country differences. This is not to say that cross-national comparisons are not important. Rather, successfully finding joint cross-country models of these constructs might give us licence to be a little more confident and creative in making contrasts between groups other than countries. And this could mean going altogether beyond comparisons of taxonomic groups, defined by socio-demographic variables, and exploring 'natural groups', defined by the attitudes, values or representations people share (Gaskell, 1994) – perhaps those defined by the latent class models in this report would be a good starting point for such analyses.

## 9.7 The value of these models in attitudinal survey research

For many years, the attitude was the defining concept in social psychology, and measuring it was the main methodological pursuit. Chapter 3 gave a brief account of some of the careful, rigorous techniques developed for this purpose by Thurstone, Guttman, and Likert – methods which represent only a portion of the great intellectual output in this field. Based, as they were, on an individualistic philosophy, they lost favour with a significant section of the social psychology community. These more interpretivist-minded psychologists developed important critiques of the individualism of the classic model of the attitude and the limitations of quantitative research. The discipline has thus diversified, and the 'quantitative imperative' (Michell, 2003), that is the insistence that the only valid research is quantitative research, has lots its hegemony. This can only be for the good; quantitative research methods cannot answer all social research questions, and in Chapter 2 I described plainly some of the limitations of survey research. However, there remain *some* research questions that we can *only* answer with probability sample surveys – specifically, those that ask about the distributions of public opinions and perceptions. Even then, surveys should be used in conjunction with other, qualitative methods, to help us ask meaningful questions and interpret them sensitively, to obtain the best possible answers to those questions. But the survey is a crucial tool in this process, and although it is by no means a perfect research instrument, there is always scope for improving it. However, many of the measures used in today's social surveys are designed and tested in a somewhat ad hoc way, with resource constraints keeping them far removed from the rigour of Thurstone, Guttman and Likert.

In this contemporary context of the need for survey data which meets the information requirements of many users, and which can be produced in a limited time frame, I hope to have given some examples of how the use of latent trait and class models can aid and improve that task. These models can provide extensive, detailed information about the associations between a set of items, with a more collective than individualistic angle of interpretation. They can help uncover response effects in the data, identify items that do not work well for our purposes, and thus inform modifications to future survey design, with consequent efficiency gains. They also constitute a valuable, sensitive tool for assessing the cross-national comparability of the measures created to represent key psychological constructs.

Latent variable models can therefore be a very useful part of, or tool in, social research. They are far from a panacea, and the more critically they can be used, the better. It is unfortunate that the methodology of attitude measurement has become such a specialist area of expertise in contemporary survey research. Too often, critiques of survey methods take the form of a wholesale rejection of the method rather than any constructive proposals for its improvement. Perhaps the nature of the subject matter goes some way towards explaining the motivation for this. In the context of developing factor analysis for the measurement of IQ, and IRT models for educational testing, very real social consequences followed, and continue to follow, from the measures created: performance in school tests directly impacts on a student's life chances. With attitudes and related constructs, the consequences of mismeasurement are less clear, and less immediate. However, drawing attention back to the first section of Chapter 2, public opinion can be of real consequence in relation to some topics. This is surely true in the case of biotechnology; a subject which does not easily lend itself to opinion polls asking referendum-style questions. Notwithstanding the limitations of survey research, it seems fair to say that there will always be a need for sensitive and rigorous methods to capture attitudes and related constructs with survey data, and always, therefore, a motivation for improving such methods.

# Appendix

## Table A.1 Standardised residuals conditional on country for logics items, GM food

| Model | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | % 2-way standardised marginal residuals >4, conditional on country | | | | | | | | | | | | | | |
| **GM food, 5 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 23.3 | 16.7 | 24.7 | 34.7 | 26.0 | 41.3 | 30.7 | 34.7 | 23.3 | 18.0 | 28.0 | 21.3 | 34.0 | 15.3 | 20.0 |
| Measurement model equal | 12.7 | 9.3 | 14.7 | 20.0 | 8.7 | 26.0 | 15.3 | 8.7 | 15.3 | 8.7 | 16.0 | 8.0 | 12.0 | 10.0 | 6.0 |
| Direct country effect for *risk* | 10.0 | 10.0 | 13.3 | 12.7 | 6.0 | 24.0 | 10.0 | 10.0 | 12.0 | 4.0 | 12.0 | 8.0 | 8.7 | 5.3 | 3.3 |
| Interaction between country, latent variable and *risk* | 4.7 | 2.7 | 10.0 | 6.0 | 4.7 | 13.3 | 6.0 | 5.3 | 10.0 | 0.0 | 4.7 | 5.3 | 5.3 | 4.7 | 5.3 |
| Direct country effect for *utility* | 12.7 | 10.7 | 10.0 | 16.0 | 7.3 | 24.7 | 11.3 | 8.0 | 14.0 | 8.7 | 12.0 | 8.0 | 9.3 | 10.7 | 5.3 |
| Interaction between country, latent variable and *utility* | 9.3 | 10.0 | 8.7 | 15.3 | 6.0 | 20.0 | 10.7 | 8.0 | 11.3 | 10.7 | 13.3 | 5.3 | 7.3 | 11.3 | 4.0 |
| Measurement model free to differ between countries | 0.0 | 0.0 | 0.7 | 0.0 | 0.7 | 0.7 | 0.0 | 1.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 | 0.7 | 0.7 |
| **GM food, 6 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 24.7 | 17.3 | 24.0 | 35.3 | 26.0 | 42.7 | 30.0 | 36.0 | 27.3 | 20.7 | 27.3 | 23.3 | 34.7 | 12.7 | 20.0 |
| Measurement model equal | 10.0 | 10.0 | 12.0 | 17.3 | 7.3 | 28.7 | 14.0 | 11.3 | 12.0 | 8.7 | 8.0 | 7.3 | 6.0 | 10.7 | 4.7 |
| Direct country effect for *risk* | 8.7 | 9.3 | 10.7 | 12.7 | 3.3 | 24.0 | 10.0 | 12.0 | 8.7 | 4.0 | 3.3 | 7.3 | 2.7 | 5.3 | 1.3 |
| Interaction between country, latent variable and *risk* | 4.7 | 2.7 | 9.3 | 6.0 | 4.0 | 15.3 | 6.7 | 7.3 | 7.3 | 0.0 | 2.0 | 5.3 | 1.3 | 4.7 | 1.3 |
| Direct country effect for *utility* | 11.3 | 9.3 | 8.0 | 15.3 | 7.3 | 26.0 | 14.7 | 10.0 | 8.7 | 9.3 | 6.7 | 6.0 | 3.3 | 9.3 | 3.3 |
| Interaction between country, latent variable and *utility* | 8.0 | 9.3 | 5.3 | 14.7 | 8.0 | 22.7 | 10.0 | 8.7 | 8.0 | 9.3 | 6.7 | 5.3 | 4.0 | 8.7 | 2.7 |
| Measurement model free to differ between countries | 0.0 | 0.0 | 0.7 | 0.0 | 0.7 | 0.7 | 0.0 | 0.7 | 0.0 | 0.7 | 0.0 | 0.0 | 1.3 | 0.0 | 0.7 |

**Table A.2 Standardised residuals conditional on country for logics items, therapeutic cloning**

| Model | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | % 2-way standardised marginal residuals >4, conditional on country | | | | | | | | | | | | | | |
| **Therapeutic cloning, 5 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 25 | 21 | 29 | 29 | 14 | 42 | 33 | 34 | 15 | 15 | 33 | 21 | 29 | 27 | 17 |
| Measurement model equal | 14.7 | 12.7 | 20.0 | 20.7 | 12.7 | 30.7 | 26.7 | 9.3 | 13.3 | 14.0 | 22.7 | 16.0 | 12.0 | 22.7 | 11.3 |
| Direct country effect for *risk* | 13.3 | 6.7 | 16.0 | 16.7 | 7.3 | 31.3 | 22.0 | 6.7 | 9.3 | 10.7 | 16.7 | 13.3 | 8.7 | 12.7 | 6.7 |
| Interaction between country, latent variable and *risk* | 11.3 | 1.3 | 10.7 | 3.3 | 2.0 | 18.7 | 20.7 | 3.3 | 6.0 | 5.3 | 11.3 | 8.0 | 6.0 | 6.7 | 5.3 |
| Direct country effect for *utility* | 12.7 | 14.7 | 12.0 | 18.7 | 11.3 | 27.3 | 20.0 | 9.3 | 11.3 | 14.0 | 20.0 | 11.3 | 10.0 | 20.7 | 8.7 |
| Interaction between country, latent variable and *utility* | 10.0 | 13.3 | 12.0 | 18.7 | 12.0 | 20.7 | 16.0 | 8.7 | 8.7 | 11.3 | 17.3 | 10.0 | 10.0 | 18.7 | 7.3 |
| Measurement model free to differ between countries | 0.0 | 0.0 | 0.0 | 0.7 | 0.7 | 2.0 | 0.7 | 6.0 | 0.7 | 0.7 | 0.0 | 0.0 | 0.0 | 0.7 | 0.0 |
| **Therapeutic cloning, 6 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 27.3 | 20.0 | 26.7 | 29.3 | 14.7 | 48.7 | 34.7 | 34.0 | 19.3 | 15.3 | 32.7 | 20.7 | 28.0 | 23.3 | 16.7 |
| Measurement model equal | 12.7 | 13.3 | 18.7 | 21.3 | 12.7 | 33.3 | 20.0 | 8.7 | 14.0 | 16.0 | 14.0 | 16.7 | 10.7 | 16.7 | 10.7 |
| Direct country effect for *risk* | 12.7 | 7.3 | 14.0 | 16.0 | 6.7 | 28.7 | 15.3 | 8.0 | 8.7 | 9.3 | 6.7 | 14.0 | 4.0 | 9.3 | 6.0 |
| Interaction between country, latent variable and *risk* | 10.7 | 2.0 | 10.0 | 3.3 | 2.1 | 18.0 | 11.3 | 3.3 | 5.3 | 4.7 | 4.7 | 8.7 | 2.7 | 4.7 | 4.7 |
| Direct country effect for *utility* | 11.3 | 14.7 | 10.7 | 18.0 | 10.0 | 24.0 | 17.3 | 8.7 | 10.7 | 16.0 | 14.7 | 12.0 | 9.3 | 16.0 | 6.7 |
| Interaction between country, latent variable and *utility* | 6.0 | 14.0 | 11.3 | 19.3 | 10.0 | 19.3 | 11.3 | 7.3 | 9.3 | 12.7 | 12.7 | 10.7 | 9.3 | 13.3 | 6.0 |
| Measurement model free to differ between countries | 0.0 | 0.0 | 0.0 | 0.0 | 0.7 | 0.0 | 0.0 | 0.0 | 1.3 | 0.7 | 0.0 | 0.7 | 0.7 | 0.0 | 0.0 |

**Table A.3 Jöreskog & Moustaki index conditional on country for logics items, GM food**

| | Jöreskog & Moustaki index, conditional on country | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Model | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
| **GM food, 5 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 3.68 | 2.00 | 6.06 | 6.30 | 3.75 | 7.33 | 7.16 | 7.28 | 3.11 | 4.01 | 7.31 | 2.60 | 4.48 | 2.06 | 2.00 |
| Measurement model equal | 1.93 | 1.76 | 2.33 | 3.37 | 1.49 | 2.91 | 2.30 | 1.62 | 2.21 | 1.80 | 1.83 | 1.42 | 1.82 | 1.66 | 1.06 |
| Direct country effect for *risk* | 1.62 | 1.40 | 2.14 | 1.81 | 1.17 | 2.29 | 2.07 | 1.49 | 1.88 | 1.10 | 1.46 | 1.41 | 1.41 | 1.24 | 0.81 |
| Interaction between country, latent variable and *risk* | 1.05 | 0.91 | 1.89 | 0.99 | 0.98 | 1.55 | 1.65 | 0.98 | 1.36 | 0.57 | 0.99 | 1.04 | 1.20 | 0.97 | 0.73 |
| Direct country effect for *utility* | 1.89 | 1.49 | 1.52 | 3.16 | 1.39 | 2.77 | 1.62 | 1.52 | 2.04 | 1.69 | 1.68 | 1.28 | 1.45 | 1.53 | 0.96 |
| Interaction between country, latent variable and *utility* | 1.56 | 1.42 | 1.24 | 3.07 | 1.26 | 2.43 | 1.54 | 1.32 | 1.70 | 1.65 | 1.49 | 1.06 | 1.28 | 1.32 | 0.94 |
| Measurement model free to differ between countries | 0.28 | 0.23 | 0.34 | 0.33 | 0.30 | 0.37 | 0.25 | 0.43 | 0.26 | 0.26 | 0.25 | 0.28 | 0.39 | 0.37 | 0.27 |
| **GM food, 6 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 3.60 | 2.00 | 6.05 | 6.25 | 3.61 | 7.57 | 7.16 | 7.11 | 3.46 | 3.97 | 7.00 | 2.80 | 4.64 | 2.02 | 1.97 |
| Measurement model equal | 1.82 | 1.74 | 2.09 | 3.25 | 1.26 | 3.01 | 2.36 | 1.63 | 1.80 | 1.78 | 1.21 | 1.29 | 1.34 | 1.51 | 0.99 |
| Direct country effect for *risk* | 1.52 | 1.39 | 1.91 | 1.79 | 0.92 | 2.39 | 2.14 | 1.47 | 1.54 | 1.06 | 0.79 | 1.30 | 0.95 | 1.11 | 0.74 |
| Interaction between country, latent variable and *risk* | 1.04 | 0.81 | 1.67 | 1.00 | 0.84 | 1.66 | 1.69 | 0.97 | 1.38 | 0.55 | 0.61 | 0.99 | 0.72 | 0.82 | 0.63 |
| Direct country effect for *utility* | 1.79 | 1.43 | 1.32 | 3.05 | 1.21 | 2.81 | 1.70 | 1.51 | 1.46 | 1.72 | 1.13 | 1.09 | 1.15 | 1.37 | 0.89 |
| Interaction between country, latent variable and *utility* | 1.52 | 1.34 | 1.12 | 2.92 | 1.16 | 2.50 | 1.50 | 1.31 | 1.34 | 1.65 | 1.06 | 1.00 | 1.08 | 1.30 | 0.84 |
| Measurement model free to differ between countries | 0.23 | 0.26 | 0.28 | 0.30 | 0.25 | 0.30 | 0.20 | 0.33 | 0.24 | 0.28 | 0.24 | 0.23 | 0.29 | 0.27 | 0.23 |

**Table A.4 Jöreskog & Moustaki index conditional on country for logics items, therapeutic cloning**

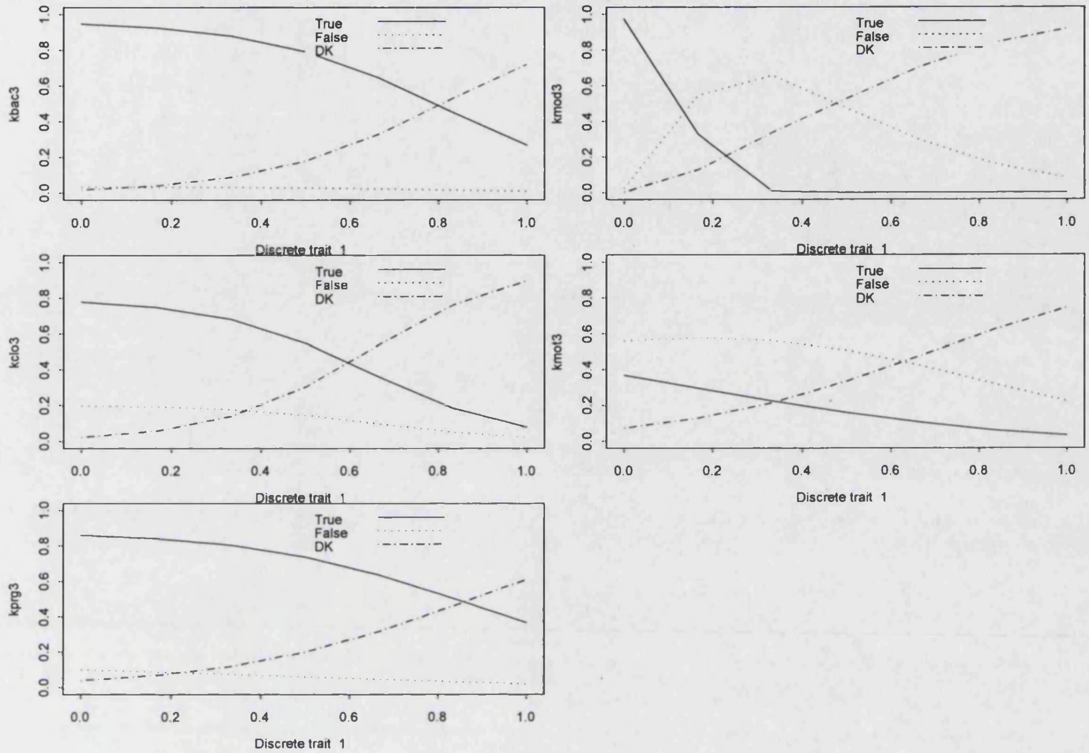
| Model | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Jöreskog & Moustaki index, conditional on country | | | | | | | | | | | | | | |
| **Therapeutic cloning, 5 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 3.86 | 2.29 | 5.27 | 6.75 | 2.50 | 9.98 | 5.91 | 9.97 | 2.09 | 2.95 | 11.24 | 2.40 | 3.13 | 3.91 | 2.34 |
| Measurement model equal | 2.14 | 1.82 | 2.56 | 5.48 | 2.13 | 4.54 | 4.44 | 1.62 | 1.77 | 2.41 | 2.92 | 2.03 | 1.71 | 3.00 | 1.60 |
| Direct country effect for *risk* | 1.95 | 1.36 | 1.96 | 2.29 | 1.15 | 3.43 | 3.88 | 1.30 | 1.52 | 1.72 | 2.11 | 1.81 | 1.46 | 1.63 | 1.29 |
| Interaction between country, latent variable and *risk* | 1.68 | 0.84 | 1.64 | 1.07 | 0.73 | 2.47 | 3.82 | 0.75 | 0.89 | 1.23 | 1.44 | 1.17 | 1.14 | 1.04 | 0.98 |
| Direct country effect for *utility* | 1.94 | 1.80 | 1.64 | 4.93 | 2.06 | 3.79 | 3.14 | 1.58 | 1.60 | 2.37 | 2.66 | 1.84 | 1.46 | 2.83 | 1.28 |
| Interaction between country, latent variable and *utility* | 1.50 | 1.69 | 1.53 | 4.89 | 2.03 | 3.18 | 2.48 | 1.48 | 1.39 | 1.95 | 2.27 | 1.71 | 1.36 | 2.65 | 1.16 |
| Measurement model free to differ between countries | 0.46 | 0.36 | 0.31 | 0.37 | 0.39 | 0.49 | 0.55 | 0.90 | 0.45 | 0.42 | 0.24 | 0.38 | 0.40 | 0.34 | 0.29 |
| **Therapeutic cloning, 6 classes** | | | | | | | | | | | | | | | |
| Country independent of latent variable and indicators | 3.86 | 2.32 | 5.30 | 6.68 | 2.56 | 10.00 | 6.07 | 9.51 | 2.27 | 2.95 | 10.67 | 2.43 | 3.12 | 3.79 | 2.33 |
| Measurement model equal | 2.10 | 1.81 | 2.58 | 5.26 | 2.15 | 4.33 | 2.87 | 1.57 | 1.77 | 2.42 | 2.03 | 2.07 | 1.42 | 2.52 | 1.55 |
| Direct country effect for *risk* | 1.94 | 1.32 | 1.98 | 2.19 | 1.17 | 3.23 | 2.40 | 1.29 | 1.49 | 1.84 | 1.18 | 1.84 | 1.15 | 1.23 | 1.24 |
| Interaction between country, latent variable and *risk* | 1.72 | 0.76 | 1.60 | 0.99 | 0.72 | 2.29 | 2.14 | 0.74 | 0.89 | 1.30 | 0.93 | 1.22 | 0.83 | 0.86 | 0.92 |
| Direct country effect for *utility* | 1.87 | 1.79 | 1.64 | 4.71 | 2.07 | 3.59 | 2.45 | 1.51 | 1.52 | 3.05 | 2.00 | 1.86 | 1.24 | 2.47 | 1.18 |
| Interaction between country, latent variable and *utility* | 1.39 | 1.70 | 1.49 | 4.68 | 2.01 | 3.02 | 2.10 | 1.40 | 1.41 | 1.93 | 1.81 | 1.70 | 1.16 | 2.38 | 1.08 |
| Measurement model free to differ between countries | 0.43 | 0.35 | 0.28 | 0.30 | 0.28 | 0.34 | 0.35 | 0.25 | 0.38 | 0.30 | 0.18 | 0.29 | 0.40 | 0.26 | 0.23 |

**Table A.5 Parameter estimates for a discrete latent trait model of knowledge items for British data; 10 polytomous items, 2 traits**

| | | Intercept | | DK trait | | 'Knowledge' trait | |
|---|---|---|---|---|---|---|---|
| Item | Category | Estimate | Standard error | Estimate | Standard error | Estimate | Standard error |
| | True | 0 | | 0 | | 0 | |
| kbac3 | False | -2.27 | 0.62 | -2.28 | 1.09 | -0.17 | 0.66 |
| | DK | -6.81 | 1.02 | 7.79 | 1.30 | -1.60 | 0.79 |
| | True | 0 | | 0 | | 0 | |
| ktom3 | False | -1.76 | 0.60 | 2.47 | 0.97 | 2.24 | 0.54 |
| | DK | -4.30 | 1.18 | 9.15 | 1.79 | -1.04 | 0.66 |
| | True | 0 | | 0 | | 0 | |
| kclo3 | False | -0.72 | 0.42 | -2.66 | 0.73 | 0.13 | 0.40 |
| | DK | -5.53 | 1.03 | 7.66 | 1.41 | -5.09 | 2.70 |
| | True | 0 | | 0 | | 0 | |
| kmod3 | False | -0.85 | 0.77 | 2.71 | 1.27 | 2.93 | 0.85 |
| | DK | -4.12 | 0.94 | 9.04 | 1.42 | -0.98 | 1.13 |
| | True | 0 | | 0 | | 0 | |
| kmot3 | False | -0.40 | 0.43 | 2.70 | 0.76 | 0.96 | 0.41 |
| | DK | -4.09 | 0.86 | 8.15 | 1.24 | -0.56 | 0.62 |
| | True | 0 | | 0 | | 0 | |
| kyea3 | False | -0.98 | 0.51 | -3.44 | 0.94 | -0.41 | 0.60 |
| | DK | -6.08 | 1.14 | 7.88 | 1.60 | -2.23 | 0.82 |
| | True | 0 | | 0 | | 0 | |
| kprg3 | False | -1.55 | 0.48 | -1.81 | 0.83 | -0.49 | 0.53 |
| | DK | -6.20 | 0.90 | 6.62 | 1.15 | -0.17 | 0.51 |
| | True | 0 | | 0 | | 0 | |
| kbig3 | False | -2.02 | 0.70 | 2.63 | 1.10 | 2.81 | 0.74 |
| | DK | -4.39 | 1.03 | 8.82 | 1.49 | -0.53 | 0.74 |
| | True | 0 | | 0 | | 0 | |
| kchim3 | False | -1.36 | 0.42 | -0.68 | 0.75 | 0.40 | 0.34 |
| | DK | -5.32 | 1.03 | 8.83 | 1.46 | -1.13 | 0.49 |
| | True | 0 | | 0 | | 0 | |
| ktrgen3 | False | -0.15 | 0.36 | -0.30 | 0.65 | 0.72 | 0.29 |
| | DK | -4.25 | 1.17 | 8.85 | 1.71 | -0.68 | 0.46 |

**Figure A.1** **Item characteristic curves for the two 'non-knowledge' traits in the final 3-trait model of 'knowledge', with measurement models equal for all traits, 15 countries**

**(a) Trait 1**




**(b) Trait 3**

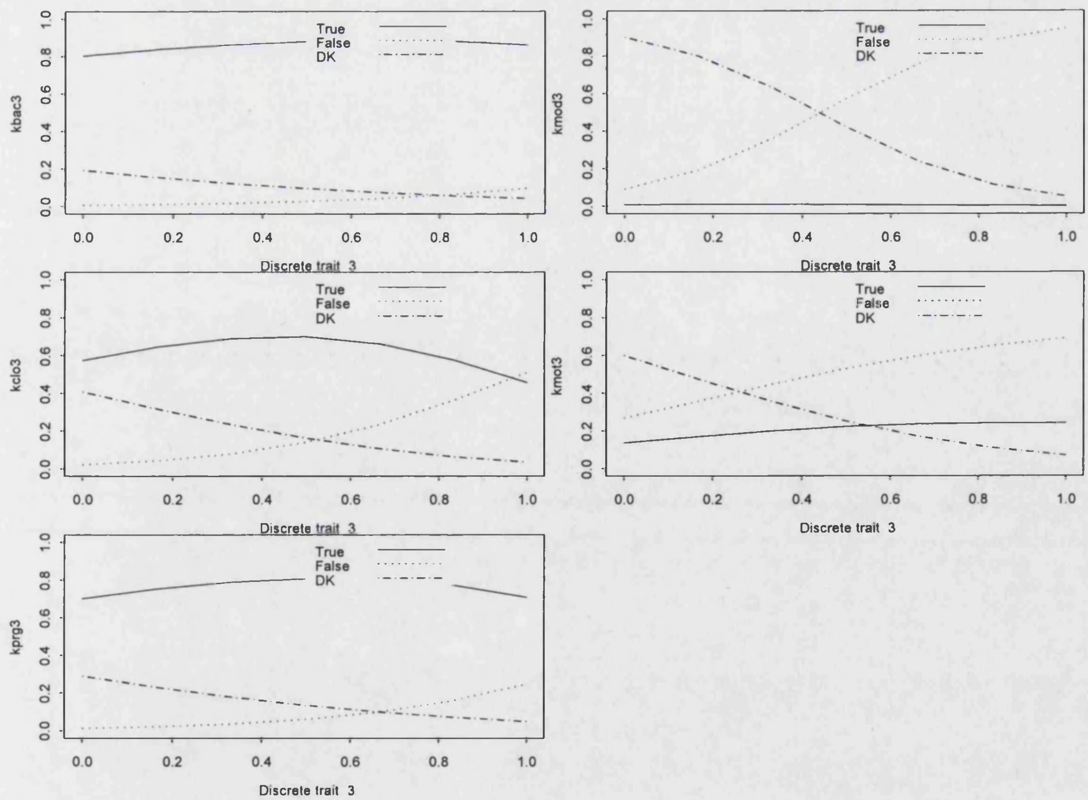


**Table A.6 Standardised residuals conditional on country for engagement with science and technology**

| Model | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | % 2-way standardised marginal residuals >4, conditional on country | | | | | | | | | | | | | | |
| **Measurement model free to differ between countries** | | | | | | | | | | | | | | | |
| 4 items (*scint, scinf, scund, sconf*), 3 classes | 1.9 | 5.6 | 1.9 | 11.1 | 3.7 | 5.6 | 1.9 | 5.6 | 3.7 | 1.9 | 7.4 | 0.0 | 1.9 | 3.7 | 5.6 |
| 3 items (*scint, scinf, scund*), 3 classes | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| **3 items, measurement model equal between countries** | | | | | | | | | | | | | | | |
| 3 classes | 19 | 33.3 | 37.0 | 40.7 | 37.0 | 51.9 | 55.6 | 25.9 | 29.6 | 18.5 | 48.1 | 29.6 | 25.9 | 66.7 | 22.2 |
| 4 classes | 14.8 | 18.5 | 25.9 | 33.3 | 25.9 | 37.0 | 40.7 | 22.2 | 33.3 | 7.4 | 44.4 | 29.6 | 18.5 | 59.3 | 11.1 |
| 5 classes | 14.8 | 14.8 | 11.1 | 25.9 | 22.2 | 11.1 | 22.2 | 14.8 | 29.6 | 7.4 | 33.3 | 11.1 | 18.5 | 33.3 | 18.5 |
| 6 classes | 3.7 | 18.5 | 3.7 | 29.6 | 11.1 | 14.8 | 22.2 | 11.1 | 14.8 | 7.4 | 7.4 | 3.7 | 3.7 | 7.4 | 11.1 |
| 7 classes | 7.4 | 14.8 | 3.7 | 7.4 | 0.0 | 3.7 | 3.7 | 11.1 | 14.8 | 3.7 | 7.4 | 0.0 | 3.7 | 0.0 | 7.4 |
| 8 classes | 14.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.7 | 7.4 | 0.0 | 0.0 | 0.0 | 3.7 | 0.0 | 0.0 | 7.4 |

**Table A.7 Jöreskog & Moustaki index conditional on country for engagement with science and technology**

| | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Jöreskog & Moustaki index, conditional on country | | | | | | | | | | | | | | |
| **Measurement model free to differ between countries** | | | | | | | | | | | | | | | |
| 4 items (*scint, scinf, scund, sconf*), 3 classes | 0.45 | 0.52 | 0.59 | 1.00 | 0.42 | 0.82 | 0.44 | 0.71 | 0.66 | 0.41 | 1.15 | 0.25 | 0.48 | 0.77 | 0.55 |
| 3 items (*scint, scinf, scund*), 3 classes | 0.09 | 0.09 | 0.11 | 0.21 | 0.08 | 0.35 | 0.14 | 0.22 | 0.33 | 0.07 | 0.13 | 0.04 | 0.08 | 0.24 | 0.05 |
| **3 items, measurement model equal between countries** | | | | | | | | | | | | | | | |
| 3 classes | 2.47 | 3.19 | 5.22 | 6.64 | 4.14 | 7.88 | 18.39 | 3.12 | 2.99 | 2.27 | 6.50 | 4.28 | 2.96 | 8.35 | 2.51 |
| 4 classes | 1.80 | 2.91 | 3.92 | 5.88 | 3.61 | 5.13 | 5.42 | 2.65 | 2.91 | 1.59 | 6.45 | 3.68 | 2.02 | 7.82 | 1.41 |
| 5 classes | 1.55 | 2.28 | 1.74 | 2.89 | 2.23 | 1.51 | 2.44 | 2.64 | 2.83 | 1.63 | 5.24 | 1.13 | 1.89 | 5.02 | 2.10 |
| 6 classes | 1.04 | 2.11 | 1.15 | 2.54 | 1.41 | 1.41 | 2.30 | 1.92 | 1.60 | 1.23 | 1.31 | 0.51 | 1.27 | 1.20 | 1.35 |
| 7 classes | 1.08 | 1.61 | 0.84 | 1.12 | 0.59 | 0.68 | 0.68 | 1.44 | 1.50 | 0.61 | 0.82 | 0.43 | 1.01 | 1.14 | 1.19 |
| 8 classes | 0.48 | 1.22 | 0.55 | 1.01 | 0.63 | 0.75 | 0.76 | 1.18 | 0.72 | 0.53 | 0.60 | 0.53 | 1.05 | 0.82 | 1.03 |

## Table A.8 Standardised residuals conditional on country for engagement with biotechnology

| Model | % 2-way standardised marginal residuals >4, conditional on country | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
| **Measurement model free to differ between countries** | | | | | | | | | | | | | | | |
| 4 classes | 0.0 | 0.0 | 0.0 | 0.0 | 2.7 | 2.7 | 0.0 | 0.0 | 2.7 | 0.0 | 0.0 | 0.0 | 5.4 | 0.0 | 2.7 |
| **Measurement model equal between countries** | | | | | | | | | | | | | | | |
| 4 classes | 35.1 | 27.0 | 54.1 | 24.3 | 29.7 | 45.9 | 45.9 | 27.0 | 35.1 | 29.7 | 24.3 | 24.3 | 45.9 | 59.5 | 40.5 |
| 5 classes | 32.4 | 24.3 | 37.8 | 18.9 | 18.9 | 32.4 | 13.5 | 13.5 | 21.6 | 24.3 | 5.4 | 27.0 | 32.4 | 43.2 | 35.1 |
| 6 classes | 37.8 | 13.5 | 35.1 | 10.8 | 18.9 | 18.9 | 2.7 | 8.1 | 2.7 | 18.9 | 0.0 | 16.2 | 37.8 | 40.5 | 32.4 |
| 7 classes | 27.0 | 13.5 | 32.4 | 0.0 | 18.9 | 5.4 | 0.0 | 8.1 | 10.8 | 21.6 | 0.0 | 0.0 | 18.9 | 21.6 | 13.5 |
| **6 classes, investigating interactions** | | | | | | | | | | | | | | | |
| Interaction between *talkbio* and latent variable | 35.1 | 5.4 | 5.4 | 10.8 | 5.4 | 10.8 | 0.0 | 10.8 | 0.0 | 5.4 | 0.0 | 10.8 | 27.0 | 35.1 | 8.1 |
| Interaction between *heardbio* and latent variable | 40.5 | 8.1 | 5.4 | 2.7 | 8.1 | 5.4 | 0.0 | 8.1 | 5.4 | 8.1 | 5.4 | 16.2 | 29.7 | 29.7 | 13.5 |
| Interaction between *discuss* and latent variable | 8.1 | 10.8 | 21.6 | 0.0 | 8.1 | 8.1 | 2.7 | 0.0 | 8.1 | 10.8 | 0.0 | 5.4 | 5.4 | 2.7 | 13.5 |
| Interaction between *readtv* and latent variable | 2.7 | 5.4 | 21.6 | 0.0 | 5.4 | 5.4 | 2.7 | 0.0 | 2.7 | 13.5 | 0.0 | 0.0 | 2.7 | 18.9 | 5.4 |

## Table A.9 Jöreskog & Moustaki index conditional on country for engagement with biotechnology

| | Jöreskog & Moustaki index, conditional on country | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Austria | Belgium | Denmark | Finland | France | Germany | Greece | Ireland | Italy | Luxembourg | Netherlands | Portugal | Spain | Sweden | UK |
| **Measurement model free to differ between countries** | | | | | | | | | | | | | | | |
| 4 classes | 0.10 | 0.18 | 0.15 | 0.38 | 0.22 | 0.60 | 0.06 | 0.17 | 1.60 | 0.12 | 0.07 | 0.24 | 0.41 | 0.04 | 1.90 |
| **Measurement model equal between countries** | | | | | | | | | | | | | | | |
| 4 classes | 5.61 | 3.69 | 7.96 | 2.99 | 3.83 | 6.20 | 6.46 | 3.59 | 8.41 | 3.84 | 2.72 | 3.44 | 6.37 | 12.31 | 4.54 |
| 5 classes | 6.41 | 3.20 | 6.57 | 2.41 | 2.76 | 3.79 | 1.81 | 2.06 | 5.90 | 3.28 | 0.78 | 2.81 | 3.93 | 8.73 | 3.90 |
| 6 classes | 6.92 | 2.42 | 6.09 | 1.54 | 2.72 | 2.16 | 0.93 | 1.22 | 1.81 | 2.51 | 0.69 | 1.98 | 4.31 | 9.15 | 3.79 |
| 7 classes | 4.73 | 2.37 | 5.97 | 0.72 | 2.67 | 1.07 | 0.87 | 0.95 | 2.14 | 2.42 | 0.60 | 0.68 | 1.94 | 2.76 | 2.70 |
| **6 classes, investigating interactions** | | | | | | | | | | | | | | | |
| Interaction between *talkbio* and latent variable | 4.20 | 0.61 | 0.70 | 1.22 | 0.69 | 1.27 | 0.51 | 1.22 | 0.54 | 0.48 | 0.48 | 1.65 | 2.36 | 5.99 | 1.49 |
| Interaction between *heardbio* and latent variable | 5.82 | 1.01 | 0.76 | 0.92 | 0.90 | 0.74 | 0.26 | 1.10 | 0.77 | 0.54 | 0.77 | 2.25 | 2.59 | 4.60 | 1.58 |
| Interaction between *discuss* and latent variable | 2.53 | 1.56 | 4.97 | 0.17 | 1.90 | 1.22 | 0.57 | 0.35 | 1.91 | 1.51 | 0.39 | 0.57 | 0.95 | 1.83 | 2.29 |
| Interaction between *readtv* and latent variable | 0.96 | 1.31 | 5.25 | 0.25 | 1.18 | 0.77 | 0.58 | 0.22 | 1.44 | 1.80 | 0.17 | 0.40 | 0.62 | 2.75 | 1.70 |

# References


Abric, J. C. (1993). Central system, peripheral system: Their functions and roles in the dynamics of social representation. *Papers on Social Representations, 2*, 75-78.

Agresti, A., & Finlay, B. (1997). *Statistical Methods for the Social Sciences, Third Edition*. Upper Saddle River, NJ: Prentice Hall Inc.

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

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behaviour*. New York: Prentice Hall.

Allport, F. H. (1937). Toward a Science of Public Opinion. *Public Opinion Quarterly, 1*(1), 7-23.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798-844). Worcester, MA: Clark University Press.

Allum, N. (2002). *Risk, Social Trust and Knowledge: Public Perceptions of Gene Technology*. London School of Economics & Political Science, London.

Allum, N., Sturgis, P., Tabourazi, D., & Brunton-Smith, I. (forthcoming). Science knowledge and attitudes across cultures: a meta-analysis. *Public Understanding of Science*.

Almond, G. A. (1950). *The American People and Foreign Policy*. New York: Harcourt, Brace and Company.

Arts, W., & Halman, L. (2002). European Values at the Turn of the Millennium: an Introduction. In W. Arts & L. Halman (Eds.), *European Values at the Turn of the Millennium*. Leiden: Brill.

Barrett, P. (2002). Beyond Psychometrics: Measurement, non-quantitative structure, and applied numerics. *Journal of Managerial Psychology, (augmented Web version of paper)*.

Bartholomew, D. J., & Knott, M. (1999). *Latent Variable Models and Factor Analysis*. London: Arnold.

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2002). *The analysis and interpretation of multivariate data for social scientists*. London: Chapman & Hall/CRC.

Bauer, M. W. (1996). Socio-demographic correlates of DK-responses in knowledge surveys: self-attributed ignorance of science. *Social Science Information, 35*(1), 39-68.

Bauer, M. W. (2000). 'Science in the Media' as a Cultural Indicator: Contextualizing Surveys with Media Analysis. In M. Dierkes & C. von Grote (Eds.), *Between Understanding and Trust - The Public, Science and Technology*. Amsterdam: Harwood Academic Publishers.

Bauer, M. W. (2003). *The vicissitudes of 'public understanding of science': from 'literacy' to 'science in society'*. Paper presented at the *Science meets Society* conference, Lisbon.

Bauer, M. W. (2005). Distinguishing *Red* and *Green* Biotechnology: Cultivation Effects of the Elite Press. *International Journal of Public Opinion Research, 17*(1), 63-89.

Bauer, M. W., Allum, N., & Miller, S. (2007). What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understanding of Science, 16*, 79-95.

Bauer, M. W., Durant, J., & Evans, G. (1994). European Public Perceptions of Science. *International Journal of Public Opinion Research, 6*(2), 163-186.

Bauer, M. W., & Gaskell, G. (2002). Researching the public sphere of biotechnology. In M. W. Bauer & G. Gaskell (Eds.), *Biotechnology - the Making of a Global Controversy*. Cambridge: Cambridge University Press.

Bauer, M. W., & Joffe, H. (1996). Meaning of self-attributed ignorance: an introduction to the symposium. *Social Science Information, 35*(1), 5-13.

Bauer, M. W., Petkova, K., & Boyadjieva, P. (2000). Public Knowledge of and Attitudes to Science: Alternative Measures That May End the "Science War". *Science, Technology, & Human Values, 25*(1, Winter), 30-51.

Beck, U. (2003). The Analysis of Global Inequality: From National to Cosmopolitan Perspective. In M. Kaldor, H. Anheier & M. Glasius (Eds.), *Global Civil Society 2003*. Oxford: Oxford University Press.

Berg, P., Baltimore, D., Boyer, H. W., Cohen, S. N., Davis, R. W., Hogness, D. S., et al. (1974). Potential Biohazards of Recombinant DNA Molecules. *Science, 185*(4148), 303.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord and N.R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Bishop, G. F. (1990). Issue Involvement and Response Effects in Public Opinion Surveys. *Public Opinion Quarterly, 54*, 209-218.

Bodmer, W. (1985). *The Public Understanding of Science*. London: Royal Society.

Bolck, A., Croon, M., & Hagenaars, J. A. (2004). Estimating Latent Structure Models with Categorical Variables: One-Step Versus Three-Step Estimators. *Political Analysis, 12*(3-27).

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203-219.

Bridgman, P. W. (1927). *The Logic of Modern Physics.* New York: Macmillan.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research.* Boston: Houghton Mifflin Company.

Chambers, R. L., & Skinner, C. J. (2003). *Analysis of Survey Data.* Chichester: Wiley.

Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and Discontent.* New York: Free Press.

Converse, P. E. (2000). Assessing the capacity of mass electorates. *Annual Review of Political Science*(3), 331-353.

Coombs, C. H. (1964). *A Theory of Data.* New York: Wiley.

Coombs, C. H., & Coombs, L. C. (1976-1977). 'Don't know': item ambiguity or respondent uncertainty? *Public Opinion Quarterly, 40*(4), 497-514.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent Variable and Latent Structure Models.* Mahwah, NJ: Lawrence Erlbaum.

Danziger, K. (1979). The positivist repudiation of Wundt. *Journal of the History of the Behavioural Sciences, 16,* 241-262.

Dawes, R. M. (1972). *Fundamentals of Attitude Measurement.* New York: Wiley.

Dawes, R. M., & Smith, T. L. (1985). Attitude and opinion measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology* (Vol. 1). New York: Random House.

de Vaus, D. (Ed.). (2006). *Research Design.* London: Sage.

Deane, J. (2004). Government to give go ahead for GM maize. *Independent,* p. 19 February (web access).

Dillman, D. A. (2000). *Mail and Internet surveys: the tailored design method.* New York: John Wiley & Sons.

Dobson, A. J. (2002). *An introduction to generalized linear models.* USA: Chapman & Hall/CRC.

Doise, W., Clemence, A., & Lorenzi-Cioldi, F. (1993). *The quantitative analysis of social representations.* New York: Harvester-Wheatsheaf.

du Toit, M. (Ed.). (2003). *IRT from SSI.* Lincolnwood, IL: Scientific Software International.

Durant, J., Bauer, M. W., Gaskell, G., Midden, C., Liakopoulos, M., & Scholten, L. (2000). Two Cultures of Public Understanding of Science and Technology in Europe. In M. Dierkes & C. von Grote (Eds.), *Between Understanding and Trust: the Public, Science and Technology.* Amsterdam: Harwood Academic Publishers.

Durant, J., Evans, G., & Thomas, G. (1989). The public understanding of science. *Nature, 340*, 11-14.

Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes.* Fort Worth: Harcourt Brace Jovanovich College Publishers.

Edwards, D. (1995). *Introduction to Graphical Modelling.* New York: Springer-Verlag.

Eiser, J. R. (2001). Attitudes, decisions and perceptions of risk: a social psychological analysis. In G. Bohm, J. Nerb, T. McDaniels & H. Spada (Eds.), *Environmental risks: perception, evaluation and management.* Oxford: Elsevier Science Ltd.

Elder, J. W. (1976). Comparative Cross-National Methodology. *Annual Review of Sociology, 2*, 209-230.

Evans, G., & Durant, J. (1995). The relationship between knowledge and attitudes in the public understanding of science in Britain. *Public Understanding of Science, 4*, 57-74.

Eysenck, H. J. (1954). *The Psychology of Politics.* London: Routledge.

Farr, R. M. (1993). Common sense, science and social representations. *Public Understanding of Science, 2*, 189-204.

Farr, R. M. (1996). *The Roots of Modern Social Psychology.* Oxford: Blackwell.

Felt, U. (2000). Why Should the Public "Understand" Science? A Historical Perspective on Aspects of the Public Understanding of Science. In M. Dierkes & C. von Grote (Eds.), *Between Understanding and Trust: the Public, Science and Technology.* Amsterdam: Harwood Academic Publishers.

Ferguson, C. (2003). *SEMNET archives.* Retrieved June, 2003, from http://bama.ua.edu/archives/semnet.html

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Fishbein, M., & Ajzen, I. (1972). *Beliefs, attitudes, intentions and behaviour: an introduction to theory and research.* Reading, MA: Addison-Wesley.

Garfinkel, H. (1967). *Studies in Ethnomethodology.* Englewood Cliffs, NJ: Prentice-Hall.

Gaskell, G. (1994). Survey research and consensuality: Statistical and natural groups. *Papers on Social Representations, 3*, 29-32.

Gaskell, G. (2001). Attitudes, social representations and beyond. In K. Deaux & G. Philogene (Eds.), *Representations of the social* (pp. 228-241). Oxford: Blackwell.

Gaskell, G., Allum, N., Bauer, M. W., Durant, J., Allansdottir, A., Bonfadelli, H., et al. (2000). Biotechnology and the European Public. *Nature Biotechnology, 18*(September), 935-938.

Gaskell, G., Allum, N., & Stares, S. (2003). *Europeans and Biotechnology in 2002 (Eurobarometer 58.0)*: Report to the EC Directorate General for Research from the project 'Life Sciences in European Society'.

Gaskell, G., Allum, N., Wagner, W., Kronberger, N., Torgersen, H., Hampel, J., et al. (2004). GM Foods and the Misperception of Risk Perception. *Risk Analysis, 24*(1), 185-194.

Gaskell, G., & Bauer, M. W. (2000). Towards Public Accountability: beyond Sampling, Reliability and Validity. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative Researching with Text, Image and Sound*. London: Sage.

Gaskell, G., Stares, S., Allansdottir, A., Allum, N., Corchero, C., Fischler, C., et al. (2006). *Europeans and Biotechnology in 2005: Patterns and Trends (Final report on Eurobarometer 64.3)*: A report to the European Commission's Directorate-General for Research.

Gaskell, G., Thompston, P., & Allum, N. (2002). Worlds apart? Public opinion in Europe and the USA. In M. W. Bauer & G. Gaskell (Eds.), *Biotechnology - The Making of a Global Controversy*. Cambridge: Cambridge University Press.

Gaskell, G., Wright, D., & O'Muircheartaigh, C. A. (1993). Measuring scientific interest: the effect of knowledge questions on interest ratings. *Public Understanding of Science, 2*, 39-57.

Gigerenzer, G. (1991). From Tools to Theories: A Heuristic of Discovery in Cognitive Psychology. *Psychological Review, 98*(2), 254-267.

Gill, S. (1947). How do you stand on sin? *Tide, 14*(March), 72.

Gould, S. J. (1981/1996). *The Mismeasure of Man*. London: Penguin.

Grabner, P., Hampel, J., Lindsey, N., & Torgersen, H. (2001). Biopolitical diversity: the challenge of multilevel policy-making. In G. Gaskell & M. W. Bauer (Eds.), *Biotechnology 1996-2000: the years of controversy*. London: Science Museum.

Gregory, J., & Miller, S. (1998). *Science in Public*. Cambridge, MA: Perseus Publishing.

Hamstra, A. (1998). *Public Opinion about Biotechnology: a Survey of Surveys*. the Hague: European Federation of Biotechnology, Task Group on Public Perceptions of Biotechnology.

Hand, D. J. (2004). *Measurement theory and practice*. London: Arnold.

Harkness, J., van de Vijver, F., & Mohler, P. P. (2003). *Cross-Cultural Survey Methods*. Hoboken, NJ: Wiley.

Harré, R. (1984). Some reflections on the concept of social representation. *Social Research, 51*, 927-938.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heinen, T. (1996). *Latent class and discrete latent trait models: Similarities and differences*. Thousand Oaks, California: Sage.

Holland, P., & Wainer, H. (Eds.). (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Hviid Nielsen, T., Jelsoe, E., & Ohman, S. (2002). Traditional blue and modern green resistance. In M. W. Bauer & G. Gaskell (Eds.), *Biotechnology - the Making of a Global Controversy*. Cambridge: Cambridge University Press.

Inglehart, R. (1990). *Culture Shift in Advanced Industrial Societies*. Princeton: Princeton University Press.

Irwin, A., & Wynne, B. (Eds.). (1996). *Misunderstanding Science? The Public Reconstruction of Science and Technology*. Cambridge: Cambridge University Press.

Jasanoff, S. (2000a). *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton, NJ: Princeton University Press.

Jasanoff, S. (2000b). The "Science Wars" and American Politics. In M. Dierkes & C. von Grote (Eds.), *Between Understanding and Trust: the Public, Science and Technology*. Amsterdam: Harwood Academic Publishers.

Jaspars, J., & Fraser, C. (1984). Attitudes and social representations. In R. M. Farr & S. Moscovici (Eds.), *Social Representations* (pp. 101-123). Cambridge: Cambridge University Press.

Jodelet, D. (1996). Commentary on meanings of self-attributed ignorance. *Social Science Information, 35*(1), 111-120.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences*. New York: Seminar.

Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research, 36*, 347-387.

Jöreskog, K. G., & Sorbom, D. (1996). *LISREL 8: User's Reference Guide*. Hove and London: Scientific Software International.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The Handbook of Social Psychology* (Vol. 1). New York: McGraw-Hill.

Kerlinger, F. N. (1984). *Liberalism and conservatism: the nature and structure of social attitudes*. Hillsdale: Lawrence Erlbaum.

Kish, L. (1965). *Survey Sampling*. New York: Wiley.

Kohn, M. L. (1987). Cross-National Research as an Analytic Strategy: American Sociological Association, 1987 Presidential Address. *American Sociological Review, 52*(6, December), 713-731.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement Volume I: Additive and Polynomial Representations*. San Diego: Academic Press.

Krosnick, J. A., & Schuman, H. (1988). Attitude intensity, importance, and certainty and susceptibility to response effects. *Journal of Personality and Social Psychology, 59*, 940-952.

Kuha, J. (2004). AIC and BIC - Comparisons of Assumptions and Performance. *Sociological Methods and Research, 33*(2), 188-229.

Lanza, S. T., Collins, L. M., Lemmon, D., & Schafer, J. L. (in press). PROC LCA: A SAS Procedure for Latent Class Analysis. *Structural Equation Modeling.*

Lazarsfeld, P. F. (1957). Public Opinion and the Classical Tradition. *Public Opinion Quarterly, 21*(1), 39-53.

Lazarsfeld, P. F., Berelson, B., & Gaudet, S. (1948). *The people's choice: how the voter makes up his mind in a Presidential election campaign*. New York: Columbia University Press.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics, 50*, 325-335.

Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 5-53.

Lord, F. M. (1953). On the Statistical Treatment of Football Numbers. *The American Psychologist, 8*(12), 750-751.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley.

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

McKennell, A. C. (1977). Attitude Scale Construction. In C. A. O'Muircheartaigh & C. Payne (Eds.), *The Analysis of Survey Data.* Chichester: John Wiley.

Michell, J. (1990). *An Introduction to the Logic of Psychological Measurement.* New Jersey: Lawrence Erlbaum Associates, Inc.

Michell, J. (1999). *Measurement in Psychology.* Cambridge: Cambridge University Press.

Michell, J. (2003). The Quantitative Imperative. *Theory & Psychology, 13*(1), 5-31.

Miller, J. D. (1983). Scientific Literacy: a conceptual and empirical review. *Daedalus, Spring*, 29-48.

Miller, J. D. (1992). Toward a scientific understanding of the public understanding of science and technology. *Public Understanding of Science, 1*, 23-26.

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science, 7*, 203-223.

Miller, J. D. (2004). Public understanding of, and attitudes toward, scientific research: what we know and what we need to know. *Public Understanding of Science, 13*, 273-294.

Miller, J. D., & Kimmel, L. G. (2001). *Biomedical Communications - Purposes, Audiences and Strategies.* San Diego: Academic Press.

Miller, J. D., & Pardo, R. (2000). Civic Scientific Literacy and Attitude to Science and Technology: A Comparative Analysis of the European Union, the United States, Japan, and Canada. In M. Dierkes & C. von Grote (Eds.), *Between Understanding and Trust: the Public, Science and Technology.* Amsterdam: Harwood Academic Publishers.

Miller, J. D., Pardo, R., & Niwa, F. (1997). *Public Perceptions of Science and Technology: A Comparative Study of the European Union, the United States, Japan, and Canada.* Madrid: BBV Foundation.

Miller, S. (2001). Public Understanding of Science at the Crossroads. *Public Understanding of Science, 10*, 115-120.

Moscovici, S. (1961). *La Psychanalyse, son image et son public.* Paris: Presses universitaires de France.

Moscovici, S. (1973). Foreword. In C. N. N. E. M. Herzlich (Ed.), *Health and illness: a social psychological analysis.* London: Academic Press.

Moscovici, S. (1984). The Phenomenon of Social Representations. In R. M. Farr & S. Moscovici (Eds.), *Social Representations*. Cambridge: Cambridge University Press.

Muthén, L. K., & Muthén, B. O. (1998). Mplus User's Guide. Los Angeles, CA.

NSB. (2004). *Science and Engineering Indicators 2004*. Retrieved 30th August, 2007, from http://www.nsf.gov/statistics/seind04/pdfstart.htm

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62*, 42-55.

Osgood, D. (2001). Dig it up: global civil society's responses to plant biotechnology. In H. Anheier, M. Glasius & M. Kaldor (Eds.), *Global Civil Society 2001* (pp. 79-107). Oxford: Oxford University Press.

Pardo, R., & Calvo, F. (2002). Attitudes toward science among the European public: a methodological analysis. *Public Understanding of Science, 11*, 155-195.

Pardo, R., & Calvo, F. (2004). The cognitive dimension of public perceptions of science: methodological issues. *Public Understanding of Science, 13*, 203-227.

Pardo, R., Midden, C., & Miller, J. D. (2002). Attitudes towards biotechnology in the European Union. *Journal of Biotechnology, 98*, 9-24.

Peters, H. P. (2000). From Information to Attitudes? Thoughts on the Relationship between Knowledge about Science and Technology and Attitudes Toward Technologies. In M. Dierkes & C. von Grote (Eds.), *Between Understanding and Trust - the Public, Science and Technology*. Amsterdam: Harwood Academic Publishers.

Pidgeon, N., Poortinga, W., Rowe, G., Horlick-Jones, T., Walls, J., & O'Riordan, T. (2005). Using surveys in public participation processes for risk decision making: the case of the 2003 British *GM Nation?* public debate. *Risk Analysis, 25*(2), 467-479.

Potter, & Litton. (1985). Some problems underlying the theory of social representations. *British Journal of Social Psychology, 1*, 81-90.

Proctor, C. H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika, 35*, 73-78.

Pühler, A. (2002). Editorial: Highlights from ECB10 - Biotechnology and Society. *Journal of Biotechnology, 98*, 1-2.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2004). *GLLAMM Manual*: U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized mulitlevel structural equation modelling. *Psychometrika, 69*, 183-206.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research.

Raza, G., Singh, S., & Dutt, B. (2002). Public, Science, and Cultural Distance. *Science Communication, 23*(3), 292-309.

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components.* New Haven, CT: Yale University Press.

Rosnow, R. L., & Rosenthal, R. (1997). *People studying people: Artifacts and ethics in behavioural research.* New York: W. H. Freeman.

Rowe, G., Horlick-Jones, T., Walls, J., & Pidgeon, N. (2005). Difficulties in evaluating public engagement initiatives: reflections on an evaluation of the UK *GM Nation?* public debate about transgenic crops. *Public Understanding of Science, 14*, 331-352.

Saffir, M. (1937). A comparative study of scales constructed by three psychophysical methods. *Psychometrika, 2*, 179-198.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: an inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum.

Schuman, H., & Presser, S. (1981). *Questions and Answers in Attitude Surveys.* New York: Academic Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin Company.

Shapiro, R. B. (1999). *Open Letter from Monsanto CEO Robert B. Shapiro To Rockefeller Foundation President Gordon Conway and Others.* Retrieved 6th August, 2007, from http://www.monsanto.co.uk/news/ukshowlib.phtml?uid=9949

Shen, B. S. P. (1975). Scientific Literacy and the Public Understanding of Science. In S. Day (Ed.), *Communication of Scientific Information.* Basel: Karger.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66*(563-575).

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal and structural equation models.* Boca Raton, FL: Chapman & Hall/CRC.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. (2002). *Risk as analysis and risk as feelings: some thoughts about affect reason, risk, and rationality.* Paper presented at the Annual Meeting of the Society for Risk Analysis, New Orleans, LA.

Smith, T. W. (2003). Developing Comparable Questions in Cross-National Surveys. In J. Harkness, F. van de Vijver & P. P. Mohler (Eds.), *Cross-Cultural Survey Methods*. Hoboken, NJ: Wiley.

Spearman, C. (1904). 'General Intelligence', Objectively Determined and Measured. *American Journal of Psychology, 15*, 201-293.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science, 103*(2684), 677-680.

Stevens, S. S. (1959). Measurement. In C. West Churchman (Ed.), *Measurement: Definitions and Theories* (pp. 18-36). New York: John Wiley & Sons, Inc.

Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. F., Star, S. A., & Clausen, J. A. (1950). *Measurement and Prediction* (Vol. IV). Princeton: Princeton University Press.

Sturgis, P., & Allum, N. (2004). Science in society: re-evaluating the deficit model of public attitudes. *Public Understanding of Science, 13*, 55-74.

Sturgis, P., & Allum, N. (2006). *A literature review of research conducted on public interest, knowledge and attitudes to biomedical science.* University of Surrey: Report prepared for the Wellcome Trust.

Sturgis, P., Allum, N., Smith, P., & Woods, A. (forthcoming). The Measurement of Factual Knowledge in Surveys. *Public Opinion Quarterly.*

Tajfel, H. (1981). *Human groups and social categories: studies in social psychology.* Cambridge: Cambridge University Press.

Tashakkori, A., & Creswell, J. W. (2007). The New Era of Mixed Methods. *Journal of Mixed Methods Research, 1*(1), 3-7.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of Differential Item Functioning Using the Parameters of Item Response Models. In P. Holland & H. Wainer (Eds.), *Differential Item Functioning.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika, 47*, 397-412.

Thomas, W. T., & Znaniecki, F. (1918-1920). *The Polish Peasant in Europe and America.* Boston: Gorham Press.

Thorndike, E. L. (1918). The nature, purposes, and general methods of measurements of educational products. In G. M. Whipple (Ed.), *The Seventeenth Yearbook of the National Society for Study of Education. Part II: The Measurement of Educational Products.* Bloomington, IL: Public School Publishing Co.

Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review, 34*, 273-286.

Thurstone, L. L. (1927b). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology, 21*, 384-400.

Thurstone, L. L. (1935). *The Vectors of Mind.* Paper presented at the American Psychological Association, Chicago.

Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude.* Chicago: Chicago University Press.

Torgersen, H., Hampel, J., von Bergmann-Winberg, M.-L., Bridgman, E., Durant, J., Einsiedel, E., et al. (2002). Promise, problems and proxies: twenty-five years of debate and regulation in Europe. In M. W. Bauer & G. Gaskell (Eds.), *Biotechnology - the Making of a Global Controversy.* Cambridge: Cambridge University Press.

Turner, J., & Michael, M. (1996). What do we know about 'don't knows'? Or, contexts of ignorance. *Social Science Information, 35*(1), 15-37.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

UN. (2005). *United Nations Declaration on Human Cloning.* Retrieved 8th August, 2007, from http://www.un.org/law/cloning/; http://daccessdds.un.org/doc/UNDOC/GEN/N04/493/06/PDF/N0449306.pdf?OpenElement

van de Vijver, F., & Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research.* Thousand Oaks, California: Sage.

Van den Wittenboer, G., Hox, J. J., & de Leeuw, E. D. (2000). Latent Class Analysis of Respondent Scalability. *Quality and Quantity, 34*, 177-191.

van der Linden, W. J., & Hambleton, R. K. (Eds.). (1997). *Handbook of Modern Item Response Theory.* New York: Springer-Verlag.

Vermunt, J. K. (1997). LEM: A general program for the analysis of categorical data. Tilburg University.

Vermunt, J. K., & Magidson, J. (2005). Technical Guide for Latent GOLD 4.0: Basic and Advanced. Belmont Massachusetts: Statistical Innovations Inc.

von Grote, C., & Dierkes, M. (2000). Public Understanding of Science and Technology: State of the Art and Consequences for Future Research. In M. Dierkes & C. von Grote (Eds.), *Between Understanding and Trust - the Public, Science and Technology.* Amsterdam: Harwood Academic Publishers.

Wagner, W., & Hayes, N. (2005). *Everyday discourse and common sense: the theory of social representations.* Basingstoke: Palgrave Macmillan.

Weakliem, D. (1999). A Critique of the Bayesian Information Criterion for Model Selection. *Sociological Methods and Research, 27*, 359-397.

Withey, S. B. (1959). Public Opinion about Science and Scientists. *Public Opinion Quarterly, 23*(3, Autumn), 382-388.

Wynne, B. (2001). Creating Public Alienation: Expert Cultures of Risk and Ethics on GMOs. *Science as Culture, 10*(4), 445-481.

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: answering questions versus revealing preferences. *American Journal of Political Science, 36*(3), 579-616.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items. Chicago: Scientific Software International.